

# The Current Population Survey

## Design and Methodology

Issued January 1978



**U.S. Department  
of Commerce**  
Juanita M. Kreps, Secretary

**BUREAU OF  
THE CENSUS**  
Manuel D. Plotkin, Director

# CONTENTS

## CHAPTER I.

Introduction .....	1
The Current Population Survey .....	1
Estimates Provided .....	1
Summary of the Current CPS Samples .....	1
History .....	2
Major Changes in the Survey: A Chronology .....	2
Impact of Research in the Labor Force Survey .....	4
The Scope of This Report .....	5

## CHAPTER II.

Design of the CPS National Sample .....	6
Introduction .....	6
Survey Specifications and Design .....	6
Survey Specifications .....	6
Survey Design .....	6
Selection of the Sample of Primary Sampling Units .....	7
Definition of the Primary Sampling Units .....	7
Stratification of Primary Sampling Units .....	8
Selection of Sample PSU's for the A-Design .....	10
Selection of Sample PSU's for the C-Design .....	12
Number of Sample PSU's in the A- and C-Design Samples .....	13
Selection of the Sample Within Sample PSU's .....	13
Calculate the Within-PSU Selection Probability .....	14
Resources for Sampling Within the PSU .....	15
Assemble the Within-PSU Sample .....	18

## CHAPTER III.

Rotation of the Sample and Replacement of PSU's .....	21
Introduction .....	21
Definitions and Purposes .....	21
The Rotation System Used in the CPS .....	21
The Operating System .....	21
The Rotation Chart .....	22
Overlap of the Sample .....	22
Introducing a New Rotating Sample .....	22
Alternative Rotation Systems .....	24
Selection of Additional Samples for Rotation .....	24
The System for Generating Additional Samples .....	24
How Additional Samples Are Generated .....	24
Assignment of Samples to Census Bureau Surveys .....	25
Replacement of PSU's .....	25
The Problem of Exhausting USU's in the PSU .....	25
Alternate Solutions .....	25
Procedure for PSU Replacement .....	25
Impact of PSU Replacement on the Survey .....	26

## CHAPTER IV.

Survey Operations .....	27
Introduction .....	27
Organizations Involved in Survey Operations .....	27
Washington, D.C. ....	27
Jeffersonville, Ind. ....	28
Data Collection Activities for a Survey Month .....	28
Identify the USU's in the Current Sample .....	28
List the Living Quarters Within the Sample Segment .....	30
Conduct the Interviews .....	33
Control of the Data Collection Operations .....	45
Questionnaire Edits .....	45
Noninterview Rate .....	46
Address Coverage Check .....	46
Reinterview .....	46
Production Standards .....	48
Observation of Field Work .....	50
Evaluation of the Interviewers' Work .....	51
Processing the Data .....	51
The Interviewer .....	51
The Regional Office .....	51
Data Processing Division, Jeffersonville, Ind. ....	51
The Data Production Cycle .....	53
Washington, D.C. ....	53

## CHAPTER V.

Preparation of Estimates from the National Sample .....	55
Introduction .....	55
Preparation of Unbiased Estimates .....	55
Weights for Rotation Groups .....	56
Special Weights .....	56
Adjustment for Type A Noninterviews .....	56
Race and Collapsed-Race Definitions .....	56
Noninterview Clusters .....	57
Noninterview Adjustment Cells .....	57
Computing Noninterview Adjustment Factors .....	57
Priorities for Combining Noninterview Adjustment Cells .....	58
Weights After the Noninterview Adjustment .....	58
Ratio Estimates .....	58
First-Stage Ratio Adjustment to 1970 Census Population Counts .....	58
Second-Stage Ratio Adjustment to Current, Independent Estimates of the Population .....	59
Composite Estimator .....	62
Seasonal Adjustments .....	63
Estimation Formulas, Monthly CPS Labor Force Data ..	63

## CONTENTS—Continued

Special Problems in Estimation .....	64
Averaging Monthly Estimates. ....	64
Family and Household Estimates for March Supplement Data .....	65
Monthly Estimates of Occupied Housing Units .....	66
Quarterly Estimates of Vacancy Rates. ....	66
Estimates for Children in the September and October Supplements. ....	67

## CHAPTER VI.

Estimates for States and Local Areas, by Supplementing the National Sample .....	68
Introduction. ....	68
Estimates for 1973 .....	68
Estimates for 1974 and 1975. ....	68
Estimates for 1976 and Subsequent Years .....	68
Design of the CETA Supplementary Sample. ....	68
Design Requirements for State Samples. ....	68
Features of the Supplemental Sample for the States. ....	69
Stratification and Selection of Supplemental Sample PSU's .....	69
Survey Operations. ....	73
Sample Introduced in Three Groups .....	73
Special Problems in Alaska .....	73
Estimation Procedure for States for 1976. ....	73
Preliminary Procedures for the Supplemented Areas. ....	73
Preliminary Procedures for the Nonsupplemented States .....	73
Combine the Sample Records for All States .....	73
Adjustment for Noninterview for States. ....	74
First-Stage Ratio Adjustment to 1970 Census Population Counts for the States .....	74
Second-Stage Ratio Adjustment to Current Independent National Estimates of the Population .....	75
Composite Estimator for States .....	75
Annual Average of State Monthly Estimates. ....	75
Ratio Estimate to Current Independent Estimates of the State Population. ....	75

## CHAPTER VII.

Accuracy of the CPS Estimates .....	77
Introduction. ....	77
Precision of the CPS Estimates. ....	77
Concept of Sampling Variability .....	77
Summary of the Response Variance Model. ....	77
Estimates of Total Variance of the CPS. ....	78
Estimates of the Response Variance Components .....	78
Bias in CPS Estimates. ....	81
Concept of Bias. ....	81
Potential Biases in the CPS .....	82

Circumstances Affecting the Objective Measurement of Labor Force Status. ....	87
Effect of Bias in Monthly Levels on Estimates of Month-to-Month Change .....	88
The Mean Square Error of the CPS Estimates .....	88

## CHAPTER VIII.

Calculation and Presentation of Sampling Errors. ....	90
Introduction. ....	90
Earlier Methods of Estimating Sampling Error. ....	90
Sampling Errors of Unbiased Estimates .....	90
Variance Estimates by the Replication Method. ....	91
Keyfitz Methods. ....	91
Current Method for Estimating Sampling Errors .....	92
Variances for State and Local Area Estimates .....	92
Complicating Factors. ....	92
Imputed Variances for States. ....	93
Generalizing Estimated Sampling Errors. ....	93
Why Generalized Sampling Errors Are Used .....	93
Method of Generalizing the Sampling Errors. ....	93
Statistical Review of Publications. ....	94
Variance Estimates to Determine Optimum Survey Design. ....	95
Variance Components Due to Stages of Sampling .....	96
Total Variance as Affected by Estimation .....	98

## CHAPTER IX.

Current Population Survey Used for Supplemental Inquiries .....	100
Introduction. ....	100
How the CPS is Used for Supplemental Inquiries. ....	100
Criteria for Acceptable Supplemental Inquiries. ....	100
Clearance by the Office of Management and Budget. ....	101

REFERENCES .....	102
------------------	-----

## APPENDIX A.

Controlled Selection of Sample PSU's .....	105
Introduction. ....	105
The Controlled Selection Problem As It Applies to the CPS. ....	105
Derive the Controls That Define Admissible Patterns .....	105
Develop and Assign Probabilities to the Admissible Patterns in a Complete Set. ....	106
Selection of the PSU's in Sample .....	110
Some Observations .....	110

## CONTENTS—Continued

## APPENDIX B.

<b>Sampling Within PSU's: Selecting the Sample of 1970 ED's</b> .....	111
Introduction .....	111
Sequence ED Records .....	111
Compute a Measure of Size for Each ED .....	111
Select the Sample of ED's .....	113
Identify the Sample USU in the Sample ED .....	113
Determine the Random Starts for the Sampling Operation .....	113
Determine Area or List Procedures for the Sample ED ..	114
Identify USU's for Subsequent (Rotating) Samples .....	114
Assign Hit Numbers .....	114
Segment Number .....	114

## APPENDIX C.

<b>Sampling Within PSU's: Selecting Sample USU's From Area ED's</b> .....	115
Introduction .....	115
The Procedure .....	115
Locate Each Address Recorded in the Address Register on a Map of the ED .....	115
Use the Addresses and Geographic Features to Map the ED into Chunks .....	115
Allocate USU's to the Chunks .....	115
Select the Sample USU .....	115
Identify the Sample Segment .....	121
Definition of Chunks Within Area ED's .....	121

## APPENDIX D.

<b>Sampling Within PSU's: Selecting Sample USU's From List ED's</b> .....	122
Introduction .....	122
The Procedure .....	122
Prepare Corrected Tape Records of Addresses of 1970 Housing Units in List Sample ED's .....	122
Sequence the 1970 Census Housing Unit Records Within the ED .....	122
Determine the Number and Size of USU's in the ED ..	123
Assign the Addresses of 1970 Census Housing Units to USU's .....	124
Identify USU's for Subsequent (Rotating) Samples ..	125
Determine the Sample in the Special Place Portion of the ED .....	125
Illustration .....	125

## APPENDIX E.

<b>Sampling Within PSU's: Reservation of USU's for Current Survey Samples</b> .....	127
---	-----

Introduction .....	127
System for Allocating USU's .....	127
Allocations, Based on Initial CPS Selections .....	127
Adjustment for Different Survey Selection Probabilities .....	127
Conflict in Designating Random Starts .....	130
Reservations in Sample PSU's, by Type .....	132
Reservations in SR PSU's .....	134
Reservations in NSR PSU's in Both A- and C-Design Samples .....	136
Reservations in NSR PSU's in the A-Design Sample ..	136
Reservations in NSR PSU's in the C-Design Sample ..	136
Computation of the Required Number of USU's Between Hits .....	136
NSR PSU's in the A-Design Sample .....	136
Self-Representing PSU's .....	137
NSR PSU's in the A- and C-Design Samples .....	137
NSR PSU's in the C-Design Sample .....	137

## APPENDIX F.

<b>Maintaining the Desired Sample Size</b> .....	138
Introduction .....	138
Sample Reduction .....	138
National Sample .....	138
Supplemental Samples for the States .....	139

## APPENDIX G.

<b>Treatment of Unusually Large USU's</b> .....	140
Introduction .....	140
The Rare Events Procedure .....	140
The Rare Events Universe .....	140
Subsampling From the Rare Events Universe .....	140
Biases in the Rare Events Procedure .....	140
Failure to Include All Large USU's in the REU .....	141
Introduction of New Redesign .....	141

## APPENDIX H.

<b>Rules for Usual Place of Residence</b> .....	142
Introduction .....	142
Rules .....	142
Families With Two or More Homes .....	142
Seamen .....	142
Other Occupants .....	142
Persons With No Usual Place of Residence Elsewhere ..	142
Citizens of Foreign Countries Temporarily in the United States .....	142
Members of the Armed Forces .....	142
Students and Student Nurses .....	143
Persons With Two Concurrent Residences .....	144



## CONTENTS—Continued

Persons in Vacation Homes, Tourist Cabins, or Trailers .....	144
Inmates of Specified Institutions .....	144

## APPENDIX I.

Supplemental Surveys Using the CPS Sample in 1975 and 1976 .....	145
Introduction .....	145
1975 Supplemental Surveys .....	145
1976 Supplemental Surveys .....	146

## APPENDIX J.

Organization and Training of the Data Collection Staff .....	147
Introduction .....	147
Regional Offices .....	147
Organization .....	147
Recruitment of Interviewers .....	147
Interviewer Training Procedures .....	147
Training of Regional Office Staff .....	149

## APPENDIX K.

Variance Estimation: CPS National Samples .....	150
Introduction .....	150
Variance of a Linear Function of Sample Totals .....	150
Estimating the Variance in SR Strata .....	150
Estimating the Total Variance in NSR Strata .....	152
Variance Components for the Stages of Sampling .....	153
Variance of Nonlinear Estimators .....	153
Theoretical Foundation, A Summary .....	154
Implications of the Assumptions .....	154

The Variance of a Nonlinear Estimator: An Illustration .....	154
The Nonlinear Estimator .....	154
Determine the Partial Derivatives .....	155
Derive the Linear Form of the Estimate .....	155
The Variance Estimator for the Linear Form .....	157
Variances and Components of Variances of Alternative Estimates .....	157

## APPENDIX L.

Reliability Statement for a CPS Report: An Illustration .....	159
Introduction .....	159
Source and Reliability of the Estimates for Educational Attainment: March 1975 .....	159
Source of Data .....	159
Reliability of the Estimates .....	159
Standard Error Tables and Their Use .....	160
Standard Error of a Difference .....	161
Standard Error of a Median .....	161
Standard Error of a Ratio .....	162

## APPENDIX M.

Location and Description of Sample Areas .....	163
Introduction .....	163
Sources .....	163
Two Four-Color Maps of the United States .....	163
One Three-Color Map Covering Selected PSU's of the Boston and New York Regional Offices .....	163
List of PSU's in Current Survey Samples .....	163
PSU Replacement Schedules .....	164

SUBJECT INDEX .....	168
---------------------	-----

## TABLES

II-1. Number and Population of Strata for 376-PSU A-Design Samples, by Census Region .....	11
II-2. Coefficient of Variation of Population of Nonself-Representing Strata for 357- and 376-Area A-Design Samples, by Census Region .....	11
II-3. Classification of New Strata in 376-Area A- and 266-Area C-Design Samples, by Status in Old A- and C-Design Samples .....	14
II-4. Sampling of PSU's Within Stratum and USU's Within PSU for PSU's 374, 412, and 433 .....	15
III-1. Rotation System (4-8-4) Proportion of USU's Common in Two CPS Samples, Separated by Selected Intervals .....	22
IV-1. Number and Distribution of Total, Noninterview and Interview Units in A- and C-Design CPS Sample .....	41
IV-2. Current Population Survey National Sample Type A Noninterview Rates, by Reason: 1975-1976, by Month; Annual Rates for 1972-1976 .....	42

## CONTENTS--Continued

## TABLES--Continued

IV-3.	Percent of Interviews Completed by Telephone, by Month in Sample, Selected Months . . . . .	44
IV-4.	Percent of Persons in Given Labor Force Status in CPS Reinterview After Reconciliation, and Percent Identically Classified in the Original Interview, by Method of Interview: July 1974-June 1975 . . . . .	45
IV-5.	Distribution of CPS Interviewer Questionnaire Edit Error Rates, National Sample: July 1974-June 1975 . . . . .	46
IV-6.	Distribution of CPS Interviewer Type A Noninterview Rates, National Sample: July 1974-June 1975 . . . . .	47
IV-7.	Rates of Net Coverage Error in the Count of Persons in the Current Population Survey National Sample, by Month: 1974-1975 . . . . .	48
IV-8.	Distribution of Interviewer Monthly Gross Coverage Error Rates in the Current Population Survey National Sample: July 1974-June 1975 . . . . .	49
IV-9.	Distribution of Interviewer Gross Content Error Rates in the Current Population Survey National Sample: July 1974-June 1975 . . . . .	49
IV-10.	Tolerance Limits for Gross Coverage and Content Error in the Current Population Survey: June 1975 . . . . .	50
IV-11.	Distribution of CPS Interviewer Production Ratios: July 1974-June 1975 . . . . .	50
IV-12.	Quarterly Evaluation of CPS Interviewers, National Sample: 1974 . . . . .	52
IV-13.	Shipment of CPS Questionnaires, Regional Offices to Jeffersonville to Washington: January 1975 . . . . .	54
V-1.	First-Stage Ratio Estimate Factors for NSR PSU's in Combined A- and C-Design National Sample: August 1974- March 1975 . . . . .	59
V-2.	Intermediate Second-Stage Ratio Estimate Factors; Ratio of Independent Estimate to CPS First-Stage Ratio Estimate of Not-White Population, by Age, Sex, and Race: June 1975 . . . . .	60
V-3.	Final Second-Stage Ratio Estimate Factors; Ratio of Independent Estimate to CPS First-Stage Ratio Estimate of Population, by Collapsed Race, Age, and Sex: June 1975 . . . . .	61
V-4.	Coverage Ratios for the CPS National Sample, by Month: 1974-1975 . . . . .	61
V-5.	Coverage Ratios, by Age, Sex, and Collapsed Race; CPS National Sample: Average for July 1974-June 1975 . . . . .	62
VI-1.	Design Effect and Percent of Population, 16 years Old and Over, for Unbiased and Combined First- and Second- Stage Ratio Estimates . . . . .	71
VI-2.	Number of Independent CPS Monthly Samples Equivalent to an Average of 3, 6, 9, and 12 Consecutive Months of the CPS . . . . .	72
VI-3.	First-Stage Ratio Estimate Factors for State Estimates, SMSA PSU's: August 1976 . . . . .	75
VI-4.	First-Stage Ratio Estimate Factors for State Estimates, Non-SMSA PSU's: August 1976 . . . . .	76
VII-5.	Gross Difference Rate for Selected Labor Force Characteristics, CPS Reinterview Unreconciled Sample, by Quarter: 1975 . . . . .	86
VII-6.	Effect of Bias on the Ratio of the Root Mean Square Error to the Standard Error . . . . .	88
VII-7.	Effect of Bias on the Probability of an Error Greater Than 1.96 Times the Standard Error . . . . .	89
VII-1.	Estimated Ratio of the Correlated Component of Response Variance to the Simple Response Variance Plus the Sampling Variance for 143 PSU's in CPS Sample: July-December 1966 . . . . .	80
VII-2.	Index of Inconsistency for Selected Labor Force Characteristics in the 20-Percent Unreconciled CPS Reinterview Sample, by Quarter: 1973-1975 . . . . .	81
VII-3.	Estimates of the Percent of Net Undercount of the Population in the 1970 Census, by Sex, Race, and Broad Age Groups . . . . .	83
VII-4.	Rotation Group Bias Indexes in the CPS for Selected Labor Force Characteristics and Type A Noninterview Rate, Average: 1973-1975 . . . . .	84
VIII-1.	Components of Estimated Sampling Variance for CPS Composite Estimates of U.S. Level, Monthly Averages: 1975 . .	95
VIII-2.	Components of Estimated Sampling Variance for CPS Composite Estimates of U.S. Month-to-Month Change, Monthly Averages: 1975 . . . . .	96
VIII-3.	Variance of Composite Estimates of Monthly U.S. Level and Variance Factors for Selected Estimators, Monthly Averages: 1975 . . . . .	97
VIII-4.	Variance of Composite Estimates of U.S. Month-to-Month Change and Variance Factors for Selected Estimators, Monthly Averages: 1975 . . . . .	98

# CONTENTS—Continued

## TABLES—Continued

A-1.	PSU Selection Probabilities, by Stratum and by Control Criteria, for Selected States of the West Region: 1970 CPS A-Design NSR Strata .....	107
A-2.	Expected Values and Control Statements for Geographic Areas and for PSU's in Old A-Design Sample, NSR Portion of the West Region: 1970 CPS A-Design .....	108
A-3.	A Hypothetical Example of the Controlled Selection of PSU's .....	109
B-1.	Selection of A-Design Sample ED's Within PSU 412: LaPorte-Starke Counties, Ind. ....	112
D-1.	1970 Census Housing Units Assigned to USU's by the Computer for a Selected Portion of List ED 091-514 in PSU 412, LaPorte-Starke Counties, Ind. ....	126
E-1.	Random Starts to Assign Matching Samples in Two Hypothetical PSU's. ....	134
F-1.	Basic Weights for A-, C-, and Combined A- and C-Design National Samples. ....	138
G-1.	Special Weights for Subsamples of Units From the Rare Events Universe, by Sample .....	141
H-1.	Summary Table for Determining Usual Place of Business. ....	143
J-1.	Monthly Sample Workloads, by Regional Office .....	149
K-1.	Partial Derivatives of the Function $x''$ at (K/23) With Respect to the Variables in (K/24) .....	156
K-2.	Constants for Use in Expression (K/37) to Approximate the Variance of Selected Estimators .....	158
L-1.	Standard Errors for Estimated Numbers of Persons Enrolled in School, CPS Estimates for Total or White Population .....	160
L-2.	Standard Errors of Estimated Percentages, CPS Estimates for Total or White Population. ....	161
M-1.	PSU Replacement Schedule for the CPS National Sample: August 1972-August 1982. ....	164
M-2.	PSU Replacement Schedule for the 461-Area Design and Supplemental Sample for State Estimates, CETA: August 1976-August 1982. ....	165

## FIGURES

II-1.	Regions and Geographic Divisions of the United States. ....	9
II-2.	Method of Sampling USU's Within PSU's, by Category of 1970 Census ED. ....	19
III-1.	Rotation Chart of CPS A- and C-Design Samples: November 1972-July 1975 .....	23
IV-1.	Sample Report (Form 11-200). ....	29
IV-2.	Area Segment Listing Sheet (Form 11-212) .....	31
IV-3.	Special Place Listing Sheet (Form 11-213) .....	32
IV-4.	Address Listing Sheet (Form 11-211A) .....	34
IV-5.	Current Population Survey Letter. ....	35
IV-6.	Control Card (Form CPS-260) .....	37
IV-7.	Current Population Survey (Form CPS-1). ....	39
IV-8.	Housing Vacancy Survey (Form HVS-1) .....	43

## CONTENTS—Continued

## FIGURES—Continued

C-1.	Census ED Map, ED 091-567, LaPorte County, Ind..	116
C-2.	Work Copy Map, ED 091-567, LaPorte County, Ind.	117
C-3.	Area ED Accumulation Worksheet (Form BC-59)	118
C-4.	ED Measure Check List, Current Surveys (Form BC 2505)	119
C-5.	Segment Map	120
E-1.	Reservation of USU's for the 461-Area Sample Design SR PSU's	128
E-2.	Reservation of USU's for the 461-Area Sample Design NSR PSU's in A- and C-Design Samples.	131
E-3.	Reservation of USU's for the 461-Area Sample Design NSR PSU's in A-Design Sample.	133
E-4.	Reservation of USU's for the 461-Area Sample Design NSR PSU's in C-Design Sample	135
J-1.	Census Field Regional Boundaries.	148
M-1.	461-Area PSU's for the Current Population Survey National Sample (Form 11-5)	166
M-2.	461 PSU's for the Area Sample and State Supplements: August 1976 (Form 11-6)	167

## Chapter I.

### INTRODUCTION

#### THE CURRENT POPULATION SURVEY

The Current Population Survey (CPS) is a household sample survey conducted monthly by the Bureau of the Census to provide estimates of employment, unemployment, and other characteristics of the general labor force, of the population as a whole, and of various subgroups of the population.

#### Estimates Provided

The CPS was initially designed primarily to produce timely estimates on a sample basis with measurable reliability for labor force data at the U.S. level each month. Although this summarizes the major objectives considered in designing the original sample program, the CPS is now used for purposes well beyond those originally envisioned. Expanding needs for additional current data by Government and other users have been met by adding additional questions to the monthly interview, in part by occasional supplementary inquiries. The survey covers the civilian noninstitutional population of the United States.

The CPS provides a large amount of detail not otherwise available on the economic status and activities of the population of the United States. It is the only source of monthly estimates of total employment, both farm and nonfarm; of nonfarm self-employed persons, domestics, and unpaid helpers in nonfarm family enterprises, as well as wage and salaried employees; and of the total unemployment, whether or not covered by unemployment insurance. It is the only comprehensive source of information on the personal characteristics of the total population both in and out of the labor force, such as age and sex, race, marital and family status, veteran status, educational background, and ethnic origin.

It provides the only available distributions of workers by the numbers of hours worked (as distinguished from aggregate or average hours for an industry), permitting separate analyses of part-time workers, workers on overtime, etc., as well as being the only comprehensive current source of information on the occupation of workers and the industries in which they work.

Information is available from the survey not only for persons currently in the labor force but also for those who are outside the labor force. The characteristics of such persons—whether married women with or without young children, disabled persons, students, older retired workers, etc.—can be determined. Information on their current desire for work, their past work experience, and their intentions as to jobseeking are available from a subsample.

Data have been provided at the national level since the inception of the CPS. After a few years, data for the census regions were also provided. More recently, funds have become available permitting expansion of the sample to increase the reliability of data for States and selected standard metropolitan statistical areas (SMSA's). (See ch. VI.) To improve the reliability of estimates tabulated at levels below the census regions, some of the monthly estimates are cumulated for publication as quarterly and annual averages.

National estimates from the CPS of the size, composition, and changes in the composition of the labor force are published each month by the Bureau of Labor Statistics in *Employment and Earnings* [75]. Estimates of the total and civilian labor force are produced in considerable detail at the national level, in most instances, by sex, race, and age. Unemployment rates are given, in addition, by marital status and relationship to the household head and by occupation and industry, duration of unemployment, whether seeking full- or part-time work, reason for unemployment, job search methods used, etc. The number of persons employed is shown by agriculture and nonagricultural industries, full- or part-time status, and hours worked by major industry group and major occupational group.<sup>1</sup> Seasonally adjusted data are provided for many of these series. In addition, the CPS is the source of periodic studies of personal and family income, migration, educational attainment, and other demographic, social, and economic topics. (App. I summarizes the subjects covered by the periodic studies during calendar years 1975 and 1976.)

Several subsets of these estimates are published in less detail as annual averages for the larger SMSA's and for States [73; 74].

Since 1968, the individual sample records from the CPS have been made available in the form of computer tape files for public use. These public use files contain all of the demographic and economic information for each person in every interviewed household in the survey; sufficient geographic information is removed, however, to insure the confidentiality of the respondent households. These microdata files have been increasingly used by private and government researchers for a variety of purposes. For example, proposed changes in various tax and public assistance programs have been simulated using these files to study the potential benefit to beneficiaries and costs to taxpayers, before the programs are enacted.

The CPS is being used increasingly to meet legislative requirements for data to administer public programs. For example, sample estimates of labor force data for the States are used to allocate Federal funds to the States under the CETA program [35]. For other programs, the Bureau of the Census now uses special statistical techniques to prepare various income and poverty estimates for the States from special tabulations of the CPS and estimates from administrative data.

#### Summary of the Current CPS Samples

The present Current Population Survey comprises two sets of samples. The national sample was originally designed to provide estimates at the U.S. level, but also has been used to make estimates for some of the larger States and SMSA's. The State supplement comprises additional samples selected within the District of Columbia and several of the smaller States and is designed to improve the reliability of estimates made for these areas from the national sample.

The present CPS national sample consists of households and persons in group quarters in 461 primary sampling units (PSU's) where each PSU consists of an independent city or county or

<sup>1</sup> A more complete summary of the tabulations is given in [70].

two or more contiguous counties. The sample of 461 PSU's comprises 923 counties and independent cities. In an average month during 1975, the number of housing units or living quarters designated for the national sample was about 58,000, of which about 3,000 were found to be nonexistent, demolished, or no longer used as living quarters. Of the remaining 55,000 units assigned for interview, about 45,000 were interviewed households, 2,000 were households at which the members were not available for interview and 8,000 were found to be vacant, occupied by persons with usual residence elsewhere, or otherwise not eligible for interview. For survey purposes, a household is defined as all the persons who occupy a housing unit. A housing unit is a room or group of rooms intended for occupancy as separate living quarters and having either a separate entrance or complete cooking facilities for the exclusive use of the occupants.

The State supplementary samples were first introduced in July 1975. At the end of 1976, the supplementary samples included about 11,000 units assigned for interview each month.

## HISTORY

The Current Population Survey had its origin in a program set up to provide direct measurement of unemployment each month on a sample basis. There were several earlier attempts to estimate the number of unemployed using various devices, ranging from guesses to enumerative counts. The problem became especially pressing during the Economic Depression of the 1930's [8].

The Enumerative Check Census, taken as a part of the 1937 unemployment registration, was the first attempt to estimate unemployment on a nationwide basis using probability sampling. During the latter half of the 1930's, the Work Projects Administration (WPA), the Works Progress Administration (WPA) prior to 1939, research staff began developing techniques for measuring unemployment, first on a local area basis and subsequently on a national basis. This research and the experience with the Enumerative Check Census led to the Sample Survey of Unemployment which was started in March 1940 as a monthly activity by the WPA [12].

### Major Changes in the Survey: A Chronology

In August 1942, responsibility for the Sample Survey of Unemployment was transferred to the Bureau of the Census and in October 1943, the sample was thoroughly revised. At that time, the use of probability sampling was expanded to cover the entire sample, and new sampling theory and principles were developed and applied to increase the efficiency of the design [20; 21, ch. 12, sec. B]. The households in the revised sample were in 68 PSU's, comprising 125 counties and independent cities. By 1945,<sup>2</sup> about 25,000 housing units were designated for the sample, of which about 21,000 contained interviewed households.

One of the most important substantive changes in the CPS sample design took place in 1954 when, for the same total budget, the sample of PSU's was expanded from 68 to 230,

without change in the number of sample households [23]. This redesigned sample, as a result of a more efficient system of field organization and supervision, made it possible to provide more information per unit of cost, and thus increase the accuracy of published statistics, and made it possible to provide more reliable regional as well as national estimates for some characteristics.

Since that time, a major revision in the sample has followed each of the decennial censuses. These and other important modifications in the Current Population Survey are included in the following chronological list [70]:

**July 1945**—Revision of CPS schedule. The questionnaire was revised to introduce four basic employment status questions. Before that time, the schedule did not contain specific question wording. Special studies showed that this and other defects resulted in the exclusion from the labor force statistics of large numbers of part-time and intermittent workers, particularly unpaid family workers. The question wording of these four items has been modified slightly since 1945, but the basic content has been unchanged.

**August 1947**—Revision in sample selection method. The method of selecting sample units within a sample area was changed so that each selected unit would have the same basic weight in the tabulations. This change simplified tabulation and estimation procedures.

**July 1949**—Introduction of special dwelling places. The sample coverage was extended to special dwelling places—hotels, motels, trailer camps, etc. This led to improvements in the statistics, since residents of these places have characteristics different from the remainder of the population.

**February 1952**—Introduction of document sensing. The CPS schedule was converted to a document-sensing card. In this procedure (replaced more recently by the FOSDIC system), entries were made by drawing a line through the oval representing the correct answer using a special pencil with electrographic lead. Punchcards were automatically prepared from the schedules by a special documenting-sensing machine.

**January 1953**—Shift to 1950 population census data for ratio estimates. Starting in January 1953, population data from the 1950 census were introduced into the Current Population Survey estimation procedure. (See ch. V for a description of these estimates.) Prior to that date, the ratio estimates had been based on 1940 census relationships for the first-stage ratio estimate and 1940 population census data brought forward to take account of births, deaths, etc., for the second-stage ratio estimate. In September 1953, "color" was substituted for "veteran status" in the second-stage ratio estimate, making it feasible to publish some separate, absolute numbers for persons by race, whereas only percent distributions had previously been provided.

**July 1953**—Change to 4-8-4 rotation system. The present sample rotation system was adopted whereby households are interviewed for 4 consecutive months one year, leave the sample for 8 months, and return for the same period of 4 months of the following year. Prior to that time, households were interviewed for 6 consecutive months and then replaced. The new system provided some year-to-year overlap in the sample, thus improving the measurement of the statistics over time. (See ch. III.)

<sup>2</sup>Before 1945, there was a larger sample with differential sampling rates among the largest sample PSU's.

**September 1953**—Conversion of tabulations to high-speed electronic equipment. This change speeded up the tabulations considerably, made possible improvements in estimation methods, and substantially expanded the scope and content of the tabulations for basic data and computation of sampling variability. A shift to more modern computers was made in 1959; this process will continue as equipment is updated and replaced.

**February 1954**—Changeover to 230-area sample. The CPS sample was expanded from 68 to 230 sample PSU's, although the overall sample size of 25,000 designated sample units was retained. The 230 areas comprised 453 counties and independent cities. At the same time, a substantially improved estimation procedure (composite estimate) was introduced which took advantage of the large overlap in the sample from month to month. These two changes improved the reliability of most of the major statistics by an amount equivalent to doubling the sample size.

**May 1955**—Addition of monthly questions concerning part-time workers. Monthly questions on the reasons for part-time work were added to the standard set of employment status items. This information had been collected quarterly or less frequently in the past and was found to be valuable in studying current labor market trends.

**July 1955**—Changes in survey week. The CPS survey week was changed to the calendar week containing the 12th day of the month for greater consistency with the time reference of other statistics in the employment field. Previously, the survey week had been the calendar week containing the eighth day of the month.

**May 1956**—Expansion to 330-area sample. The CPS was expanded from a 230-PSU to a 330-PSU sample. The overall sample size was increased by roughly two-thirds to a total of about 40,000 designated units (about 35,000 occupied units). The expanded sample was located in 638 counties and independent cities with at least some households in the 48 contiguous States. All of the former 230 areas were continued in the expanded sample. The expansion increased the reliability of the major statistics by around 20 percent and made possible publication of greater detail.

**January 1957**—Change in employment status definition. Two relatively small groups of persons, formerly classified as employed "with a job but not at work," were assigned to different classifications as a result of a comprehensive interagency review of the Government's employment and unemployment data. These groups were persons on layoff with definite instructions to return to work within 30 days of the layoff date and persons waiting to start new wage and salary jobs within 30 days of interview. Most of the persons in these two groups were shifted to the unemployed classification. The only exception was the small subgroup in school during the survey week and waiting to start new jobs that was transferred to "not in labor force." The changes in definition did not affect the basic questions or enumeration procedures.

**June 1957**—Seasonal adjustment. Limited seasonally adjusted data on unemployment were introduced early in 1955. Some extension of the data—using more refined seasonal adjustment methods programmed on electronic computers—was instituted in July 1957, including a seasonally adjusted rate of unemployment and charting of seasonally adjusted total employment and

unemployment. Significant improvements in methodology grew out of research conducted at the Bureau of Labor Statistics (BLS) and Census Bureau in the ensuing years [69].

**July 1959**—Transfer of functions. Responsibility for the planning, analysis, and publication of the labor force statistics from the Current Population Survey was transferred to the Bureau of Labor Statistics as part of a major exchange of statistical functions between the Commerce and Labor Departments. The Bureau of the Census continued to have (and still has) responsibility for the collection and tabulation of these statistics, for maintenance of the CPS sample, and related methodological research. Interagency review of policy and technical issues continues to be conducted by committees under the aegis of the Statistical Policy Division, Office of Management and Budget.

**January 1960**—Addition of Alaska and Hawaii to the population estimates and the CPS sample. Upon achieving statehood, Alaska and Hawaii were introduced into the independent estimates of the population and into the sample survey, thereby increasing the number of PSU's in the sample from 330 to 333. The addition of these two States affected the comparability of population and labor force data with previous years. This inclusion resulted in an increase of about half a million in the noninstitutional population of working age and about 300,000 in the labor force, four-fifths of this in nonagricultural employment. The levels of other labor force categories were not appreciably changed.

**October 1961**—Conversion to the FOSDIC system. The CPS questionnaire was converted to the FOSDIC type used in the 1960 census; this system is still currently in use. Entries are made by filling small circles with an ordinary lead pencil. Microfilms of these questionnaires are scanned by a special mechanical reading device which transfers the information directly to computer tape. This system permits a larger-size form and a more flexible arrangement of items than the previous document-sensing procedure and does not require the preparation of punchcards.

**March 1963**—Updating of sample and population data used in ratio estimates. During the period December 1961 to March 1963, the CPS sample was revised gradually. This change was made to reflect the changes in population size and distribution revealed by the 1960 census and in the industrial mix among areas. The overall sample size was unchanged, but the number of sample areas was increased slightly to 357 PSU's to provide for greater coverage in fast-growing portions of the country. Also in a major part of the sample, selection of units from census lists, developed in the 1960 census, was introduced to replace area sampling [44, p. 69]. These changes resulted in a further gain in reliability of about 5 percent for most statistics. The use of updated population information from the census was introduced in April 1962 into the first- and second-stage ratio estimates used in the CPS.

**January 1963**—New descriptive information. In response to recommendations of a special review committee [30], two new items were added to the monthly questionnaire. The first was an item, formerly carried only intermittently, on whether the unemployed were seeking full- or part-time work. The second was an expanded item on household relationship, formerly included

only annually, to provide more detail on the unemployed persons by household relationship and marital status.

**January 1967—Expansion to 449-PSU sample.** The CPS was expanded from a 357-PSU to a 449-PSU sample. As a result of an increase in total budget, the overall sample size was increased by roughly 50 percent to a total of about 60,000 housing units (52,500 occupied units). The expanded sample had households in 863 counties and independent cities with at least some coverage in every State. This expansion increased the reliability of the major statistics by about 20 percent and made possible the publication of greater detail.

**January 1967—Change in the concepts of employment and unemployment.** In line with the basic recommendations of the President's Committee to Appraise Employment and Unemployment Statistics [30], an experimental program was conducted for several years to develop and test proposed changes in the concepts. The principal improvements resulting from this research were put into effect in the household survey in January 1967. The changes included a revised age cutoff in defining the labor force; added probing questions to improve the information on hours of work, on the duration of unemployment, and on the self-employed; and a slightly revised definition of unemployment. The differences arising in the estimates of level and month-to-month change because of the revised unemployment definition are relatively small. Further information on the nature of these changes is given in [70].

**March 1968—Separate age/sex ratio estimate cells introduced for Negro and for other races.** There are three categories for race reported on the CPS questionnaire as white, Negro, and other. After the separate ratio estimates for Negro and for other, a second set of factors, by age and sex, are applied separately for the total of all persons with race reported as not-white and for those reported as white; this set of factors use a larger number of age cells than the first. The procedure used previously did not apply separate factors for Negro and for other. This change amounts essentially to an increase in the number of ratio estimate cells from 68 to 116. (See ch. V.)

**January 1971 and January 1972—Introduction of 1970 census occupational classification.** The questions on occupation were made more comparable to those used in the 1970 census by adding a question on major activities or duties on that job. The new classification was introduced into the CPS coding procedures in January 1971. The tabulations were produced in the revised version beginning in January 1972 [76].

**December 1971-March 1973—Expansion to 461-area sample and updating of sample and population data used in ratio estimates.** During the period December 1971 to March 1973, the CPS sample was revised gradually to reflect the changes in population size and distribution shown by the 1970 census. As part of an overall sample optimization, the sample size was reduced slightly (from 60,000 to 58,000 total units designated), but the number of sample areas was increased to 461 PSU's. Also, a change was made from clusters of six nearby (but not contiguous) households to four households that are usually contiguous. This change was instituted after Census Bureau studies indicated that a smaller cluster size would result in a more efficient sample. Thus, even with the reduction in sample size, there was a small gain in reliability for most characteristics due to this change. The noninterview adjustment and first-stage ratio

estimate adjustment were also changed slightly to improve the reliability of estimates for central cities and the balance of SMSA's.

The independent estimates of the civilian noninstitutional population by age, race, and sex, used for the second-stage ratio estimation procedure, were changed over to the 1970 census base in January 1972.

**January 1974—Use of inflation-deflation method for deriving independent estimates of population.** The derivation of independent estimates of the civilian noninstitutional population by race, sex, and age, used in the second stage of ratio estimation in preparing the monthly labor force estimates, was changed over to the inflation-deflation method. (See ch. V.)

**September 1975—Introduction of State supplementary samples.** An additional sample, comprising about 14,000 assigned interviews each month, was introduced beginning in July 1975 to supplement the national sample in 26 States and the District of Columbia. A total of 165 new PSU's was involved. The supplement was added to meet a specific reliability standard for estimates of the annual average number of unemployed persons for each State. Effective August 1976, by introducing an improved estimation procedure and modifying the reliability requirement, the supplemental PSU's were dropped from three States (ch. VI) and the size of the supplemental sample was reduced to about 11,000 assigned interviews in 155 PSU's (app. J).

## Impact of Research in the Labor Force Survey

Because the labor force survey has had major impact on Government policy from its inception, all aspects of the sample design and survey procedures have been of intense concern. Substantial effort is routinely directed at the analysis and control of CPS quality; for example, by the observation and reinterview of the interviewer's work and by the analysis of the reinterview results, by allocating a substantial portion of the budget to the initial and continuing training of interviewers and other staff, by analyzing the estimates of components as well as the total sampling variance; and by other means. (See ch. IV.)

Major research efforts are also directed at specific survey problem areas. Examples of some of the more recent research include—

1. A series of studies conducted in three PSU's, independent of the CPS, to investigate alternative interviewing techniques, the use of self-respondents, and the collection of labor force data for more than 1 month on a single visit.
2. Studies on the characteristics of households for which the CPS is unable to obtain an interview and the impact on the survey results by the method of adjusting for noninterviews in the estimation process.
3. Research in the optimum size of the cluster of addresses used in the final stage of sampling, taking into account the cost as well as the variance of alternate cluster sizes.
4. Measurement of the contribution of the interviewer to the variance of the survey.
5. Studies of the difference in the response of persons interviewed in the CPS when compared to their response in a decennial census.
6. The apparent difference in estimates when based on households that have been interviewed once before in the CPS, twice before, ..., etc.



DESIGN

More than half of the major changes listed in the chronology relate to improved methods of sample selection, estimation, or processing of the data. Only three of these involved an expansion in the number of households in the sample—the May 1956 expansion to the 330-area sample, the January 1967 expansion to the 449-area sample, and the 1975 addition for State data. However, many of the other changes increased the precision of the survey results and thus had the same effect as enlarging the sample at a much smaller cost. If the same sampling and estimation methods were used in 1975 as in 1943, when probability sampling methods were first introduced in the CPS, a sample from one and one-half to three times the current size would be necessary to produce estimates with the present level of reliability. The increase in efficiency varies somewhat from item to item. Among major labor force categories, the gain has been greatest for estimates of agricultural employment, for which the current reliability is equivalent to that of a sample more than two and one half times as large, using the methods employed in 1943. For nonagricultural employment and unemployment, the gains are equivalent to increases of more than 80 and 60 percent in sample size, respectively.

## THE SCOPE OF THIS REPORT

This report is a review of the CPS produced by the redesign after the 1970 census that resulted in the 461-area national sample and the 155-area State supplemental sample. The sample design and survey procedures are described as of January 1977; explanations are provided to show how and, where possible, why a specific design decision or procedure has been chosen. The major interest in preparing this report has been to document the CPS survey mechanism; less emphasis has been given to the subject matter. The survey design and methodology are discussed in some detail with supplementary references indicated where appropriate; major aspects of the program are discussed in further detail with illustrations by the use of appendixes. In general, the report does not consider future plans or changes in methodology that are or have been under consideration; such topics have been left for treatment in future publications. An earlier publication, Bureau of the Census Technical Paper 7 [44], describes the CPS as redesigned after the 1960 census.

There are nine chapters in this report. Chapter II describes the design and the selection of the 461-area national sample.

Chapter III describes the rotation system used to replace the units in the sample and explains how this replacement can be accomplished in an unbiased manner.

Chapter IV gives an account of the regular monthly survey operations in four sections:

1. A description of the various organizations within the Census Bureau that regularly apply their specialized knowledge to the design and implementation of the CPS.
2. The data collection activities involved in a survey month.
3. A discussion of the controls on the quality and progress of the data collection and processing system.
4. A summary of the separate data processing steps that are carried out each month.

The steps in the routine monthly estimation procedure for the national sample are discussed in chapter V, along with a summary of the special estimation procedures to tabulate additional inquiries in selected months.

Chapter VI, describes the sample design for a supplement to the national sample used in preparing estimates for certain States. This chapter also describes how estimates are prepared for each of the States, including those States covered by the national sample only.

Chapter VII discusses the accuracy of the CPS estimates dealing separately with biases and the components of variance. The mean square error of the CPS estimates is also discussed.

The methods of calculation and the presentation of sampling errors appear in chapter VIII.

Finally, chapter IX describes how the CPS sample, interviewing staff, and processing organization are being used as a facility for conducting other surveys which may be more or less independent of the content of the CPS. All studies, which propose to use the CPS facilities, must meet the guidelines discussed in this chapter.

The report includes 13 appendixes, each of which deals with a particular aspect of the CPS in detail. There are three large maps (approx. 30 x 40 in.) and separate publications which deal with the identification and the location of the sample PSU's; these materials are discussed in appendix M and may be purchased for a cost, using the order form on the last page of this publication.

The index begins on page 168.

## Chapter II.

# DESIGN OF THE CPS NATIONAL SAMPLE

## INTRODUCTION

Because the CPS national sample for over three decades has been one of the major sources for current measurement of demographic statistics, the reliability of the sample design has been of continuing concern. The design has been under constant review to find ways of responding to demands for new data and to take advantage of new survey design developments and new information (especially census results) that would increase the reliability of the sample estimates. Changes in the design flowing from such review are introduced with concern for minimizing the cost and the impact on accuracy. In the discussion which follows, the design of the CPS national sample is described as it existed after the major design revisions undertaken to incorporate new information from the 1970 census. These revisions were largely completed as of March 1973.

A history of prior versions of the CPS is discussed in chapter I. The design of the supplemental samples selected to improve the reliability of estimates made from the national sample for the smaller States is discussed in chapter VI.

## SURVEY SPECIFICATIONS AND DESIGN

### Survey Specifications

The specifications characterizing the present version of the CPS national sample are summarized as follows:

1. The CPS is a probability sample. As a consequence of this specification, it is possible to estimate most of the components of the survey error from data produced by the sample.
2. The sample is designed primarily to produce estimates of the major components of the labor force, in general, estimated levels with minimum variance for a fixed cost.
3. The major statistics of interest to be produced by the CPS are labor force characteristics of the U.S. population. The problem of maximizing the reliability for a wide range of other demographic statistics and for other tabulation areas is recognized as an important although secondary requirement. The monthly survey covers the civilian noninstitutional population.
4. The sample is selected with probability controls to insure that sample areas are designated in each of the 50 States and in the District of Columbia. This requirement insures an additional geographic spread of the sample and facilitates understanding the validity of the sample by a wider range of users.

### Survey Design

The following summarizes the survey design which has evolved to meet these specifications.

The CPS sample design, which is a multistage stratified sample of the U.S. population, is made up of two independent

national samples called the A-design and the C-design. The full sample is used for the regular monthly series of estimates of labor force and related statistics while the A-and C-subsamples can be used for other national investigations which do not require the full sample. Furthermore, the existence of two independent samples makes it possible to provide better estimates of some of the survey errors.

Each of the A-and C-designs is selected in stages. For the A-design, the first stage involves the definition of the United States in terms of counties or groups of counties called primary sampling units (PSU's) which are assembled into homogeneous groups called strata; one PSU is selected for the sample from each stratum. A sample of addresses within the sample PSU's is obtained in a second stage. Most of the sample addresses are selected from census lists in a single stage of sampling within the selected PSU; for a relatively small proportion, an additional stage of selection within the PSU is necessary. For the C-design, a sample of A-design strata is first selected and then an independent sample of PSU's and sample addresses is selected within these strata, using the same principles as in the A-design.

The multistage plan, despite its apparent complexity, is roughly equivalent to the simple plan of dividing the entire U.S. into ultimate sampling units (USU's), each containing about four neighboring housing units and selecting clustered samples of these USU's for interview.

The variables chosen for grouping PSU's into strata reflect the primary interest of the CPS in maximizing the reliability of estimates of labor force characteristics. However, many of these variables are also correlated with other demographic characteristics which are the subject of other surveys conducted as part of the CPS program. A byproduct of the system is the designation of other samples for demographic data collected by survey processes that are otherwise independent of the CPS.

The CPS sample is interviewed monthly. To improve the reliability of estimates of month-to-month change, three-fourths of the USU's in sample, in a given month, are retained in sample the next month; to improve the reliability of estimates of year-to-year change, one-half of the USU's in sample, in a given month, are in sample a year hence. (See ch. III.)

The following statements pertain to the A- and C-design samples:

1. The number of sample USU's in the A-design is twice the number in the C-design.
2. Each of these samples is self-weighting. With a few exceptions, all sample USU's in the A design are designated with the same overall probability and an estimate can therefore be made by applying a constant weight to each of the sample USU's. Because sampling the USU's is controlled to insure selection without bias, we call these unbiased estimates even though such estimates are known to have some bias. (See ch. VII.) A similar statement applies to the C-design.
3. Unbiased estimates from the full CPS are weighted averages of the separate A-and C-design unbiased estimates. If  $x_a$  and  $x_c$  represent the unbiased estimates from

the A- and C-designs respectively, the estimate  $x$  for the CPS is given by

$$x = (2/3)x_a + (1/3)x_c \quad (11/1)$$

The fractions 2/3 and 1/3 are weights (approximately the optimum weights) used to average the two estimates. The product of each of these fractions and the selection probabilities for the A- and C-design samples mean the same constant inflation factor can be applied to all USU's in the combined sample.

4. The A- and C-designs are very nearly independent in the statistical sense; that is, the sampling procedures used maintain the validity of the following expression involving expected values of the estimates

$$E(x_a x_c) = (E x_a) (E x_c).$$

This statement also implies that the variance of the estimate in (11/1) is approximated by

$$\text{Var}(x) = (2/3)^2 \text{Var}(x_a) + (1/3)^2 \text{Var}(x_c) \quad (11/2)$$

## SELECTION OF THE SAMPLE OF PRIMARY SAMPLING UNITS

The selection of PSU's is carried out in three major steps:

1. Definition of the PSU's.
2. Stratification of PSU's.
3. Selection of the sample PSU's.

Each of these steps will be described in the following paragraphs. The CPS has undergone major revisions to incorporate more up-to-date information available from each of the last three decennial censuses of population and housing. At each revision, each of the changes proposed for the redesign has been reviewed, not only for its impact on the reliability of estimates from the new sample, but also for the cost and nonsampling error aspects of the change.

The desire to minimize the change in sample areas in each updating or redesign is reflected in the definition, stratification, and selection of sample PSU's. In each instance, the procedures used are designed to be unbiased and to produce the same expected value as if all sampling operations were performed independent of other existing CPS samples.

### Definition of the Primary Sampling Units

Usually a group of counties rather than only a single county is used as a PSU. For a sample design characterized by a fixed overall sampling fraction (hence a fixed total number of sample USU's), the effect of increasing the size of PSU's by combining counties is to spread the sample of USU's over a larger number

of counties (to reduce the amount of clustering) and, thus, to reduce the sampling errors of most estimates.<sup>1</sup>

However, increasing the size of PSU's also increases the cost of interviewer travel between USU's within the PSU. The optimum solution, therefore, lies somewhere between a random or systematic sample of USU's from all counties (where the PSU's are made so large that all must be in sample) and a sample of the same number of USU's selected from PSU's made up of single counties (or from even smaller geographic areas).

The rules adopted for defining primary sampling units are listed in some detail. After some empirical research in the late 1940's to help establish rules, the PSU's were first determined in late 1949 and early 1950 and have been modified subsequently, as necessary, to conform to these rules.

Rules for defining PSU's—The effect of the initial definitions and such modifications, as have occurred, is to define *primary sampling units* (PSU's) under the following rules:

1. Each standard metropolitan statistical area (SMSA) is defined as a separate PSU. However, the two standard consolidated areas (SCA's)<sup>2</sup> are each counted as one PSU, although both SCA's consist of more than one SMSA.
2. Each PSU comprises one county or two or more contiguous counties. This rule does not apply for certain states in New England<sup>3</sup> where minor civil divisions (towns or townships) are used to define the PSU's. In some states, county equivalents are used: Cities, independent of any county organizations, appear in Maryland, Missouri, Nevada, and Virginia; in Louisiana, the county equivalents are known as parishes; and, in Alaska, where there are no counties, census divisions are used.
3. The PSU's are defined within the boundaries of the four census regions; that is, Northeast, North Central, South, and West. The States in the four census regions are shown in figure 11-1. Exceptions to this rule have been made for SMSA's which cross regional boundaries. Those SMSA's which lie in more than one State are counted as single PSU's for sampling purposes, but the parts within the individual States are separately identified for convenience in tabulation and control. There are 30 of these SMSA's, 7 of which lie in more than one census region.

<sup>1</sup> The variance of an estimate, based on a two-stage sample, has the general form

$$\sigma_x^2 = \frac{1}{m}(\sigma_b^2 + \frac{\sigma_w^2}{\bar{n}}) \quad (11/3)$$

where

$\sigma_b^2$  = variance between first-stage units, e.g., PSU's.

$\sigma_w^2$  = variance between second-stage units, e.g., USU's, within first-stage units.

$m$  = number of PSU's in the sample.

$\bar{n}$  = average number of USU's in the sample per PSU.

The sum of  $\sigma_b^2$  and  $\sigma_w^2$  is not affected by changing the size of the PSU's. An increase in the size of the PSU's will normally result in a decrease in  $\sigma_b^2$ , in which case  $\sigma_w^2$  must be increased by the same amount. However, the net result will be a decrease in  $\sigma_x^2$ , since the coefficient of  $\sigma_b^2$  is greater than that of  $\sigma_w^2$  ( $1/m$  versus  $1/m\bar{n}$ ). (For further discussion, see [21, ch. 9, sec. 7].)

<sup>2</sup> The New York-Northeastern New Jersey SCA and the Chicago-Northwestern Indiana SCA.

<sup>3</sup> Towns and cities are used in defining PSU's in Connecticut, Massachusetts, Rhode Island, and parts of New Hampshire and Maine.

4. The area of a PSU does not exceed 2,000 square miles in the West and 1,500 square miles in the other regions except in cases where a single county exceeds the maximum area.
5. The 1970 census population of each PSU is at least 7,500 in the west and 10,000 in the other regions except where this would have required exceeding the maximum area specified in rule 4.
6. In addition to meeting the limitations on total area, PSU's are formed to avoid extreme length in any direction.

PSU's defined for earlier CPS designs—The PSU's currently in use are closely related to those defined shortly after the 1950 census. The principal items of information used to define PSU's at that time included the following:

**Economic area [66]**

Principal industry—used primarily in urban areas

Value of agricultural products—used primarily in rural areas

The Hagood Index—an index designed to measure the level of living in rural areas [18;19]

Proportion of the population having race reported as not white—used in areas where there was appreciable variation between counties in the proportion, primarily in the South

Rate of population increase 1940-1950

Counties are grouped in such a way as to make the PSU's heterogeneous; that is, to make the counties within each PSU differ substantially with respect to as many as possible of the principal items mentioned. In most multicounty PSU's, for example, each county comes from a different economic area.

The operation of combining counties to form PSU's was originally carried out State by State. Subsequently, adjustments were made when it seemed desirable to combine counties in different States within the same census region.

The composition of the PSU's has been reviewed as a part of the several later revisions in the CPS design. However, changes in the composition of the PSU's have been made almost entirely to reflect changes in definitions of SMSA's<sup>4</sup> rather than in the characteristics of the counties comprising the original PSU's.

Combining Counties into PSU's for the current design—The 3,141 counties, county equivalents and independent cities of the United States, existing during the 1970 census, have been combined into 1,924 PSU's. The steps employed in combining counties to conform to the above rules are roughly as follows:

1. The counties involved in SMSA's that are new or redefined since the previous revision are incorporated into the PSU definitions. Parts of PSU's remaining after such an operation are reviewed according to the criteria to see whether further combination of counties into PSU's is necessary.
2. Any single county exceeding the maximum area limitations is automatically classified as a separate PSU, regardless of population.
3. The characteristics of all other counties within the same State are then examined to determine whether they might advantageously be combined with contiguous counties, keeping in mind the objective of heterogeneity and the

<sup>4</sup> Formerly standard metropolitan areas (SMA's).

population and area limitations. Combinations are made for—

Counties with less than the minimum population that are combined with other counties, providing this does not conflict with other requirements.

Counties with more than the minimum population that are combined with other counties only if the maximum area is not exceeded and if the combination substantially increases the heterogeneity within the PSU. To illustrate this, a proposed combination involving two counties in different economic areas is generally considered desirable even if both counties exceed the minimum population. The maximum area limitation is adhered to as strictly as possible, since excessively large PSU's add to operating costs through increased interviewer travel expense.

4. The characteristics of counties in contiguous States within regions are then compared to determine whether better combinations can be obtained by combining counties from neighboring States.
5. The proposed combinations are reviewed to make the final determination of PSU's. The reviews of this operation (and of the operation grouping PSU's into strata) represent an effort to apply the competent judgments of several persons to these processes. Personal judgment can have no place in the actual selection of sample units as known probabilities of selection for all units in the population can only be achieved by the application of random processes of selection. However, there are a large number of ways in which a given population can be structured and arranged prior to the application of the random selection processes, and personal judgment can legitimately play an important role in devising an optimum arrangement, i.e., one that will minimize the variances of the sample estimates.

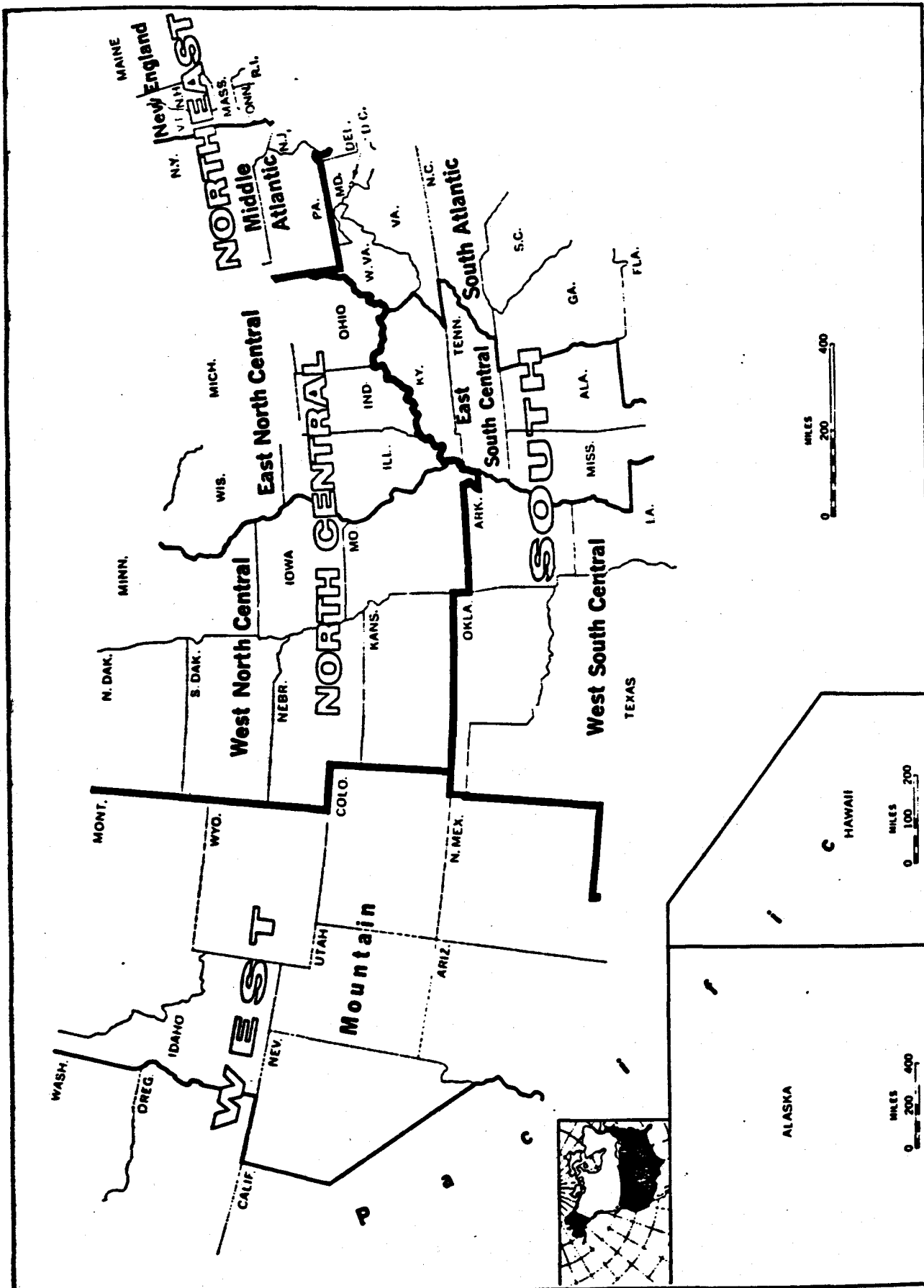
### Stratification of Primary Sampling Units

The sample design adopted for the CPS calls for the combination of PSU's into strata with the selection of one PSU from each stratum. For samples of this design, sampling theory indicates the desirability of forming strata with approximately equal measures of size (total population). With a self-weighting sample, equal stratum sizes also have the advantage of providing equal interviewer workloads. The objective of the stratification, therefore, is to group PSU's with similar characteristics into homogeneous strata with approximately the same 1970 populations in each stratum. The 1970 census preliminary population figures for the PSU's are the measure of size used in the stratification and selection procedures.

The theory also requires that PSU's with the largest populations must be in the sample with certainty. PSU's with the largest populations, therefore, are designated as self-representing (SR), i.e., each of the SR PSU's is treated as a separate stratum and is included in the sample with certainty.

Number of strata—The number of strata has been determined by variance and cost considerations.

If the redesign of the sample is considered as a problem independent of the CPS sample in use at the time of the redesign, one could approach the question of determining the number of strata as in the following illustration: First, for purposes of



discussion, assume the budget is sufficient to fund a total sample of about 40,000 housing units designated each month for the A-design portion of the CPS. Second, assume that the interviewer assignments are about 65 designated sample housing units. (Earlier CPS experience has shown this is an efficient assignment size.) Ideally, a total sample of 40,000 designated units should, therefore, consist of roughly 615 interviewer assignments of this size. This implies each interviewer should work in a stratum of roughly 330,000 population (about 1 in 615 of the total U.S. population). However, there are many SMSA's, and some other PSU's, having populations in excess of this figure that should, therefore, be designated as self-representing PSU's. With this criterion of stratum size (and using other necessary rules for stratification discussed in the following sections), just over 60 percent of the population of the United States would reside in PSU's determined to be self-representing. These SR PSU's (i.e., strata) would account for about 400 of the total 615 assignments, with the balance of the assignments allocated to NSR strata. This analysis suggests that once the SR PSU's have been defined, the balance of the PSU's would be classified into just over 200 NSR strata.

Although the foregoing is rough and oversimplified,<sup>5</sup> it suggests the approximate number of strata and sample PSU's appropriate for a new design. However, the process of defining the sample cannot be made without considering the existing sample.

At each redesign of the CPS, it is important to minimize the economic impact of the selection of a new set of PSU's. A substantial investment has been made in the hiring and training of interviewers in the existing sample PSU's. For each PSU dropped from sample, and replaced by another in the new sample, the expense of hiring and training a new interviewer must be accepted. Furthermore, there is a temporary loss in accuracy of the results produced by new and relatively untrained interviewers. Concern for these factors is reflected in the rules for reviewing and modifying existing strata for use in the redesign.

**Need to redefine strata**—The review of the stratification of PSU's has, in each instance, been done primarily by searching out misfit PSU's (no longer proper members) of the existing stratum rather than undertaking an entirely new stratification. An independent stratification might prove to be more efficient in terms of information per unit in the sample but would likely require introducing a larger number of new sample PSU's and retiring a larger number of current sample PSU's than the system used.

Changes in the size of PSU's have accounted for a fairly large number of the changes in the CPS samples. When one or more counties are added to an SMSA in a different stratum or when a PSU should be made SR because of growth, the size of the remainder of the stratum is often reduced below the minimum size considered desirable. Also, some strata have either increased or decreased in population considerably since the previous decennial census. Substantial changes in the characteristics of the population occur in a few of the PSU's, and reassignment of these PSU's to strata becomes justified to increase the homogeneity within some of the NSR strata.

The most important of the characteristics which provided the basis of the original PSU stratification that have been considered for the restratification are as follows:

- SMSA or not
- Rate of population change
- Percent of population living in urban area
- Percent of population employed in manufacturing
- Principal industries
- Average per capita value of retail trade
- Proportion of population with race reported as not white

Limits are established for each of the characteristics; when a stratum contains a PSU with characteristics outside these limits, the PSU is considered for removal from the stratum. However, extreme change in a single characteristic does not always justify removing a PSU from a stratum. In some cases other characteristics of the PSU indicate that stratification would not be improved by making a change. In other cases, no other stratum is found containing PSU's with characteristics similar to the PSU being considered for removal. Ordinarily, a change in several population characteristics, beyond the established limits, is required before a PSU is removed from a stratum. The total number of revisions, based on extreme characteristic changes, is usually quite small.

**Rules for redefining strata**—In general, the stratification derived for the earlier CPS design has been accepted; exceptions for PSU's and strata are given by the following rules:

1. The largest of the 1,924 PSU's in the United States, as of the 1970 census, are designated as self-representing PSU's. Except in special cases, all PSU's with population greater than about 250,000 population in 1970 are made SR. The remaining PSU's are combined with other PSU's (with similar characteristics) to form a number of NSR strata. Strata conform to regional lines, Northeast, North Central, South, and West. A summary of the primary strata is given in table II-1.
2. Strata with NSR PSU's are made to have approximately equal total 1970 populations. Average sizes for these strata vary from 321,000 in the South to 357,000 in the Northeast. (The coefficients of variation of census NSR-stratum populations are given in table II-2).
3. NSR SMSA PSU's are grouped into strata which include only SMSA PSU's.
4. A basic rule applies to rearranging PSU's into NSR strata. The restratification is done strictly on the basis of the characteristics of the PSU's within a stratum, without regard or any reference to the sample PSU's of an existing CPS design.

### Selection of Sample PSU's for the A-Design

Each self-representing PSU is in sample by definition. As shown in table II-1, there are 156 SR PSU's in the A-design; these same PSU's are also SR in the C-design. In each of the remaining 220 NSR strata in the A-design, one PSU is selected for the sample.

**Objectives of the selection procedure**—The selection of the sample of nonself-representing PSU's is carried out within the revised strata using 1970 census population totals as measures of size. A selection system has been devised to accomplish the following objectives:

1. Select one sample PSU from each stratum with probability proportionate to 1970 census population.

<sup>5</sup> Interviewer workloads and average stratum sizes cannot, in practice, always conform to the values cited.

TABLE II-1. Number and Population of Strata for 376-PSU A-Design Samples, by Census Region

Census region	Total		Stratum size	SMSA		Non-SMSA	
	Strata	Population <sup>1</sup>		Strata	Population <sup>1</sup>	Strata	Population <sup>1</sup>
Representing (SR) <sup>2</sup>							
Total .....	156	130,495,743	Minimum size (X)	146	126,450,045	10	4,045,698
Northeast .....	50	41,854,236	<sup>3</sup> 259,754	42	38,265,766	8	<sup>4</sup> 3,588,470
North Central .....	32	32,824,092	256,843	32	32,824,092	—	—
South .....	48	30,049,022	227,222	46	29,591,794	2	457,228
West .....	26	25,768,393	213,358	26	25,768,393	—	—
Nonself-Representing (NSR) <sup>2</sup>							
Total .....	220	72,681,943	Average size (X)	35	12,932,619	185	59,749,324
Northeast .....	20	7,143,875	357,194	2	879,979	18	6,263,896
North Central .....	72	23,752,566	329,897	13	4,838,426	59	18,914,140
South .....	102	32,745,594	321,035	16	5,607,432	86	27,138,162
West .....	26	9,039,908	347,689	4	1,606,782	22	7,433,126

Not applicable.

Entry represents zero.

Preliminary 1970 population; figures differ slightly from published 1970 census counts.

In the C design, the same self-representing strata are used, and the nonself-representing areas within each region are grouped into half as many strata as in the A design.

Under a special rule, 28 areas in 4 New England States are defined and counted as self-representing PSU's; several of these areas have a smaller population.

<sup>1</sup>Includes 782,185 non-SMSA population in the New York Standard Consolidated Area.

TABLE II-2. Coefficient of Variation of Population of Nonself-Representing Strata for 357- and 376-Area A-Design Samples, by Census Region

Census region	357-area design			376-area design		
	Number of NSR strata	Coefficient of variation		Number of NSR strata	Coefficient of variation	
		1960 population	1970 population <sup>1</sup>		1960 population	1970 population
Northeast .....	30	0.2323	0.3853	20	0.1607	0.1120
North Central .....	76	.2263	.2948	72	.1214	.1298
South .....	110	.2333	.3010	102	.1441	.1106
West .....	29	.2069	.2472	26	.1847	.1592
Total .....	245	(X)	(X)	220	(X)	(X)

X Not applicable.

<sup>1</sup> Preliminary 1970 population

2. Retain in the new sample the maximum number of sample PSU's from the old design sample.
3. For each State, have the number of PSU's in sample reflect the population of the State.

**Keyfitz method**—The use of a method proposed by Keyfitz [27] accomplishes the first and second objectives in those NSR strata comprised of the same PSU's in the old (357-PSU) design and in the new (376-PSU) design. Perkins [41:42] extended the procedure for use in the 1960 revision as well as the current one to cover the conditions arising when the strata are comprised of different PSU's on the two occasions. In the extended procedure, probabilities of selection are calculated for each PSU in the revised strata.

Consider a revised stratum in the new design made up of parts from more than one stratum in the old design. One of these parts can be selected with probability proportionate to 1970 size. If the selected part does not include a sample PSU in the old A-design, a selection can be made from PSU's in that part with probability proportionate to 1970 size. On the other hand, if the selected part contains a sample PSU from the old design, the selection from this part is made by the Keyfitz method, which maximizes the chance of retaining the previously selected PSU. In actual practice, the sample PSU is not selected from the part at this point, but the conditional probability of its selection (given the part has been selected) is computed and used in a controlled selection procedure to reflect the concerns of the third objective. The conditional probabilities can range between 0.0000 and 1.0000.

**Controlled selection**—The third objective is accomplished by controlling the number of sample PSU's selected from each State to approximately the number expected to be selected from the State. The expected number is determined by summing the selection probabilities (determined by the Keyfitz method) of all PSU's in the State.<sup>6</sup> For some of the larger States, selection areas within the States are defined and treated as though they are separate States in the selection process. By controlling the selection of the expected number of sample PSU's from each State to areas within the State, a further geographic spread of sample PSU's is assured. As the CPS strata do not cross the boundaries of the census regions, the objectives can be realized by controlling the selections separately, by region.

This procedure for selecting PSU's is quite complicated and has limited benefit in terms of reducing the sampling variability for national estimates. The procedure was incorporated largely to improve the probability of having sample PSU's appear in each of the States—a desirable attribute when obtaining congressional support. It is necessary to contend with the complications only once, however, when the sample PSU's for the design are selected; moreover, the desired outcome did occur—A-design sample PSU's appeared in each of the States. A further benefit became apparent when additional emphasis was later given to estimates for the States. (See ch. VI.) Controlled selection of the PSU's gives improved estimates for the States when based on the national sample.

The actual process of selecting the specific sample PSU's operates in two stages. In the first stage, a computer operation

designates selection patterns of sample PSU's (one from each stratum) that enable the number of sample PSU's in each State selection area to be one of the two integers closest to the expected number. The computer further restricts the selection patterns to a set of admissible patterns so that each admissible pattern has approximately the expected number of sample PSU's also in the old (357-area) CPS design sample; in a few cases, this latter restriction could not be completely satisfied. Probability is assigned to each of the admissible patterns which sums over the set to 1.0000.

At the second stage, using these probabilities, one of the admissible patterns is randomly selected to represent the new 376-area sample of PSU's for the A-design.<sup>7</sup>

Each PSU in the new sample design actually has a probability of selection in proportion to its size, a policy which has also been applied to previous CPS designs. Probability proportionate to size is not biased by the technique used to calculate probabilities of selection for PSU's in the 376-area strata, as can be shown by a mathematical proof. Controlled selection does not affect the probabilities of selection either. It can be said, therefore, that the 376-sample PSU's for the A-design are selected from their strata with probability proportionate to their 1970 census population of the PSU's.

The procedure for selection of PSU's in the new (376) A-design used 1970 census based probabilities; the selection of sample PSU's for the old (357-area) CPS design used 1960 census based probabilities with the Keyfitz probabilities calculated and the controlled selection performed in essentially the same way [44].

Number of PSU's in the 1970 and 1960-based A-designs—Comparison of the new 376 area sample PSU's<sup>8</sup> with sample PSU's in the old 357-area design shows that the PSU's can be classified in five groups; these five groups add to 376 PSU's. There are 51 PSU's in the old 357-area sample completely missing from the 376-area sample, and there are—

1. 250 PSU's in the old sample retained in their entirety in the new sample
2. 44 PSU's (mostly SMSA's) in the old sample and in the new sample but with additional area added
3. 3 PSU's in the old sample and retained in the new sample but with some area added to them and some area dropped from them
4. 9 PSU's in the old sample and retained in the new sample but with some area dropped from them
5. 70 PSU's which are entirely new to the 376-area sample

Strata in 1970-based design, by status in 1960-based design—Table 11-3 summarizes, in part, the classification of the A-design strata and population for the new (376-area) sample according to the status of the stratum in the old (357-area) sample.

### Selection of Sample PSU's for the C Design

**Relation of A- and C-design PSU selection methods**—The sample PSU's for the C-design consist of the same self-representing PSU's designated for the A-design plus an independent sample of PSU's to represent the balance (the NSR

<sup>6</sup> The sum of these conditional selection probabilities for a State may be substantially different from the sum of the directly calculated probabilities proportionate to size.

<sup>7</sup> The controlled selection procedure is illustrated in detail in app. A; some experience in selecting the sample of NSR PSU's in the West Region is also summarized.

<sup>8</sup> The location of the sample PSU's is discussed in app. M.



area) of the United States. The same definitions of PSU's and strata developed for the A-design are used for the C-design. The number of NSR sample PSU's in the C-design is one-half the number chosen for the A-design. The selection of the C-design NSR PSU's is accomplished by selecting an equal probability sample of one-half the A-design NSR strata and selecting a single PSU with probability proportionate to the 1970 population from the selected NSR stratum. The same series of subsequent steps as for the A-design sample defines the sample of units for interview within the selected sample PSU's.

The sample selection procedures at each step are developed to produce A- and C-design samples independent in the statistical sense. The following observations apply:

1. The probabilities used to select the C-design NSR PSU's maximize the number of old design C-sample PSU's retained in the new C-sample and are computed by the same methods used for the similar purpose in the A-design sample.
2. A selection procedure is employed to achieve a controlled geographic spread of sample PSU's to the States and to have approximately the expected number of old design C-sample PSU's retained.
3. When PSU's have been selected for both samples, a given NSR PSU can be in sample: in the A-design, the C-design, or in both A- and C-designs.
4. A self-weighting subsample of USU's is designated with the overall selection probability for the C-design equal to one-half the A-design probability.

**Determine the C-design NSR sample PSU's**—The first step in determining NSR C-sample PSU's requires the grouping of the 220-NSR A-design strata into pairs within each census region. The strata are combined with the objective of producing homogeneous pairs; however, as in reviewing the stratification of PSU's into A-design strata, concern for maximizing the ability to retain old C-sample PSU's in the new sample is observed. In constructing pairs in which A-strata are comprised of the same PSU's as in the old design, the same pairing of A-strata for the new C-design is desirable, since this increases the likelihood of selection of the same sample of PSU's as in the old design. Where the A-design strata have different PSU's in the old and new designs, pairing of the new A-strata attempts to maximize the 1970 population for PSU's remaining in the same pair in the old and new C-designs.

Efforts to retain the old PSU's in the C-design are, in general, subordinated to the need to achieve homogeneity among the A-design strata in each pair, particularly with respect to total 1970 census population. Other characteristics for A-design strata considered in defining pairs include—

SMSA or not  
Percent of population urban  
Percent of population with race reported as not white  
Number of farms  
Per capita retail sales

The second step in designating the C-sample PSU's chooses one A-design stratum out of each pair with equal probability. This is done independently from pair to pair.

The third step chooses, with probability proportionate to

1970 census population, one sample PSU out of each of the design strata which were selected in the second step for the design. This selection is independent of the choice of sample PSU's for the A-design. As with the PSU selection process in the A-design, the selection is accomplished by calculating assigning probabilities computed by the modified Keyf procedure, previously described. Controlled selection is used to improve the geographic spread of the sample PSU's and govern the number of selected PSU's also in the old C-sample.

Table II-3 shows, in part, the classification of the C-design strata for the new (266-area) sample, according to the status the stratum in the old (235-area) sample.

### Number of Sample PSU's in the A- and C-Design Samples

The number of sample PSU's selected for the A-design and for the C-design samples can be summarized as follows:

Sample PSU's	Number of sample PSU's	
	A-design	C-design
<i>Self-representing PSU's, in sample in both designs, total</i>	156	156
<i>Nonself-representing PSU's, total</i>	220	110
<i>In A-design or in C-design samples but not in both</i>	195	85
<i>In both A- and C-design samples</i>	25	25
<i>Number of PSU's in sample, total</i>	376	266

There are 461 different sample PSU's in the combined A- and C-design sample, i.e.,  $156 + 195 + 85 + 25$ .

### SELECTION OF THE SAMPLE WITHIN SAMPLE PSU's

The object of subsampling within each of the sample PSU's to obtain a self-weighting probability sample of housing units and special places. In a self-weighting sample, every sample unit has the same overall probability of selection so that estimates of population totals may be made by applying a single expansion factor to totals from the sample.

Optimum sampling considerations indicated that for most of the characteristics considered, the variance would be close to the minimum for a fixed cost if the sample was selected in compact clusters comprising on the average about four adjacent housing units (or the equivalent) [58]; these clusters are called ultimate sampling units (USU's).

Since part of the sample is replaced each month, it is necessary to provide for the selection of additional samples of USU's from time to time. In this section, however, the discussion is restricted to the methods of selecting the first:

\* The location of the sample PSU's is discussed in app. M.

TABLE II-3. Classification of New Strata in 376-Area A- and 266-Area C-Design Samples, by Status in Old A- and C-Design Samples

Composition of strata <sup>1</sup>	A-design		C-design	
	Number of strata	1970 population (millions)	Number of strata	1970 population (millions)
(1)	(2)	(3)	(4)	(5)
Self-representing (SR) strata in new design				
Total .....	156	130.7	156	128
Entirely from SR strata in old design .....	78	86.6	78	86.6
Entirely from NSR strata in old design .....	37	9.0	37	9.0
In sample in old design .....	(X)	7.5	(X)	7.5
Not in sample in old design .....	(X)	1.5	(X)	1.5
Part SR, part NSR strata in old design .....	41	35.1	41	35.1
Population from SR strata in old design .....	(X)	31.0	(X)	31.0
Population from NSR strata in old design .....	(X)	4.1	(X)	4.1
Nonself-representing (NSR) strata in new design				
Total .....	220	72.7	110	72.7
New and old design strata are identical .....	39	12.9	4	12.9
Sample PSU's identical in new and old designs .....	37	12.2	3	12.2
Sample PSU's different in new and old designs .....	2	.7	1	.7
New design strata lying wholly within one old design stratum .....	53	17.8	22	17.8
Sample PSU's identical in new and old design strata .....	40	13.6	15	13.6
Sample PSU's different in new and old design strata .....	13	4.2	7	4.2
New design strata comprising parts of two or more old design strata .....	128	42.0	84	42.0
Sample PSU's identical in new and old design strata .....	96	31.5	46	31.5
Sample PSU's different in new and old design strata .....	32	10.5	38	10.5

X Not applicable.

<sup>1</sup> There were 357 strata in the old (1960-based) A-design sample and 235 strata in the old C-design sample.

sample of USU's designated for the 461-area design. The method of using this initial sample to generate additional samples needed subsequently is discussed in chapter III and appendix E.

### Calculate the Within-PSU Selection Probability

Plans for the CPS design after the 1970 census initially indicated the budget would support a CPS sample of about 60,000 designated housing units per month; two-thirds, or about 40,000 would come from the A design. A sample of 40,000 out of the 70.6 million living quarters presumed to exist as of the 1970 census represents a sampling fraction of roughly 1 in 1,800. (These figures were changed during the planning process, and the fraction actually used was 1 in 1,968). The 70.6-million figure is the sum of the 1970 census count of 68.6 million total housing units plus an allowance of one unit for every three members of the census population not residing in housing units.

As the sample is self-weighting, the overall probability of selection in the sample for the A-design.

The sampling fractions at each stage of selection are determined in order to meet this requirement. Thus, if  $P_{hi}$  represents probability of selection of the  $i$ -th PSU in the  $h$ -th stratum,  $1/W_{hi}$ , the probability of selecting sample USU's within  $i$  then  $W_{hi}$  must satisfy the relationship

$$P_{hi} \times 1/W_{hi} = \frac{1}{1,968}$$

A similar requirement applies in selecting the sample A-design which has an overall selection probability of 1/12 and, therefore, a sample of one-half the number of USU's.

Table II-4 illustrates the computation of the probabilities used at each stage of selection for the two samples and one C-design sample chosen from a pair of strata in the north-central region. The sampling probabilities shown on line II-4 are used to select the sample of USU's within the PSU's.

## Resources for Sampling Within the PSU

A number of resources are available from which samples are drawn to represent separate portions of the current population and housing in the PSU's. However, these portions are not mutually exclusive; new construction, for example, may be represented in more than one of the resources. This section describes how each resource is sampled. The integration of the resulting samples to represent the current population and housing is discussed on p. 18.

The discussion of the resources requires amplification of the terms describing living quarters, USU, and segment:

A *housing unit* is a group of rooms, or a single room, occupied or intended for occupancy as separate living quarters. Separate living quarters are those in which the occupants do not live and eat with any other person in the structure and that have either direct access from the outside of the building or through a common hall, or complete kitchen facilities for that unit only.

All persons not living in housing units are classified as living in *group quarters (GO)*. The two categories of persons in group quarters are persons under care or custody in institutions and

classified as *inmates of institutions* (such persons are outside the scope of the Current Population Survey) and all other persons living in group quarters who are not inmates of institutions (their living quarters are called *other units*). Living quarters are called group quarters if there are five or more persons unrelated to the head or when no head is designated, if six or more unrelated persons share the unit. Rooming and boarding houses, communes, workers' dormitories, and convents or monasteries fall into this category. Persons living in certain other types of living arrangements are classified as living in group quarters regardless of the number or relationship of people in the unit. Included are crews of vessels, persons residing in college dormitories (or in sorority and fraternity houses), staff members in institutional quarters (not in housing units), and persons in missions, flophouses, etc.

A *household*, for the purpose of the CPS, includes the occupants of a housing unit or an other unit whose usual place of residence is the unit.

The result of all subsampling processes within a PSU is a set of *ultimate sampling units (USU's)*. Each USU comprises a cluster of housing units selected with a known probability.

TABLE II-4. Sampling of PSU's Within Stratum and USU's Within PSU for PSU's 374, 412, and 433

Line No.	Action	A-design samples		C-design sample
		Stratum 374	Stratum 412	
1	Overall selection probabilities .....	1	1	1
		1,968	1,968	2(1,968)
2	Stratum			
3	Identification code .....	374	412	374 and <sup>1</sup> 412 pair
	Stratum population 1970 census = $N_h$ .....	342,507	361,903	2(361,903)
	Sample PSU			
4	Ident code .....	374	412	433
5	Sample area (all are counties in State of Indiana) .....	Greene-Monroe	LaPorte-Starke	Franklin-Union-Wayne
6	Population of PSU, 1970 census = $N_{hi}$ .....	111,743	124,622	102,634
7	Probability of selecting sample PSU			
	$P_{hi} = N_{hi}/N_h$ .....	0.32625	0.34435	(0.5) (.28360)
8	Probability of selecting sample USU's in sample PSU = $1/W_{hi}$ where			
	$(1) = P_{hi} \times 1/W_{hi}$ .....	1	1	1
		642.06	677.68	588.12
9	Expected sample of 1970 housing units in PSU			
10	Total 1970 HU's in sample PSU .....	37,420	41,724	34,102
	Expected designated sample = (8) X (9) .....	58	62	61
11	Weights used for inflation estimates			
12	For separate A-and C-estimates .....	1,968	1,968	(2) 1,968
	For CPS, weighted average of separate A-and C-estimates .....	(2/3)(1,968) = 1,312	(2/3)(1,968) = 1,312	(1/3)(2)(1,968) = 1,312

<sup>1</sup> The stratum selected from the pair, for the C-design sample is stratum 412.

Sample estimates are based on interviews to be conducted for every housing unit (or nonhousing unit person) in each USU in the sample. Most USU's consist of about four housing units (or the equivalent in terms of GQ population), although the size depends on the source from which the USU is selected. For the most important statistics measured by the CPS, the optimum size of USU is about four housing units (58). The USU's are defined using the addresses of the living quarters as recorded in the 1970 census or as determined in a more recent listing operation.

A *segment* is selected at an intermediate sampling stage prior to designating the USU. The character of the segment depends on the source it is selected from; it may, for example, be a small area, a place, or an enumeration district. Segments do not have a common size but can always be expressed as an integral number of USU's. Each USU belongs to a specific segment and, in general, only one USU in a segment is in sample at a given time. In practice, for the purpose of administrative control, segments are given *segment numbers* and the USU's in a given sample are identified by listing the numbers of the segments involved.<sup>10</sup>

Use of resources from the 1970 census—Extensive use is made of data and geographic materials available from the 1970 Census of Population and Housing.

The 1970 Census procedures required the definition of about 229,000 *enumeration districts (ED's)*. The ED's are relatively large geographic areas defined in advance of the census and expected to contain on the average about 350 housing units. Each housing unit enumerated in the 1970 census is associated with a specific ED.

The 1970 census enumeration used two kinds of census procedure, mail and conventional (nonmail). The *mail procedure* was used for about 60 percent of the enumeration and involved preparation of a list of addresses, called an address register, to control the mail-out and the check-in of census questionnaires. There were about 121,000 ED's in the mail census which covered most of the larger urban centers of the country. For about 95,000 mail-census ED's, the address register was prepared before the start of enumeration by the purchase and correction (through the cooperation of the Postal Service and local agencies) of a computer tape copy of a commercial mailing list. These ED's are known as *tape address register (TAR) ED's*. TAR ED's are located in the Post Office city delivery areas within 145 selected SMSA's. The remaining ED's in mail census areas had their address registers constructed by a listing procedure conducted by Census Bureau personnel in advance of the census. These ED's are known as *prelist ED's*; tape copies of these address registers were not developed as a part of the decennial census operations.

For the *nonmail or conventional* census procedure used in the balance of the country, the enumerator prepared an address register for each ED as a part of the enumeration process.

In these various ways, the census enumeration produced an *address register* for each ED in the country. The registers show the addresses of housing units and of group quarters enumerated in the ED. The census enumerators were instructed to correct errors detected in the addresses of units in the register during the census enumeration.

<sup>10</sup> The term "segment," as used within the Census Bureau and in some publications, is often used in different contexts to mean USU or segment; this publication consistently uses the terms defined as above.

For each ED, except TAR ED's, a map was provided to the enumerator for use in the listing and enumeration. Where street names and house numbers were not present, primarily in rural areas, the enumerator was asked to record the serial number from the proper location on the ED map. The housing unit serial number was assigned to the household during the listing operation. TAR ED's were defined by the computer, and maps were not provided for each individual ED.

Each ED is identified by a state-county code and an ED number. For conventional and prelist ED's, the number assigned by mapping the ED's and serializing them in serpentine fashion within a larger area, typically within a county and within the balance of county. By using this code system for sorting, ED's are arranged in a systematic order, and ED's which are neighbors on the ground tend to be adjacent in the sort. The serpentine numbering system is used in assigning identification numbers to tracts within cities or counties, to city blocks within tracts, and in the same general way to TAR ED's within tracts.

The completeness of the address information in the register determines the way the census information can be utilized for sampling. The CPS design uses address registers, where possible, as the source for address samples selected to represent the population and housing units currently existing at addresses recorded in the census. In other ED's, where the recorded address information is incomplete or insufficient to identify specific addresses without ambiguity, samples of small areas are used. These are called *area samples*. Area sample ED's are primarily rural although occasionally ED's in urban areas receive this treatment.

It is necessary to set a standard to determine whether the quality of information in the address register for a given ED is adequate for an address sample or whether an area sample must be used. ED's are classified "list" or "area" for this purpose.

1. A *list ED* has at least 90 percent of the 1970 census addresses recorded in the register with complete street name and house number. A further condition, discussed below, requires the ED to be geographically compact within a jurisdiction issuing building permits for construction as of January 1970. TAR ED's are classified "list ED's." Address samples are selected from list ED's.
2. All other 1970 census ED's are defined "area ED's." Address samples are selected from area ED's.

**Area sampling**—An area sample could theoretically be developed within an area ED by subdividing the ED into small areas, each expected to contain the number of living quarters desired for a USU. Of course, this cannot be done without knowledge of the precise location of each unit in the ED. In practice, it is usually not possible to define land area small enough to contain an expected four housing units. Further analysis of sampling and interviewer cost and sample size shows that areas as small as this are relatively inefficient. Instead, somewhat larger areas, called *blocks* or *sample blocks*, are defined, having geographic features which are shown on a map (and actually exist) as boundaries and contain a population and housing units equivalent to 2 to 5 USUs. It is easier to identify the location of census units in these areas using the available information in the 1970 address register. A sample block is chosen and becomes the *area segment*.

sampling operations are explained in greater detail, with an illustration, in app. C.)

A separate visit is made to list all current living quarters (housing units and special places) in the area segment. The list is subsampled to identify the living quarters in a particular USU. The selection probabilities at each step are computed to keep the sample self-weighting.

Experience has shown that skilled technicians are required to prepare area samples properly. The operation is expensive and time consuming, and few of the preparatory steps are, at present, subject to mechanization. In addition, because it is difficult to associate 1970 census units to the blocks, estimates based on area samples alone are sometimes subject to relatively high variance. On the other hand, estimates based on area USU's have the advantage of covering the population residing in all living quarters existing at the time of interview. This advantage applies even though the method of associating living quarters to the blocks is not completely accurate and irrespective of whether units existing at the time of the survey were, at the time of the census, enumerated, missed, used for other (e.g., nonresidential) purposes, or came into being since the last census.<sup>11</sup>

**Address sampling**—A sample of living quarters recorded in the census address register and called an *address sample* is the major means of representing current population and housing in list ED's; as indicated below, this source does not represent the current population completely.

The sampling process within a selected list ED is equivalent to first removing all addresses with incomplete street name and house number information from the register (up to 10 percent of the 1970 addresses in the register may be removed), combining the remaining addresses into USU's, and then selecting one of the USU's for the sample. The intention is to make up geographically contiguous clusters of four housing units as USU's. The ED containing the sample USU is the *address segment*. (This sampling process is illustrated and explained in more detail in app. D)

The size of the address sample USU is subject to more control than for an area sample (if one were to be designated in the same ED) so that an estimate based on address samples should have the relatively lower variance. A substantial part of preparing the address sample can be performed by the computer, but a significant capital investment is required to adapt the address registers and the sampling operations to the computer (app. D). Once accomplished, however, the additional sample USU's regularly needed for the CPS are designated by the computer in list ED's at less expense, more quickly, and subject to better control than for sampling performed manually. When the capital costs are considered, however, it is not clear that the computer sampling process is more efficient. (This will be reviewed prior to the next CPS redesign.)

The instructions for the CPS interviewer insure that all current units at a sample 1970 census address are correctly represented, even though the number of units differs from the number reported in the census. However, the sample of 1970 addresses in a list ED does not represent all current units in the ED; the remaining units are represented by other means, as

explained in the following paragraphs. The address sample<sup>12</sup> does not represent—

1. Census units not recorded in the address register (presumably missed in the census).
2. Census units at addresses recorded with incomplete address information in the register (these have, in effect, been removed from the register before sampling).
3. Units constructed in the ED since the census.
4. Units at addresses converted from nonresidential use.

**Use of building permits issued for new construction**—Currently, there are about 14,000 jurisdictions (places) which issue building permits for new construction. These account for about 85 percent of all new residential housing units constructed. A continuing Census Bureau survey of these places, the Survey of Construction, is the source of a regular publication series providing estimates of housing authorized by building permits [54].

The CPS uses this survey to represent units constructed since the census in roughly 13,000 places issuing permits, as of January 1970, that are in list ED's; the balance of the new construction, by definition, occurs in area ED's and is covered by the area sample.

The sample of building permits representing new construction in list ED's is selected from all permit issuing places in the 461-sample PSU's in the A- and C-designs that were issuing permits, as of January 1970. Places that started to issue permits after that date are represented in the area sample.

The permit places are considered in two groups for sampling purposes. Places with a history of issuing permits for 50 or more residential units annually make monthly reports to the Census Bureau for the Survey of Construction. For the CPS, the number of units authorized by each permit is used to select a self-weighting sample representing new construction units in places having large permit volume. For such places, all permits issued in a given month at the sample permit issuing place comprise the *permit segment for larger places*, and the USU is an expected four housing units selected from these permits.

Other permit places, for sampling purposes, are assumed to issue permits at an average of two units per month (i.e., four units, the expected size of a USU in 2 months). From this group, a self-weighting sample of places is selected. All permits issued during a 2-month period comprise the sample USU; the *permit segment for smaller places* is identical to the USU.

**The Cen-Sup sample**—Housing units in list ED's at addresses missed in the census or inadequately described in the address register are represented by a separate sample developed for this purpose called the *census supplemental (Cen-Sup)* sample.

Representation of these units has been developed as follows: An independent sample of list ED's has been selected in the A- and C- design sample PSU's. A probability sample of city blocks (or, in noncity areas, sample areas about equivalent in size to city blocks) within these ED's has been intensively canvassed since the census. The results of the canvass, when matched

<sup>11</sup> The samples for CPS designs used prior to 1963 were based entirely on area samples.

<sup>12</sup> In addition to the units described, there are small groups of current units not adequately represented in the CPS sample. (See section on potential bias from coverage loss in ch. VII, p. 82.)

against the census address register for the appropriate census ED's, identify two categories of address—

1. Addresses for units built before the census and found in the canvassed area but not appearing in the address register; this defines a sample representing units at addresses missed in the census.<sup>13</sup>
2. Addresses found in the canvassed area and listed in the ED address register but with street name and house number information too incomplete or ambiguous to adequately specify the address location.

As long as good address information appears in the register for at least one housing unit at an address, the CPS address sample properly represents all current units at the address. (See the listing of living quarters in the sample segment in ch. IV, p. 30.) The Cen-Sup sample represents units at addresses completely missing or inadequately recorded in the register.

On a per-unit basis, the Cen-Sup sample is much more expensive to prepare than any of the other samples. To increase efficiency, some of the sample has been selected at lower probability and, therefore, requires an additional weight factor to reflect the lower sampling rate. To take advantage of work already completed to evaluate the 1970 census and differential costs, the additional weight factor is allowed to vary by USU. The expected number of housing units in the USU is established at one. The small area canvassed after the census, and referred to above, is the *Cen-Sup segment*.

As indicated in chapter III, a USU, once designated for the CPS sample, is not reassigned to a sample during the life of the current design. This restriction does not apply for Cen-Sup units; the same USU's are reintroduced every five samples, i.e., a new series of eight monthly interviews begins every 40 months.

### Assemble the Within-PSU Sample

The samples described in the section, "Resources for Sampling within the PSU," represent separate sectors of the current population. However, the sectors are not mutually exclusive; some categories of the population are represented by more than one of the samples. This section shows how the sample resources are integrated to provide correct representation of the current population.

Select the sample of census ED's—The first step in subsampling within the sample PSU is to select a sample of census ED's. This is an appropriate place to start, because the complete universe of 1970 ED's completely exhausts the land area of the 50 states and the District of Columbia. Every current housing unit is, therefore, associated with a specific census ED. The sample resources previously discussed are associated with census procedures which differ by ED. A measure of size for each ED (census counts of housing and group quarters population), correlated with current population, can be assigned.

The selection of sample ED's within the sample PSU and the determination of the number of USU's to be taken from the

selected ED's are accomplished in the following (these steps illustrated in app. B by the selection performed in an A-design sample PSU):

1. Computer summary records for ED's are sorted so that ED's of the following geographic categories in the PSU are placed together—

- C — The central city of an SMSA
- B — Urbanized area not in category C<sup>14</sup>
- U — Urban place not an urbanized area and not category C
- R — All other ED's

Within each of these categories, the ED's are sorted by identification number, a geographic code which is used to place geographically contiguous ED's together in sort.

2. Data from the 1970 census are used to develop a measure of size (i.e., the number of USU's to be assigned) for each ED by the formula—

$$M_e = [H_e + P_e/3]/4$$

where  $H_e$  is the total (occupied and vacant) housing units in the  $e$ -th ED. The value of  $P_e$  is the population (excluding institutional units) in housing units in the  $e$ -th ED. The value of  $M_e$  is rounded to the nearest integer but is assigned a value of at least one for each ED.

3. The ED's are sampled systematically (with a random start), with probability proportionate to the measure of size ( $M_e$ ) assigned. The probability of selection of the  $i$ -th ED in the  $h$ -th stratum, having a measure of  $M_{hie}$  USU's, is

$$M_{hie}/W_{hi}$$

where  $W_{hi}$  is the within PSU sampling interval (see app. B, table II-4). If the ED falls in sample, one USU is selected from it if

$$M_{hie} \leq W_{hi}$$

and one or more USU's are selected if

$$M_{hie} > W_{hi}$$

The actual number of selected USU's is determined by the sampling procedure. The random start for the sampling within the ED is determined by the selection from the sampling operation performed by the computer in the prior sample PSU. Independent sampling operations are conducted in each of the categories of ED (C, B, U, and R), with random starts carried forward independently from the previously sampled PSU.

<sup>13</sup> The 1970 Census Evaluation Program (E6) [67] was designed in this way to measure the number of housing units at missed addresses in mail areas; the Cen-Sup sample is an extension of this program to cover the two categories of address in mail and nonmail areas.

<sup>14</sup> An urbanized area consists of one or more cities comprising a community of 50,000 or more population and including the closely settled territory, urban in character [57].

FIGURE II-2. Method of Sampling USU's Within PSU's,  
by Category of 1970 Census ED's

1970 census ED's, by quality of address information recorded in the census address register	
1970 census ED's, by location within jurisdiction issuing building permits: January 1970	Good address ED's: Have at least 90 percent of addresses with complete street name and house number
	All other 1970 census ED's
In a jurisdiction requiring a building permit for new private residential construction	(1) Address sampling for units recorded in 1970 Census with good addresses  Census supplement (Cen-Sup) for units with poor addresses or missed in 1970 Census  Permit sample to represent units constructed since the census
	(3) Area sample for all except units constructed since the census  Permit sample to represent units constructed since the census
All other ED's	(2) Area sample to represent all existing units
	(4) Area sample to represent all existing units

<sup>1</sup> All tape address register (TAR) ED's are put in quadrant 1.

Sampling method determined by resources in the ED—The next step in the operation ascertains how to use the sample resources to represent current units in the selected ED's.

Four classifications of census ED are determined in response to yes or no answers to the following two questions:

1. Does the ED have at least 90 percent of the units recorded at addresses in the register with complete street name and house number?
2. Is the ED within the jurisdiction of an authority issuing building permits, as of January 1970?

A "yes" to both questions defines a list ED; also, all TAR ED's are defined as list ED's. Any combination of "no" answers defines an area ED. The complete universe of 1970 ED's is illustrated in figure II-2 that shows four quadrants associated with the four possible classifications of ED's, by answers to the two questions above. The figure also summarizes the samples used to represent current population in each category of ED in the quadrant.

Integrate the sampling methods—The methods of selecting the within-PSU sample are considered separately for the ED's in each of the quadrants and are dealt with in order of increasing complexity. The interviewing procedures for the USU's are described in further detail in chapter IV.

In quadrants 2 and 4 of figure II-2, only area samples are used. (The area sample selection procedure is illustrated, in detail, in app. C.) New construction is a major cause of variation in USU size; but, in the current CPS sample, this is of less concern, because only a small portion (recently, about 15 percent) of the total new construction appears in these quadrants.

In quadrant 3, a permit sample is used in addition to the area sample. As part of interviewing area USU's in this quadrant, the interviewer asks for the year the structure was built. Units built since the census are omitted when found in area segments in this quadrant, since they are represented by the separate sample of USU's in the permit sample. All population in this quadrant could be represented by an area sample alone, but, by using the permit sample to represent new construction, improved control

in USU size and reduction in variance results. The area sample USU's in quadrants 2, 3, and 4 comprise about one-fourth of all USU's in the CPS national sample.

In quadrant 1 all ED's are list ED's. Their current population is represented by three list samples, representing current units as follows:

1. The address sample, selected from the completely recorded addresses in the census register. The method of designating an address sample within a list ED is illustrated, in detail, in appendix D. Special places (dormitories, rooming houses, and other living quarters for the

nonhousing unit population) that have address sample USU's selected from them are termed "*special place segments*," and somewhat different procedures are followed in interviewing them. (See the listing of living quarters in the sample segment in ch. IV, p. 30.)

2. The Cen-Sup sample, representing units at addresses not recorded as enumerated in the census or units enumerated in the census but at locations with bad addresses.
3. The permit sample, to represent units built since the census.



## Chapter III.

# ROTATION OF THE SAMPLE AND REPLACEMENT OF PSU'S

## INTRODUCTION

### Definitions and Purposes

*Rotation of the sample* is the partial replacement of the USU's in sample in successive months in accordance with a planned, continuing design. In addition, the rotation system used in the CPS occasionally requires *replacement of PSU's* which involves substituting all USU's in one NSR sample PSU by USU's sampled from another PSU in the same stratum. In the CPS, the rotation of the sample and replacement of PSU's are carried out in such a way that each month's sample is a probability sample of the population covered by the survey.

When the CPS rotation system was first incorporated in the design, the major purpose was to avoid the undue reporting burden expected of survey respondents, if they were to constitute a permanent panel. Substantial reductions in sampling costs, as well as the sampling error of estimates of month-to-month change, could result if it were feasible to conduct interviews of the same sample of households each month. However, it was believed that asking the same respondents to supply information once every month for an indefinite period might create administrative problems due to complaints and, by substantially increasing the refusal rate, introduce biases and reduce the sampling efficiency. To avoid this difficulty, while preserving some of the advantages of overlap from month to month, a system of partial rotation of the sample each month was adopted.

Once the rotation system had been adopted, further study of its properties showed that it could be used to reduce sampling error in two ways. First, it is possible to use a composite estimation procedure that results in significant reduction in the sampling error of estimates of level and change for most items. This estimation procedure, which depends on the fact that data from previous months are usually highly correlated with the corresponding estimates for the current month, was adopted in July 1953. (The estimation procedure is further described in ch. V.)

Secondly, rotation of the sample makes it possible, at slight additional cost, to reduce sampling error by increasing the representation in the sample of USU's with unusually large numbers of housing units. Although changes in the sampling system have since lessened the impact of such USU's, it still remains as a real advantage. (The procedure for handling unusually large USU's is described in app. G.)

The overlap of the sample from month to month (and over other periods) also makes possible certain specialized kinds of analysis. By matching specific individuals between consecutive months, for example, it is possible to cross-classify status in one month compared to another and obtain some insights into the dynamics of the labor market. Although subject to certain bias, this type of analysis provides greater understanding of the changes taking place which are often obscured in the net month-to-month differences [77].

Further experience with the CPS rotation system has shown that the expected value of the response may be related to the number of times interviews have been conducted at the unit. This problem, called rotation group bias, is discussed in chapter VII.

## THE ROTATION SYSTEM USED IN THE CPS

The following describes how the CPS rotation system functions once it is in full operation. Problems connected with starting up such a system in a new design are discussed on page 22; some of the alternative rotation systems that have been considered are discussed on page 24.

For this discussion, it is necessary to define the terms "sample" and "rotation group." As used in the discussion of rotation, a *Sample*<sup>1</sup> is a set of USU's selected according to the procedures described in chapter II, with a specified overall sampling fraction; thus an A-Sample could be a set of USU's selected from A-design PSU's at the overall rate of 1 in 1.968 and a C-Sample selected at 1 in  $2 \times 1.968$  from C-design PSU's. (See table II-4.) If all of the units in the survey are replaced each month, a new Sample would be needed each month. Actually, since, in the CPS, each new unit introduced is included in the survey for 8 months, a new Sample is needed only once every 8 months after the system is functioning.

Since each new Sample is introduced into the survey over a period of 8 months, it must be divided into eight approximately equal parts or subsamples, each a probability sample of the population covered by the survey. Each of these subsamples is known as a *rotation group*.

It is possible to describe and operate the system by treating each rotation group as a separate and independent sample and numbering the groups in a single continuing series; however, there are good reasons for working with Samples and identifying each rotation group by Sample and rotation group number. For example, rotation groups one through eight of Sample A35 refer to the eight rotation groups in the 35th A-design Sample used in the CPS. (The first Sample designated in the current 376-area A-design is Sample A29.) Samples selected from the C-design are similarly identified with the letter-prefix C and are also comprised of eight rotation groups. (The first C-design Sample for the present design is C13.) The complete CPS for a given month uses rotation groups from specific combinations of A- and C-design Samples.

In the CPS, a particular rotation group is identified by its Sample number and followed by its rotation group number in parentheses; e.g., A30(5) and C14(5) would denote rotation group five of Sample A30 and Sample C14, respectively. Samples for other programs, drawn from the same sampling frames, have similar identification systems. For example, the national samples conducted by the Bureau of the Census for the National Center for Health Statistics, known as the Health Interview Survey, are denoted by the letter prefix Y. (See app. E.)

### The Operating System

The rotation system used in the CPS, once in operation, may be fully described, as follows:

<sup>1</sup>In this report, Sample is capitalized to distinguish this specific meaning; hence, the sample (uncapitalized), interviewed in March 1973, comprises parts of Samples A30, C14, A31, and C15 (the word is capitalized in the latter case)

1. Each Sample is comprised of eight subsamples called rotation groups that are implicitly specified by the designation of the USU's for Sample A29 and for Sample C13. The selection procedure (described more completely in app. B, C, and D) is that the PSU's are first arranged in a specific order—essentially by SMSA, non-SMSA groups within the four census regions—and then ED's are sorted within the categories C, B, U, and R<sup>2</sup> within PSU. Once in this order, four systematic samples of USU's are selected for Sample A29 (one in each of C, B, U, and R) and another four for Sample C13. In the same operation, the USU's in each of the Samples A29 and C13 are assigned systematically to eight rotation groups, using a random starting point.
2. One new rotation group is introduced into the survey each month.
3. The rotation groups are introduced in order, i.e., A30(1), A30(2), . . . A30(8), A31(1), A31(2), . . . A31(8), A32(1), . . . etc., and similarly for the corresponding C-design samples, C14(1), . . . C14(8), C15(1), . . . etc.
4. Each new rotation group is included in the survey for 4 months, is excluded for 8 months, is returned for an additional 4 months, and is then retired from the CPS sample. Because of this feature, the plan is referred to as the 4-8-4 rotation system.

### The Rotation Chart

The chart presented in figure III-1 summarizes how the rotation system operates. The chart begins with December 1972; for the purpose of the present discussion observe the entries for March 1973, the first month the complete 461-PSU design is in effect. Examination of this chart will identify the following important characteristics of the 4-8-4 rotation system.

1. After the first month, six of the eight rotation groups will have been in the survey for the previous month; i.e., there will always be a 75-percent month-to-month overlap.
2. When the system has been in full operation for 1 year, (March 1974<sup>3</sup>), four of the eight rotation groups in any month will have been in the survey for the same month, 1 year ago; i.e., there will always be a 50-percent year-to-year overlap.
3. As of March 1974, figure III-1 shows that of the two new rotation groups that come into the survey each month, one is completely new, and one is returning after an absence of 8 months.

### Overlap of the Sample

Table III-1 shows, for the fully operating system, the proportion of overlap between any 2 months as a function of the interval between them. The proportion of overlap has a strong effect on the sampling errors of estimates of change for most characteristics.

<sup>2</sup>The categories C, B, U, and R are discussed in the selection of the sample of census ED's in ch. II, p. 18.

<sup>3</sup>The new sample was introduced in phases by the system described later on this page so that the date of November 1973 actually applies for these statements.

TABLE III-1. Rotation System (4-8-4) Proportion of USU's Common in Two CPS Samples, Separate by Selected Intervals

Interval (In months)	Percent of USU's common in the two samples
1	75
2	50
3	25
4-8	—
9	12.5
10	25
11	37.5
12	50
13	37.5
14	25
15	12.5
16	—
.	—
.	—
.	—

— Entry represents zero.

### Introducing a New Rotating Sample

Assume a new survey incorporating a 4-8-4 rotation system begins with interviews conducted for the first time in all eight rotation groups in March 1973. Clearly the 75-percent month-to-month overlap would be achieved in the second month of operation. There would be no year-to-year overlap, however, until March 1974, the 12th month of operation. At that time there would be an overlap of 50 percent (four-eighths). After this time, the full 50-percent year-to-year overlap is maintained. Note that while the going system would require a new sample only once every 8 months, the inauguration of the new system, under this process, requires, as a minimum, parts of two samples: the first month of operation and part of a third in the second month. Thereafter, the requirement drops off to one new sample every 8 months.

New CPS design introduced in one month—When a redesigned sample is to be introduced into the ongoing program, there are a number of reasons that inhibit discarding the old CPS and introducing a new complete sample in the space of 1 month. The redesigned sample will involve new areas and, therefore, a number of new and less experienced interviewers. Also, since the new sample of USU's is drawn from a different set of sampling frames, all units would be interviewed for the first time in the first month of the new sample. Modifications in survey procedures usually are also made with a new design. These factors increase the risk of a discontinuity in the estimates when the entire transition is made at one time.

New CPS design introduced in phases—To reduce the problems of this nature, a gradual transition from the old to the new sample design is employed. When the 461-area design replaced the old 449-area sample, two groups of areas (called "phases") were used with separate systems for introducing the new sample and retiring the old sample operating in each group.

FIGURE III-1. Rotation Chart of CPS A- and C-Design Samples: November 1972-July 1975

Year and month	Sample and rotation					
	A 29 C 13	A 30 C 14	A 31 C 15	A 32 C 16	A 33 C 17	A 34 C 18
1972 NOV...	...5 6 7 8		1 2 3 4			
DEC...	...6 7 8	1 .....	2 3 4 5			
1973 JAN...	...7 8	1 2 .....	3 4 5 6			
FEB...	...8	1 2 3 .....	4 5 6 7			
MAR...		1 2 3 4 .....	5 6 7 8			
APR...		2 3 4 5 .....	6 7 8 1			
MAY...		3 4 5 6 .....	7 8 1 2			
JUNE...		4 5 6 7 .....	8 1 2 3			
JULY...		5 6 7 8 .....	1 2 3 4			
AUG...		6 7 8 1 .....	2 3 4 5			
SEPT...		7 8 1 2 .....	3 4 5 6			
OCT...		8 1 2 3 .....	4 5 6 7			
NOV...		1 2 3 4 .....	5 6 7 8			
DEC...		2 3 4 5 .....	6 7 8 1			
1974 JAN...		3 4 5 6 .....	7 8 1 2			
FEB...		4 5 6 7 .....	8 1 2 3			
MAR...		5 6 7 8 .....	1 2 3 4			
APR...		6 7 8 1 .....	2 3 4 5			
MAY...		7 8 1 2 .....	3 4 5 6			
JUNE...		8 1 2 3 .....	4 5 6 7			
JULY...		1 2 3 4 .....	5 6 7 8			
AUG...		2 3 4 5 .....	6 7 8 1			
SEPT...		3 4 5 6 .....	7 8 1 2			
OCT...		4 5 6 7 .....	8 1 2 3			
NOV...		5 6 7 8 .....	1 2 3 4			
DEC...		6 7 8 1 .....	2 3 4 5			
1975 JAN...		7 8 1 2 .....	3 4 5 6			
FEB...		8 1 2 3 .....	4 5 6 7			
MAR...		1 2 3 4 .....	5 6 7 8			
APR...		2 3 4 5 .....	6 7 8			
MAY...		3 4 5 6 .....	7 8			
JUNE...		4 5 6 7 .....	8			
JULY...		5 6 7 8 .....				
Year and month	A 29 C 13	A 30 C 14	A 31 C 15	A 32 C 16	A 33 C 17	A 34 C 18

*Phase I* consisted of land areas that were self-representing in both the old and new designs. The areas did not necessarily comprise full PSU's. Approximately 57 percent of the U.S. population in 1970 was in phase I areas. In these areas, the sample was introduced, one rotation group at a time, beginning in December 1971. Interviews conducted entirely from the new design were first accomplished in March 1973. The percent of overlap has conformed to the figures given in table III-1 for each of the old and new designs during the transition period.

All new PSU's, and parts of PSU's not in phase I, were included in *phase II*. In these areas, the new sample was introduced, one rotation group at a time, for 4 months (August through November 1972) and two rotation groups at a time, beginning December 1972. By this system, all phase II areas were also completely on the new design, as of March 1973.

### Alternative Rotation Systems

Prior to July 1953, when the current rotation system was introduced, the samples were divided into five rotation groups, with households interviewed for 6 successive months, and then dropped from the sample. Each month, for 5 months, one-fifth of the units in sample were interviewed for the first time and, then in the sixth month, no new units were added. In that rotation system, 80 percent of the sample units were common month-to-month for 5 months and 100 percent for the sixth month, but there was no overlap for the same month year-to-year.

The 4-8-4 rotation system presently used in the CPS is one of several that could have been adopted; a major factor in its choice was to save costs. The system currently used increases the period households remain in sample from six to eight interviews but masks the impact of the additional contacts on the household by spacing the interviews over a 16- rather than a 6-month period. The current system reduced the month-to-month overlap in the sample from either 80 or 100 percent to 75 percent and provides a 50-percent overlap in the sample year-to-year.

While the 4-8-4 system has been in effect, the size of the sample has been increased, estimation procedures have been improved, and better controls on survey error have improved the accuracy of the CPS results. As a consequence, survey errors, formerly not separately recognizable in the mean square error, have become more important in evaluating the accuracy of the CPS estimates. One of these is the phenomenon of rotation group bias that is further discussed in chapter VII, page 83. This problem must be considered when proposing modification in the existing rotation system, because the extent of the rotation group bias is related to the number of times the household is interviewed, and its impact and causes are not well understood. More complete knowledge of this problem might allow other options for rotation. (See, for example, the design requirements for State samples in ch. VI, p. 68.)

## SELECTION OF ADDITIONAL SAMPLES FOR ROTATION

The use of a rotation system requires the periodic selection of additional samples. It would be possible to select these samples, each completely independent of the other, according to the procedures described in chapter II; however, this would obviously be extremely costly and inefficient.

## The System for Generating Additional Samples

The plan adopted for development of additional samples has three major features

1. The 4-8-4 rotation system used in the CPS involves replacement of a portion of the sample USU's each month. Replacing a portion of the sample, by an equivalent and independent sample of USU's in general, will increase the sampling error of month-to-month change. But, this increase could be minimized by using replacement USU's with characteristics similar to those being replaced. To approximate this, the first feature of the system is to assign adjacent or nearby USU's in the same set of PSU's as replacements; that is, the replacement USU's are made up of addresses next door, or nearly so, to addresses in the original USU.
2. The assignment procedure permits the orderly use of the sampling frames; thus, work involved in selecting samples for a number of different surveys in the same PSU's can be coordinated to reduce costs.
3. The method of assigning replacement USU's enables the Bureau of the Census to avoid contacting the same household for different surveys that have their samples selected from the same sampling frame. We cannot avoid the possibility of again contacting some of the same households in surveys having samples selected independently, or from different sampling frames, or if households move to another address.

In practice, the sample selection plan conforms approximately to these features. However, the third of these must be ignored during the changeover from an old to a new CPS design, because the sample within each PSU is selected on the basis of the most recent census information and is independent of any sampling frames previously used. Once the new design is introduced, however, the selection process avoids a second use of HU's designated for an earlier CPS sample from the current frame. It is Census Bureau practice to extend this control, where possible, to avoid a second use of a USU by any of the Census Bureau sample programs.<sup>4</sup>

## How Additional Samples Are Generated

The plan used for the generation of an additional sample from the initial one is illustrated by means of a hypothetical example. The illustration has been simplified by using an integer for the sampling interval, making the total measure of size in the sample exactly divisible by this integer, and by dealing with only one of the several ED categories. The procedure actually used considers the complications which appear in practice while conforming to the concepts given in this illustration.

Consider a sample PSU in which ED's of, for example, category B have a total measure of size of 1,100 USU's and within-PSU sampling fraction is 1 in 100. Assume that all of category B in this PSU have been completely divided into USU's and numbered, serially, from 1 to 1,100, using a graphic order (the actual procedure used for sampling is

<sup>4</sup>Housing units in Cen-Sup segments (less than one percent of CPS units) are exceptions; these units, by design, begin an additional use in sample every 44 months. (See ch. II, p. 17.)

PSU's is theoretically equivalent to using this). The initial Sample of 11 USU's is selected by the random choice of one of the first 100 USU's (by the process described in app. B, p. 113) and by taking every 100th USU thereafter.

Under the plan used, the second Sample consists of the 11 USU's immediately following the USU's in the first Sample. The third would consist of 11 USU's immediately following those in the second Sample, etc. Thus, if number 90 is the first USU selected, the successive Samples would be, as follows:

Sample	Serial number of USU's in sample					
First	90	190	290	...	990	1090
Second	91	191	291	...	991	1091
Third	92	192	292	...	992	1092
.	.	.	.	.	.	.
.	.	.	.	.	.	.
10	99	199	299	...	999	1099
11	100	200	300	...	1000	1100
12	1	101	201	...	991	1001
.	.	.	.	.	.	.
.	.	.	.	.	.	.
19	8	108	208	...	908	1008
20	9	109	209	...	909	1009
.	.	.	.	.	.	.
.	.	.	.	.	.	.
99	88	188	288	...	988	1088
100	89	189	289	...	989	1089

This system of generating successive Samples, by substituting the next USU for each USU in the current Sample, completely identifies the additional Samples needed for subsequent use by the act of selecting the initial Sample. Furthermore the next USU is usually located in the same sample ED. For list ED's, all USU's in the ED are identified by the computer when the initial USU is designated (app. D) so that the next USU is typically comprised of addresses that are neighbors of the first USU. For area ED's, the next USU is usually in the same block (See app. C.) In either case, the amount of preparatory work required to designate succeeding samples and locate the units on the ground is considerably less than required for the initial sample or for a sample selected independently of the previous one.

The same rotation group numbers assigned to the initial selections are assigned to the adjacent (next) USU's as they rotate into sample.

### Assignment of Samples to Census Bureau Surveys

The preceding procedure suggests a method of assigning the USU's needed for the CPS for the life of the sampling frame. For the 461-PSU design, a total of 19 USU's adjacent to each USU in the initial Sample have been reserved for the needs of the CPS national sample in the current decade.

The requirements for other current surveys conducted by the Census Bureau using the same sampling frames as the CPS have been estimated, and USU's have also been reserved for this purpose. Because the various surveys considered have different

sample sizes and sample designs (some are in only the A-design or the C-design PSU's and some are in both sets of PSU's), the pattern of reservations is not the same for each CPS. PSU. The assignment of USU's to the expected needs over the decade for a number of survey programs is described in app. E.

## REPLACEMENT OF PSU's

### The Problem of Exhausting USU's in the PSU

For some of the smaller PSU's, the number of USU's needed for all surveys planned for the life of the sample design can turn out to be more than the number in the PSU. As a practical matter, this problem should not arise for an SR PSU where the basic sampling interval (1 in 1,312) means there are 1,312 different systematic samples of USU's in the PSU. Within some NSR PSU's, however, the sampling interval may not designate a sufficient number of different systematic samples to accommodate all of the surveys. For example, an NSR PSU in the A-design only with a sampling interval of 1 in 80 will have just enough USU's for the surveys planned; if the initial selections are separated by less than 80 USU's, the sampling frame will be exhausted before the end of the decade. (See app. E, p. 136.)

### Alternate Solutions

There are two major alternative methods for handling the problem. The simpler of the two would be to drop the requirement that each new Sample contain only unused USU's. Thus, in the previous illustration where the hypothetical PSU is divided into 100 different Samples of 11 USU's, a 101st Sample (if needed) could be designated by reusing the first Sample, a 102nd by reusing the second Sample, etc. From the sampling point of view, this would be a completely unbiased system.

A second alternative, and the procedure used for the CPS, is to develop an unbiased system of gradual replacement of sample PSU's over the life of the sample design, e.g., the current decade. However, a system which moves to a different PSU only when the sample PSU is exhausted eventually would exclude the smaller PSU's. To avoid this bias, clusters of PSU's are first established within each NSR stratum that contains a PSU unable to supply the sample USU's for the entire decade. Such clusters, called rotation clusters, are determined before the sample PSU is designated. By the procedure given in the following section, each PSU in the rotation cluster containing the sample PSU is assigned the portion of the decade it is to be in sample so that each PSU has its proper expected sampling rate. The sample is shifted from one PSU to another within the cluster to conform to these assigned portions.

### Procedure for PSU Replacement

Establish rotation clusters within the NSR strata—If a stratum contains one or more PSU's that would not have enough USU's to cover survey needs for the decade, the PSU's in the stratum are grouped into rotation clusters. A *rotation cluster* is a group of PSU's, all members of the same A-design stratum, that,

taken together, will have enough USU's to serve expected current survey needs over the decade. In strata made up entirely of PSU's with a sufficient number of USU's, each PSU comprises a rotation cluster by itself. In considering this problem, all PSU's in the stratum (sample and nonsample) must be examined.

Rotation clusters are made to consist of a minimum number of PSU's and to be sufficient in combined size to provide the needed USU's. These criteria are designed to require the least change in the PSU structure and, thus, to maintain minimum impact on the survey program.

The NSR PSU's can be in sample in the A-design or C-design alone or in both designs simultaneously, and this influences the number of USU's that the PSU must furnish over the decade. These requirements are 80, 56, and 128 USU's, respectively (see app. E, p. 00); thus, defining the most efficient rotation cluster requires knowing the sample PSU. However, the rotation clusters must be determined before the sample PSU's are selected. To resolve this apparent impasse, rotation clusters for A-only PSU's are considered first and are used when the PSU is an A-only or C-only sample. If these clusters are not sufficient to serve the needs of both the A-and C-samples simultaneously, A-and C-clusters are prepared as combinations of A-only clusters.

In practice, the criteria for forming rotation clusters are expressed in terms of  $P_{hi}$ , the proportion of the 1970 census population of the h-th A-design stratum in the i-th PSU. From table II-4, we have  $W_{hi} = 1.968P_{hi}$ . Rotation clusters must be identified, if necessary, by combining PSU's within the same A-design stratum so that

$$1.968P_{hi} = 1.968 \sum_i^c P_{hi} \begin{cases} \geq 80 & \text{is required for the c-th} \\ & \text{cluster supplying an A-} \\ & \text{only or C-only sample PSU,} \\ & \text{or} \\ \geq 128 & \text{is required for the c-th} \\ & \text{cluster supplying an A-and} \\ & \text{C-sample, simultaneously.} \end{cases}$$

**Replacement when sample PSU's are identified**—After the sample PSU's are identified, PSU replacement is required if a sample PSU has been combined with others into a rotation cluster. The number of months each PSU in the cluster is to be in sample is derived as in the following illustration.

Assume an A-design stratum includes several PSU's too small to provide the required number of USU's if they should fall in sample. Each of these small PSU's is combined with other small or large PSU's into, for example, two rotation clusters X and Y, such that sufficient USU's can be supplied for all surveys in the decade if a PSU in either cluster were selected for the A-or C-sample (but not necessarily large enough to accommodate both simultaneously). Let the sum of the  $P_{hi}$  for the PSU's in the clusters, be  $P_{hx}$  and  $P_{hy}$ , respectively. To satisfy requirements of both designs, if one of the clusters were to supply USU's for the A-and C-samples simultaneously, clusters X and Y are combined into a single cluster XY. For simplicity, let all PSU's in the stratum (other than those in cluster XY) be large enough to supply the required USU's if they should fall in sample. The illustration assumes the A- and C-sample PSU's fall in cluster X.

It is not difficult to determine the total number of USU's that the cluster XY could contribute to the combined requirements of

the A-and C-samples if the sample design were to be used indefinitely. The expected proportion of this number to be assigned to each PSU in cluster XY is determined in two steps.<sup>5</sup>

First, determine the expected proportion of the total USU's in cluster XY that cluster X should contribute to the sample (the complement of this proportion applies to cluster Y). It can be shown that the proportions are given by

$$P_{hy}^2 / (P_{hx}^2 + P_{hy}^2)$$

and

$$P_{hx}^2 / (P_{hx}^2 + P_{hy}^2)$$

for clusters X and Y, respectively.

The second step determines the proportion of USU's to be contributed from each of the PSU's in cluster X and cluster Y to expressions of the form:

$$P_{hi} / P_{hx}$$

for the i-th PSU from stratum h in cluster X.

The derived proportions are used in a random process to determine the number of months out of the 120 in the decade that the sample USU's are to be supplied by each PSU in the cluster designated by the sampling process.

## Impact of PSU Replacement on the Survey

Usually the replacement of a sample PSU with another requires training a new interviewer and the loss of an experienced one. Also, there is an extra one-time cost associated with sampling in a PSU. The gradual rotation system, therefore, does involve an increase in cost, but, in relation to the other costs over the decade, it is very small, primarily because the number of PSU's involved in the replacement problem is small.

The impact of the replacement problem during the life of the 461-area design can be judged by the following facts. The life of the design, for this purpose, is taken to be the decade defined by the 10-year period beginning August 1972, the starting date for introducing the NSR sample in the current design.

1. There are 17 NSR PSU's, originally designated for the sample, which will be replaced during the decade (13 are in the A design, 3 in the C-design, and one in both designs)
2. The replacement process used will involve an additional 14 A- and 5 C-design PSU's at some time during the decade.

The replacement of a relatively small number of PSU's over the 10-year period is expected to have a minimal effect on monthly estimates. The full sample for each replacement PSU is therefore, introduced in a single month.

The PSU's involved and their scheduled dates for entering and leaving the A- and C-design samples are shown in appendix table M-1.

<sup>5</sup> The theory supporting this process is derived in [62].

## Chapter IV.

# SURVEY OPERATIONS

### INTRODUCTION

The survey operations for the design, based on the 1970 census, incorporate a number of changes in the procedure employed in earlier CPS programs. These modifications are primarily a reflection of changed methods in using decennial census data and other materials in selecting the sample and in the increased use of the computer. However, the information collected, as well as the principles of quality control and emphasis on economy of operation, remain essentially unchanged.

The survey operations for the CPS have been developed with the following aims in view—

1. To produce reliable statistics on the composition of the labor force related to a specific week in each month of the year. Interviews are conducted during a 9-day period, beginning on Monday of the week containing the 19th, and relate to activity during the week containing the 12th day of the month.
2. To produce these statistics rapidly. The estimates for a given month are generally made public on the first Friday of the following month. This is the 19th day of the survey cycle, counting from the first day of interview.<sup>1</sup>
3. To produce these statistics economically.

The CPS is one of a number of surveys conducted by the Census Bureau. Insofar as possible, Bureau programs have been designed so that the same survey materials, survey procedures, personnel, and facilities can be applied to as many surveys as possible. This objective has obvious advantages in terms of cost and in utilizing skills of survey personnel for other programs. For example, CPS interviewers can be (and often are) employed on non-CPS activities, because the sampling materials, interviewer instructions pertaining to listing and coverage, and, to a lesser extent, questionnaire content are designed to be the same for a variety of programs.

The survey operations are described under three major headings: Monthly data collection activities, control of the field operations, and processing of the data. The organization of the personnel and facilities to perform these functions is discussed first.

The operations are discussed in terms of the 461-area CPS national sample; however, the same survey operations apply also to the monthly increments which have been added to the national sample to improve the reliability of estimates for certain States. Figure J-2 indicates the impact of the workload from these increments by regional office area. (See also the section on survey operations in ch. VI, p. 73.)

### ORGANIZATIONS INVOLVED IN SURVEY OPERATIONS

This discussion of organization is concerned primarily with groups directly involved each month in producing CPS results.

<sup>1</sup>The schedule is lengthened by a day or two in selected months to reflect the occurrence of holidays in either the data collection or processing periods; the scheduled dates are announced in advance each month in the *Statistical Reporter* (10).

However, other Census Bureau staffs, not necessarily involved each month, influence the program in important ways, including the Associate Director for Demographic Fields, the Associate Director for Statistical Standards and Methodology, the Chief Mathematical Statistician, and the Chief and Staff of the Statistical Research Division.

The Bureau of Labor Statistics-Census Labor Force Committee, chaired by the Bureau of the Census, meets regularly (usually each month) to keep key personnel of the BLS and Census Bureau informed of future plans and survey activities. The CPS Policy Committee and the Interagency Technical Committee on Labor Force Statistics, both chaired by the Office of Management and Budget, may be convened to consider major issues affecting the CPS. The Census Advisory Committee of the American Statistical Association (ASA) meets regularly (usually twice each year) to advise on Census programs, including the CPS.

The Census Bureau Staff, associated with CPS activities, operates out of 14 physical locations, the headquarters in the Washington, D.C., area, the processing center in Jeffersonville, Ind., and 12 regional field offices.

#### Washington, D.C.

The Bureau of the Census, under the U.S. Department of Commerce, has its headquarters in the Washington, D.C., area. Census Bureau offices are located about 1 mile southeast of the District of Columbia, in Suitland, Md.

Major organizational units within the Census Bureau are called divisions, and they report to the Director, Bureau of the Census, through the appropriate associate director. The CPS is within the purview of the Associate Director for Demographic Fields.

The Demographic Surveys Division is responsible for coordinating the functions of all organizations within the Census Bureau associated with the CPS. It also coordinates subject-matter needs in the labor force area with the Bureau of Labor Statistics, the prime user and a participating sponsor of the CPS program, and the Statistical Policy Division of the Office of Management and Budget, which has responsibility for approving all statistical inquiries undertaken by the Federal Government. The Chief of the Demographic Surveys Division is the contact for inquiries concerning the CPS from the public and other governmental units. The division, in addition to coordinating the Census Bureau's CPS activities, is responsible for the design of the monthly questionnaire, including development of questions for the supplemental surveys; the procedures for processing the survey in the processing center and the regional office each month; and the preparation of the monthly instructions and training materials for the interviewing staff. The computer programming staff in this division is responsible for the programs necessary to edit, weight, and tabulate data from the monthly CPS and the supplemental surveys conducted in conjunction with the CPS. This division also performs the final review of the CPS results prior to their delivery to the Bureau of Labor Statistics, U.S. Department of Labor. (See the section on weighting and tabulation, p. 54.)

The following organizations, associated with the regular processing of the monthly CPS, are also located in the Washington

area. Their major functions are indicated, but it is not possible to enumerate all contributions of each division nor the way each division assists the others in the performance of assigned responsibilities.

The Statistical Methods Division is responsible for the sample design, the selection of the sample, the estimation procedure, developing quality control systems, measurement of survey errors, and maintenance of quality standards in census publications based on sample data. These responsibilities extend to all demographic surveys, as well as for the decennial census. Some aspects of the sampling activities are performed in Jeffersonville.

The Field Division headquarters staff is responsible for the administration of the interviewer and enumeration staffs maintained under the 12 regional offices, as well as serving as liaison between other divisions in Washington and the regional offices. (The organization of the regional offices and the training of their personnel is described in more detail in app. J.)

The Computer Services Division provides computer facilities and supplies services to process the CPS information. However, the computer programmer staff responsible for the CPS is part of the Demographic Surveys Division.

Each month, the Population Division supplies projections of the total U.S. population for the age-sex-race categories used in the second stage of ratio estimation. (See ch. V, p. 59.) Technical direction for the industry and occupation (I and O) coding operation is also provided by this division. The Population Division sponsors a number of repetitive supplements included in the CPS and prepares a wide variety of analytic reports based on CPS data.

In collecting and processing data for the construction industry, the Construction Statistics Division collects counts of residential building permits by permit issuing jurisdiction. These counts are used by the Statistical Methods Division to select a sample representing units built since the last census.

#### Jeffersonville, Ind.

The Data Preparation Division in Jeffersonville has two organizations regularly involved in preparing materials for the survey.

The Statistical Methods Section performs a number of functions requiring large-scale technical and clerical input leading to designation and control of sample USU's. Specifications for these functions are provided by the Statistical Methods Division in Washington. These activities include evaluating the quality of address information in the Census address registers (see the discussion of resources from the 1970 census in ch. II, p. 16); preparing ED's for computer segmentation (app. D); designating USU's in new construction segments, in special places, and in Cen-Sup segments; preparing the segment folders; and reviewing the field listings for address segments.

The Geographic Operations Branch performs the mapping work required in selecting area sample segments.

The Data Preparation Division is also involved in the processing of the monthly information collected by the field staff.

The General Operations Branch checks in and processes the CPS questionnaire and does the I- and O-coding.

The Support Services Staff and the Data Processing Systems Branch provide microfilm and FOSDIC services and transmit copies of FOSDIC tapes by wire to the Computer Services Division in Washington for processing by the computer.

## DATA COLLECTION ACTIVITIES FOR A SURVEY MONTH

The process of collecting information for the survey in a typical month is described under three topics. (1) Identify the USU's in sample. (2) list the living quarters in the segments and, where necessary, subsample the listing to identify the USU, and (3) conduct the interview. Specific instructions for these operations are given, in detail, in the *CPS Office Manual* [40] and the *CPS Interviewer's Reference Manual* [39].

### Identify the USU's in the Current Sample

The USU's to be interviewed for a CPS sample are identified by segment and rotation group using CPS form 11-200, the sample report. (See figure IV-1.) Normally, segments in two rotation groups are listed on this form in the "incoming" and the "returning" rotation groups. For a given month, the rotation group entering the sample for the first of the eight interviews is called the *incoming group*; the *returning rotation group* is entering the sample for the fifth interview after having an 8-month period out of the sample. The incoming and returning rotation group segments are listed in sections I and II on the form 11-200; these sections are labeled "initial sampling" and "auxiliary sampling," respectively. The segments shown on this form for the current month and on the forms for each of the 3 preceding months provide a complete list of all sample segments for the current month. (See for example, the rotation chart in fig. III-1 where, in April 1973, rotation group A32(1), rotation group 1 of Sample A32, is the incoming rotation group, and A30(5) is the returning group.)

Sample report, form 11-200—Segments for incoming and returning rotation groups are presented in three sets of sample report forms 11-200.

1. The first set of forms is generated in Washington by the computer and shows all segments except permit and Cen-Sup. This set is sent to the regional offices about 2 months before the interview month to allow time for listing units in the segments before interview.
2. The second set is manually prepared showing Cen-Sup segments and is transmitted with the first set.
3. The third set of forms lists permit segments and is prepared about a month later to allow the sample of building permits to be more current. This set of sample reports is prepared in Jeffersonville and is based on sampling conducted by a clerical operation.

Completed sample report illustrated—Figure IV-1 illustrates a completed sample report for the incoming and returning rotation groups of PSU 412, LaPorte-Stärke, Ind., for August 1973. This is an illustration of the first set of forms showing segments for the two rotation groups.

Column 1 shows whether the segment (identified in col. 2) is in the A- or C-sample. The sampling instructions to designate the USU for area segments are given in columns 3 and 4. Entries similar to those given in columns 1 to 4 for the incoming rotation group segments are shown in columns 9 through 12 for the returning rotation group segments. The entries in other columns



U.S. DEPARTMENT OF COMMERCE SOCIAL AND ECONOMIC STATISTICS ADMIN. BUREAU OF THE CENSUS										PSU - Name LA PORTE - STARKES, INDIANA		Code 4-12		Sheet 1 of 1	
SAMPLE REPORT										Section I - INITIAL SAMPLING		Section II - AUXILIARY SAMPLING			
Sample		Relation		Week		Sample		Beginning initial interview month (or quarter)		Beginning auxiliary interview month (or quarter)					
A 32/C 16		5		AUGUST 1973		A 32/C 15		1		AUGUST 1973					
Type of sample	Segment number	Serial numbers	Number of units		Sampling instructions	Segment number	Type of sample	Serial numbers	Segment also in	Sampling instructions	Number of units		Serial numbers of all units NOW in sample specified above	Update segment	
			Listed	In sample							Listed	In sample			
(11)	(22)	MICHIGAN CITY, CITY (PERMIT AREA)	4	(6)		(10)	A	(7)		ADDRESS	5	01-05	(15)	(16)	
A	5014	KNOX, CITY (PERMIT AREA)	4	(6)		1015	A	01-04		AREA	4	01-04		X	
A	5025	NORTH BEND TWP STARKES CD (PERMIT AREA)	4	(6)		1026	A	01-03							
A	5036		19	3											

Section III - SAMPLE REPORT ON SPECIAL PLACES (Enter "X" in appropriate box if NO special places located in section I or if components: <input type="checkbox"/> Section I <input type="checkbox"/> Section II)										Section IV - REMARKS			
Selected (C or R)	Segment number	Name and address of special place (If street number, give it; if not, give location.)	Type of special place and TYPE CODE	Sampling instructions		Total number of units		Number of units in sample		Serial numbers			
				Start with	Take every	Staff	Un-classified	Staff	Un-classified				
(17)	C 1026	Rocky Hill Hall, Route 420 Springville, Indiana	School	4	4	(21)	(22)	(23)	(24)	(25)	(26)	(27)	05-07

of sections I, II, and III have been completed by the field staff after the listing operation and a control copy is returned to Washington.

**Segment folder**—The rotation of the sample over time introduces new USU's (and new segments) into sample. When a new segment is introduced, the regional office is supplied a new segment folder and, for area segments, a new map. The segment folder is used to—

1. Record the segment identification and the survey processes to be applied to the segment in a given month (list, interview, etc.).
2. Record the transmittal of questionnaires and control cards for completed interviews and for noninterviews.
3. Organize the exchange of survey materials between the regional office and the interviewer.

New segments appearing through rotation of the sample are identified by appending an alphabetic suffix to the 4-digit number of the segment being replaced. New segments appear in the rotating sample depending on the type of segment:

1. When the last USU in an area segment is replaced, the next USU will be in a different segment (in the same or a different ED).
2. When the last USU in an address ED is replaced, the next USU will be in a new segment and will be in a different ED.
3. New permit segments appear when rotation brings in a new address or when more recently issued building permits are included by extending the sampling interval. A separate segment folder is typically used for each of the different geographic locations introduced by the sample of building permits.

### List the Living Quarters Within the Sample Segment

The living quarters to be listed are usually housing units. In group quarters and in special places (such as a transient hotel, rooming house, dormitory, trailer camp, convent) where the occupants have special living arrangements, the living quarters listed may be housing units, rooms, beds, etc. In the discussion of listing, all of these living quarters are included in the term "unit" when it is used in the context of listing or interview.

All units in the sample segments are listed prior to interview. For address, permit, Cen-Sup segments and TA places in special place segments,<sup>2</sup> both the listing and interviewing steps are performed by the interviewer on the first visit to the segment. For segments of this type, or addresses within such segments that have more than one USU in the segment, the units in each of the USU's are defined by predesignated lines on the listing sheet.

For all other segments (i.e., area segments and NTA places in special place segments), the listing is usually done a month before interview, and the specific units to be interviewed are sampled from the listing by a clerk in the regional office. Performing

those two steps separately allows each to be verified and allow more complete control of sampling to avoid bias in designating the units to be interviewed. The separation of the two operations is particularly important for area segments; for example since they are, by design, made up of a larger number of USU than other segments, require more preparatory work, and are sampled over a longer time.

The units to be interviewed within the segment are selected in an objective manner to avoid bias. However, the sample rotation system used in the CPS tends to remove biases that may have been introduced in the selection process. For example, if address USU is to be defined within a large structure where it will be assigned to several USU's, the interviewer could conceivably manipulate the order of listing the housing units to be the selection for a given USU. However, as the CPS rotation system eventually replaces each USU by others made up of units adjacent on the listing at the address, the unknown error from the selection bias in this and all other USU's of this character is replaced by a (measurable) variance. It is important to recognize that the first sample selected for a new design could have unknown selection biases if units for all USU's are designated by the process used for address USU's in the CPS.

**Area segments and NTA places in special place segments**—When the first USU from an area segment is to be sampled for any program which uses the A- or C-design sample, an interviewer visits the area to establish a list of living quarters in the segment. Units within the segment are recorded on an area-segment-listing sheet, form 11-212. (See fig. IV-2.) Heading entries on the form are filled by the office clerk, and items 2, 3, and 4 are used by the interviewer to record a thematic list of all occupied and vacant housing units within segment boundaries. All structures are accounted for even if there are no living quarters within them. Nonresidential structures on the same property as a housing unit are recorded in column 4 of the area segment listing sheet. Nonresidential structures on a property with no living quarters are recorded on a separate form printed on the inside cover of the segment folder.

If the area sample segment is within a jurisdiction that is building permits as of April 1970, it is necessary to determine the year each listed unit was built. This is required because only units constructed since that date are represented by permit segments. To determine "year built," the interviewer inquires each listed unit and circles the appropriate code in column 5 of the time of listing.<sup>4</sup> The inquiry at the time of listing is omitted in areas with low new construction activity; in such cases, construction units are identified later in completing the CC card during the interview and deleted from the sample at that time (i.e., a type C noninterview). This process avoids any contact at units built before 1970.

If the area segment is not in a building permit issuing jurisdiction, then all units within the segment are eligible for sample and the interviewer does not determine "year built."

The "special place" information at the top of form 11-2 is used to inform the interviewer of the presence of a special place area segment. Listing of the special place is performed on form 11-213. (See fig. IV-3.) The listing instructions apply

<sup>2</sup> The abbreviations TA and NTA, for take all and nontake all, respectively, distinguish assignments at which the interviewer is to interview all units (TA) or a sample of units (NTA).

<sup>3</sup> The same instructions for listing apply for a number of surveys using samples. Some of these programs are mentioned in app. E.

<sup>4</sup> Line 9, fig. IV-2, illustrates a unit constructed after the census and from the listing.

FIGURE IV-2. Area Segment Listing Sheet  
(Form 11-212)

Form 11-212 7-22-75		U.S. DEPARTMENT OF COMMERCE BUREAU OF THE CENSUS		PSU code <b>412</b>	Segment number <b>5036</b>	Survey <input checked="" type="checkbox"/> CPS <input type="checkbox"/> AHS <input type="checkbox"/> HIS <input type="checkbox"/> NCS <input type="checkbox"/> QNS		Year built codes
AREA SEGMENT LISTING SHEET				Post office (City, town, or village) <b>NORTH BEND TWP STARKE Co, IND</b>				
SPECIAL PLACE INFORMATION								
Name, address, ZIP code, and type of special place				<input type="checkbox"/> STAFF <input type="checkbox"/> UNCLASSIFIED		Before 4/1/70 A After 4/1/70 ND - Not to be determined		
				Person to contact				
				Name				
				Title				
				Telephone				
Line No. (1)	Street name or road designation (2)	House number or side of road (3)	Description or location of unit (4)	Non-residential structures on same property Farm bldgs. Store Other	Sample designation (5)	Serial number (6)	Remarks and date of change (7)	Year built (8)
1	Cory Road	L or R 1220	..... Name	..... Farm bldgs. Store Other				Ⓟ A ND
2		L or R 1338	..... Name	..... Farm bldgs. Store Other				Ⓟ A ND
3		L or R 307	Opt. A ..... Name	..... Farm bldgs. Store Other	A30	01		Ⓟ A ND
4	14th Avenue	L or R 307	Opt. B ..... Name	..... Farm bldgs. Store Other	A31	01		Ⓟ A ND
5		L or R 313	..... Name	..... Farm bldgs. Store Other	A32	01		Ⓟ A ND
6		L or R 315	..... Name	..... Farm bldgs. Store Other				Ⓟ A ND
7	Delany Road	L or R 1359	..... Name	..... Farm bldgs. Store Other				Ⓟ A ND
8		L or R 1357	Front ..... Name	..... Farm bldgs. Store Other	A30	02		Ⓟ A ND
9		L or R 1357	Rear ..... Name	..... Farm bldgs. Store Other	A31	<del>X</del>		Ⓟ A ND
10		L or R 1351	..... Name	..... Farm bldgs. Store Other	A32	02		Ⓟ A ND
Remarks								

the form and further detail is given in the *Interviewer's Reference Manual* [39].

After the listing of living quarters in the area segment has been completed, the listing forms are returned to the regional office. A clerk then applies the sampling instructions, shown in columns 3 and 4 on the sample report, form 11-200, to identify the units in the sample USU. (See fig. IV-1.) The result of the listing and sampling operations are summarized in columns 5, 6, and 7 of form 11-200. The entry in column 6 determines the number of serial numbers to be assigned to the sample units.

To reflect changes in the number and the use of the living space for units in the area sample segment, it is desirable to periodically review and correct (i.e., update) the listing so that the current USU, and later USU's to be selected from the segment, will represent the current status. The following rule is used: The USU interviewed in the first or fifth month in sample for CPS must be selected from a listing updated within the previous 8 months. A given segment, however, may be involved in more than one survey, but the rules for listing area segments are the same for all surveys using the current survey sample. The updating of the segment for any of these surveys can be considered in satisfying the foregoing rule.

The area segment is updated by retracing the path of travel used in the original listing, verifying the existence of each unit on the listing, and adding any unlisted housing units or other living quarters that are found at the end of the list. By extending the original sampling pattern, the unlisted units are given their chance of being in sample.

**Address, Cen-Sup and permit segments, and TA places in special place segments**—The listing in these USU's is conducted during the first visit to the segment for an interview. When the units in a segment have been listed, interviewing at units in the appropriate USU (identified as units falling on predetermined lines on the listing sheet) can be done on the same visit. The rule for type of interview (i.e., a personal visit, rather than telephone interview, is required for first month in sample) is a reflection of this procedure.

The computer prepares an address listing sheet, form 11-211A for each address segment as a part of the sampling operations described in appendix D. (See fig. IV-4.) The computer enters the address information in columns 1, 2, and 3, and identifies the USU for the unit on each line of the listing sheet in column 5. The number of housing units existing at that address in the 1970 census is recorded in column 7 and informs the lister concerning the number of current units to expect at the address. The regional office clerk has also used this figure to make an initial assignment of household serial numbers (col 6) to the units at that address.

For TA addresses (where all units at the address are assigned to the same USU), the listing sheet records the address separated by a sufficient number of lines to allow for additional units found at the address. The interviewer is required to list and interview all units at TA addresses.

One page of the listing sheet is used for each NTA address. The lister records all living quarters presently existing at the address and interviews those recorded on the lines for the appropriate USU (col. 5). Figure IV-4 shows an example of a completed address listing sheet for a nontake-all (NTA) address segment.

A regular system of updating the listings in address segments is also followed.

## Conduct the Interviews

Each month during interview week, the interviewers attempt to contact and interview a responsible person living in each sample unit to complete a CPS questionnaire, form CPS-1 (See fig. IV-7.) The week containing the 19th day of the month is the *interview week*. An introductory letter is sent each household in sample for the first and for the fifth month. The letter describes the CPS and announces the forthcoming visit. Figure IV-5 shows the introductory letter sent to sample units in the area administered by the Detroit Regional Office.

**Prepare roster of persons at sample unit**—At the time of the first interview, the interviewer prepares a list of all persons who are staying in the sample unit, including visitors and others temporarily away from the unit who usually live there. The roster is constructed using the field control card, form CPS 260. (See fig. IV-6.) The roster and questions on the control card are used to identify the living space making up the separate housing units or other units<sup>5</sup> (if any) including vacant living space, comprising the sample unit. A control card is prepared for each housing unit or other unit found at the sample unit whether occupied or vacant. Thus, a control card exists for each housing unit or other unit in the interviewer's assignment, including noninterview units of all types.

The control card provides for recording the personal characteristics of each person who is determined to be a member of a sample household, i.e., a person for whom the sample unit is the usual place of residence. (The rules for usual place of residence are discussed in app. H and are summarized in the table in that appendix. Examples of the personal characteristics entered on the control card are: Relationship to the head, date of birth and age, marital status, race, and sex. This record of members is brought up to date at each subsequent interview to take account of new or departed residents, changes in age, marital status, etc., and constitutes the complete sample of persons from which subsamples, having specified characteristics, are selected for specific studies. For example, information for the monthly labor force survey is collected for all civilian members 14-years old and older. (Persons in the Armed Forces and inmates of institutions are excluded.)

**Noninterviews**—For various reasons, interviewers are unable to obtain an interview at every housing unit designated for the sample. A CPS-1 and an updated control card are nevertheless required for each unit in the assignment. Table IV-1 shows the average monthly number of assigned units in 1975 and the number and proportion of noninterviews, by type. The type of noninterview is defined by the units listed on the CPS questionnaire, form CPS-1. (See fig. IV-7, item 15.)

The units not interviewed are classified in three categories, as follows:

*Type A noninterviews* are occupied units eligible for the survey for which the interviewer was unable to complete the interview. Table IV-2 provides, monthly for 1975-76 and annually for 1972-1976, the proportion of eligible sample households that are type A noninterviews by type. As can be seen, the overall rates range from about 3 to 5 percent each

<sup>5</sup> See ch. II, p. 15, for the definitions of housing unit and other unit.

FIGURE IV-4. Address Listing Sheet  
(Form 11-211A)

Form 11-211A 4-60-101		U.S. DEPARTMENT OF COMMERCE BUREAU OF THE CENSUS		PSU code 412	Segment Number 5014		Survey Type ADDRESS	CPS	
ADDRESS LISTING SHEET				Tract 20	1970 EO 32-091-0514		Address type NTA		
Line No.	Block (1)	Basic address Street name or road designation (2)		House No. (3)	Unit address (Location of unit or apartment No.) (4)	Sample designation (5)	Serial No. (6)	Number of Census units (7)	Remarks
1	125	HALL AV			Basement	A31			
2					Apt. 101	A31			
3					102	A31			
4					103	A32	01		
5					201	A32	02		
6					202	A32	03		
7					203	A32	04		
8						A31			
9						A31			
10						A31			
11						A32			
12						A32			
13						A32			
14						A32			
15									
16									
17									
18									
19									
20									

ADDRESS COVERAGE CHECK						Footnotes SAMPLE A31 ALSO IN TA ADDRESS
Name of enumerator	Listed		Updated			
	R. Brown	RB				
Year 1972	73					

UNITE - Regional Office

SHEET 1 OF 1 SHEETS

FIGURE IV-5. Current Population Survey Letter  
(Front)CPS-263  
(7-76)UNITED STATES DEPARTMENT OF COMMERCE  
Bureau of the Census  
Washington, D.C. 20233

OFFICE OF THE DIRECTOR

Dear Friend:

You may have read in the newspaper--or heard on the radio or television--the official Government figures on total employment and unemployment issued each month. These figures, as well as information about persons not in the labor force, are obtained from the Current Population Survey, for which your address and about 60,000 others throughout the United States will be visited during the next few months.

This survey, which is conducted by the Bureau of the Census under the authority of title 13, United States Code, section 181, provides vital up-to-date information on the status of the economy. A Census interviewer, who will show an official identification card, will call on you during the week in which the 19th of the month falls. The interviewer will ask questions about the members of your household--their ages, employment, occupations, and similar important facts. All information given to the interviewer is, by law, held in strict confidence. Your answers may be used only to prepare statistical summaries from which no information about you as an individual can be identified.

Because this is a sample survey, your answers represent not only yourself and your household, but also hundreds of other households like yours. For this reason, your participation in this voluntary survey is extremely important to ensure the completeness and accuracy of the final results. Although there are no penalties for failure to answer any question, each unanswered question substantially lessens the accuracy of the final data. Your cooperation will be a distinct service to our country.

On the other side of this letter are the answers to questions most frequently asked about this survey.

Sincerely,

VINCENT P. BARABBA  
Director  
Bureau of the CensusFurther information  
may be obtained from:Mr. Robert G. McWilliam  
Regional Director  
Bureau of the Census  
Washington Blvd. Bldg., Room 2100  
234 State Street  
Detroit, Michigan 48226

Telephone: (313) 226-7742



**FIGURE IV-5. Current Population Survey Letter  
(Back)**

**What is this survey all about?** This survey — called the Current Population Survey — is taken each month to provide an up-to-date estimate of the number of persons working, the number who are unemployed, and many other related facts. Occasionally we ask additional questions on education, health and disability, family income, housing, and other important subjects.

**Who uses this information? What good is it?** In a country as big as ours, and one that is changing so rapidly, up-to-date facts are needed by people in government, business, and other groups in order to plan programs intelligently. It is important to know how many people are working or out of work (to help direct programs which would contribute to an expanding economy and provide new jobs), how many children will be attending school (to plan for schools and the training of an adequate number of teachers), how many new families are being formed (to plan for adequate housing to meet their needs), and so on. The Current Population Survey is one of the ways these and many other facts are collected between the major Censuses taken every 10 years.

**How was I selected for this survey?** Actually, your living quarters, rather than you personally, was selected for this survey. About 15,000 groups of addresses, relatively close together, are scientifically selected to be representative of the country. There are about four dwellings in each group of addresses included in the survey.

**If you should move away while your dwelling is still in the survey, we would interview the family that moves in.**

**How many times will I be visited?** Occupants of a selected dwelling will be visited for eight months — four consecutive months the current year and the same four months in the following year. In addition, a small number of households is visited twice during one of the eight months to insure the validity of our statistics and verify that our interviewers are doing the best job possible.

**I hesitate to tell some things about myself. What protection do I have?** All information given by individuals to the Census Bureau is held in the strictest confidence by law (Title 13, United States Code, Section 9). The interviewer has taken an oath to this effect and is subject to a jail penalty and a fine if he discloses any Census information given him. Only statistical totals, in which the information for individuals cannot be identified, are ever published.

**Is this survey authorized by law?** Yes. Title 13, United States Code, Section 181, and Title 29, United States Code, Sections 1 through 9, authorize the collection of most of the information requested in this survey. In addition, portions of the survey in any one month may be authorized by one of the following: Title 7, United States Code, Sections 1621–1627, Title 38, United States Code, Section 219, and Title 42, United States Code, Sections 247 and 2825, and Public Laws 89-10, 92-318, and 93-380. In some months, the survey may contain questions authorized under laws other than those cited; further information concerning the authority for any particular portion of the survey can be obtained from the interviewer who contacts your household.

**Why do you include me — I'm retired?** Some retired persons may feel that their activities are not important to this type of survey and may wonder why they are included. However, in order to have a cross-section of the entire population, it is necessary to include persons in all age groups as well as persons in all kinds of geographic areas. Our statistics show that almost a third of our retired citizens do work in the course of a year and it is important to reflect this information in our figures. Moreover, as mentioned above, information is collected in this survey from time to time on subjects of special interest to retired citizens, such as family income, housing, social security coverage, disability and health insurance.

NAME (Last, first, middle initial) If married, include husband's name		DATE OF BIRTH (Month, day, year) If married, include husband's date		PLACE, STATE, ZIP CODE		RELATIONSHIP TO HEAD OF HOUSEHOLD		MARITAL STATUS		EDUCATION		OCCUPATION		INCOME		MILITARY SERVICE		REMARKS	
1. NAME	2. DATE OF BIRTH	3. PLACE, STATE, ZIP CODE	4. RELATIONSHIP TO HEAD OF HOUSEHOLD	5. MARITAL STATUS	6. EDUCATION	7. OCCUPATION	8. INCOME	9. MILITARY SERVICE	10. REMARKS	11. NAME	12. DATE OF BIRTH	13. PLACE, STATE, ZIP CODE	14. RELATIONSHIP TO HEAD OF HOUSEHOLD	15. MARITAL STATUS	16. EDUCATION	17. OCCUPATION	18. INCOME	19. MILITARY SERVICE	20. REMARKS
21. NAME	22. DATE OF BIRTH	23. PLACE, STATE, ZIP CODE	24. RELATIONSHIP TO HEAD OF HOUSEHOLD	25. MARITAL STATUS	26. EDUCATION	27. OCCUPATION	28. INCOME	29. MILITARY SERVICE	30. REMARKS	31. NAME	32. DATE OF BIRTH	33. PLACE, STATE, ZIP CODE	34. RELATIONSHIP TO HEAD OF HOUSEHOLD	35. MARITAL STATUS	36. EDUCATION	37. OCCUPATION	38. INCOME	39. MILITARY SERVICE	40. REMARKS
41. NAME	42. DATE OF BIRTH	43. PLACE, STATE, ZIP CODE	44. RELATIONSHIP TO HEAD OF HOUSEHOLD	45. MARITAL STATUS	46. EDUCATION	47. OCCUPATION	48. INCOME	49. MILITARY SERVICE	50. REMARKS	51. NAME	52. DATE OF BIRTH	53. PLACE, STATE, ZIP CODE	54. RELATIONSHIP TO HEAD OF HOUSEHOLD	55. MARITAL STATUS	56. EDUCATION	57. OCCUPATION	58. INCOME	59. MILITARY SERVICE	60. REMARKS
61. NAME	62. DATE OF BIRTH	63. PLACE, STATE, ZIP CODE	64. RELATIONSHIP TO HEAD OF HOUSEHOLD	65. MARITAL STATUS	66. EDUCATION	67. OCCUPATION	68. INCOME	69. MILITARY SERVICE	70. REMARKS	71. NAME	72. DATE OF BIRTH	73. PLACE, STATE, ZIP CODE	74. RELATIONSHIP TO HEAD OF HOUSEHOLD	75. MARITAL STATUS	76. EDUCATION	77. OCCUPATION	78. INCOME	79. MILITARY SERVICE	80. REMARKS
81. NAME	82. DATE OF BIRTH	83. PLACE, STATE, ZIP CODE	84. RELATIONSHIP TO HEAD OF HOUSEHOLD	85. MARITAL STATUS	86. EDUCATION	87. OCCUPATION	88. INCOME	89. MILITARY SERVICE	90. REMARKS	91. NAME	92. DATE OF BIRTH	93. PLACE, STATE, ZIP CODE	94. RELATIONSHIP TO HEAD OF HOUSEHOLD	95. MARITAL STATUS	96. EDUCATION	97. OCCUPATION	98. INCOME	99. MILITARY SERVICE	100. REMARKS



**FIGURE IV-6. Control Card  
(Form CPS-260)  
(Back)**

FOR UNITS IN HOUSING VACANCY SURVEY (HVS)										Name and telephone number of interviewer responsible (including HVS)											
<p>Transcribe from HVS schedule the first month unit is vacant. Transcribe to HVS in subsequent months unit is vacant during same 6-month period of enumeration. If unit is vacant in both the first and second 6-month periods of enumeration, repeat above procedure for second 6-month period of enumeration.</p>																					
<p>10. Interviewer Check Item</p> <p>Is this housing unit on a piece of - 1st 2nd</p> <p>Less than 10 acres . . . . . O</p> <p>10 acres or more . . . . . O</p> <p>1970 or earlier . . . . . O</p> <p>1970 or later . . . . . O</p> <p>1965 to 1969 . . . . . O</p> <p>1960 to 1964 . . . . . O</p> <p>1950 to 1959 . . . . . O</p> <p>1940 to 1949 . . . . . O</p> <p>1930 to 1939 . . . . . O</p> <p>1920 or earlier . . . . . O</p>										<p>11. In what year was this structure built?</p> <p>1970 or later . . . . . O</p> <p>1965 to 1969 . . . . . O</p> <p>1960 to 1964 . . . . . O</p> <p>1950 to 1959 . . . . . O</p> <p>1940 to 1949 . . . . . O</p> <p>1930 to 1939 . . . . . O</p> <p>1920 or earlier . . . . . O</p>		<p>12. How many units are in this structure?</p> <p>(Count occupied and vacant)</p> <p>1st 2nd</p> <p>1 . . . . . O</p> <p>2 . . . . . O</p> <p>3 . . . . . O</p> <p>4 . . . . . O</p> <p>5 or more . . . . . O</p>		<p>13. How many rooms are in this unit (count the kitchen but not the bathroom)?</p> <p>1st 2nd</p> <p>1 . . . . . O</p> <p>2 . . . . . O</p> <p>3 . . . . . O</p> <p>4 . . . . . O</p> <p>5 or more . . . . . O</p>		<p>14. How many bedrooms are in this unit?</p> <p>1st 2nd</p> <p>None . . . . . O</p> <p>1 . . . . . O</p> <p>2 . . . . . O</p> <p>3 . . . . . O</p> <p>4 or more . . . . . O</p>		<p>FOR VACANT SEASONAL HOUSING UNITS AND FOR MIGRATORY WORKERS</p> <p>Transcribe from CPS-1 item 17 first month unit is vacant. Transcribe from Central Card to CPS schedule item 17 in subsequent months unit is vacant during same 6-month period of enumeration. If unit is seasonally vacant in both the first and second 6-month periods of enumeration, repeat above procedure for second 6-month period of enumeration.</p>			
<p>15. Is there hot and cold piped water inside structure for this unit?</p> <p>1st 2nd</p> <p>Yes . . . . . O</p> <p>No, only cold piped water . . . . . O</p> <p>No piped water . . . . . O</p>										<p>16. Is there a flush toilet inside structure for this unit?</p> <p>1st 2nd</p> <p>Yes, for this unit only . . . . . O</p> <p>Yes, but also used by a household in another unit . . . . . O</p> <p>No flush toilet . . . . . O</p>		<p>17. Is there a bathtub or shower inside structure for this unit?</p> <p>1st 2nd</p> <p>Yes, for this unit only . . . . . O</p> <p>Yes, but also used by a household in another unit . . . . . O</p> <p>No both tub or shower . . . . . O</p>		<p>CPS-1 Item 17.</p> <p>In this unit usually occupied?</p> <p>1st 2nd</p> <p>Summers only . . . . . O</p> <p>Winters only . . . . . O</p> <p>Other (Describe in notes) . . . . . O</p>							
<p>18. Is there hot and cold piped water inside structure for this unit?</p> <p>1st 2nd</p> <p>Yes . . . . . O</p> <p>No, only cold piped water . . . . . O</p> <p>No piped water . . . . . O</p>										<p>19. Is there a flush toilet inside structure for this unit?</p> <p>1st 2nd</p> <p>Yes, for this unit only . . . . . O</p> <p>Yes, but also used by a household in another unit . . . . . O</p> <p>No flush toilet . . . . . O</p>		<p>20. Is there a bathtub or shower inside structure for this unit?</p> <p>1st 2nd</p> <p>Yes, for this unit only . . . . . O</p> <p>Yes, but also used by a household in another unit . . . . . O</p> <p>No both tub or shower . . . . . O</p>		<p>CPS-1 Item 17.</p> <p>In this unit usually occupied?</p> <p>1st 2nd</p> <p>Summers only . . . . . O</p> <p>Winters only . . . . . O</p> <p>Other (Describe in notes) . . . . . O</p>							

TABLE X - LIVING QUARTERS DETERMINATIONS AT LISTED ADDRESS									
No.	LOCATION OF UNIT	USE OR CHARACTERISTICS				ALL QUARTERS		CLASSIFICATION	
		1. If inside Area Segment boundary, mark box below, STOP and - Go to next line of Table X, if additional quarters determined, or - Continued with interview for original unit	2. If outside Area Segment boundary, mark box below, STOP and - Go to next line of Table X, if additional quarters determined, or - Continued with interview for original unit	3. If inside Area Segment boundary, mark box below, STOP and - Go to next line of Table X, if additional quarters determined, or - Continued with interview for original unit	4. If outside Area Segment boundary, mark box below, STOP and - Go to next line of Table X, if additional quarters determined, or - Continued with interview for original unit	5. Do these quarters in (Specify location) have sleeping facilities (Specify location) through a common hall?	6. Do these quarters in (Specify location) have sleeping facilities (Specify location) through a common hall?		
(1)		(13)	(14)	(15)	(16)	(17)	(18)	MU } Separate unit-Interview on a separate Central Card (9)	
1		5	1	Yes	No	Yes	No	N HU OT	
2		5	1	Yes	No	Yes	No	N HU OT	
3		5	1	Yes	No	Yes	No	N HU OT	

NOTE: BE SURE TO CONTINUE INTERVIEW FOR ORIGINAL UNIT AFTER COMPLETING TABLE X FOR ALL LINES.

NOTES

19. CONTROL NO. OF ORIGINAL SAMPLE UNIT

20. In case of separate unit, list as many as possible in separate listing sheet.

Sheet - Line

**FIGURE IV-7. Current Population Survey  
(Form CPS-1)  
(Front)**

[illegible]

(Form CPS-1)  
(Back)

[illegible]

TABLE IV-1. Number and Distribution of Total, Noninterview and Interview, Units in A- and C-Design CPS Sample

(Monthly average: 1975)

Line No.	Sample <sup>1</sup>	Number of units <sup>2</sup>	Percentages	
			A	B
1.	Total sample units designated .....	57.667	100.0	(X)
2.	Type C noninterviews detected in previous interview months .....	2.589	4.5	(X)
3.	Units assigned for interview (1-2) .....	55.078	95.5	(X)
4.	Type C noninterviews detected in current interview month .....	480	.8	(X)
5.	Type B noninterview .....	7.589	13.2	(X)
6.	Occupied units (households) eligible for interview (3-4-5) .....	47.009	81.5	100.0
7.	Type A noninterview, total .....	1.982	(X)	4.2
8.	No one home .....	(NA)	(X)	.9
9.	Temporarily absent .....	(NA)	(X)	.9
10.	Refusal .....	(NA)	(X)	2.2
11.	Other, occupied .....	(NA)	(X)	.2
12.	Interviewed households (6-7) .....	45.027	(X)	95.8
13.	Total persons interviewed, age 14 years old and over .....	99.281	(X)	(X)

NA Not available.

X Not applicable.

<sup>1</sup>See text, p. 33, for definition of noninterviews, by type<sup>2</sup>Housing units and group quarters listing units. figures include an additional sample of about 1,800 households with Spanish head interviewed in March

month. The results from the interviewed sample require adjustment for these noninterviews to reflect the entire population more adequately. The noninterview adjustment (described in ch. V) requires that the race of the household head and the farm-nonfarm status of the unit be obtained.

Type B noninterviews comprise vacant units, vacant sites for tents or mobile homes or units occupied by persons ineligible for the survey. Type B noninterviews are visited each month to interview any that may have become occupied by persons eligible for the survey. About 13 percent of the units designated for the sample in a given month are type B noninterview. The main categories are—

1. Nonseasonal regular housing units that are vacant, temporarily occupied by persons with usual residence elsewhere or used for storage of household furniture. These are also reported on a Housing Vacancy Survey questionnaire, form HVS-1. (See fig. IV-8.)
2. Units unfit for occupancy or to be demolished, those under construction, converted to temporary business or storage, occupied by members of the armed forces or solely by persons under age 14, unoccupied tent or trailer sites, or units for which a building permit has been issued but construction has not started.

Type C noninterviews are units demolished, converted to permanent storage or business use, moved from site, or found

to be in sample by mistake. Once detected, these units are not visited again by the CPS interviewer. The number of type C noninterviews, relative to the total designated sample, increases with the age of the sample design.

Complete the CPS questionnaire—For each household and for each civilian household member 14 years old and over, the interviewer completes a CPS questionnaire that asks the household respondent a series of structured questions concerned with economic activity during the week containing the 12th day of the month. This week, referred to as the *survey week*, is the week preceding interview week. These questions appear as items 19-24 on the CPS questionnaire. (See fig. IV-7.) The primary purpose of these questions is to classify the sample population into three basic groups—the employed, the unemployed, and those not in the labor force. Interviewers are trained to ask the questions as they appear on the questionnaire.

Some examples of the information collected include—for the employed, the hours worked during the survey week together with a description of the current job and, for those temporarily away from their jobs, the reason for not working during the survey week. For the unemployed, information is obtained on the length of time looking for work and a description of the previous job. For those outside the labor force, the principal activity during the survey week, whether keeping house, going to school, or doing something else, is recorded. For one-fourth of the respondents reported as not in the labor force (that is, for two rotation groups), the respondents are also asked when they last

TABLE IV-2. Current Population Survey National Sample Type A Noninterview Rates, . . .  
by Reason: 1975-1976, by Month; Annual Rates for 1972-1976

Month	Type A noninterview rate, by reason <sup>1</sup>				
	Total	No one home	Temporarily absent	Refusal	Other
1976 .....	4.4	0.8	0.8	2.6	0.2
January .....	4.4	.9	.6	2.6	.3
February .....	4.7	.9	.7	2.8	.3
March .....	5.2	1.0	.8	3.1	.2
April .....	4.6	.9	.6	2.9	.2
May .....	4.6	.9	.6	3.0	.2
June .....	4.7	.8	1.0	2.6	.2
July .....	4.7	.9	1.3	2.3	.2
August .....	4.8	.8	1.3	2.4	.2
September .....	4.2	.8	.7	2.4	.3
October .....	3.8	.7	.5	2.4	.2
November .....	4.0	.9	.5	2.4	.2
December .....	3.7	.8	.5	2.3	.2
1975 .....	4.2	.9	.9	2.2	.2
January .....	3.4	.7	.6	1.8	.2
February .....	3.8	.8	.7	2.1	.2
March .....	4.6	.9	.9	2.5	.3
April .....	4.0	.8	.6	2.4	.2
May .....	4.4	1.0	.9	2.3	.2
June .....	4.5	.8	1.2	2.2	.3
July .....	4.7	.9	1.4	2.2	.2
August .....	4.9	1.0	1.5	2.2	.2
September .....	4.1	.9	.8	2.1	.2
October .....	3.9	.8	.7	2.2	.3
November .....	4.0	1.0	.6	2.2	.2
December .....	4.5	1.0	.8	2.5	.2
1974 .....	4.1	.9	.9	2.0	.3
1973 .....	4.3	1.1	1.0	1.9	.4
1972 .....	4.0	1.0	.0	1.8	.3

<sup>1</sup>Percent of eligible sample households noninterview.

worked, the reasons for leaving work, and, for those who worked during the past 5 years, a description of the last job is also recorded.

**Labor force concepts**—The concepts of the labor force, employment, and unemployment were introduced into the field of labor market analysis in the late 1930's. Earlier measurements dealt with the concept of the gainful worker [44, p. 32].

The presently used concepts are intended to be as objective as possible in an imprecise field and depend principally upon each person's actual activity, such as working, looking for work, or doing something else within a designated time period[70]. The official measures relate to persons 16 years old and over, although separate data are published for 14 and 15 year olds. The

criteria used in classifying persons on the basis of their activity are described in the following.

**Employed persons** comprise (1) all civilians who, during survey week, did any work at all as paid employees or in own business, or on their own farm or who worked 15 hours or more as unpaid workers on a farm or in a business operation, a member of the family, and (2) all those who were not working but had jobs or businesses from which they were temporarily absent because of illness, bad weather, vacation, or labor-union dispute, or because they were taking time off for personal reasons.

**Unemployed persons** are those civilians who had no employment during the survey week, were available for work

[illegible]

(1) had engaged in any specific jobseeking activity within the past four weeks, (2) were waiting to be called back to a job from which they had been laid off, or (3) were waiting to report to a new wage or salary job scheduled to start within the following 30 days.

The *civilian labor force* consists of all civilians classified as employed or unemployed. The *total labor force* comprises the civilian labor force plus members of the armed forces stationed in the United States or abroad.

All persons not classified as employed, unemployed, or in the armed forces are defined as *not in the labor force*.

**Basic and reimbursable programs**—Additional supplementary questions are usually asked as part of the CPS interview each month. (App. I shows the programs conducted during calendar years 1975 and 1976.)

In March of each year, the CPS interviewer asks additional questions on migration, work experience, and income; in May, on multiple job holding and usual weekly earnings; and in October, on school enrollment. These data, together with the labor force and other demographic information regularly supplied by the CPS, provide valuable information on changes in the American economic and social structure. The joint funding provided by Census Bureau and the Bureau of Labor Statistics to conduct the CPS includes these additional studies that are considered part of the basic CPS program.

Other supplementary questions, dealing with a wide variety of subject matter, are included as space and time permit. Some of these are one-time additions others are added periodically. These questions are included, at cost, and are sponsored by the Government as well as private study organizations. (See ch. IX.)

**Personal and telephone interviews**—Interviewers obtain information, by personal visit, or by telephone. The rules governing telephone use for interviews were originally established in 1954; the rules are that personal interviews are required for

first, second, and fifth month in sample households, all other units may be interviewed by telephone, provided the respondent has previously been contacted by personal interview and is given permission for telephone interviews. The telephone is used to complete a noninterview in the second month after personal interview has been attempted.

The proportion of interviews completed by telephone has increased, reflecting the greater availability of telephones in private households and the pressures to decrease automobile usage because of the energy crisis. (Table IV-3 illustrates change in the proportion of telephone interviews in selected months of recent years.) Although the rules do not encourage interviewers use the telephone in a limited fashion in the first and fifth months, largely to complete noninterviews. Interviewers have also increased telephone interviews in the second month. The evidence currently available indicates that telephone use does not materially affect the results.

Before instituting the telephone interview procedure, a test was conducted to determine whether the quality of results obtained by telephone differed in any significant way from results obtained by direct interview. The results of that test, conducted in a limited number of PSU's in the early 1950's, indicated there did not appear to be material differences in the labor force data obtained by the two methods of interviewing. The results of the reinterview operation (described on p. 46), available at the time, were also studied to determine differences between errors of response in telephone interviews versus personally interviewed households. Neither of these tests was a completely controlled experiment, since the PSU's involved in the first test were not randomly chosen, and the reinterview studies were based on tabulations of households in the reinterview sample that were interviewed originally by telephone or by personal interview. The choice of the method used for the original interview was not randomly assigned.

TABLE IV-3. Percent of Interviews Completed by Telephone,  
by Month in Sample, Selected Months

(National sample)

Month in sample	Percent of interviews completed by telephone <sup>1</sup>					
	July 1969	December 1970	December 1973	December 1974	December 1975	December 1976
<i>All months</i> .....	45.5	49.9	58.1	58.7	58.7	59.2
1 .....	2.9	3.7	3.6	3.7	3.7	3.0
2 .....	35.6	45.3	51.9	51.0	49.8	48.9
3 .....	63.1	67.2	79.4	79.5	80.0	81.3
4 .....	63.0	67.4	80.5	79.5	81.2	82.0
5 .....	10.6	11.9	12.6	13.7	12.2	12.6
6 .....	58.5	68.0	76.5	76.0	78.1	79.1
7 .....	64.8	66.1	81.5	80.7	81.2	82.8
8 .....	65.1	68.1	80.5	82.3	80.9	82.6

<sup>1</sup>Includes telephone callbacks where at least one personal visit has been made to the household, but an eligible respondent was not available.

Table IV-4 shows a tabulation of the reinterview sample, by method of original interview, for July 1974-June 1975. The table shows the percentage of households having identical responses on both the original and the reinterview visit, by method of original interview. The table also shows how the population interviewed, by the two methods, is distributed. It must be remembered that the choice of the method of original interview is influenced by the respondent's willingness to be interviewed by telephone. (See the reinterview sample on p. 46.)

### CONTROL OF THE DATA COLLECTION OPERATIONS

The data collection activities for a survey month involve a substantial number of participants, and much of the work is performed by personnel out of immediate reach of direct supervision. As the nature of survey work requires that the various steps must be performed in a specific sequence, the impact of deviations from correct procedure becomes cumulative. Moreover, the time allotted for monthly survey activities is so restricted that there is little time available for correcting or repeating incorrect work. For these reasons, the procedures include a number of checks to control the quality and progress of the survey operation.

### Questionnaire Edits

The information recorded on the control card and questionnaires is reviewed for completeness and accuracy to insure that instructions are being followed, insofar as can be determined, and to detect obvious errors and omissions. All interviewers systematically review their own work before transmitting it to the regional office. The regional office reviews all work received from relatively new interviewers and from experienced interviewers who have not maintained a prescribed level of quality as measured by the review. Finally, during the processing of the information in Washington, the computer edits all questionnaires for consistency and completeness and, for each interviewer, provides a listing of all omissions and inconsistencies detected by questionnaire item number.

The regional office edit checks that questionnaires have been returned for each designated sample unit, that FOSDIC markings are properly made, that correct forms have been filled and that questionnaires are in good condition. In addition, as indicated below, some of the CPS questionnaires are reviewed and errors found are corrected before being sent on to Jeffersonville for further processing.

The interviewer's total error rate (errors detected in the regional office as well as by the computer edit) is computed reflecting the number of errors per questionnaire page filled (one page

TABLE IV-4. Percent of Persons in Given Labor Force Status in CPS Reinterview After Reconciliation, and Percent Identically Classified in the Original Interview, by Method of Interview: July 1974-June 1975

(Monthly average)

Item	Method of interview			
	Personal		Telephone	
	Distribution	Responses identical (percent)	Distribution	Responses identical (percent)
<i>Total reinterview sample (persons aged 14 years old and over) . . . . .</i>	<i>17,554</i>	<i>96.7</i>	<i>24,617</i>	<i>97.1</i>
<i>Total distribution . . . . .</i>	<i>100.0</i>	<i>(X)</i>	<i>100.0</i>	<i>(X)</i>
Employed . . . . .	51.5	96.8	55.8	97.2
Agriculture . . . . .	3.1	97.0	2.1	94.9
Nonagriculture . . . . .	48.4	97.2	53.8	97.3
At work . . . . .	45.5	97.2	50.2	97.8
Full time . . . . .	33.6	97.6	37.9	98.4
Part time . . . . .	11.9	96.1	12.3	96.0
Not working . . . . .	2.9	90.4	3.5	90.7
Unemployed . . . . .	4.7	95.1	3.8	94.8
Not in labor force . . . . .	43.8	96.8	40.4	97.1

X Not applicable.



is filled for each household and one page for each person in the household). Prior to the next month's enumeration, each interviewer is notified of his error rate and given a listing of the errors found by the regional office and the computer edit. An interviewer having excessive error rates is scheduled for a special needs observation (see p. 50) and retrained in the editing of completed questionnaires. Periodically, edit procedures are reviewed in interviewer group training sessions so that all interviewers are kept up to date.

The *qualified edit (QE)* system is devised to take advantage of the computer's ability to review and edit a large amount of data quickly and allows regional office personnel to devote more time to a thorough edit of the questionnaires from new or less qualified interviewers. Currently, an interviewer maintaining errors at less than seven and one-half per hundred completed pages (i.e., an error rate of 7.5 or less) for 3 consecutive months achieves QE status. If the error rate exceeds 7.5 for 3 consecutive months or exceeds 12.5 in any 1 month, the interviewer loses QE status. Work for QE interviewers is not subject to an item-by-item review by the regional office. However, all work by all interviewers is completely edited by the computer. Approximately 78 percent of all interviewers are on QE status in any given month. For most months, the average error rate for all interviewers does not exceed five and one-half errors per hundred pages.

When supplemental inquiries are added to the CPS, the regional office edits all of the supplemental data for nonqualified edit interviewers and for a specified number of interviewed households (usually 10) for each QE interviewer. The computer edit does not include errors in supplemental inquiries in the tabulation of interviewer errors.

Table IV-5 shows a frequency distribution of the average monthly interviewer errors per 100 pages filled. The rates include all errors from the computer and from the regional office edit.

#### Noninterview Rate

The computer prepares the type A noninterview rate for all work of the interviewer in each month. A rate in excess of 5 percent of the occupied eligible households in sample is considered below standard. Table IV-6 provides a distribution of the interviewer type A noninterview rates for the period July 1974 through June 1975.

#### Address Coverage Check

Each month the regional office reviews copies of the listing sheets for segments or addresses that have been listed and prepares them for further operations. One copy of each address listing sheet is used for sampling the listing and is returned to the interviewer for conducting the interview. A second copy of each address listing sheet is sent to Jeffersonville for a second review to identify unexplained differences between the number of units expected and the number found in the listing, evidence of inappropriate listing patterns which might reflect operational biases, accuracy in extending or modifying the sample pattern when required, etc. The regional office is advised of corrective action required by this review.

Sample reports are reviewed in Washington to insure that the regional office has properly applied the sampling instructions in area segments and NTA places in special place segments.

TABLE IV-5. Distribution of CPS Interviewer Questionnaire Edit Error Rates, National Sample: July 1974-June 1975

(Monthly average)

Edit error rate and rating for interviewer <sup>1</sup>	Interviewers (percent)
<i>Total interviewer assignments</i> <sup>2</sup> 992 .....	100.0
Interviewers rated .....	98.6
<i>Excellent</i> .....	48.3
0 to 0.4 .....	13.9
0.5 to 1.4 .....	19.6
1.5 to 2.4 .....	14.8
<i>Satisfactory</i> .....	32.0
2.5 to 3.4 .....	10.4
3.5 to 4.4 .....	7.5
4.5 to 5.4 .....	6.0
5.5 to 6.4 .....	4.6
6.5 to 7.4 .....	3.5
<i>Needs improvement</i> .....	10.1
7.5 to 8.4 .....	3.0
8.5 to 9.4 .....	2.5
9.5 to 10.4 .....	1.7
10.5 to 11.4 .....	1.6
11.5 to 12.4 .....	1.3
<i>Unsatisfactory</i> .....	8.2
12.5 to 14.4 .....	1.6
14.5 to 16.4 .....	1.3
16.5 and over .....	5.1

<sup>1</sup>Number of errors per 100 questionnaire pages

<sup>2</sup>Average number of interviewer assignments per month

#### Reinterview

A continuing program of reinterview on subsamples of households is carried out every month during the week following the week of interview (43: 60). The reinterview procedures designed to provide the regional office supervisor and the Washington staff with knowledge of the specific types of differences and errors made by interviewers in listing and in interview. This allows some determination of how misunderstandings on the part of interviewers or respondents or inadequacies in materials or shortcomings in the procedures.

The major objective in any program of quality control is to keep the overall rate of error within some predetermined acceptance numbers. In the reinterview program, this objective is achieved in two ways. First, since the reinterview pattern is random rather than systematic, the interviewers know their work may be checked at any time; this is believed to have a powerful effect. Second, the program detects enumerators with below standard performance who require more training.

**TABLE IV-6. Distribution of CPS Interviewer Type A Noninterview Rates, National Sample: July 1974-June 1975**

(Monthly average)

Noninterview rate and rating for interviewer <sup>1</sup>	Interviewers (percent)
<b>Total interviewer assignments<sup>2</sup> 992</b> .....	<b>100.0</b>
<b>Interviewers rated<sup>3</sup></b> .....	<b>98.6</b>
<b>Excellent</b> .....	<b>51.5</b>
0 to 0.9 .....	26.8
1.0 to 1.9 .....	9.1
2.0 to 2.9 .....	15.6
<b>Satisfactory</b> .....	<b>21.1</b>
3.0 to 3.9 .....	11.4
4.0 to 4.9 .....	9.7
<b>Needs improvement<sup>4</sup></b> .....	<b>16.8</b>
5.0 to 5.9 .....	7.3
6.0 to 6.9 .....	5.6
7.0 to 7.9 .....	3.9
<b>Unsatisfactory</b> .....	<b>9.2</b>
8.0 to 8.9 .....	2.7
9.0 to 9.9 .....	1.6
10.0 to 10.9 .....	1.5
11.0 and over .....	3.4

<sup>1</sup> Percent of occupied eligible households noninterview

<sup>2</sup> Average number of interviewer assignments per month

<sup>3</sup> Interviewer must have 6 months' CPS experience to be rated

<sup>4</sup> On basis of local conditions (e.g., bad weather) or if interviewer is directed to complete another interviewer's assignment with large number of unconfirmed refusals, the regional director may accept a type-A rate in excess of 5 percent as satisfactory

**Reinterview sample**—The monthly reinterview sample covers about 1 in 18 of the units in a monthly CPS sample and comprises about one-third of the USU's in about one-sixth of the interviewer assignments. The sample is selected in the regional offices after the interviewer assignments for the month have been determined. Each month, the PSU's in which reinterviews are to be conducted and the method of choosing the interviewers to be reinterviewed is specified by Washington. Care is exercised to avoid disclosing the interviewers or the households to be checked in advance of the original interview.

The reinterview is performed by a reinterviewer who makes contact by a personal visit or by telephone at the designated sample households within the reinterview segment. About 80 percent of the reinterviews are conducted by senior interviewers

who are thoroughly trained on CPS procedures and on conducting the reinterview and reconciliation. It is assumed they can list and interview better than the initial interviewer and thus their work can serve as a norm for judging the work of individual interviewers. The remaining reinterview (about 20 percent) is performed by the CPS supervisors. The reinterview is conducted by telephone, when feasible, at all reinterview units aside from the exceptions cited below. Currently, about 80 percent of the households in reinterview are contacted by telephone.

**The components of the reinterview**—The three main parts of the reinterview are the listing and household coverage checks, noninterview classification checks, and content reinterview and reconciliation.

1. The listing check consists of canvassing the area segments (or structures in other types of segments) to determine whether units found in the segment or at the address appear on the listing sheet. This is a check for omissions, duplicate listings, and errors in identifying residential units. For the household coverage check, conducted for each unit occupied at the time of reinterview, the roster of household members on the control card is checked for accuracy and completeness. The listing check for an area segment requires the reinterviewer to visit the segment.
2. Units originally reported as noninterview are checked for classification errors, e.g., a unit originally reported as type B, vacant noninterview which was, in fact occupied by a household temporarily absent during interview week should have been reported as a type A noninterview. Units reported as type C noninterview have their classification checked by a visit to the unit.
3. Content reinterview is conducted for all eligible household members in the reinterview sample using a questionnaire having content items identical to those used on the original interview. Whenever possible, the reinterview is conducted with the person for whom the information is being obtained; the second choice is the original respondent. If neither of these respondents is available, any knowledgeable household member is accepted.

**Conditioning of responses**—It would be desirable to have the responses at the original and the reinterview visits obtained under the same circumstances, i.e., have the reinterview independent of the original. However, this is not possible as the reinterview respondent, or at least the respondent household, has been conditioned in some degree by the original visit. The conditioning has some effect on comparisons of the results of the reinterview with the original responses.

To determine the causes of differences which occur between the original and the reinterview responses, the respondent is informed of differences, if any, in the two answers recorded for the sample person. Differences are reconciled on a form containing the original responses for the household. The reinterviewer has the original responses in his possession during the reinterview but is instructed not to consult them until the respondent has been reinterviewed lest the reinterviewer's questioning process also be conditioned. An indication of how the reinterviewer's behavior is influenced by the presence of the original interviewer's results is made possible by withholding the original information for a 20-percent sample of the reinterview house-

holds. This permits comparison of the original vs. reinterview separately for the 80-percent sample (where the reinterviewer has the original responses) and for the 20-percent sample (where the reinterviewer does not have the original responses).

**Origins of coverage error**—The reinterviewer discusses the results of the reinterview assignment with the interviewer. To evaluate these results, it is necessary to consider the origins of coverage error:

1. Erroneous omissions of units occur when the original lister fails to canvass an area segment thoroughly or to verify the listing in other types of segments, from missing units in multiunit structures, and from recording multiunit structures as single-unit structures.
2. Erroneous omissions of persons arise from omission of units, from failure to list usual residents on the control card, from treating an occupied unit as vacant, and from misclassifying a household member as a nonmember in applying the definition of usual residence. Most of these latter cases involve persons with loose attachment to the household, i.e., persons who spend part of their time elsewhere or persons who, for their own reasons, do not wish to be recorded as living within a particular household.
3. Erroneous inclusion of units occurs from listing units located outside area segment boundaries or from reporting a single unit as two units through faulty application of the housing unit definition.
4. Erroneous inclusion of persons arises from errors in the listing of units and in treating persons with a usual residence elsewhere as household members of the sample unit.

**Reinterview results to control and measure quality**—The results of the reinterview program are used to control quality and to measure quality. (Further information, relative to quality measurement, is given in ch. VII.) The quality control procedures are designed to provide the field staff with feedback on the types of differences and errors made by interviewers in listing and interviewing. This permits some determination of the nature of misunderstanding by the interviewer or the respondent and problem areas in the materials or procedures used.

The most important measure of quality of coverage is the net coverage error in the count of persons of all ages for CPS as a whole. *Net coverage error* is defined as the number of erroneous inclusions minus the number of erroneous omissions. The net coverage error rate is the ratio of the net coverage error to total population of all ages as measured by reinterview. The estimated net coverage error rates are given in table IV-7 for 1974 and 1975, by month.

The principal index used to control the quality of coverage, interviewer by interviewer, is the "interviewer gross coverage error rate." *Interviewer gross coverage error* is obtained from the sum of the erroneous omissions and inclusions. The distribution of estimated monthly gross coverage error rates, interviewer by interviewer, for the 12-month period ending June 1975 is given in table IV-8. As estimated by reinterviews, less than 4 percent of the interviewers accounted for about half of the errors in coverage, and 10 percent of the interviewers accounted for about 87 percent.

The results of the measurements of content error are summarized in the same manner as for coverage error. Table IV-9 shows the results for gross content errors for the period July

TABLE IV-7. Rates of Net Coverage Error in the Count of Persons in the Current Population Survey National Sample, by Month: 1974-1975

Month	Net error as percent of total persons in reinterview <sup>1</sup>
<b>1975</b>	
January .....	- 22
February .....	- 34
March .....	- 04
April .....	- 22
May .....	- 16
June .....	- 10
July .....	- 10
August .....	- 16
September .....	-
October .....	- 33
November .....	- 15
December .....	- 09
<b>1974</b>	
January .....	- 42
February .....	- 29
March .....	- 06
April .....	- 19
May .....	- 35
June .....	- 07
July .....	- 04
August .....	- 13
September .....	- 40
October .....	- 25
November .....	- 30
December .....	- 19

— Entry represents zero.

<sup>1</sup>Net coverage error as a percent of total persons measured by reinterview, minus sign indicates net undercount

1974 through June 1975. Although not so concentrated as coverage errors, a small proportion of interviewers (perhaps percent) account for about half of the content errors.

The results of the quality check program identify interviewers with gross-content or gross-coverage error rates above established tolerances. The tolerances in effect as of June 1 are shown in table IV-10

### Production Standards

Production standards have been developed to assist in limiting costs within budget, to maintain an acceptable level of efficiency in the program, to help analyze activities of individual interviewers, and to assist the supervisor in identifying interviewers needing corrective action.

TABLE IV-8. Distribution of Interviewer Monthly Gross Coverage Error Rates in the Current Population Survey National Sample: July 1974-June 1975

Gross coverage error rate (percent)	Interviewers checked		Gross errors	
	Number	Cumulative percent	Number	Cumulative percent
<i>Total</i> .....	1,460	—	762	—
0 .....	1,243	100.0	—	100.0
0.1 to 0.9 .....	21	14.8	30	100.0
1.0 to 1.9 .....	54	13.4	79	96.1
2.0 to 2.9 .....	31	9.7	54	85.7
3.0 to 3.9 .....	28	7.6	82	78.6
4.0 to 4.9 .....	21	5.7	81	67.8
5.0 to 5.9 .....	12	4.3	39	57.2
6.0 to 6.9 .....	9	3.5	32	52.1
7.0 to 7.9 .....	5	2.9	29	47.9
8.0 to 8.9 .....	2	2.5	2	44.1
9.0 to 9.9 .....	7	2.4	54	43.8
10.0 to 14.9 .....	9	1.9	46	36.7
15.0 to 24.9 .....	10	1.3	96	30.7
25.0 and over .....	8	.6	138	18.1

— Entry represents zero

TABLE IV-9. Distribution of Interviewer Gross Content Error Rates in the Current Population Survey National Sample: July 1974-June 1975

Gross coverage error rate (percent)	Interviewers checked		Gross errors	
	Number	Cumulative percent	Number	Cumulative percent
<i>Total</i> .....	2,004	—	4,012	—
0 .....	805	100.0	—	100.0
0.1 to 0.9 .....	105	59.8	116	100.0
1.0 to 1.9 .....	309	54.6	465	97.1
2.0 to 2.9 .....	234	39.2	559	85.5
3.0 to 3.9 .....	142	27.5	496	71.6
4.0 to 4.9 .....	122	20.4	539	59.2
5.0 to 5.9 .....	89	14.3	447	45.8
6.0 to 6.9 .....	69	9.8	410	34.6
7.0 to 7.9 .....	35	6.3	233	24.4
8.0 to 8.9 .....	29	4.6	221	18.6
9.0 to 9.9 .....	10	3.2	89	13.1
10.0 to 14.9 .....	43	2.7	345	10.9
15.0 to 24.9 .....	8	.6	61	2.3
25.0 and over .....	4	.2	31	.8

— Entry represents zero

**TABLE IV-10. Tolerance Limits for Gross Coverage and Content Error in the Current Population Survey: June 1975**

Number of units <sup>1</sup>	Acceptable number of errors
1 to 10 .....	1
11 to 20 .....	2
21 to 30 .....	3
31 to 40 .....	4
41 to 50 .....	5
51 to 60 .....	6
61 to 70 .....	7
71 to 80 .....	8
81 to 90 .....	8
91 to 100 .....	9
101 to 150 .....	12
151 to 200 .....	15
201 to 250 .....	18
251 to 300 .....	22
301 and over .....	23

<sup>1</sup>Number of opportunities for error. For listing, coverage and household composition checks, this is number of housing units originally reported in the reinterview assignment. For labor force content check, this is five times the number of persons 14 years old and older originally reported in the 80-percent (reconciled) portion of the reinterview assignment.

The standards used in the program are designed to measure the efficiency of individual interviewers and the regional office field functions. Efficiency is measured as the ratio of the standard allowable time per assignment to the actual interview time. The standard allowable time for interviewers varies with the characteristics of the PSU and the total number of cases in the workload.<sup>6</sup> As the interviewers' workload is fairly constant, the standard allowable time is known by the interviewer. An interviewer efficiency ratio of less than 0.8, based on the work of one quarter, is considered substandard, and the CPS supervisor takes appropriate remedial action. In the same way, standards are observed for an entire regional office in order to permit the regional office staff to judge whether their activities are under control.

The system helps to identify interviewers producing at a low rate of efficiency because they are not exerting enough effort or because of inefficient procedures. On the other hand, high efficiency ratios may indicate interviewers working extremely rapidly or taking shortcuts in doing the work. Therefore, some checking is also in order in the case of highly unusual ratios.

<sup>6</sup>Several hypothetical cost functions have been developed to study the cost relationships of the various field data collection activities (e.g., number of segments in the interview assignment and number of households per USU, distance between segments and between the interviewer's home to a segment, times required for travel between and within segments, times required for listing and for interviewing a household, etc.). These studies have influenced the planning and supervision of the field activities; one of the functions (dating from 1956), derived by Webber and Silver at the Census Bureau, is summarized in [31, p. 91-94].

Table IV-11 shows a distribution of interviewer production ratios for the period July 1974-June 1975.

### Observation of Field Work

Field observation is one of the methods used by the supervisor to check and improve performance of the interviewer staff. There are three situations calling for an observation: Observation is an integral part of the initial training given to all interviewers (see app. J); each interviewer is regularly observed once each year; *special needs observations* are made where there is evidence of interviewer problems or poor performance detected by other checks on the interviewer's work. Most observation is of the latter type and is generated by interviewers having a high questionnaire edit error rate (more than 7.5), a high type A noninterview rate (more than 5 percent), a low production ratio (less than 0.8), a failure on reinterview, having an unsatisfactory report on a previous observation, a request for help, or for other reasons relative to the interviewer's performance.

An observer accompanies the interviewer for a minimum of 4 hours during an actual work assignment. The observer takes notes

**TABLE IV-11. Distribution of CPS Interviewer Production Ratios: July 1974-June 1975**  
(Monthly average)

Production ratio and rating for interviewer <sup>1</sup>	Interviewers (percent)
<i>Total interviewer assignments<sup>2</sup> 992 .....</i>	<i>100.0</i>
Interviewers rated .....	88.3
<i>Excellent<sup>3</sup> .....</i>	<i>35.1</i>
1.50 or more .....	3.1
1.40 to 1.49 .....	2.8
1.30 to 1.39 .....	5.5
1.20 to 1.29 .....	9.7
1.10 to 1.19 .....	14.0
<i>Satisfactory .....</i>	<i>45.7</i>
1.00 to 1.09 .....	17.5
0.90 to 0.99 .....	16.4
0.80 to 0.89 .....	11.8
<i>Needs improvement .....</i>	<i>5.1</i>
0.70 to 0.79 .....	5.1
<i>Unsatisfactory .....</i>	<i>2.4</i>
Less than 0.70 .....	2.4

<sup>1</sup>Ratio of standard allowable time per assignment to the actual interviewer time.

<sup>2</sup>Average number of interviewer assignments per month.

<sup>3</sup>If checking shows no short-cuts are being taken.

of the interviewer's performance including how the interview is conducted and how the survey forms are completed. The observer stresses good interviewing techniques: Asking questions as worded and in the order presented on the questionnaire, adhering to instructions on the forms and in the manuals, knowing how to probe, recording answers in the correct manner and in adequate detail, developing and maintaining good rapport with the respondent conducive to an exchange of information, avoiding questions or probes that suggest a desired answer to the respondent, and determining the most appropriate time and place for the interview.

The observer reviews the interviewer's household-by-household performance and discusses the interviewer's strong and weak points with an emphasis on correcting habits which interfere with the collection of reliable statistics. In addition, the interviewer is encouraged to ask the observer to clarify survey procedures not fully understood and to seek the observer's advice on solving other problems encountered.

### Evaluation of the Interviewers' Work

At the end of each quarter, the interviewer is evaluated on the basis of scores on the questionnaire edits, type A noninterview rates, and production ratios. The interviewer output, as reflected by these checks, is actually under review each month so that the supervisor can take immediate action to correct substandard work. However, the quarterly evaluation requires action by the supervisor if the interviewer is rated as "needs improvement" or "unsatisfactory." For interviewers rated "needs improvement," the supervisor is to conduct a special needs observation. Occasionally, the supervisor may determine that the low rating is caused by events beyond the interviewer's control and the observation is omitted; for example, the interviewer may have had a high noninterview rate because of an emergency assignment to complete the work of an interviewer who resigned. If an interviewer is rated "unsatisfactory," in addition to a special needs observation, the interviewer is given a letter warning that the work is substandard and that the interviewer is subject to separation if the work does not show improvement. For "unsatisfactory" cases, the regional office director must send the warning letter to the interviewer or inform the chief of the Field Division in Washington of the special reasons the letter has not been sent.

In general, the supervisor schedules an observation with the interviewer as a first response to evidence the interviewer is producing substandard work. Additional training directed at the problem area is given during the observation. Further training, coupled with observation, is usually required before the next regular assignment if the interviewer fails a reinterview check.

If conferences on specific problem areas, observation, and training do not produce improvement to expected minimum standards, the interviewer is separated.

Table IV-12 summarizes the quarterly evaluation of the interviewers in the national sample averaged over the four quarters of 1974.

## PROCESSING THE DATA

Information collected for the CPS in a given month moves through several stages: From the interviewer, to the appropriate

regional office, to the Data Preparation Division in Jeffersonville Ind., and to the Census Bureau in Washington.

The process moves information through the system on a flow basis. Each day during the processing period, the interviewer, the regional office and Jeffersonville send forward all work completed for the day. Work received daily at the Census Bureau in Washington is moved through an early computer process to insure the information can be read by the computer and to identify possible problem areas in the collection or processing system. The following summarizes the major data processing operation at each location.

### The Interviewer

Interviewers begin contacting their assigned units on Monday of interview week. All of the assignment is to be completed by the following Saturday. New sample units and units known from earlier months to be difficult are contacted in the early portion of the week to give more time to complete type A noninterviews. For some noninterview units, the interviewer may arrange to contact the respondent as late as Tuesday of the following week and telephone the information to the regional office.

### The Regional Office

**Control of materials**—Daily shipments from each interviewer are checked in and compared with the expected units in the assignment. A fairly precise count of the expected units is known for each assignment, since 75 percent of the segments were contacted the previous month and good estimates of units in the remainder are available from census records or subsequent listings. Extra units turned up at the time of interview are separately reported. Each interviewer must account for all units in the assignment.

**Edit of questionnaires and control cards**—All questionnaires and control cards from interviewers not on QE status are edited. (See p. 45.) Errors suggesting omission of sample unit or of persons are dealt with immediately by telephoning the interviewer.

**Prepare for reinterview**—Information from questionnaire for units in the reinterview sample is transcribed to the reinterview forms used in the reconciliation. (See p. 47.)

### Data Processing Division, Jeffersonville, Ind.

**Control of materials**—The CPS questionnaires received from the regional offices are checked in by comparing the count of documents in the shipment with the number entered on transmittal by the regional office. The transmittals report the count of documents, by PSU, in each shipment. The regional offices occasionally send late shipments directly to Washington where the complete processing is performed to save time.

**Industry and Occupation Coding**—The industry and occupation (I and O) coding operation is performed for all schedule A clerk assigns industry codes by reference to a list of company names carrying the code recorded in the 1970 census. Usually 60 to 70 percent of the industry codes can be assigned from the list. The remainder are assigned by reference to the coding

TABLE IV-2. Quarterly Evaluation of CPS Interviewers, National Sample: 1974

Category of interviewer	Interviewers	
	A (percent)	B (percent)
Interviewers with CPS assignments (average of 993 interviewers to be evaluated each quarter) .....	(X)	100
Quarterly evaluation not performed <sup>1</sup> .....	(X)	10
Quarterly evaluation performed <sup>2</sup> .....	100	90
Lowest rating "Excellent" .....	14	(X)
Lowest rating "Satisfactory" .....	57	(X)
Lowest rating "Needs improvement" .....	22	(X)
Special needs observation .....		
Given .....	(X)	100
Not given <sup>3</sup> .....	(X)	57
Lowest rating "Unsatisfactory" .....		
Given .....	(X)	100
Not given <sup>3</sup> .....	(X)	65
Warning letter .....		
Sent to interviewer .....	(X)	100
Not sent <sup>3</sup> .....	(X)	26
Warning letter information not available .....	(X)	67
		7

X Not applicable.

<sup>1</sup>Largely new interviewers who must have at least two full quarters of experience before being rated<sup>2</sup>Interviewer evaluation based on questionnaire edit error rates, type A noninterview rates, and production ratios<sup>3</sup>Regional director may determine normal criteria are not applicable to interviewer during the rating period

manuals used in the 1970 census. After assigning the industry code, the clerk assigns the occupation code from the written entries on the CPS-1 form, again using the 1970 census coding manual. These codes are assigned at the 3-digit level.

Coding occupation and industry for the national sample for a typical month requires 40 or more coders for a period of just over 1 week. At other times, these same coders are available for similar activities on other surveys where their skills can be maintained; they also form the cadre from which coders and supervisors have been drawn for this operation in the decennial census of population.

This is a demanding and error-prone clerical task requiring skills attained only with training and experience. A substantial effort is directed at supervision and control of quality of this operation. A 10-percent quality control sample of schedules is selected to be coded independently by two additional coders after the current month processing has been completed. Discrepancies among the three codes are resolved by taking the majority code as correct. The sample is selected so that the work of each coder can be evaluated. A coder whose error rate is greater than 7.5 percent of the total codes processed in any 2 out of 3 months is called unqualified. Unqualified coders have

their work verified by a second coder who examines and corrects the codes previously entered. The dependent verification continues for subsequent months to determine whether a coder has been able to reach qualified status. Coders are reassigned to other work if qualified status is not reached in four months.

Transmit information to Washington—After coding, the questionnaires are microfilmed, the film is developed and read by the FOSDIC (film optical sensing device for input to computers). A microfilm copy of the CPS information, which has been recorded on the CPS-1 (fig. IV-7) by blackening the appropriate small circles, is read by the FOSDIC machine and translated to codes on a tape which can be read by the census computers. The tape is transferred to Washington through an electronic transmission system.

Computer edit rejections—All questionnaires (except for the first 3000 and all of the last shipment) are kept in Jeffersonville for reference in responding to computer edit rejections identified later in Washington. Questionnaires for the last shipment are mailed to Washington so that edit rejections for these last cases can be handled there. The first 3,000 questionnaires are mailed to Washington to verify the data-acceptance operation, discussed in the following section.

Cycle day	Week	Procedure
<b>First week</b>		
1	Monday .....	Interviewing begins on Monday of the week containing the 19th day of the month. Regional Office begins monitoring of interviewer assignments. Interviewers ship completed work to the Regional Office daily.
2	Tuesday .....	Regional Offices begin processing. Completed work is shipped to Jeffersonville daily.
6	Saturday .....	Interviews should be completed except for special noninterview problems.
<b>Second week</b>		
8	Monday .....	All interviewer work (except for special noninterview problems) should be received in Regional Office.
9	Tuesday .....	Regional Offices close out. The last shipment of processed CPS questionnaires is mailed to Jeffersonville.
11	Thursday .....	Jeffersonville closes out. The last shipment of CPS data is transmitted to Washington.
12	Friday .....	Final computer processing begins in Washington.
<b>Third week</b>		
15	Monday .....	Tabulations delivered for review within Census. Census releases results to Bureau of Labor Statistics (BLS). BLS performs seasonal adjustment of data and begins analysis and preparation of press release.
19	Friday 1/31 .....	Results released to the public by the BLS.

### The Data Production Cycle

In a typical month, 19 days elapse from the start of interviewing to the public announcement of the results. This includes the work in the field, in Jeffersonville, and in Washington. The highlights of the 19-day data production cycle are summarized above.

Table IV-13 shows the progress of the data collection and processing for the month of January 1975 in terms of dates of shipment of materials from the regional offices to Jeffersonville, Ind., to Washington, D.C.

### Washington, D.C.

**Data acceptance check**—Each transmission, as received, is checked in and then immediately subjected to a computer data-acceptance operation which checks the readability of the data tapes and rejects questionnaires for specific failures, e.g., poor FOSDIC marks, evidence of malfunction of the camera, film or FOSDIC machine, missing questionnaire pages, or unclear non-interview status. The computer removes all of these questionnaires from the work unit; Jeffersonville is asked to locate and correct the offending questionnaires and include them in a later transmission. The reject and error listings generated by the first data acceptance runs (which include the first 3000 questionnaires mailed from Jeffersonville) are reviewed, item by item, with the questionnaires. This checks the data acceptance program (which is frequently changed to accommodate added questionnaire items) and the operation of the cameras and the

FOSDIC machines. (The added questions arise from the supplemental topics such as those listed in app. I.)

An unedited tally of the number of employed and unemployed persons is observed during these data acceptance runs to get a preliminary reading on the most crucial result of the survey, i.e., the unemployment rate. Because these results are not based on weighted or edited data, they will differ from the final result; however, the preliminary reading may warn of technical problems in the procedures that can be corrected early in the processing cycle. These preliminary tallies are subject to the same security precautions as those applied to the final results.

This program also counts the errors (inconsistencies and missing information) in the accepted data to produce the computer tallies of interviewer errors. (See p. 45.)

**Computer edit and recode**—A computer edit and recode program reviews each record to prepare it for weighting and tabulation. Responses are imputed for all missing and/or inconsistent entries. The questionnaire entries for employment status, occupation groups, family relationship, and age-sex categories are assigned recodes more suitable for processing and tabulation. Recent experience shows that a missing or inconsistent entry is allocated four to six times per 100 records in a typical month. During processing, a record is prepared for each interviewed person and each housing unit (interviewed and noninterviewed).

**Close-out for final processing**—In the national sample, for a typical month, there are about 55,000 units assigned for interview. (See table IV-1.) The control forms, listing sheets, and questionnaires for each of these units must, in a period of 2 weeks or less, pass through several processes at widely separated locations including, for most of it, two trips through the



TABLE IV-13. Shipment of CPS Questionnaires, Regional Offices to Jeffersonville to Washington:  
January 1975

Cycle day	Week	Mailed from regional office		Received in Jeffersonville		Received in Washington <sup>1</sup>	
		Number <sup>2</sup>	Cumulation (percent)	Number <sup>2</sup>	Cumulation (percent)	Number <sup>2</sup>	Cumulation (percent)
	<b>Interview week<sup>3</sup></b>						
2	Tuesday 1/21 .....	2,789	4.9	—	—	—	—
3	Wednesday .....	8,914	20.5	2,413	4.2	1,800	3.2
4	Thursday .....	11,014	39.8	8,758	19.6	6,900	15.3
5	Friday .....	11,234	59.5	10,584	38.1	7,200	27.9
6	Saturday 1/25 .....	1,841	62.8	11,273	57.9	—	—
	<b>Second week</b>						
8	Monday 1/27 .....	14,279	87.8	2,783	62.8	15,100	54.4
9	Tuesday .....	6,813	99.7	11,058	82.2	14,100	79.1
10	Wednesday .....	414	100.0	9,560	98.9	9,617	95.9
11	Thursday .....	—	—	455	99.7	2,296	99.9
12	Friday .....	—	—	—	—	15	100.0
	<b>Total</b>	<b>57,028</b>	<b>(X)</b>	<b>56,884</b>	<b>(X)</b>	<b>57,028</b>	<b>(X)</b>

— Entry represents zero

X Not applicable

<sup>1</sup> Most information arrives by wire from Jeffersonville (See p. 52.)<sup>2</sup> These figures are the number of CPS-1 questionnaires, as one questionnaire has pages for only four persons aged 14 and over, some households will have more than one questionnaire<sup>3</sup> Interview week for January 1975 Sunday 1/19-Saturday 1/25<sup>4</sup> Documents received in regional offices after close out (Tuesday, cycle day 8) are mailed directly to Washington

mail. Although close control of the materials is exercised, it is not unusual for some of the materials to be misplaced or delayed. When the number of missing questionnaires is reduced to 75 or less, the final processing is allowed to go forward. The allowance of 75 covers questionnaires of any kind which have not been accounted for,<sup>7</sup> including those still in the processing pipeline, or missing questionnaires for occupied units or for types A, B, or C noninterview, etc. When emptying the processing pipeline does not reduce the number below 75 and the outstanding questionnaires are concentrated in a few PSU's, the information may be obtained by immediately contacting the necessary households by telephone. To the extent possible, the missing data are treated by the most appropriate method (e.g., missing units may be made type A or B noninterview according to their occupancy status, if known).

**Weighting and tabulation**—When the acceptable proportion of all questionnaires for the month has been received, the results

are weighted and tabulated. (The method of weighting the records for tabulation is described in ch. V. The regular tabulations are summarized in ch. 1, page 1. See also [70].) In addition, special tabulations are produced as required.

When weighted tabulations for a CPS month become available, the results of the computer edit and weighting programs are checked by the Demographic Surveys Division to insure the processing has proceeded normally. The checks cover the number of data records processed, error and allocation rates for separate important data items, and the weighting factors produced during the various steps in the estimation procedure. A number of consistency checks are performed on the tabulated data results. The data are released to the Bureau of Labor Statistics only when the Census staff is confident all processing has been performed correctly.

**Security**—The tabulations are provided to the Bureau of Labor Statistics for seasonal adjustment, analysis, and publication in the press release. All preliminary and final tabulation results are subject to strictly enforced security rules to avoid unauthorized disclosure of the new labor force figures before the scheduled day for their announcement. The date of the press release is announced in advance [10].

<sup>7</sup> The 75 questionnaires represent about 0.13 percent of all questionnaires in an A- and C-design monthly sample; the 75 tolerance was not changed when the D-design sample was added. (See ch. VI.)

## Chapter V.

# PREPARATION OF ESTIMATES FROM THE NATIONAL SAMPLE

## INTRODUCTION

This chapter describes how the CPS monthly labor force estimates are prepared from the interviews in the combined A- and C-sample designs in the 461-PSU national sample. These estimates cover civilian noninstitutional population 14 years old and over residing in the 50 States and the District of Columbia. Estimates covering other characteristics and sectors of the population are discussed as special problems and estimation on page 64 of this chapter and in chapter VI.

Several distinct processes are included in the transformation of raw survey data into the estimates published each month from the CPS procedure. Although some of these processes have been combined in the operations needed to make in estimates, each one has its own specific objective. These processes are—

- Preparation of simple unbiased estimates.
- Adjustment for noninterview.
- First-stage ratio estimates, based on ratios of 1970 census totals for selected population categories to estimates of these totals based on 1970 census totals for the sample PSU's.
- Second-stage ratio estimates, based on ratios of independent estimates of the current population for the month of interview for age-sex-race groups to the first-stage ratio estimates of totals of these groups from the sample. The independent estimates are not necessarily more precise in all instances; however, it is believed that, in general, ratio estimates of this kind have both smaller sampling variances and smaller biases than the simple unbiased or first-stage ratio estimates.
- Composite estimates taking into account survey data available from previous months.
- Seasonal adjustment for selected key labor force statistics which is applied to adjust for normal seasonal variations.

The processes in the estimator described here have remained essentially unchanged since March 1968 when the second-stage ratio estimate was expanded to include independent estimates separately for persons with race reported Negro and other, by age and sex. Seasonal adjustment for selected labor force categories has been a part of the estimator since June 1957. However, after each decennial census, when the CPS sample is restratified and a new set of PSU's is identified, and at other times, some of these processes have been modified; e.g., the method of applying the noninterview adjustment, the source and scope of the independent estimates of current population, and others.

Estimation procedures for producing annual averages for the States are described in chapter VI.

## PREPARATION OF UNBIASED ESTIMATES

The importance of the ability to prepare simple unbiased estimates from survey data is very great even though the estimates

finally used may be of a different kind. *Simple unbiased estimates* have the property that their expected or mean value for repeated samples is equal to the value that would be obtained by applying the same data collection procedures to the entire survey population; they can be made only when probability samples are used. Any one of the other estimation steps described here could be (and most of them frequently are) applied to a nonprobability sample. One could for example, obtain labor force data from a sample of individuals chosen by any convenient method whatever, classify these individuals by age, sex, and race and make estimates of, for example, the number of employed persons by applying the proportion employed in the sample for each age-sex-race group to an independently derived estimate of the total population in that group and then summing the products over all groups.

Superficially, such a process resembles that used in the CPS; however, the results could be very different. No amount of adjustment and refinement of data from a nonprobability sample can obscure the two basic weaknesses of the resulting estimates; i.e., they are subject to unknown biases, and a nonprobability sample does not provide any basis for an objective evaluation of its precision [21, ch. 2].

By starting with unbiased estimates from a probability sample, various kinds of estimation and adjustment procedures (as for noninterview) can be applied with reasonable assurance that the overall accuracy of the estimates will be improved; in some cases, it is possible to measure the degree of improvement. Some of the adjustment procedures which are used may be unbiased, some biased but consistent, and some biased and not consistent [22, ch. 3, sec. 7b].

In the strictest sense, the CPS sample for any given month comes close to being, but is not in all respects, a probability sample of the population covered by the survey. For one thing, there is always a small amount of noninterview averaging, about 3 to 5 percent. Other factors, such as occasional errors occurring in the sample selection procedure, households missed by interviewers, etc., make it impossible to determine the exact probability of selection for every unit in the population. Nevertheless, a survey such as CPS, where every attempt is made to keep departures from true probability sampling to a minimum, is fundamentally different from a data collection operation which tolerates the use of judgment or quota sampling and high non-interview rates.

A simple unbiased estimate of the population total, for any characteristic investigated in the survey, may be made by multiplying the value of that characteristic for each sample unit—person or household—by the reciprocal of the probability with which that unit was selected and summing the products over all units in the sample. Strictly speaking this would give an estimate of the population total for units responding; adjustments for noninterview are described on page 56.

If all units in a sample have the same probability of selection, the sample is called *self-weighting* and simple unbiased estimates can be made by multiplying sample totals by the reciprocal of this probability.

## Weights for Rotation Groups

The first Sample designated for the 461-area design was selected with overall probability of 1/1,312; however, as time goes on, the number of households and the population increases, and continued use of this same probability leads to a larger sample size (with an increase in costs). Using the process described in appendix F, the probability is regularly reduced as the Samples are selected in order to maintain a fairly constant size (and cost) when the samples are interviewed. Usually the reduction applies to all eight rotation groups in a Sample and a constant weight, called the *basic weight* for the Sample, is maintained for all rotation groups. For example, the eight rotation groups of Sample A30 all have the same weight which differs from the weight common to all eight rotation groups of Sample A31. As a result, a single weight will typically not apply to all rotation groups in sample in a given month, although (except for special weighting situations) all USU's in a given rotation group are selected with the same probability. The differences in basic weights for rotation groups in a particular month's sample are not great. (See app. table F-1.)

## Special Weights

The CPS sample is largely self-weighting. However, a limited amount of special weighting is required before an estimate can be obtained by applying the appropriate overall weight to the sample totals. Each month the Washington staff prepares a listing of the sample households requiring special weights. This listing, which identifies each household and gives the special weight (a multiple of the basic weight) to be applied, is introduced into the computer at an appropriate stage of the weighting program. The computer applies the special weights to the data for the indicated households, and it is then possible to make a simple unbiased estimate for any item by multiplying the sample total by the appropriate basic weight (given in app. table F-1). One could, for example, make an estimate of the total civilian noninstitutional population 14 years old and over in this manner.

The special weighting is necessary, because different overall sampling fractions were used for certain parts of the population in the following two situations.

**Subsampling in large USU's**—Housing units in USU's that were subsampled in the field, because their observed size was much larger than the expected four housing units receive special weights. For example, an area-sample USU, expected to have four housing units but found, at the time of interview, to contain 36 HU's, would be subsampled at the rate of 1 in 3 to reduce the interviewer workload. Each of the 12 designated HU's in this case would be given an extra weight factor of 3 [40, ch. 4E].

USU's with a very large number of dwelling units (100 or more), receive special treatment; they are retained in a survey beyond the normal period, and do not in general require special weights (these are the rare event universe cases and are discussed in app. G).

**Weights for Cen-Sup USU's**—In general, USU's from the Cen-Sup sample require special weights. As described in chapter II, page 18, the Cen-Sup sample USU's are sometimes selected with lower probabilities in order to save cost.<sup>1</sup>

<sup>1</sup>These variable weights have relatively little impact on the sampling error of the estimates. For example, in April 1975, there were 491 Cen-Sup segments (and therefore an expected 491 units) in sample. The net effect of the special weights adjusted this number to a total of 552 expected sample units.

## ADJUSTMENT FOR TYPE A NONINTERVIEWS

Of the three types of noninterview discussed in chapter IV page 33, only the type A noninterviews represent population and occupied housing units eligible for the survey which is not represented by information from completed interviews. As used in this discussion, *noninterview* refers to the failure to obtain a usable response from an eligible household. This is not to be confused with another element of response error arising from an incomplete or inaccurate reply to a given item on a questionnaire referred to as *item nonresponse*, and discussed in chapter VII page B2.

Some noninterview is virtually certain to occur in any survey of a human population. In the CPS, holding noninterview at a low level requires a special effort of the field staff because survey data must be collected within about 1 week and results must be tabulated and released to the public within a few weeks after the period to which they refer.

We do not know of any unbiased or even consistent method of making adjustments for noninterview. In general, the magnitudes of the biases resulting from the adjustment procedure used in the CPS are not known for each month's tabulation. However, we have conducted research on this problem in special studies carried out in the mid-1960's. An intensive field followup of noninterview households was conducted during the 3 week following the CPS survey week in September 1965 [29]. Interviews were obtained for only about one-half of the households in the scope of the study. Because such a large proportion of these households remained inaccessible, the resulting data are biased estimates for the noninterview households. However, the data are consistent with what one would logically expect from CPS noninterview households but probably understate the true differences between interviewed and noninterviewed households. These data have been used to evaluate the bias due to noninterview adjustment in CPS. Overall, the results suggest that noninterview adjustment procedures used in CPS do not seriously distort the labor force distribution, at least for the major labor force categories such as employed and unemployed. Some of the subcategories of the distribution are more seriously affected, notably the categories with a job, not at work (these essentially represent persons on vacation or on temporary layoff from a job), parttime workers, school attendees, and others not in labor force (primarily representing retired workers). (See ch. VII.)

To adjust for type A noninterview in tabulations, based on the 461-area sample, noninterview adjustment factors are applied to interviewed households within clusters of PSU's, separately by rotation group. Housing unit population in SMSA and non-SMSA PSU's are adjusted separately; there is also a separate adjustment for the nonhousing unit population.

## Race and Collapsed-Race Definitions

For the purpose of discussing the separate steps in estimation in detail, it is necessary to define and assign terms to categories and combinations of categories of race. The population interviewed on the CPS questionnaire is classified by the question on race into three categories, White, Negro, and other. (See item 1 of the questionnaire, fig. IV-7.) in this report, when we state that the population is classified by race, we refer to the three-w

classification of the population, according to the responses to item 29 on the CPS questionnaire. It is also convenient to combine the categories, Negro and other, into one category called *not-white*. When we say the population is classified by *collapsed race*, we refer to the two-way classification of the total population into *white* vs. *not-white*.<sup>2</sup>

#### Noninterview Clusters

Clusters for making the noninterview adjustment have been formed by grouping sample PSU's that are similar in population characteristics, such as the percent of population not-white, percent urban, and the rate of population change since the last census. Also, all PSU's in a given cluster are either from SMSA areas or non-SMSA areas and are, with minor exceptions,<sup>3</sup> from the same census region. The number of PSU's (more precisely, the expected sample size) in each cluster is related to the proportion of the population not-white and is recognized as one of the priorities for combining noninterview adjustment cells. (See p. 58.)

#### Noninterview Adjustment Cells

The noninterview adjustment cells are defined by rotation group within each of the clusters and depend on the type of PSU, the SMSA-non-SMSA and urban-rural locations of the housing units and GQ's are defined by the 1970 census ED containing the units. The white or notwhite classification of interviewed households is given by the collapsed race of the principal person in the unit<sup>4</sup>; for noninterviewed units it is given by the race of the head, as determined by the interviewer on the current visit. (See fig. IV-7, item 14.) The farm or nonfarm classification of households in the rural portion of non-SMSA PSU's is determined by the interviewer.

Each noninterview cluster for housing unit population has 48 noninterview adjustment cells: Three residence categories, by collapsed race, by eight rotation groups.

Housing unit population in SMSA's—Housing unit population in SMSA's is classified into six collapsed race-residence adjustment cells, by rotation group, in each cluster.

Collapsed race	Central city of SMSA	Balance of SMSA	
		Urban	Rural
White	a	b	c
Not-white	a	b	c

<sup>2</sup> This treatment of the race categories, for estimation and tabulation, has a long history in the CPS: the terms *nonwhite* and *color*, for not-white and collapsed race, respectively, have been used in many earlier reports and references cited in this technical paper.

<sup>3</sup> An SMSA in more than one region is assigned to the region of its largest component for the purpose of noninterview adjustment.

<sup>4</sup> The principal person is defined as the wife of the head in husband-wife households or the head in other households. (See p. 66.)

Housing unit population not in SMSA's—The housing unit population in non-SMSA PSU's is classified into six race-residence adjustment cells, by rotation group, in each cluster.

Housing unit population not in SMSA's—The housing unit population in non-SMSA PSU's is classified into six collapsed race-residence adjustment cells by rotation group in each cluster:

Collapsed race	Urban	Rural	
		Nonfarm	Farm
White	d	e	f
Not-white	d	e	f

Nonhousing unit population—Nonhousing unit (GQ) population is combined separately for SMSA and non-SMSA PSU and then divided into 10 adjustment cells for each rotation group at the U.S. level.

Nonhousing unit population—Nonhousing unit (GQ) population is combined separately for SMSA and non-SMSA PSU's and then divided into 10 adjustment cells for each rotation group at the U.S. level.

Collapsed race	SMSA's			non-SMSA's	
	Central city	Balance	Urban	Rural	
				Non-farm	Farm
White	g	h	k	m	n
Not-white	g	h	k	m	n

#### Computing Noninterview Adjustment Factors

The numbers of interviewed and noninterviewed households are tabulated separately into each of the noninterview adjustment cells. Note that to perform the adjustment, the collapsed race of the principal person for interview households and the current race of the head for noninterview households, as well as residence of all households, must be available. Unweighted counts of households are used, except for the special weight discussed previously in this chapter.

For each of the noninterview cells in each rotation group, a following ratio is computed:

$$\frac{\text{Interviewed households} + \text{noninterviewed households}}{\text{Interviewed households}}$$

These ratios are applied to data for each person in interviewed households in the corresponding groups except in cells where the ratio is found to equal or exceed 2.0. In this case, the counts are combined for all races within the residence category in the cluster. A common factor is computed and applied for all races unless the new factor exceeds 3.0; if it does, a noninterview adjustment factor of 3.0 is used. The combinations within the clusters are indicated by common letters in the cell descriptions.

### Priorities for Combining Noninterview Adjustment Cells

The rules for combining cells within a noninterview cluster given in the preceding paragraph treat residence as more important than collapsed race since the noninterview adjustment always preserves the residence distinction. This selection of priorities, based on correlations of collapsed race and of residence with labor force and on tabulation objectives, is considered in determining the number and characteristics of PSU's to be combined into a cluster. PSU's having high proportions of not-white population have been combined into clusters with a larger expected sample size (generally an expected number of 80 or more households per rotation group). This is to increase the chance of applying noninterview adjustment factors separately by collapsed race and residence in the cluster. Clusters with a low proportion of the population not-white are expected to have only the three residence factors applied.

### Weights After the Noninterview Adjustment

At the completion of the noninterview adjustment procedure, the records for each person should bear a weight reflecting the product of—

(basic weight)  
(adjustments for special weighting)  
(noninterview adjustments).

At this point, records for all individuals in the same household should have the same weight (unless the household has members in different collapsed-race categories). A household with wife and husband of different races may be counted in different noninterview adjustment cells in separate months if the interview status differs; this can occur, because collapsed race of the household is determined by the principal person for interviewed households but, for noninterviewed households, by race of the head, as determined by the interviewer. (Basic weights are shown in table F-1.)

## RATIO ESTIMATES

The distribution of the population designated for the sample in any month will differ somewhat from that of the Nation in such basic characteristics as age, race, sex and farm-nonfarm residence. These particular population characteristics are closely correlated with labor force and other characteristics estimated from the sample. Therefore, the precision of some of the sample estimates can be increased when, by the use of appropriate ratio estimates, the sample population is brought as closely into agreement as possible with the known distribution of the entire

population with respect to these characteristics. Such weighting is accomplished by means of ratio estimates which are treated two stages.

### First-Stage Ratio Adjustment to 1970 Census Population Counts

**Purpose of the first-stage ratio estimate**—The purpose applying the first-stage estimate is to reduce the contribution the variance arising from the sampling of PSU's. That is, variance that would still be associated with the estimates if included in the survey each month all households in every sample PSU.

There are several factors to be considered in determining what information to use in applying this ratio adjustment. First, information has to be available for each PSU from the 1970 Census. Second, the information should be correlated with many of the statistics of importance published from the CPS possible. Third, the information should be reasonably stable over time so that the gain from the estimation procedure does not deteriorate. The basic labor force categories (unemployed, not agricultural employed, etc.) could be considered; however, information could badly fail the third consideration.

Race and residence categories were chosen even though these variables are not highly correlated with a number of published statistics. By using these categories, the first-stage ratio estimate compensates for the fact that race and residence are not fully considered in selecting the sample PSU to represent each stratum.

The 1970 census proportions of the population in each F for race and residence categories and whether the PSU is SMSA or non-SMSA were important criteria considered in grouping PSU's into strata. However (as pointed out in discussing need to redefine strata in ch. II, p. 10), considerable variation among PSU's within a stratum was accepted for these items. The NSR-sample PSU's were selected using probability proportionate to total 1970 population and with a selection procedure which controlled the geographic distribution of PSU's among States; neither of these is necessarily related to race or residence within the PSU's.

**Computation of factors**—The first-stage ratios are based on 1970 census data and are applied only to sample data for nonself-representing PSU's. Separate ratios are used for each collapsed race and residence category as those used in the noninterview adjustment. A ratio could be computed and applied for each of the six collapsed race-residence cells for each PSU. However, this would result in a very large number of ratio estimate cells and could result in substantial biases [22, ch. sec. 23]. Thus, ratios are applied only for two groups of strata (in SMSA and not in SMSA) for each of the four census regions. Applying the ratio estimate in this way is consistent although still technically biased, but the sample per cell is large enough and the number of cells is small enough, that the bias is negligible. The following description of the ratio indicates the computation:

1970 Census population in the collapsed race-residence for NSR strata in the SMSA or non-SMSA group in the region  
Estimate of this population, based on the 1970 census population for sample PSU's in the same group.

The estimate used in the denominator of each of the ratios is obtained by multiplying the census population in the appropriate collapsed race-residence cell for each sample PSU by the reciprocal of the probability of selection for that PSU and summing over all nonself-representing PSU's in the stratum group in the region.

Factors used for the A- and C-designs—A single set of first-stage ratio estimate factors is used for the combined A- and C-design CPS samples. The factors in use August 1974 through March 1975 are shown in table V-1. The factors are changed when PSU's are replaced. (See ch. III, p. 25.) Separate sets of factors are required when the A- or the C-design samples are used alone.

The factors are computed by the foregoing expression with the denominator, computed by formula (V/4). In general if a large ratio results, a maximum factor of 1.3 is assigned to the cell with an appropriate adjustment to the factor computed for the other collapsed race category in the same residence cell.

Information needed to apply factor—Two sets of information are required for each sample person to distinguish the first-stage ratio estimate factor:

1. The census region, SMSA-non-SMSA designation of the PSU and the location of the USU within the PSU, are determined when the sample segment is designated; i.e., central city, balance urban or balance rural for SMSA PSU's and urban or rural for non-SMSA PSU's.
2. The collapsed race category for the sample person and, for units in the rural portion of non-SMSA PSU's, the rural farm or rural nonfarm designation of the household are determined by the interviewer.

Weights after the first-stage ratio estimate—At the completion of the first stage of ratio estimation, the records for each person should bear a weight reflecting the product of—

(basic weight)  
(adjustments for special weighting)  
(noninterview adjustment)  
(first-stage ratio estimate factor)

At this point, records for all individuals in the same household should have the same weight, except for a household having members in different collapsed-race categories. The first stage of ratio estimation, as well as the noninterview adjustment factors, are applied to individuals rather than by applying a single factor to all members of the household.

Meaning of the term first-stage ratio estimate—In discussing the CPS estimation problem, we may refer to the *first-stage ratio estimate* for a given item. This is the estimate produced by tabulating the appropriate sample records bearing the results of special weights, noninterview adjustments, basic weight (the inverse of the probability of selection), and the first-stage ratio estimate factors.

## Second-Stage Ratio Adjustment to Current, Independent Estimates of the Population

The second-stage ratio estimate adjusts U.S. sample estimates of population in a number of age-sex-race groups to independently derived current estimates of the population in each of these groups.

TABLE V-1. First-Stage Ratio Estimate Factors for NSR PSU's in Combined A- and C-Design National Sample: August 1974-March 1975

Ratio-estimate cell	Region <sup>1</sup>			
	Northeast	North Central	South	West
<b>SMSA PSU's</b>				
Central cities				
White .....	0.9212	1.1601	1.0358	1.15
Not-white .....	1.3000	1.0028	1.0743	1.09
Balance urban				
White .....	1.3000	.8403	.9398	.91
Not-white .....	1.3000	1.1772	1.0420	.77
Balance rural				
White .....	.9379	.9189	.9308	.93
Not-white .....	.8088	.7996	1.1531	.94
<b>Non-SMSA PSU's</b>				
Urban				
White .....	1.0004	.9719	1.0369	1.01
Not-white .....	1.0687	.9162	1.0294	1.20
Rural nonfarm				
White .....	.9913	.9949	.9962	.96
Not-white .....	1.1122	1.2343	.9516	1.30
Rural farm				
White .....	.9424	1.0206	.8999	.91
Not-white .....	.7983	1.0884	1.0044	1.30

<sup>1</sup>First-stage ratio estimate factors require change with replacement PSU's (See ch. III and app. table M-1.)

The independent estimates are prepared by inflating the recent census counts to include the estimated net census undercount by age, sex, and race, aging this population forward each subsequent month and later age by adding births and migration, and subtracting deaths. These postcensal population estimates are then deflated to census level to reflect the pattern of net undercount in the most recent census, by age, sex, race. This *inflation-deflation* method of preparing postcensal estimates preserves the actual pattern of population change time in any age group [49; 50]. The civilian noninstitutional population is obtained by subtracting estimates for the institutional population and for the armed forces. The institutional population is derived by assuming that the proportion of the total population that is institutional is the same as in

1970 census by sex, race, and single year of age. The armed forces population is provided by the Department of Defense.

The CPS sample returns (taking into account the weights determined after the first-stage of ratio estimation) are, in effect, used only to determine the percentage distribution within each age-sex-race group, by employment status, and various other characteristics. To obtain estimates of totals, the percentage distributions are multiplied by the independent population estimates for the appropriate age-sex-race groups and the results are summed over groups.

How the adjustments are applied—Second-stage ratio estimate factors are applied separately, by rotation group, sex, and age groups associated with various race categories.

Intermediate ratio estimate for Negro and other races—Intermediate ratio estimates are first performed separately for sample persons with race reported as Negro and other. The age and sex categories used in these two groups are given in table V-2. For most of these cells, it is unlikely that the denominator of the ratio (the population estimated from CPS sample persons) will be based on zero or very few cases. However, it can occur in some of the cells involved in the following list. When this does happen, cells are combined and a common ratio estimate factor is computed and used so that the following categories are maintained:

Categories maintained	Number of age categories combined <sup>1</sup>
<b>Race reported as Negro</b>	
20 to 24 years, by sex .....	2
60 to 64 years, by sex .....	2
70 years and over, by sex .....	2
<b>Race reported as other</b>	
<b>First combination</b>	
14 to 34 years, by sex .....	3
35 to 54 years, by sex .....	2
55 years and over, by sex .....	2
<b>Final combination<sup>2</sup></b>	
Males, 14 years and over .....	7
Females, 14 years and over .....	7

<sup>1</sup> See table V-2 for definition of age categories.

<sup>2</sup> Used if first combination provides insufficient sample cases.

The other ratio estimate cells for Negroes, by sex, are large enough not to need combination.

Final second stage ratio estimate—A final second-stage ratio estimate is next applied to all sample records (including those involved in the intermediate ratio estimate), with separate factors computed and used by sex, collapsed race, and 17 age groups, for a total of 68-ratio estimate cells, by rotation group. (The age-sex-collapsed race groups are given in table V-3.) The

expected sample size is considered large enough to provide stable estimates of population in each of the cells, by rotation group. The ratio estimate factors would be exactly equal to 1:0 for sample persons reported as not-white if the age categories used for the separate Negro and other adjustments were the same as for this adjustment. Since this is not the case, these factors differ slightly from 1:0.

TABLE V-2. Intermediate Second-Stage Ratio Estimate Factors; Ratio of Independent Estimate to CPS First-Stage Ratio Estimate of Not-White Population, by Age, Sex, and Race: June 1975

(Average for all rotation groups)

Age	Race reported as Negro <sup>1</sup>	
	Male	Female
<b>Total</b> .....	1.1314	1.0562
14 to 15 years .....	1.0040	1.0159
16 to 17 years .....	1.0467	.9673
18 to 19 years .....	1.0910	1.0282
20 to 21 years .....	1.3585	1.1090
22 to 24 years .....	1.2829	1.0420
25 to 29 years .....	1.0608	1.0181
30 to 34 years .....	1.2611	1.1509
35 to 39 years .....	1.0996	1.1180
40 to 44 years .....	1.1480	1.0875
45 to 49 years .....	1.2411	1.0565
50 to 54 years .....	1.1481	1.0848
55 to 59 years .....	1.1526	.9739
60 to 61 years .....	1.0556	.9762
62 to 64 years .....	1.0550	1.1238
65 to 69 years .....	1.0920	1.1600
70 to 74 years .....	1.1051	.8696
75 years and over .....	1.0320	1.1277
Age	Race reported as other <sup>1</sup>	
	Male	Female
<b>Total</b> .....	1.1830	1.1815
14 to 17 years .....	1.4321	1.1303
18 to 24 years .....	1.1514	1.3696
25 to 34 years .....	1.1827	1.0751
35 to 44 years .....	.9508	1.0232
45 to 44 years .....	1.1754	1.2163
55 to 64 years .....	1.3258	1.4932
65 years and over .....	1.3476	1.3858

<sup>1</sup> Ratio estimate factors are computed and applied for each rotation group separately.

Reduction of the mean square error—The second-stage ratio estimate is not a consistent estimate, since the denomina

Table V-3. Final Second-Stage Ratio Estimate Factors; Ratio of Independent Estimate to CPS First-Stage Ratio Estimate of Population, by Collapsed Race, Age, and Sex: June 1975

(Average for all rotation groups)

Age	White <sup>1</sup>		Not-white <sup>1</sup>	
	Male	Female	Male <sup>2</sup>	Female <sup>2</sup>
<i>Total</i> .....	1.0383	1.0228	1.0000	1.0000
14 to 15 years .....	.9985	1.0024	.9882	1.0035
16 to 17 years .....	.9949	.9901	1.0126	.9964
18 to 19 years .....	1.0321	1.0296	1.0201	1.0158
20 to 21 years .....	1.0677	1.0507	.9813	.9613
22 to 24 years .....	1.1111	1.0732	1.0142	1.0176
25 to 29 years .....	1.0603	1.0205	.9950	.9966
30 to 34 years .....	1.0432	1.0148	1.0060	1.0041
35 to 39 years .....	1.0141	1.0143	.9921	.9774
40 to 44 years .....	1.0748	1.0539	1.0081	1.0249
45 to 49 years .....	1.0280	1.0023	1.0021	1.0019
50 to 54 years .....	1.0554	1.0175	.9977	.9979
55 to 59 years .....	1.0274	1.0185	.9918	1.0133
60 to 61 years .....	.9929	.9432	1.0554	.9890
62 to 64 years .....	.9888	1.0174	1.0288	.9832
65 to 69 years .....	1.0688	1.0069	.9587	.9749
70 to 74 years .....	.9913	.9909	.9597	1.0111
75 years and over .....	1.1038	1.0969	1.0489	1.0229

<sup>1</sup>Ratio estimate factors are computed and applied for each rotation group separately.

<sup>2</sup>Estimates of the not-white population include effect of intermediate second-stage ratio estimate (See table V-2.)

tors of the ratios are not unbiased estimates of the numerators (the independent population projections, by age, sex, and race). The CPS sample estimates of population, by age, sex, and race (the denominators of the ratios) have the biases introduced by noninterview and by the first-stage ratio estimate. Nevertheless, the second-stage ratio estimate is believed to be an effective method of reducing mean square errors of most estimates because the correlations of labor force characteristics with population in age-sex-race groups tend to be higher than the correlations of these characteristics with total population 14 years old and over. As a result, ratio estimates based on age-sex-race groups can reduce the sampling variability component of the mean square error. Secondly, the unbiased and first-stage ratio estimates from the 461-area sample consistently produce estimates of total population 14 years old and over that are smaller than independently derived estimates. (See table V-4.) Also, comparisons of first-stage ratio estimates for age-sex-collapsed race groups with the corresponding independent estimates (for example, see table V-5) indicate the possible existence of biases in the sample estimates which vary in size and direction from one group to another. We believe that the adjustments based on the independent projections, by age, sex, and collapsed race, have the effect of reducing these biases. (Ch. VII discusses the extent of these biases.)

Table V-4. Coverage Ratios for the CPS National Sample, by Month: 1974-1975

(Average for all rotation groups)

Month	Coverage ratio <sup>1</sup>	
	1974	1975
January .....	0.955	0.965
February .....	.964	.962
March .....	.958	.958
April .....	.952	.964
May .....	.947	.963
June .....	.958	.963
July .....	.959	.965
August .....	.960	.967
September .....	.962	.968
October .....	.967	.971
November .....	.967	.969
December .....	.964	.965

<sup>1</sup>Ratio of sample population 14 years old and over (after first-stage estimation) to independent estimates of the population.



TABLE V-5. Coverage Ratios, by Age, Sex, and Collapsed Race; CPS National Sample:  
Average for July 1974-June 1975  
(Average for all rotation groups)

Age	Coverage ratios <sup>1</sup>			
	Males		Females	
	White	Not-white	White	Not-white
<i>Total</i> .....	<i>0.963</i>	<i>0.873</i>	<i>0.978</i>	<i>0.927</i>
14 to 19 years .....	.979	.947	.996	.935
14 to 15 years .....	1.001	.981	1.000	.964
16 to 17 years .....	.984	.984	1.022	.966
18 to 19 years .....	.949	.859	.965	.870
20 to 24 years .....	.935	.776	.941	.901
20 to 21 years .....	.941	.733	.947	.852
22 to 24 years .....	.930	.808	.937	.936
25 to 44 years .....	.952	.843	.979	.914
25 to 29 years .....	.938	.856	.977	.942
30 to 34 years .....	.953	.786	.982	.881
35 to 39 years .....	.958	.921	.975	.912
40 to 44 years .....	.963	.819	.981	.918
45 to 64 years .....	.968	.874	.987	.950
45 to 49 years .....	.970	.825	.979	.949
50 to 54 years .....	.959	.867	.986	.934
55 to 59 years .....	.967	.883	.988	.984
60 to 61 years .....	.993	.946	1.037	.952
62 to 64 years .....	.966	.941	.971	.924
65 years and over .....	.988	.944	.973	.937
65 to 69 years .....	.970	.935	.972	.861
70 to 74 years .....	1.006	1.099	1.020	1.184
75 years and over .....	.995	.830	.943	.860

<sup>1</sup>Average of monthly ratios of population estimated from the sample (after first-stage estimation) to independent population projections.

### COMPOSITE ESTIMATOR

The next step in the preparation of estimates is the derivation of a composite estimate. A *composite estimate* is a weighted average of two estimates for the current month for any particular item. The first of these two estimates is the two-stage ratio estimator that includes all estimation steps given above, including the noninterview adjustment and the first and second stages of ratio estimation. The second estimate consists of the composite estimate for the preceding month to which has been added an estimate of the change from the preceding month to the present month, based on that part of the sample which is common to the 2 months (about 75 percent). Operationally, all that is needed is the composite estimate for the preceding month, an estimate of change from the preceding month, and an estimate of level for the current month. Hence, it is not necessary to store data for more than 2 months in order to make this estimate.

For most statistics, there is a moderate to high correlation over time for data from the same segments. The composite esti-

mate takes advantage of this by using accumulated information from earlier samples, as well as the information from the current sample.

In general, for such a composite estimate to be unbiased, weights for the two components must add to one; how they need not necessarily be equal. In CPS, the weights used combining these components are each one-half. Equal weights satisfy the condition that, for virtually all items, the composite estimate will be somewhat more reliable than the two-stage estimate. It would be possible to use approximations to optimum weight for each statistic in the averaging process to maximize the gain from the composite estimate [16]. However, such estimates would be internally inconsistent. For example, better estimates of civilian labor force could be made that would not necessarily equal the sum of the better estimates for employed and unemployed. Using the same weight for all statistics avoids these inconsistencies.

The gains in reliability from the use of the composite estimator are greater in estimates of month-to-month change, although

gains are also realized for estimates of levels in a given month, a change from year to year, or over other intervals of time [4].

In contrast to the previous estimation steps, composite estimation modifies the results of a tabulation rather than adjusting the weights of individual sample records. This statement also applies to the seasonal adjustment procedure.

## SEASONAL ADJUSTMENTS

Seasonal adjustment is applied to estimates of level for the major labor force categories by age, sex, race, class of worker, marital status, etc.<sup>6</sup> The system adjusts for normal seasonal variations to help distinguish the underlying economic situation in month-to-month changes. For example, the unadjusted unemployment rate in June is generally much higher than for May because of the influx of students into the labor force; after seasonal adjustment, the unemployment rates for the 2 months should be about the same if there are no significant changes in the economy during that time period.

Many economic statistics reflect a regularly recurring seasonal movement which can be estimated on the basis of past experience. By eliminating that part of the change that can be ascribed to usual seasonal variation, it is possible to observe the cyclical and other nonseasonal movements in the series. However, in evaluating deviations from the seasonal pattern, i.e., changes in a seasonally adjusted series, it is important to note that seasonal adjustment is merely an approximation, based on past experience. Seasonally adjusted estimates may have a broader margin of possible error than the original data on which they are based, since they are subject not only to sampling and other errors but, in addition, are affected by the uncertainties of the seasonal adjustment process itself. However, some studies indicate the variances are about the same for seasonally adjusted data as for nonadjusted data. Seasonally adjusted series of selected labor force data are given in the monthly publication *Employment and Earnings*; the seasonal adjustment factors are usually provided in the October issue [75].

The seasonal adjustment methods used for these series are an adaptation of the standard ratio-to-moving average method, with a provision for moving adjustment factors to take account of changing seasonal patterns. A detailed description of the method is given in [69].

Data for the household series are seasonally adjusted by the Census Bureau's X-11 Method. Each of three major labor force components (agricultural employment, nonagricultural employment, and unemployment) for four age-sex groups (male and female workers, 16-19 years old and 20 years old and over) are separately adjusted for seasonal variation and then added to give seasonally adjusted total figures. In order to produce seasonally adjusted total employment and civilian labor force data, the appropriate series are aggregated. The seasonally adjusted rate of unemployment for all civilian workers is derived by dividing the figure for total unemployment (the sum of four seasonally adjusted age-sex components) by the figure for the civilian labor force (the sum of 12 seasonally adjusted age-sex components). Other series, such as unemployment by duration or employment by major occupational groups, are adjusted independently.

<sup>6</sup> Seasonal adjustment of the data is done by the Bureau of Labor Statistics, U.S. Department of Labor, using the method discussed in [69].

The seasonal adjustment factors applying to current data are based on a pattern shown by past experience. Once each year (January), these factors are revised.

## ESTIMATION FORMULAS, MONTHLY CPS LABOR FORCE DATA

The unbiased, ratio, and composite estimation procedure described can be stated algebraically.

Let  $x_aSR$  be an estimate of the number of persons having certain characteristic, prepared by tabulating the weight of each sample person in the  $a$ -th age-sex-race category having that characteristic in SR PSU's. The weights used (see p. 58), in practice, reflect the probability of selection (including the special weights from subsampling and the noninterview adjustment). (The estimation formulas given here do not separately indicate these details.) Let  $x_{ac}$  be an estimate prepared similarly from  $c$  sample persons in NSR PSU's having the characteristic with the  $a$ -th age-sex-race in the  $c$ -th collapsed race-residence category. Let  $y_aSR$  and  $y_{ac}$  be estimates prepared similarly for the total number of persons in these groups. Each of these estimates is based on a weighted average of the separate estimates from A- and C-design samples.

With the foregoing definitions, the estimate, after two stages of ratio estimation<sup>7</sup> for the CPS, is

$$x^* = \sum_a \frac{x_a}{y_a} \hat{y}_a \quad (V)$$

where

$$x'_a = x_aSR + \frac{NSR}{Z_c} x_{ac} \frac{Z_c}{Z_c} \quad (IV)$$

$$y'_a = y_aSR + \frac{NSR}{Z_c} y_{ac} \frac{Z_c}{Z_c} \quad (IV)$$

and  $\hat{y}_a$  is the current independent estimate of the number of persons in the  $a$ -th age-sex-race group.

The expressions  $x'_a$  and  $y'_a$  are the weighted averages of the A- and C-design estimates after the first-stage ratio adjustment.  $x'_a$  is the number of persons with the characteristic in the  $a$ -th age-sex-race group and the total number of persons in that group respectively.

The subscript SR denotes values for the self-representing strata; estimates for NSR are indicated without a subscript.

$Z_c$  is the 1970 Census count of population in nonself-representing strata for the  $c$ -th collapsed race-residence-region group.

$$Z_c = \left(\frac{2}{3}\right) \sum_h \frac{1}{h} \sum_i \frac{N_{Ah}}{N_{Ah}} Z_{Ach} - \left(\frac{1}{3}\right) \sum_h \frac{1}{h} \sum_i \frac{2N_{Ah}}{N_{Ch}} Z_{Ch} \quad (IV)$$

is an estimate of  $Z_c$  obtained by inflating census counts for sample PSU's to stratum totals separately for the A- and

<sup>7</sup> The estimation formula given here treats the intermediate and final second-stage ratio estimates as one combined ratio estimate.

design samples and computing a weighted average of the estimates from the two sample designs. Weights of two-thirds and one-third are used for the A- and C-design sample estimates, respectively.

- $Z_{Aci}$  is the 1970 census population for the  $c$ -th collapsed race-residence group in the  $i$ -th sample PSU in NSR stratum  $h$  of the A-design sample.
- $Z_{Chi}$  is the same for the  $i$ -th sample PSU in NSR stratum  $h$  of the C-design sample.
- $N_{Ahi}$  is the measure of size (1970 census population) for the  $i$ -th sample PSU in the  $h$ -th A-design stratum.
- $N_{Chi}$  is the measure of size for the C-design sample PSU.
- $N_{Ah}$  is the measure of size of the  $h$ -th A-design stratum.
- $L$  is the number of nonself-representing A-design strata in the region.

The composite estimate may then be written as follows:

$$\hat{x}_t = (1-K)x_t^* + K[\hat{x}_{t-1} + d_{t,t-1}] \quad (V/5)$$

where

- $\hat{x}_t$  is the composite estimate for the current month and  $x_t^*$  is the second-stage ratio estimate for the current month as given by expression (V/1).
- $\hat{x}_{t-1}$  is the composite estimate for the preceding month.
- $d_{t,t-1}$  is an estimate of the difference in level between the current month and preceding month, based only on the USU's included in the sample for both months. It has the form

$$d_{t,t-1} = x_t^{**} - x_{t-1}^{**}$$

- $x_t^{**}$  for month  $t$  is the same form as  $x_t^*$ , except that the sample estimates  $x_a$  and  $y_a$  are based on USU's that are in sample for both months  $t$  and  $t-1$ .
- $x_{t-1}^{**}$  is the estimate for month  $t-1$ , based on the same USU's used to prepare the estimate  $x_t^{**}$ .
- $K$  in general, is a number between 0 and 1 and, for the CPS, has been assigned a value of 0.5.

## SPECIAL PROBLEMS IN ESTIMATION

Although the CPS program objectives originally formulated deal with national estimates for labor force data, there is continuing interest in expanded scope of subject matter (see app. I) and tabulation detail using the existing national sample. Even without special attention directed at such objectives in designing the sample, the present CPS can provide valid estimates for a variety of population characteristics, including some that are relatively rare and difficult to obtain. These data are often desired for combinations of States, the larger individual States, and SMSA's.

These requirements are reflected in a number of alternative survey and estimation procedures discussed in the following paragraphs.

## Averaging Monthly Estimates

A device frequently used to reduce the sampling error of estimates based on the national sample is to average CPS estimates for a number of months. For example, an annual average monthly CPS estimates of a given statistic at the national level would be based on 12 times the number of interviews that produce a monthly estimate. However, since 75 percent of the C' sample for a given month is also interviewed in the following month and the remaining 25 percent of the sample is comprised of USU's in geographic proximity, the estimates from the two samples are not independent. Annual averaging, therefore, reduces the sampling variance by a factor substantially less than 12. Also (with minor exceptions<sup>7</sup>), the same first-stage sample units are involved over the period so that averaging monthly estimates does not reduce the component of variance arising from the sampling of PSUs or from sampling strata in the design portion of the sample.

**Variance reduction by averaging estimates**—The reduction in sampling error by averaging CPS estimates over adjacent months has been studied using 18 months of data collected beginning January 1969 [4]. The CPS sample in use at that time was the 449-PSU 1960 census based design.

The study showed that characteristics for which the month-month correlation is low, such as unemployment, were helped considerably by averaging monthly data, while items for which the correlation is high show less benefit, such as rural residence and agricultural items. Variances of national estimates of unemployment were reduced by factors of about 2.0, 3.2, and when averaged over 3, 6, and 12 months, respectively.

For national estimates of unemployment, the largest component of variance is due to sampling within the PSU; the study showed it to be in excess of 95 percent of the total sampling variance for the composite estimator (V/5). (See ch. VIII-1 and VIII-2.) This relationship also indicates the variances to be expected from averaging for areas that are comparable with combinations of CPS sample PSU's, where the variance of the estimate would have no component from the sampling of PSU's. For most States, however, there are a number of strata, and the gains from averaging would be less. (See ch. VI.)

**Estimates for States and SMSA's by averaging**—The following estimation procedures have been used to prepare averages of monthly CPS data for selected States (and SMSA's), based on the 461-area national sample. Two estimation procedures have been used for State estimates for calendar years 1973, 1974, and 1975. (Estimates for subsequent years are discussed in ch. VI.)

**Annual averages for calendar year 1973 were provided States<sup>8</sup> by the following process.** The procedure begins with the result of all stages of estimation as given above for national estimates; the composite estimation and seasonal adjustment of national estimates do not affect the weights for the final design, and, therefore, have no bearing on the following process.

A tabulation of the weighted file for the State (or SMSA) is performed for each month's data. A composite estimator, to (V/5), but incorporating State estimates is used. An arithmetic

<sup>7</sup>Due to the replacement of PSU's, see ch. III, p. 25.

<sup>8</sup>The 19 States are identified in ch. VI, p. 68.

mean of the 12 monthly estimates is computed. Finally, the level of the mean is adjusted by a ratio estimate using an independently established estimate of the population, 16 years old and over, for each State, as of July 1973.

The estimate of the annual average of the 12 monthly composite estimates during calendar year 1973 for the  $j$ -th State is written:

$$\bar{x}_j = \frac{\sum_{t=1}^{12} \hat{x}_{tj}}{\sum_{t=1}^{12} \hat{y}_{tj}} \hat{Y}_{1973j} \quad (V/6)$$

where

- $\hat{x}_{tj}$  is the composite estimate of the population having the characteristic  $x$  in State  $j$  in month  $t$ , as produced by the A- and C-design national sample.
- $\hat{y}_{tj}$  the same as  $\hat{x}_{tj}$  for the estimate of population, 16 years old and over.
- $\hat{Y}_{1973j}$  the independent control estimate of population 16 years old and over for State  $j$  as of July 1973.

Annual averages for calendar 1974 and 1975 were provided for 27 States (8 States in addition to the 19 mentioned) by an estimator, otherwise similar to estimator (V/6) but employing an additional first-stage ratio estimate involving decennial census totals for NSR-sample PSU's and the total NSR portion of the State. (The 23 States are identified in ch. VI.)

### Family and Household Estimates for March Supplement Data

The March sample is expanded to include children under 14 years old, and male members of the armed forces living in civilian housing not on a military base or with the family on a military base.<sup>9</sup> In addition, the March questionnaire includes questions on individual income, work experience, migration, tenure, and expanded questions on relationship and marital status. The March sample is used for the preparation of family and household tabulations incorporating these statistics.

The CPS weighting operation regularly used for monthly national labor force estimates (outlined in the first portion of this chapter, pp. 55-64) produces tabulations that very nearly agree with independent population controls, by age, sex, and race. However, when household or family characteristics are tabulated from CPS records weighted in this manner, some inconsistencies become apparent. A weighting system which reduces the observed inconsistencies has been developed and is used when tabulating data for households.

**Consistency with CPS estimates**—The system has been developed by observing relationships in the regular CPS tabulations, exploring reasons for these relationships, and adopting a model from which modifications in the weighting procedure be-

come plausible. The weighting is based on the following observations and assumptions.

When records are weighted by the regular monthly CPS system, a tabulation of persons by marital status, by sex, will typically show different estimates of males married with spouse present (Mmsp) and females married with spouse present (Fmsp). Both of these figures represent estimates of husband-wife households, but the male-based figure usually turns out to be the larger. As all adult household members are included every CPS sample unit,<sup>10</sup> the number of sample records for the two categories should be the same. The difference occurs, because different weights can be assigned to different members of a sample household.

Of the two estimates of married spouse present households, the female-based figure is believed to be the better. This conclusion is based on the following:

- Coverage ratios for ages associated with household heads (ages 20-64) show more complete coverage for females than for males; table V-5, for example, shows the female coverage ratios are usually closer to 1.0 for these ages than for males. Tabulations for Fmsp before the second-stage ratio estimate to population controls have, therefore, been subjected to less adjustment than for Mmsp, by the weighting. Coverage ratios are essentially the inverse of the second-stage ratio estimate factors.
- Coverage loss in population can be considered in categories: Persons in missed sample units and mispersons in interviewed sample units. There is some evidence<sup>11</sup> to indicate that the missed rate for females about the same in these two categories and the misrate in interviewed households is higher for males than for females. This suggests that the regular CPS second-stage factor for females has less effect on the Fmsp counts than the male ratios affect Mmsp counts.
- Second-stage ratio estimate factors are actually computed and applied separately, by rotation group. These factors are observed to be less variable for females than for males among the rotation groups in a monthly sample and, in a similar way, among the monthly averages month-to-month.

**March family weighting system**—The foregoing observations are the basis of the *March family weighting procedure* summarized in this section.

Because tabulations of the supplementary data collected in the March are produced by labor force status, the family weighting uses an additional stage of ratio estimation to avoid showing different sets of labor force estimates for the same month. An additional ratio estimate does not necessarily improve the reliability of tabulations prepared for the supplementary data. The regular March CPS results, by age, sex, and collapsed race, for four labor force status categories, become the controls in the additional ratio estimate: Total unemployed, employed in agriculture, employed in nonagriculture, and not in the labor force. The age-sex-collapsed race categories are the same as in 1970 V-3 except that the age classes 22-23 and 24-29 years old are used instead of the age cells shown in the table. The weights

<sup>9</sup>Estimates for members of the Armed Forces within the scope of the March survey are prepared without the intermediate and final second stages of ratio estimation; the current independent population controls do not include members of the Armed Forces.

<sup>10</sup>This is approximately true for all months, but actually true only for the 1970 CPS sample, because of the treatment of family members in the Armed Forces.

<sup>11</sup>Unpublished findings in the 1970 Census CPS March Study.

produced by the regular monthly estimation procedure, described in this chapter, are modified by this ratio estimate and other estimation procedures.

The steps outlined in the following are performed for all rotation groups combined. The process begins using the weights established in the regular CPS processing on page 60, just prior to performing the composite estimator.

1. The first step repeats the intermediate ratio estimates for Negro and other races to the separate independent estimates. (See p. 60 of this ch.<sup>12</sup>) The weights from the estimators for the not-white population are modified by this ratio estimate and used in the subsequent estimation steps. The age categories used in this estimate are the same as table V-2 for race reported Negro, but, for race reported other, the age categories are 14-19, 20-29, 30-39, 40-49, 50-59, and 60 years old and over.

2. The weight for Fmsp in husband-wife households is assigned to the male partner. The result of this is usually to substitute a smaller weight for the Mmsp.

3. The weights for spouses in husband-wife households are tabulated to form the following ratios for each age, collapsed-race, labor force status category:  $\frac{Fmsp}{Mmsp}$ . These

ratios (usually less than 1.0) are applied to the weights carried by male heads other than Mmsp; this group of males is defined as Omh (other male heads).

Multiplying by these ratios represents a proper change (usually a reduction) in the second-stage ratio estimate to adjust the effect of undercoverage of Omh by the same factor as for Mmsp.

4. If all males are now tabulated (with the two weight adjustments for male heads), the total would not agree with the control figures. (In general, both adjustments are expected to reduce the numbers of males.) The control figures used here are the age, collapsed-race, labor force status totals produced in the regular March CPS. Consistency with these male control totals is established by a ratio adjustment applied to the remaining males (i.e., males other than Mmsp and Omh). The numerator of the adjustment ratio is the difference between the original CPS-labor force status result, by age and collapsed-race, and the sum of the figures newly established with the adjusted weights for Mmsp and Omh. Some collapsing of cells is done in this ratio adjustment so that complete agreement is not achieved in all subordinate categories.

5. All tabulations of the March supplemental data use the weights as adjusted by the March family weighting procedure; household and family characteristics use the adjusted weight of the family head.

**Estimates for children in March**—March estimates for children under 14 years old are prepared using the following steps:

1. The process begins with the weights produced by the March family weighting system (See p. 65.) The weight derived for the wife of the head (if present), otherwise the weight of the head, is assigned to the child's record.

<sup>12</sup>The terms used to describe race and collapsed race are discussed in the section on adjustment for type A noninterviews in this chapter, p. 68.

2. Separate ratio adjustments are applied using independent established population controls for children for Negro for other. There are 12 age cells, by sex, for Negro and age cells, by sex, for other. A total of 34 age-sex cells used in the ratio estimate for all rotation groups combined. The age categories are for—

Race reported Negro	Race reported other
under 1 year old	2 years old and under
1	3-4
2	5-6
3	7-9
4	10-13
5	
6	
7	
8	
9	
10-11	
12-13	

3. A second-stage ratio estimate is next applied in the 48 estimate cells given by age, sex, and collapsed race age categories are the same as the 12 groups given for reported as Negro in item (2).

### Monthly Estimates of Occupied Housing Units

National estimates of total housing units, by occupancy tenure, are used to provide consistency for estimated housing data in Census Bureau programs (as, for example, in the A Housing Survey [51]).

**Estimation procedure**—Estimates of occupied housing for this purpose are produced from the CPS as follows:

1. The process begins using the weights developed in regular monthly processing for the CPS at the completion of the second stage of ratio estimation. (See p. 60.)
2. For married-spouse-present households, the weight for the female partner (wife) is assigned to the household head; weights for other household heads are unchanged.
3. A tabulation of the weight of household heads (as adjusted for all households) provides the estimated number of occupied housing units. (Type A, B, and C noninterview omitted from this tabulation.)

**Commonly used terms**—The weight of the head, including those adjusted by item (2) is referred to as the *household indicator weight*. The term "principal person" has been given to the wife in husband-wife households and to the head in all households. The *principal person weight* is another name for the household indicator weight. Note that using the March weight, a tabulation of the males and females will yield consistent with the independent population controls; this does not result when using the household indicator weight.

### Quarterly Estimates of Vacancy Rates

The type B noninterview units detected each month in the CPS provide a basis for estimating the number of current vacancies. These units, along with the estimates of occupied housing units, are used to compute the quarterly vacancy rates.

units, described in the preceding section, are the source of Census Bureau publications on estimated rental and owner vacancy rates.<sup>13</sup>

**Source of information on vacancies**—Each month the vacant living quarters and those occupied entirely by persons ineligible for the CPS are recorded as type B noninterviews. (See items 15 and 16 in forms CPS-1, fig. IV-7.) The interviewer fills a vacancy survey form, HVS-1 (fig. IV-8) to obtain housing characteristics for those type B noninterview housing units that are classified as "year-round" and "vacant-regular," "vacant storage of household furniture," or "temporarily occupied by persons with usual residence elsewhere." (See the section on noninterviews in ch. IV, p. 33.)

**Computing the vacancy rates**—The vacant units are accumulated for the 3 months of a quarter without weights except for special weights previously described on page 56 of this chapter. However, the occupied housing unit estimates needed to compute vacancy rates are weighted estimates. To remove these weights, the weighted sum of the occupied housing unit estimates for the 3 months of a quarter are divided by an average weight so that the occupied and vacant sample counts can be combined to compute vacancy rates. The average weight is computed to recognize that the different A- and C-design Samples used in a CPS month have slightly different selection probabilities. (See app. F, table F-1.)

**Owner and renter vacancy rates**—The tenure for occupied units (that is, whether occupied by owner, renter, etc.) is determined in 1 month of each quarter in the CPS sample. The owner and renter occupied units, used in determining the owner and renter vacancy rates, are derived by allocating total occupied units to owner and renter based on the proportions observed in the 1 month this information is obtained by interview.

### Estimates for Children in the September and October Supplements

In September and October of each year, the CPS interview includes a series of questions for children under 14 years old. In September, the questions concern the immunization history for

children under 14 years old at households in six of the rotation groups. In October, questions on school enrollment asked of children 3 through 13 years old at sample households in all eight rotation groups. Counts of children are also obtained each January, April, and July.

The procedure used to prepare estimates of children in each of these months is, as follows:

**Step 1. Assign household indicator weights to children**—The process begins with the weights produced in the monthly CPS weighting at the completion of the second stage ratio estimation. (See p. 60.) A weight is assigned to the child record equal to the weight derived for the wife of the head (if present), otherwise the weight of the head is used. This is equivalent to assigning the household indicator weight to the child.

**Step 2. Ratio estimate by age and sex for Negro and other**—Separate ratio adjustments are applied, using independent established population estimates for children with race reported as Negro and other.<sup>14</sup> The procedure is similar to the estimate for children in March, given in this chapter on page 66, except that the ratio estimates are made separately by rotation sets (pairs of rotation groups), according to the month in sample status of the rotation groups.

Rotation sets for September comprise rotation groups are—

2nd and 6th month in sample.

3rd and 7th month in sample.

4th and 8th month in sample.

For the other months, the additional set, comprising the rotation groups that are 1st and 5th month in sample, is also used.

The ratio adjustment is computed in each of the 34 cells of the ratio estimate by tabulating the children's records with the household indicator weight. Although the estimate is performed by rotation set, the ratio estimate cells are considered large enough to require an acceptable ratio adjustment factor for all of the 24 cells of children reported as Negro. However, for race reported as other, a ratio adjustment factor larger than 3.0 or with a zero denominator can occur and will cause the cells to be combined as follows:

Initial age class	If necessary, combine counts to use ages—	If necessary, further combine to use ages—
2 years and under 3 to 4 years	4 years and under	4 years and under
5 to 6 years 7 to 9 years 10 to 13 years	5 to 9 years and 10 to 13 years 7 to 13 years 7 to 13 years	5 to 13 years

**Step 3. Ratio estimate, by age, sex, and collapsed race**—A final ratio estimate adjusts the weights using 48 independent control figures (12 age cells, by sex, separately for white and not-white children). The age categories are the same as those

specified for children in March on page 66 of this chapter. The weights developed in step (1) are used for white children, and the weights produced in step (2) are used for not-white children in determining the final weight adjustment.

<sup>13</sup> Vacancy rates, by characteristics of the unit, incorporate estimates of occupied units from another survey, the Quarterly Household Survey (QHS) (55).

<sup>14</sup> The race and collapsed race definitions are given on p. 56 of this chapter.

## Chapter VI.

# ESTIMATES FOR STATES AND LOCAL AREAS, BY SUPPLEMENTING THE NATIONAL SAMPLE

### INTRODUCTION

Although in the original design of the CPS the primary objective was the preparation of labor force estimates at the national level, there has always been interest in developing reliable estimates of labor force and other data for SMSA's, States, smaller areas, and combinations of these areas. Because the CPS is a probability sample, it can provide estimates for such areas with measurable sampling error. However, for smaller areas, the reliability may be so unsatisfactory that an estimate based on the national sample for a single month has limited use.

The effective sample size can be increased by averaging the monthly CPS results over time; this system has been used to prepare tabulations for areas below the national level. (See ch. V.) Where such averaging does not provide sufficient reliability or when the length of the period needed to obtain sufficiently reliable averages is excessive, a system of supplementing the sample must be employed.

The Comprehensive Employment and Training Act of 1973, referred to as CETA, provides for distribution of Federal funds to the States with the allocation for each State based on the "relative number of unemployed persons within the State as compared to such numbers in all States" [35].

Through this Act, administered by the Secretary of Labor, funds have been provided to the Census Bureau to improve estimates of unemployment for the States and the District of Columbia by probability sampling methods. Under the reliability criteria specified by the Department of Labor, estimates of annual average labor force data have been published for States and SMSA's.

#### Estimates for 1973

For calendar year 1973, estimates were published for 19 States for which it was determined the CPS national sample could meet the following reliability requirement: Provide estimates of the annual average number of the monthly unemployed with a coefficient of variation of 10 percent or less, assuming an unemployment rate of 5 percent.

The procedure used for these estimates is described in chapter V, page 64.) In general, the 19 States having the largest population (and largest CPS national sample sizes) are—

California	Massachusetts	Ohio
Florida	Michigan	Pennsylvania
Georgia	Minnesota	Texas
Illinois	Missouri	Virginia
Indiana	New Jersey	Washington
Maryland	New York	Wisconsin
	North Carolina	

#### Estimates for 1974 and 1975

For calendar years 1974 and 1975, the national sample was used to provide estimates for 8 States in addition to the 19 States mentioned. The additional States were included partly

because an improved estimation system was introduced and partly by adopting a revised reliability standard for the State estimates. The improved estimator resembles the estimator for calendar year 1973, except that it includes a first stage ratio estimator to the 1970 census population counts for the NS portion of the State. (This State first-stage ratio estimator is described on p. 74.) With the revised estimation procedure, the CPS national sample in these 27 States is considered sufficient to meet the following standard: Provide estimated annual averages of the number of monthly unemployed with a coefficient of variation of 10 percent or less, assuming an unemployment rate of 6 (rather than 5) percent. These eight States are—

Alabama	Louisiana	South Carolina
Connecticut	Oklahoma	Tennessee
Kentucky	Oregon	

#### Estimates for 1976 and Subsequent Years

For calendar years 1976 and later, under the CETA, annual averages are to be provided for the 50 States and the District of Columbia with the same reliability standard as given above for calendar year 1974 and 1975 estimates. For the 27 States the estimates are based on the national sample; for the 23 remaining States and the District of Columbia, a supplemental sample is designated each month to bring the reliability to the required standard. The design of the CETA supplementary sample (D-design sample) is explained in this chapter, and the procedure for preparing estimates for each of the 50 States and the District of Columbia for calendar 1976 annual averages is discussed on page 73. (The location of the supplemental PSU's is discussed in app. M.)

The D-design sample for the CETA was prepared originally to provide estimates consistent with the reliability requirement specified for calendar year 1973 estimates on this page. As of September 1975, supplemental samples were being interviewed each month in the District of Columbia and in the States that required the supplement to meet this condition. Effective Aug. 1976, the supplemental sample was revised to provide estimates under the reliability standard given for calendar year 1974 and 1975 estimates given above; as a result, the CETA sample could be reduced in each of the States and the District of Columbia and the supplemental sample PSU's could be eliminated entirely in three States (Oklahoma, Oregon, and South Carolina). The number of units in the CETA sample under these alternatives is given in appendix table J-1.

### DESIGN OF THE CETA SUPPLEMENTARY SAMPLE

#### Design Requirements for State Samples

Alternative designs that could be used to supplement the national sample within the States were evaluated on the basis of reliability, cost, and the need for consistency in the estimates.

over all States. Three principles have guided the development of the sample design for CETA.

1. The CETA sample, when combined with the CPS sample in the State, must provide estimates of the annual average number of unemployed with the reliability requirement given for calendar year 1973 estimates on page 68.
2. The sample design and survey procedures for the CETA sample must be compatible with the corresponding features of the CPS national sample. This is to allow use of the interviewer staff, training, sampling, processing, and estimation procedures for both samples.
3. The two specifications given suggested that the supplemental sample should have a rotation system consistent with the CPS. This was given careful consideration because of the impact on cost and consistency of estimates from the national and supplemental parts of the combined sample. As discussed elsewhere (ch. VII), experience has shown that panels of households, interviewed at monthly intervals, have reported levels of unemployment in the first month which differ from that reported in succeeding months. Estimated numbers of unemployed, based on a CPS rotating sample which has only one-eighth of the households interviewed for the first time, would, therefore, have a different expected value than a sample all from first month interviews. If the supplement has an expected value differing from the CPS, the expected value of the State estimate would be related to how large the supplement is, relative to the total State sample. In practice, the amount of supplementation varies from State to State, and an inequitable distribution of funds could result.

#### Features of the Supplemental Sample for the States

The design of the sample selected to supplement the CPS within the 23 States and the District of Columbia is summarized, as follows:

1. For most of the States, a two-stage sample design has been used. In the first stage, additional sample PSU's are selected proportionate to 1970 census population from strata defined within the State (controlled selection is not used in selecting PSU's); in the second stage, a sample of USU's is selected within the sample PSU. For a few of the States requiring supplementation, the first of these stages, the sampling of PSU's, is omitted either because the entire State is SR in the national sample (Rhode Island and the District of Columbia) or because all, or nearly all, PSU's are large enough to be SR in the State sample (Delaware, New Hampshire, and Vermont).
2. The PSU's (both sample and nonsample) defined for the national sample design have been maintained for the State sample with two exceptions —
  - In those few PSU's originally made up of counties from different States, each State part is treated as a separate PSU for the State samples.
  - Maui County, Hawaii, treated as a separate PSU in the national sample design, has been divided into two PSU's for administrative convenience for the State sample.

3. The stratification and selection of supplementary sample PSU's in a State take account of national sample PSU's in the State, that is, the supplementary PSU's are sample dependent on the national sample PSU's. A different selection of national sample PSU's could generate a different set of strata and supplementary sample PSU's within the State, although the procedures for stratification and sampling would be the same.

This system was chosen over an alternative which would select an independent sample of supplemental PSU's for a restratification of all PSU's within the State. The chief consideration in making this decision is that the dependent system is considered to be more flexible than the alternative if future changes are called for in the survey objective and is also more sensitive to the difference between the expected and the actual national sample in the State. The latter consideration is important in determining the supplemental sample requirements.

4. The sample units are identified in eight rotation groups which rotate in the same way as the national sample. Sample USU's are identified as "D" with a numeric descriptor in the same way as the A- and C-Samples are identified as A and C (see fig. III-1 and app. E).
5. USU's are selected from the same sampling frames and by the same systematic sampling procedures as for the national sample. Where supplemental USU's are sampled within a PSU that is also in the national sample, the selections are made to avoid choosing the same USU in both samples.
6. In general, the supplemental sample is selected so that when combined with the existing national sample, a self-weighting sample is produced for the State.
7. A separate PSU-replacement schedule is established for the supplemental sample. (See table M-2, app. M.)

#### Stratification and Selection of Supplemental Sample PSU's

The number of strata and sample USU's required is determined by assuming that a simple inflation estimate is used to prepare the published figures for the State. The variance of the estimate, from the combined CPS and supplemental samples for the State, has been treated as having two variance components (from the two stages of sampling): The variance between PSU within strata and the variance among USU's within PSU's. To approximate the sample size required to yield the given reliability estimates of the between- and within-variance components, the estimates are necessary; in both instances the variance estimates have been derived from the variance of simple random sampling, adjusted by a design effect<sup>1</sup> to reflect the variance of the estimate.

Iterative procedures for each State—The procedure for stratification and selection of the supplementary sample PSU's

<sup>1</sup> The design effect is a factor which multiplies the variance of a simple random sample of the same size to approximate the increase (occasionally the decrease) in variance from the features of the sample design, e.g., clustering, stratification, selection with probability proportionate to size, estimation procedures, etc. Estimates of sampling errors actually produced for the interviewed sample are discussed in ch. VIII.



summarized in the following steps that are applied separately for each State. The process begins by approximating the sample size that will produce the required reliability in the State. There is a problem in deriving this approximation. The within-PSU variance component of the State sample is assumed to be inversely proportional to the number of sample households (and is treated as a continuous function in this analysis). However, the number of strata in the State is relatively small and the variance between PSU's changes with the number of strata and by amounts difficult to predict. The success of a stratification can be measured only after the stratification has been done. As a result, the sample is determined by an iterative process with the steps of the following paragraphs repeated in sequence until an allocation is obtained which is expected to produce State estimates with the required sampling error.

1. The first trial estimate of the total number of sample housing units needed to produce the required reliability is made, assuming an overall design effect of 1.75 over simple random sampling of persons 16 years old and over. This design effect is supposed to cover the combined effect of both stages of sampling and is taken as a starting point for the iteration.
2. The estimated number of sample units needed, the total units in the State and the self-weighting feature of the sample within the State, determine a sampling fraction that defines stratum sizes consistent with the desired interviewer workload (ranging from 40 to 70 designated-sample units). PSU's with populations larger than this are made SR for the State sample. The sample allocated to the remainder of the State and the desired interviewer workload lead to an estimated number of NSR State strata (and hence NSR sample PSU's) to be defined in the balance of the State.

The national sample strata in the balance of the State are reviewed to insure proper representation in the dependent sampling process. Those A-design NSR strata having a national sample PSU within the State are retained for the State sample and are represented by sample USU's selected from the national sample PSU's, provided the implied sample workload is not too large; if so, the stratum is divided into parts that would yield acceptable workloads with the State stratum becoming the part containing the national sample PSU. The remaining PSU's in such strata and all other PSU's from national sample strata without representation in the State are pooled and grouped into the required number of strata by the computer process for stratification, described later on this page.

3. Estimate the sampling error from the sample determined in (2). The required sampling error is prepared by first approximating the variance among USU's within PSU's and adjusting this result to reflect the estimated annual average. This approximation is then combined with an estimate of the variance between PSU's and substituted for the variance determined in the previous iteration or for the initial approximation in (1).

The between-PSU variance component is calculated by considering all possible samples of PSU's, using the 1970 decennial census population for probabilities and the 1960

census unemployment data for each of the PSU's in the strata.

The within-PSU variance component is estimated by computing the variance of a simple random sample of persons 16 years old and over and multiplying this result by design effect to reflect the impact of the CPS sample design. A design effect of 1.4 was used for estimates of unemployment in every iteration. This figure is consistent with an earlier study of design effects over a period of 3 months of the CPS (calendar years 1967 through 1969 [4]). Selected results from this study are shown in table VI-1.

The impact of annual averages (that affect the within PSU variance only) is assumed consistent with other results of the same study, displayed in table VI-2. These data suggest for example, that the effective sample size from 1 interview during a calendar year is equivalent to 5 independent interviews for unemployment, about 2.2 for the total civilian labor force, etc.

4. If the revised variance estimate from (3) suggests that the coefficient of variation of annual average estimates from the sample derived in the current iteration will exceed 1 percent or will be substantially less than 10 percent, derive a new sample size for the next iteration by calculating new overall design effect, using the revised variance estimate from (3). The process is then repeated from (2). The computer ends the search when it determines the stratification leading to a sample that has a coefficient of variation closest to, but not exceeding, 10 percent.

The sample size, determined from the final iteration specifies the sampling ratios needed to sample within the PSU's to make the sample self-weighting. Where a PSU the CPS sample is also selected for the State, units also designated within the PSU for the CPS are used to satisfy the State sample requirements. Where the CPS sample too small, an additional sample is selected; where the CPS sample is larger than required, it is retained and used in the State estimate with appropriate weight adjustments.

**Computer process for stratification**—The definition of the NSR State strata, other than those portions of the national sample strata retained for the State sample, and all of the steps outlined in (1) thru (4) are performed by the computer. The NSR stratum size is established to provide the desired interview workload indicated. The computer regresses various 19 census social characteristics and economic data [37] for the NSR PSU's in the State to derive equations for predicting unemployment rates, sorts the PSU's by their predicted unemployment rates, and divides the PSU's into the required number of strata approximately equal size, by population [7]. The characteristics considered in the stratification differ by State, depending on observed ability of the regression to predict unemployment. Characteristics frequently among the highest in importance include the proportions of workers in 1970 that were farm, white collar, blue collar, and service workers; the worker-nonworker ratio; and the relative number of families in poverty. The strata produced by the computer were reviewed to insure that characteristics that are apparently important for stratification indeed relevant to the situation in the State.

Table VI-1. Design Effect and Percent of Population, 16 years Old and Over, for Unbiased and Combined First- and Second-Stage Ratio Estimates

(CPS: Average for calendar years 1967-1969)

Table 3. Average for calendar years 1967-1969			
Characteristic	Percent of population 16 years old and over	Design effect <sup>1</sup>	
		Unbiased estimate	After first- and second-stage ratio estimates
<b>CIVILIAN LABOR FORCE</b>			
Total.....	60		
Female.....	22	9.82	1.02
		2.71	.96
<b>EMPLOYED</b>			
Male.....	36	4.22	.29
Negro and other races.....	6.1	5.79	.53
Self-employed.....	5.4	1.85	1.41
<b>Nonagriculture</b>			
Total.....	55	8.80	1.15
Male.....	34	4.29	.44
Wage and salary.....	50	7.99	1.26
Working 35 or more hours.....	40	5.76	1.14
With a job, not at work.....	3.1	1.66	1.37
<b>Agriculture</b>			
Total.....	2.8	4.29	3.15
Male.....	2.3	3.33	2.54
Wage and salary.....	.96	3.25	3.06
Working 35 or more hours.....	1.9	3.47	2.67
<b>UNEMPLOYED</b>			
Total.....	2.3	1.54	1.33
White male.....	.93	1.28	1.19
White female.....	.94	1.26	1.15
Not-white male.....	.22	1.55	1.22
Not-white female.....	.26	1.51	1.23
<b>FARM POPULATION</b>			
Total.....	5.4	7.64	5.59
Male.....	2.8	4.03	2.96
<b>MALE</b>			
Total.....	47	5.21	( <sup>2</sup> )
<b>FEMALE</b>			
Total.....	53	5.91	( <sup>2</sup> )
<b>NEGRO AND OTHER RACES</b>			
Total.....	11	9.72	( <sup>2</sup> )
<b>HOUSEHOLD HEADS</b>			
Total.....	46	3.81	.55

<sup>1</sup> Ratio of the variance of CPS to the variance of simple random sample of persons 16 years old and over for the same sample size (4, table 1).<sup>2</sup> The estimates for independent control figures have no variance.

**TABLE VI-2. Number of Independent CPS Monthly Samples Equivalent to an Average of 3, 6, 9, and 12 Consecutive Months of the CPS**

Characteristic	Variance ratio for average of <sup>1</sup> -			
	3 months	6 months	9 months	12 months
<b>CIVILIAN LABOR FORCE</b>				
Total .....	1.3	1.7	2.0	2.2
Female .....	1.3	1.7	2.1	2.3
<b>EMPLOYED</b>				
Male .....	1.6	2.0	2.4	2.7
Negro and other races .....	1.4	1.9	2.3	2.5
Self-employed .....	1.3	1.7	2.0	2.1
<b>Nonagriculture</b>				
Total .....	1.3	1.7	2.0	2.0
Male .....	1.4	1.7	2.0	2.1
Wage and salary .....	1.3	1.6	1.9	2.0
Working 35 or more hours .....	1.5	2.1	2.4	2.6
With a job, not at work .....	2.3	4.0	6.2	7.7
<b>Agriculture</b>				
Total .....	1.2	1.4	1.6	1.7
Male .....	1.2	1.4	1.6	1.7
Wage and salary .....	1.3	1.6	1.9	2.1
Working 35 or more hours .....	1.3	1.6	1.8	2.0
<b>UNEMPLOYED</b>				
Total .....	2.0	3.2	4.3	5.0
White male .....	2.1	3.3	4.3	5.3
White female .....	2.0	3.3	4.8	6.2
Not-white male .....	2.0	3.4	4.8	5.6
Not-white female .....	2.1	3.6	5.0	5.9
<b>FARM POPULATION</b>				
Total .....	1.1	1.4	1.5	1.6
Male .....	1.2	1.4	1.6	1.6
<b>HOUSEHOLD HEADS</b>				
Total .....	1.4	1.9	2.3	2.5

<sup>1</sup>Rate of variance of the CPS estimate for 1 month to variance of averages for 3, 6, 9, and 12 consecutive months, first- and second-stage combined ratio estimates, averaged for the January 1969-June 1970 CPS (4, table 5)

**Selection of sample PSU's**—When the NSR strata have been defined within the State, a sample PSU is selected with probability proportionate to the 1970 population size in each stratum. The rules for stratifying and selecting the sample PSU's dependent on the PSU's in the national sample allow the following generalizations: All PSU's in the national sample are retained in the State sample of PSU's; all CPS SR PSU's (or portions thereof) remain SR in the State sample; a CPS NSR PSU (whether in the national sample or not) becomes a SR PSU in the

State sample if it is large enough in population to yield an interviewer workload of about 50 to 55 designated sample units.

Over all States, the supplemental sample originally selected to meet the reliability standard, given for calendar year 1973, estimates on page 68, included about 165 new PSU's in a 9 month not in the national sample. In the States where PS were not sampled (Delaware, New Hampshire, and Vermont) the count includes all PSU's in the State not in the national sample.

When it was first introduced, the supplemental sample each month added about 14,000 units assigned for interview to the approximately 55,000 units per month for the national CPS sample. In August 1976, to meet the reliability standard, given for calendar years 1974 and 1975 on page 68, the supplementary sample was reduced to about 11,000 units assigned for interview in 155 PSU's. For a given sample design, the number of PSU's will vary slightly over time because of the replacement of PSU's. (See the replacement schedule, table M-2, app. M.) Appendix table J-1 shows how the national and supplemental samples are distributed by regional office.<sup>2</sup>

## SURVEY OPERATIONS

Survey procedures for the State supplemental samples conform essentially to the processes developed for the national sample. (See ch. IV.) The field interviewer and supervisory staff have been expanded and similar personnel adjustments elsewhere permit the same data collection and processing schedules to be followed.

### Sample Introduced in Three Groups

To reduce the impact on the survey organization that would result from a large sudden increase in workload, the additional sample units were introduced in three groups of States, roughly equal in size, in July, August, and September 1975. The data produced each month from the new sample were accumulated for the annual averages, beginning in January 1976; data collected from the added sample in 1975 were used to verify the expanded sample system and to provide experience for the new interviewers.

### Special Problems in Alaska

Only one PSU from the CPS national sample falls in Alaska (the Anchorage Census Division). The additional State sample PSU's in Alaska that are required for CETA introduce a number of problems because of extreme weather conditions and communications. This concern is reflected in some changes made in sampling as well as survey procedures.

Although every effort is directed at selecting and interviewing the sample by the principles specified for the CPS sample, special situations have occasionally required modifications of the procedures in Alaska. The following illustrates procedures employed in some areas.

Problems in designating the sample and conducting interviews have arisen because of the sparse population, the large area of some 1970 census ED's, and difficulty in associating ED boundaries with the geographic features on the ground. Fortunately, in many of these areas, the population tends to be clustered into small communities that lend themselves to sampling. In these instances, the communities and the living quarters, located in a specific radius around them, are used as blocks or chunks in the sampling procedure discussed in appendix C. In this situation, the community is defined as the segment.

<sup>2</sup> The location of PSU's in which interviews are conducted for the State estimates is discussed in app. M.

The listing of units in a segment, particularly in the situation above, may involve local inquiry for hard to reach units in addition to visiting the location of accessible units.

## ESTIMATION PROCEDURE FOR STATES FOR 1976

Each month, the units in the A-, C-, and D-design samples are interviewed. The estimation procedures, described in ch. V are applied to the results from the A- and C-design samples for 2 States to provide the regular monthly national labor force estimates.<sup>3</sup> The estimator, described in this section, is then applied to the full A-, C-, and D-design samples in the 23 States and the District of Columbia where supplementation is required and to the A- and C-design samples in the remaining 27 States to provide the monthly input to the estimated annual averages for the individual States.

These weighting operations require the reprocessing of a sample records to produce the weights used for State estimates.

### Preliminary Procedures for the Supplemented Areas

The A- and C-design samples from PSU's in the 23 supplemented States and the District of Columbia are combined with the D-design sample from these areas for the following weighting operations.

**Apply basic weights**—Apply the basic weights (i.e., the inverse of the probability of selection) for units in the State samples. For the A- and C-design samples, a new weight replaces the weight entered in the operations of chapter V. Although the combined A-, C-, and D-samples are intended to be self-weighting for each State, preserving the national sample in the A- and C-design PSU's means that the basic weights usually differ by PSU. The weights also vary somewhat by rotation group. (See app. I.)

**Apply special weights**—Apply the appropriate special weights for sample units required because of subsampling. For the A- and C-portion of the sample, these special weights are generally the same as for the national sample, except for the special situations described in step (4) on page 70.

### Preliminary Procedures for the Nonsupplemented States

The A- and C-design sample records from the 27 nonsupplemented States are assigned weights representing the inverse of the overall probability of selection of the national sample including the special weights required. (These weights are described in ch. V, p. 56.)

### Combine the Sample Records for All States

After completing the preliminary procedures for the supplemented areas and the nonsupplemented States, the A-, C-, and

<sup>3</sup> A revised estimation procedure is being considered that would incorporate data from the D-design sample in the monthly national estimates; the additional information is expected to make some small reduction in the sampling error of the national estimates.

D-design records for all States are combined. All subsequent weighting applies to the combined samples for all States.

### Adjustment for Noninterview for States

A noninterview adjustment is applied to represent the type A noninterview units in the sample for the States. The procedure differs from the noninterview adjustment described for national estimates in chapter V, page 56. The State and national noninterview-adjustment procedures take advantage of relationships involving response patterns by residence and collapsed race of the respondent. However, the objectives of the State estimates suggest these relationships should be exploited in different ways to increase the number of interview cases within the State noninterview-adjustment cells.

**Noninterview clusters for States**—These are defined within the State. The States have two sets of noninterview clusters: One set is made up of clusters of SMSA PSU's within the State and the other from the non-SMSA PSU's. Some States have additional noninterview clusters depending on the number of sample units. There are a total of 113 noninterview clusters in the 50 States and the District of Columbia.

**Noninterview adjustment cells for States**—These are defined within each cluster by the same six collapsed race-residence categories described for national noninterviews. For the State system, however, the cells are defined in four rotation sets (pairs of rotation groups), rather than by separate rotation groups. Rotation groups in sample the first and fifth, second and sixth, third and seventh, and fourth and eighth months define the four rotation sets. The State system also combines and does not separately distinguish the housing unit and group quarters population.

There are a total of 48 noninterview adjustment cells for a typical State: Twenty-four cells for SMSA PSU's (central city of SMSA, balance of SMSA urban, and balance rural, by collapsed race, for four rotation sets) and 24 cells for non-SMSA PSU's (urban, rural nonfarm, and rural farm, by collapsed race, for four rotation sets).

**State noninterview adjustment factors**—The factors are computed within each of the noninterview adjustment cells. The weighted tallies of the interviewed and noninterviewed units are used to compute—

$$\frac{\text{Interviewed households} + \text{noninterviewed households}}{\text{Interviewed households}}$$

If the computed factor for a cell is based on 30 or more unweighted sample cases (interview plus noninterview)<sup>4</sup> and the computed factor is less than 2.0, the factor is considered acceptable. The weight of each interviewed case in the cell is replaced by the product of the factor and the weight.

If either of these conditions is not met, noninterview adjustment cells are combined in a specific sequence with the resulting factor checked at each step until an acceptable adjustment factor

<sup>4</sup>This condition is specified for the State noninterview adjustment but has not been specified for the national sample

is computed. The weight of each record in the combined cells is replaced by the product of the weight and the noninterview-adjustment factor.

The order of combining the cells differs from the system used for the national sample (See ch. V, p. 56.) In general, for State estimates, the noninterview-adjustment cells are collapsed, when necessary, by first combining rotation sets, then residence, and then collapsed-race categories. The changed order of combining cells for State estimates reflects the desire to base adjustments on interviews completed within the State, by collapsed race: where possible, by collapsed race and residence, and further, by collapsed race, residence, and month in sample.

### First-Stage Ratio Adjustment to 1970 Census Population Counts for the States

Ratio estimates, similar in form and objectives to those used for national estimates, are used to improve the estimates for each of the States. Some adaptations of the national procedure are necessary to accommodate the more restricted sample sizes in the States.

The first-stage ratio adjustment is introduced to reduce the component of variation arising out of sampling NSR PSU's. The estimator is described in the same terms as the first-stage ratio adjustment for national estimates (ch. V, p. 58) with the following modifications.

**Computation of factors**—A separate factor is computed for each of the 12 ratio estimate cells in the State. The factor for a cell is the ratio of the 1970 census population of the cell in the NSR portion of the State to the corresponding estimate of the population, based on the census figures for the sample PSU's. The A- and C-design and, where appropriate, the D-design sample PSU's are used to compute the denominator of the ratio.

**Conditions for acceptable factors**—To be accepted for use in the weighting for the State, each ratio estimate factor must satisfy two conditions: The denominator of the factor must have nonzero contributions from at least three sample PSU's, and the computed factor can be no larger than 2.0. If either condition is not met, ratio estimate cells are combined (numerators and denominators of the ratios are combined) until an acceptable factor is produced for the combined cell. Finally, the weight of each sample person (or household) in the given ratio-estimate cell is replaced by the product of the weight and the appropriate ratio estimate factor.

**Combining ratio estimate cells**—The order of combining ratio estimate cells to arrive at acceptable factors treats as most important the retention of separate first-stage adjustments by urban and rural, then further by SMSA and non-SMSA, further by the separate urban and rural classifications within these areas, and further by collapsed race within each of these categories. Tables VI-3 and VI-4 show the first-stage ratio adjustment factors for the States in effect as of August 1976; the factors are changed to reflect the PSU replacements. (See the PSU-replacement schedule in app. table M-2.) Identical factors for the separate cells show how the ratio estimate cells have been combined. The factors for some of the States in the Northeast Census Region exhibit adaptations of the ratio adjustment procedure to resolve special estimation problems.

TABLE VI-3. First-Stage Ratio Estimate Factors for State Estimates, SMSA PSU's: August 1976

State <sup>1</sup>	Central city		Balance of SMSA			
	White	Not-white	Urban		Rural	
			White	Not-white	White	Not-White
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Alabama .....	0.8758	0.8745	0.8758	0.8745	0.7896	0.9790
California .....	.8955	.8047	.8955	.8047	1.0659	.8894
Georgia .....	1.0422	.8731	1.0422	.8731	.8524	.9905
Illinois .....	1.2134	1.1767	1.2134	1.1767	1.1121	.8838
Indiana .....	.8457	.7494	.7025	1.6380	.6442	.4900
Iowa .....	.9756	.7606	.9756	.7606	1.0271	1.1239
Kentucky .....	.8415	.7638	.8415	.7638	.9708	.6682
Louisiana .....	.9695	.7807	.9695	.7807	.7067	.7174
Michigan .....	1.0659	1.2653	1.1618	1.1618	.9629	1.1744
Missouri .....	1.2583	1.0187	1.2583	1.0187	.7230	.5790
New Jersey .....	.9613	1.5364	.9613	1.5364	.9441	.7643
North Carolina .....	1.0904	.9894	1.0307	1.0802	.8507	1.1160
Ohio .....	.8381	.8607	.6910	.5551	.6854	.8270
Oklahoma .....	1.1480	1.4271	1.1480	1.4271	1.3856	1.0622
Oregon .....	.8995	1.1489	.8995	1.1489	.6417	.8148
Pennsylvania .....	1.0662	.7239	1.0662	.7239	1.0039	.6550
Texas .....	1.0769	1.0306	.8979	1.0362	1.0729	1.3436
Virginia .....	.4888	.6135	.6154	1.5779	.4290	.5290
Wisconsin .....	.6690	.6831	.4456	.4003	.7355	.3823

<sup>1</sup>States not shown do not have MSR PSU's of this type in the State sample

### Second-Stage Ratio Adjustment to Current Independent National Estimates of the Population

Unlike the first-stage ratio estimates, which are made on a State-by-State basis, the second-stage ratio estimate is prepared for the Nation as a whole. This ratio adjustment has the same objectives, uses the same current independent national estimates of the population, and the steps are executed in essentially the same way as the corresponding adjustments for the national sample, described in chapter V, page 58. The two differences from the A- and C-design adjustment procedure are that sample records from the A-, C-, and D-design samples are used, and the weights developed in the preceding sections of this chapter are used to compute the factors.

### Composite Estimator for States

The composite estimator (as described in ch. V, p. 62, for national estimates) is applied separately each month for each of the 50 States and the District of Columbia

### Annual Average of State Monthly Estimates

The arithmetic mean of the 12 monthly composite estimates in a calendar year is computed for each of the States. When required, 6-month and quarterly averages are also computed as arithmetic means.

### Ratio Estimate to Current Independent Estimates of the State Population

The final step in the estimation process is a ratio adjustment to independent estimates of the civilian noninstitutional population of the State 16 years old and over. The independent estimates are established for the population as of July of the calendar year. At present, independent estimates are available only for the total population 16 years old and over. Further detail, by sex, age, and race, may be incorporated if sufficiently reliable control figures for the additional categories become available by State. The ratio estimate is applied to the annual averages by formula (V/6).

Table VI-4. First-Stage Ratio Estimate Factors for State Estimates, Non-SMSA PSU's:  
August 1976

State <sup>1</sup>	Urban		Rural nonfarm		Rural farm	
	White	Not-white	White	Not-white	White	Not-white
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Alabama	0.8758	0.8745	0.7896	0.9790	0.7896	0.9790
Alaska	.6485	.6205	1.0704	1.1950	1.0704	1.1950
Arizona	.9916	1.3420	.8514	1.4844	.7487	1.3060
Arkansas	.9104	.9887	1.0391	1.1692	1.0391	1.1692
California	.8955	.8047	1.0659	.8894	1.0659	.8894
Colorado	.9934	1.6453	.9934	1.6453	.9934	1.6453
Connecticut <sup>2</sup>	.3565	.3565	.3565	.3565	.3565	.3565
Florida	.9915	1.0427	.8208	.8361	1.0357	1.0357
Georgia	1.0422	.8731	.8524	.9905	.8524	.9905
Hawaii	.7928	1.1056	.7928	1.1056	.7928	1.1056
Idaho	1.3698	1.5610	.8921	.8331	.8872	.7639
Illinois	1.2134	1.1767	1.1121	.8838	1.1121	.8838
Indiana	.8660	.8430	.8576	.9485	.9283	.9386
Iowa	.9756	.7606	1.0271	1.1239	1.0271	1.1239
Kansas	1.0599	.9019	.9200	.9324	1.0293	.7920
Kentucky	.8415	.7638	.9708	.6682	.9708	.6682
Louisiana	.9695	.7807	.7067	.7174	.7067	.7174
Maine	.8872	1.4096	1.0416	1.0416	1.0416	1.0416
Maryland	1.4350	1.8329	1.1767	1.1838	1.1906	.5406
Massachusetts <sup>3</sup>	1.1801	1.3366	1.3227	1.6372	1.2693	1.2693
Michigan	.7970	.8504	.9400	.9891	.7883	1.1378
Minnesota	1.0021	.6649	1.0284	.2946	1.0284	.2946
Mississippi	1.0560	1.0481	.9097	.9957	.9805	1.4192
Missouri	1.2583	1.0187	.7230	1.5790	.7230	.5790
Montana	.7746	1.0890	1.3177	1.0923	.8269	1.0923
Nebraska	1.0466	.5022	.9752	1.3959	.9752	1.3959
Nevada	1.0131	.8854	1.0131	.8554	1.0131	.8554
New Jersey	.9613	1.5364	.9441	.7643	.9441	.7643
New Mexico	1.1472	.7761	.9514	.7676	.9112	.5687
New York	1.1311	1.5792	1.3422	1.6300	1.2317	.8496
North Carolina	1.0418	1.0781	1.1485	.9628	1.0008	.8457
North Dakota	.8944	1.2526	.9969	.6600	1.1380	.7545
Ohio	1.0352	.8853	1.0741	.8915	1.2056	.8847
Oklahoma	1.1480	1.4271	1.3856	1.0622	1.3856	1.0622
Oregon	.8995	1.1489	.6417	.8148	.6417	.8148
Pennsylvania	1.0662	.7239	1.0039	.6550	1.0039	.6550
South Carolina	.9876	1.1740	.9304	1.0310	1.0294	1.1124
South Dakota	.8838	.4471	1.0734	1.0680	1.1110	.7758
Tennessee	1.3659	1.6654	1.0888	1.1663	.9034	1.3474
Texas	.8243	.8998	.9248	.7689	.7297	.8318
Utah	1.6954	1.6954	.8760	.3029	1.3002	.2565
Virginia	1.7858	1.1243	1.6404	.9320	1.5056	.7969
Washington	1.4964	1.3148	.8159	.5286	.8278	.5525
West Virginia	1.3540	1.8014	.9499	.8170	.8243	.7216
Wisconsin	1.1233	.7452	1.5996	1.5996	1.8671	1.8671
Wyoming	.8143	1.0249	1.1218	1.0324	1.4075	1.0324

<sup>1</sup> States not shown do not have NSR PSU's of this type in the State sample<sup>2</sup> All of Connecticut is in sample, and all PSU's except Windham County are self-representing. Windham County is in the A-design sample and is non-SMSA, the factor shown is 1.5 times the probability of selecting the PSU from the A-design stratum<sup>3</sup> All sample PSU's in Massachusetts are self-representing, all of the State is in sample except Barnstable, Dukes, Franklin, and Nantucket Counties (all of these are non-SMSA PSU's). For this State only, the factors shown are applied to non-SMSA sample records in all PSU's (SR as well as NSR)

## Chapter VII.

# ACCURACY OF THE CPS ESTIMATES

### INTRODUCTION

The accuracy of a given estimate for a characteristic is measured by the expected size of the difference between the estimate and the true value of the characteristic in the population. Accordingly, it seems desirable to summarize first what is known about the source and size of the various components of error that affect CPS estimates, and follow this by indicating how these various errors combine into the total mean square error. The mean square error is the technical name for the statistical measure of accuracy. Not considered in this chapter and not part of the mean square error concept are considerations of relevance—that is, how well the particular concepts used in the CPS meet the needs of analysts—for example, the definition of employment and unemployment.

The *precision* of estimates from the CPS is measured by the total variance. The total variance reflects the variability in the estimates that would result by carrying out the CPS in each of all possible samples that could occur within the CPS survey design. All aspects of random variation are considered, including variation in the particular sample of individuals selected, the assignment of interviewers to these individuals, who the respondent is, and the responses given. Each source of variation operates within the constraints of the CPS survey design, i.e., the sample selection method; the questionnaire used; the training, selection, and supervision of interviewers; the extent of coverage loss and noninterview rates; the method of processing the data; and the estimation procedure. *Bias*, on the other hand refers to how closely the average result from all of these possible samples would agree with the truth.

As a matter of convenience in presentation, the various components of error that affect CPS estimates have been grouped into two categories: Those errors that contribute to the reliability of the estimates (alternatively referred to as the precision of the estimates), and those errors that contribute to bias. The section on precision of the CPS estimates in this chapter (pp. 77-80) deals with errors as they affect reliability. The section on pages 81-88 deals with errors that produce bias, and the section on pages 88-89 discusses how measures of reliability and bias combine mathematically to produce a total mean square error. Given the fact that the CPS survey is based on a probability sample, the overall measures of precision are entirely objective. At the present time, the measures for the national sample are computed from the sample itself, and, since the sample is quite large and the computations can be routinely repeated over the months, the measures are fairly accurate. (See ch. VIII, p. 80.)

Whereas measures of the reliability of CPS estimates are relatively easy to determine, it is not so easy to measure the bias in CPS data. The measurement of bias implies knowledge or assumptions regarding truth that go beyond data from the sample survey. The section on bias in CPS estimates discusses sources of potential bias in the CPS and summarizes results from special research and evaluation studies conducted for the purpose of measuring some of these biases in the CPS data. Since measures of the CPS bias are incomplete and inexact, this

section also discusses related statistics that put outside limits on the effects of bias and indicate the relative importance of these biases.

The measure of accuracy of a CPS estimate, the mean square error, is computable from the formula provided on page 88, given the reliability of the estimate and the amount of the bias. Typically, the amount or direction of bias is not known precisely; however, the section on the mean square error of the CPS estimates on pages 88-89 shows that using the measure of precision as the measure of accuracy does not appreciably affect the interpretation of the survey results, provided the bias is not large relative to the measure of precision.

### PRECISION OF THE CPS ESTIMATES

#### Concept of Sampling Variability

Assuming the measurement of a characteristic for each individual included in a particular sample is free of error and the interviewers assigned to complete the CPS schedules for the sample individuals exert no influence on the responses, the *sampling variance* is defined as

$$\text{Var}(x) = E(x_k - E x_k)^2 \quad (\text{VII/1})$$

Here  $x_k$  denotes the estimate of characteristic  $X$  derived from the  $k$ -th possible sample. The operator,  $E$ , denotes the *expected value*: The sum of the values for each of all possible samples, each of which is weighted by the probability that it can occur. The procedure used to estimate  $x_k$  can encompass any or all of the stages of estimation discussed in chapter V; the variance obtained will reflect the stages of estimation considered.

If all of the estimates  $x_k$  are known, formula (VII/1) will measure the precision of the estimate derived from a particular sample. This follows, because, under the assumptions, the variability of the estimates results only from the chance selection of a particular sample from among all possible samples. The square root of the sampling variance is the measure most commonly used to define the sampling error and is alternatively called the *standard error* of the estimate.

#### Summary of the Response Variance Model

To assume that the measurement processes are error-free and that interviewers exert no influence on the responses obtained is not reasonable, in view of the substantial body of research data to the contrary. Thus, in discussing the precision of CPS estimates, one must recognize other sources of error in addition to those introduced by the variability arising from the selection of the particular sample of individuals.

To account for the variability in responses of the sample person to a given question, we postulate a survey procedure (not attainable in the real world) in which the process of recording a response for a sample person is repeatable and gives rise to a random variable whose value at trial  $t$  is independent of its value at any other trial. That is, conceptually we imagine that a large



number of random, independent repetitions (trials) of the measurement on a sample person are defined, and, as a result, a population of responses for each sample person is generated. We also postulate that each of the conceptually repeatable trials is conducted under the same set of general survey conditions that include the questionnaire used, the training given to the interviewers, the method of recording the observations, etc. [24].

As a result, we deal with a random variable whose value equals the measurement obtained for the  $i$ -th sample person by the  $j$ -th interviewer at the  $t$ -th trial under the same general set of survey conditions  $G$ . Under this model, a survey estimate can be viewed as the result of a probability sample of the population of persons, interviewers, and responses of each sample person. Viewed in this way, the estimate of characteristic  $X$ , assuming a simple random sample of  $n$  persons out of a population of  $N$ , is

$$x_{tG} = \frac{N}{n} \sum_{j=1}^b \sum_{i=1}^{\bar{n}} x_{ijtG} \quad (VII/2)$$

where

- $x_{ijtG}$  = The measurement value for characteristic  $X$  obtained for the  $i$ -th sample person by the  $j$ -th interviewer at the  $t$ -th trial under the set of general survey conditions  $G$ .
- $b$  = The number of interviewers selected.
- $\bar{n}$  = The average number of sample persons randomly assigned to each interviewer.
- $n$  =  $b\bar{n}$ .

The total variance of this estimate can be shown (see [6, sec. 13.14]), under certain assumptions, to be approximately

$$\text{Var}(x_{tG}) \doteq \frac{N^2}{n} \left[ \sigma_R^2 + (\bar{n}-1) \rho_R \sigma_R^2 + \sigma_S^2 \right] \quad (VII/3)$$

In expression (VII/3),  $\sigma_R^2$  is the *simple response variance*. Conceptually, it is the variance of responses for a given person generated by repetition of the survey and averaged over all persons in the population.

The term  $\rho_R \sigma_R^2$  is referred to as the *correlated component of response variance*, and it measures the contribution to the total variance of any correlation between the measurement values obtained for different persons in any interviewer's assignment. This term reflects the extent to which interviewers influence the recorded observations.<sup>1</sup> For example, a CPS interviewer who inquires about labor force status and implies, by his manner, that he does not expect to find a housewife with children gainfully employed may tend to record fewer employed women than an interviewer who presumes that every adult should be gainfully employed.

The term  $\sigma_S^2$  is referred to as the *sampling variance*. It is the variance among expected responses for all persons in the population where, for a given person, the expected response is the

<sup>1</sup>Additional correlation between measurement values could be introduced at other stages of data collection and processing, for example, by field supervisors or coders [25]. For simplicity, the potential impact of these on the total variance has been ignored.

average of all responses obtained from the conceptual repetition of the survey.

The model discussed here does not conform to the actual CF sample design given in chapter II. For example, the assumption of simple random sampling of persons and the assumption that sample persons are randomly assigned to interviewers are not met. However, the point of the discussion is to introduce the concept of simple response variance and of correlated response variance which, along with sampling variance, contribute to the total variance of the survey estimate. One considers all the components in attempting to achieve maximum precision with the resources available.

### Estimates of Total Variance of the CPS

A major advantage of probability sampling methods is that an objective measure of the precision can be computed from sample data itself. The method of calculating estimates of variance for the CPS has taken many forms over the life of the survey as more sophisticated methods have become available. A discussion of the current procedures is given in chapter VIII.

The formulas used for estimating variances for the CPS include the components due to simple response variance and sampling variance. However, the method currently in use reflects the correlated response variance component only from NSR strata; this component is not reflected in the estimates of variance for the SR strata. For this reason, the estimates of variance calculated from the CPS itself tend to underestimate the total variance. However, the variance computation methods described in chapter VIII tend to overestimate the sampling variance components and may offset, to some extent, the inherent understatement of total variance. Tables VIII-1 through VIII-4 show variance of the composite estimate for the U.S. level and month-to-month change for selected characteristics for calendar year 1975.

It is not possible to provide separate estimates of each variance component as a part of the routine variance estimation procedure. The CPS survey procedures are designed to minimize the reliability of estimates for a specific class of data within a fixed budget. The actual measurement of the reliability is treated as an important, although secondary, consideration. Specially designed experimental studies are usually required to estimate the magnitude of each of the components of variance relative to the others. Separate studies and control procedures instituted to further the primary objective of the CPS, some offer a means of investigating the components of error.

### Estimates of the Response Variance Components

The following discussion deals with two separate investigations of the response variance components that allow general conclusions to be drawn about the relative magnitudes of these variance components.

Estimates of the correlated component of response variance—Unbiased estimates of the correlated component of response variance in the CPS could be computed from repetitions of the CPS sample, each having the same survey procedures independently applied. To do this with two repetitions would require selecting two independent national samples and assigning the units to a random sample of interviewers.

conducting the identical CPS survey process on the two samples under the same survey conditions (e.g., all interviewers use the same questionnaires with interviews conducted during the same period of time). All interviewers should have undergone the same procedure for selection, training, and supervision and have equal familiarity with CPS procedures.<sup>2</sup>

The existence of the Monthly Labor Survey (MLS), carried out in the last 6 months of 1966 concurrently with the CPS, provided a basis for making estimates of the correlated components of response variance in the CPS [61]. The field procedures for the two surveys were basically the same, and the questionnaires were essentially the same in all aspects except for some modification in definitions and questions pertaining to unemployment. There were 143 primary sampling units in which both surveys operated, independent samples of USU's were selected within these PSU's, and the interviewers in the two surveys were independently selected, trained, and supervised. It seemed reasonable to assume that, in each of the 143 PSU's, the two surveys provided nearly independent estimates with roughly the same expected values for a number of different characteristics. Actual differences in the expected values could be anticipated since the CPS interviewers were more experienced on the whole than the MLS interviewers, and the questionnaires used in the two surveys were slightly different, especially with respect to the labor force items.

This study provided estimates of the ratio of the correlated component of response variance to the sum of the sampling variance and the simple response variance. That is, in terms of formula (VII/3), the study estimated a ratio of the form

$$\frac{(n-1)\rho_R \sigma_R^2}{\sigma_R^2 + \sigma_S^2} \quad (\text{VII/4})$$

For example, if the ratio for a particular characteristic is 0.4, the correlated component of response variance is estimated at 40 percent of the simple response variance plus the sampling variance.<sup>3</sup> The results of this study are somewhat limited due to the differences between the surveys, particularly for labor force characteristics, and the large sampling errors associated with many of the estimated ratios. There are, however, estimated ratios with sampling errors small enough to be considered reliable which indicate that interviewers may contribute substantially to the total variability of the CPS estimates. The estimated ratios and their associated standard errors are given in table VII-1.

<sup>2</sup>More efficient experimental designs exist for measuring the correlated component, e.g. [11].

<sup>3</sup>It can be seen from equation (VII/3) that the simple response variance and sampling variance components are decreasing functions of  $n$ , the number of persons in the sample but the correlated component is not. The contribution of the correlated component to the total variance can be expressed as

$$N^2 \frac{(K-1)}{bK} \rho_R \sigma_R^2$$

This component is an essentially decreasing function of the number of interviewers,  $b$ , used in the survey. It follows that, when the number of interviewers is increased so that each is assigned only one person (i.e., when  $b = n$  and  $K = 1$ ), the correlated component of the variance vanishes.

Estimates of the simple response variance component—To obtain an unbiased estimate of the simple response variance, it is necessary to have at least two independent measurements of the characteristic for each sample person using the identical measurement procedure on each trial. It is also necessary that responses to a second interview not be affected by the response obtained on the first interview.

Two flaws occur in every attempt to measure the simple response variance by means of a reinterview. The first is lack of independence between the original interview and the reinterview; a person visited twice within a short period of time and asked the same questions may tend to remember his original responses and repeat them. A second flaw is that the data collection methods used in the original interview and in the reinterview are seldom the same. Both of these flaws exist in the CPS reinterview program data used to measure simple response variance; the latter flaw is exaggerated because the CPS reinterview program is conducted primarily for the control and monitoring of the data collection process rather than for measurement of error. For example, the interviews in the reinterview sample are conducted with self-respondents, where possible, and are obtained largely by telephone by senior interviewers or supervisory personnel. Also, some defects exist in the reinterview survey process itself which introduce unknown biases in the estimates (e.g., a high noninterview rate and exclusion of some interviewers from coverage. (See ch. IV, p. 47.)

Studies attempting to measure the simple response variance component from the CPS reinterview have shown that the use of a dependent reconciliation (as is used in 80 percent of the CPS reinterview) introduces additional correlation between the original and reinterview responses. Consequently, to reduce this source of bias, estimates of the simple response variance are based on units selected for reinterview which are not subject to reconciliation. All data presented in this section are based on this subset of the reinterview sample.

It is useful to consider a schematic representation of the results of the original CPS interview and the reinterview in estimating, say, the number of persons reported as being unemployed.

Results of Original and Reinterview for Identical Persons

Reinterview	Original interview		Total
	Unemployed	Not unemployed	
Unemployed	a	b	a + b
Not unemployed	c	d	c + d
Total	a + c	b + d	n = a + b + c + d

In this schematic representation, the number of identical persons having a difference between the original interview and the reinterview for the characteristic unemployed is given by (b + c). If

TABLE VII-1 Estimated Ratio of the Correlated Component of Response Variance to the Simple Response Variance Plus The Sampling Variance for 143 PSU's in CPS Sample: July-December 1966

Characteristic	Estimated ratio <sup>a</sup>	Estimated standard error of ratio	Characteristic	Estimated ratio <sup>a</sup>	Estimated standard error of ratio
Mental status			Class of worker		
Married with spouse present	0 77	0 12	Private wage worker	0 64	0 10
Single	57	10	Self-employed worker	53	08
Sex relationship			Hours worked		
Male head	94	15	1 to 34 hours, usually full time	66	10
Wife of head	74	11	35 to 40 hours	67	09
Race			41 hours or more	72	10
White	66	10	Nonworker	67	10
Employment status			Living quarters		
At work—full time	80	13	Dwelling unit	1 0	14
At work—part time	60	09	Total population	1 0	14
Housework	65	10			
Other	80	09			

<sup>a</sup>Ratio of variance components, see formula (VII/4).

Source: [61, p. 15-16]

these responses are two independent measurements for identical persons using identical measurement procedures, then an estimate of the simple response variance component is given by  $(b + c)/2n$  and  $E\left(\frac{b + c}{2n}\right) = \sigma_R^2$  as in formula (VII/3) [24].<sup>4</sup> We know that  $(b + c)/2n$ , using CPS reinterview data, underestimates  $\sigma_R^2$ , chiefly because the two responses for identical persons are correlated, so that fewer differences (i.e., entries in the b and c cells) occur than would be expected if the responses were in fact independent.

Apart from its impact on the total variance, the simple response variance is potentially useful to evaluate the precision of the survey measurement methods. It can be used to evaluate the underlying difficulty in assigning individuals to some category of a distribution on the basis of responses to the survey questionnaire; as such, it is an aid in determining whether a concept is sufficiently measurable by a household survey technique to serve the purpose for which the survey data is intended. For characteristics having ordered categories (e.g., number of hours worked last week), the simple response variance is helpful in determining if the detail of the classification is too fine. To provide a basis for this evaluation, the estimated simple response variance is expressed, not in absolute terms, but as a proportion of the estimated total population variance. This ratio is called the *index of inconsistency* and has a theoretical range of 0.0 to 100.0 when

expressed as a percent [24]. The denominator of this ratio, the population variance, is the sum of the simple response variance ( $\sigma_R^2$ ) and the population sampling variance ( $\sigma_S^2$ ). When identical responses are obtained from trial to trial, the simple response variance is zero and the index of inconsistency has a value of zero. As the variability in the classification of an individual over repeated trials increases (and the measurement procedures become less reliable), the value of the index increases until, at the limit, the responses are so variable that the simple response variance equals the total population variance. At the limit, if a single response from N-individuals is required, equivalent information is obtained if one individual is randomly selected and interviewed N-times, independently.

In general, a single value of the index of inconsistency for a characteristic obtained from a set of survey measurement procedures has no meaning by itself. Its meaning is deduced either by comparing it to the value of the index that could be obtained for the characteristic by using the best (or preferred) set of measurement procedures that could be devised or by considering whether the precision of the measurement procedures indicated by the level of the index is still adequate to serve the purposes for which the survey is intended. In the CPS, the index is used for the latter purpose; as a result, the index is used primarily to monitor the measurement procedures over time. Substantial changes in the indexes that persist for several months result in review of field procedures to determine and remedy the cause.

Table VII-2 provides estimates of the index of inconsistency shown as percents for selected labor force characteristics, by quarter, for 1973, 1974, and 1975.

<sup>4</sup>The expression  $(b + c)/n$  is referred to as the gross difference rate, thus, the simple response variance is estimated as one-half the gross difference rate

**TABLE VII-2. Index of Inconsistency for Selected Labor Force Characteristics in the 20-Percent Unreconciled CPS Reinterview Sample, by Quarter: 1973-1975**

Year and quarter	Employed in agriculture		Working, nonagricultural full time		Working, nonagricultural part time	
	Sample estimate	95-percent confidence limits	Sample estimate	95-percent confidence limits	Sample estimate	95-percent confidence limits
<b>1973</b>						
January to February	13 04	7 38 to 23 05	9 72	7 95 to 11 88	25 68	21 50 to 30 69
April to June	9 54	6 05 to 15 03	11 14	9 64 to 12 88	27 20	23 70 to 31 22
July to September	23.63	17.25 to 32.38	12 92	11 18 to 14 92	35 52	31 02 to 40 66
November			9 56	7 08 to 12 91	26 76	20 69 to 34 62
<b>1974</b>						
January to February	11 37	5 43 to 23 82	12 05	9 57 to 15 17	25 95	20 70 to 32 53
April to June	18 43	12 97 to 26 20	11 15	9 64 to 12 90	27 86	24 47 to 31 72
July to September	19 22	13 20 to 27 99	13 48	11 53 to 15 75	32 88	28 05 to 38 55
October to December	28 24	19 52 to 40 86	9 75	8 04 to 11 81	28 44	24 16 to 33 48
<b>1975</b>						
January to February	16 99	10 36 to 27 88	12 27	10 37 to 14 53	30 60	26 34 to 35 55
April to June	14 54	9 43 to 22 41	13 25	11 25 to 15 59	31 10	26 49 to 36 51
July to September	13 27	8 09 to 21 77	15 60	13 29 to 18 30	37 96	32 35 to 44 55
October to December	15 61	10 22 to 23 83	12 87	11 18 to 14 81	32 02	28 24 to 36 31
	With a job—nonagricultural		Unemployed		Not in labor force	
	Sample estimate	95-percent confidence limits	Sample estimate	95-percent confidence limits	Sample estimate	95-percent confidence limits
<b>1973</b>						
January to February	50 98	37 75 to 68 84	26 56	19 23 to 36 69	10 56	8 77 to 12 72
April to June	44 43	35 44 to 55 70	36 63	29 04 to 46 19	10 11	8 71 to 11 74
July to September	25 30	20 13 to 31 81	34 77	27 35 to 44 22	11 83	10 20 to 13 71
November	35 68	21 04 to 60 52	42 60	28 40 to 63 90	10 98	8 35 to 14 44
<b>1974</b>						
January to February	51 97	34 35 to 78 63	31 84	20 20 to 50 18	1 70	5 84 to 10 15
April to June	42 51	33 76 to 53 52	39 93	32 46 to 49 11	10 45	9 00 to 12 12
July to September	27 52	21 36 to 35 45	38 31	30 08 to 48 80	11 23	9 48 to 13 29
October to December	32 16	21 61 to 47 85	44 70	35 61 to 56 11	11 26	9 45 to 13 40
<b>1975</b>						
January to February	51 64	39 83 to 66 95	36 52	29 59 to 45 07	13 76	11 79 to 16 06
April to June	44 61	33 75 to 58 98	32 63	26 03 to 40 90	12 99	11 07 to 15 26
July to September	30 22	24 30 to 37 57	30 25	24 26 to 37 31	12 27	10 30 to 14 61
October to December	36 17	26 69 to 49 01	33 03	27 11 to 40 25	12 52	10 91 to 14 38

<sup>1</sup> Less than 40 sample cases in the category.

## BIAS IN CPS ESTIMATES

### Concept of Bias

As discussed on pages 77-80, the precision of an estimate measures how much the value of any particular estimate is likely to differ from the average or expected value of all possible esti-

mates that could result from samples selected and surveyed under the same basic conditions. If the expected value of all sample estimates corresponds exactly to the true population value of the statistic being estimated, the sample estimate is unbiased and a valid measure of its precision is also a valid measure of its accuracy. If the expected value of all sample estimates differs from the true population value, the estimate is biased and, as the amount of bias increases, the overall accuracy of the estimates

tends to decrease. However, the accuracy of the estimate is fairly insensitive to the amount of bias so long as the bias remains small, relative to the standard error. Where bias is larger than the standard error, on the other hand, bias is the dominant factor in accuracy; in this situation, accuracy is very sensitive to changes in the amount of bias. How this works mathematically is discussed in the section on the mean square error on page 000 of this chapter.

Bias has direction as well as size. Thus, a given estimate may be biased positively (its expected value is greater than the true population value) or negatively (its expected value is smaller than the true population value). The effect of all sources of bias is given by their sum with signs taken into account. This is called the *net bias*. A small bias in one direction partially offsets a larger bias in the opposite direction in the net bias.

An individual source of bias for a given estimate will, if it goes contrary to the general direction of bias error, work toward improving the given estimate. The total effect of all biases, however, regardless of whether the net bias is positive or negative, can only add to total mean square error, thus, decreasing the overall accuracy of the estimate.

### Potential Biases in the CPS

Substantial effort has been directed at minimizing the impact of bias in the CPS sample design and survey processes. These efforts are not always completely successful, however, and potential for bias remains in several aspects of the survey.

**Potential bias from coverage losses**—A small part of the total U.S. population is not represented in the survey, either because of imperfections in the sampling frame, so that some persons do not have a chance to be selected for the sample, or errors in the field that result in completely overlooking certain households or individuals that belong in sample. A potentiality for bias exists if the characteristics of the unrepresented population differ from the rest of the population.

Part of the population omitted from the CPS sampling frame includes a portion of households in mobile homes put in place in last ED's after the 1970 census. In addition, the permit lists used for the new construction sample do not cover housing built in 1970 or later and covered by permits issued before January 1970; consequently, population living in new construction is somewhat underrepresented [52, p. app.-41]. An estimate, made in late summer 1974, suggests that about one million of the 78 million housing units in the inventory did not have a chance of selection in the sample. An estimated 400 thousand of these were mobile homes, and 600 thousand were newly constructed housing units. In addition, homeless individuals with no attachment to any address are also not completely represented.

A small part of the population is lost in field operations, for example, from interviewer error in canvassing only part of the assigned USU, failing to recognize residential structures, and mistakenly classifying some residences as vacant. In addition, some persons are lost within interviewed households either because they deliberately avoid being covered in the interview or are inadvertently omitted from the household roster. The net impact of these coverage losses is indicated by the coverage ratios given in table V-4.

There are, of course, occasional errors of overcoverage, e.g., an interviewer works beyond the boundaries of his assigned USU

or includes in his listings persons whose usual residence is elsewhere, but such errors are quite insignificant in the total picture. Throughout CPS history, the effect of coverage errors has consistently been a net loss in population.

Although the second stage ratio estimate inflates CPS-mated totals to independent controls for population,<sup>8</sup> by age, sex, and race, this adjustment should be regarded as possibly ameliorating but certainly not as removing the potential bias involved in coverage losses. In the final analysis, the assumption must be made that the labor-force characteristics of noncovered portions of the population may be imputed from the characteristics of the interviewed CPS sample. This assumption is inherently biased, no matter how detailed the age-race-sex cross-classification. Moreover, even with respect to the total U.S. population, adjustment is only partial, since the independent control figures are projections from the latest census count. These control counts understate the true population, as of the census date, owing to a net coverage loss resulting from census process [48]. Table VII-3 shows the estimated percent of net undercount in the population in the 1970 census, by race, sex, and broad groups. Each of these percentages has the corrected population as its base; that is, the census population, plus an estimate of net undercount in the population.

**Potential bias from nonresponse**—We group two kinds of error under the term "nonresponse": noninterviews and nonresponse. Coverage losses from noninterview and nonresponse deal with data that belong in the CPS tabulations but are missing from them. Missing data may modify (i.e., bias) estimates, even though the program recognizes their absence and substitutes for what is missing by imputation.

Some of the households in the group, identified by the interviewer as belonging in sample, are not interviewed. This occurs when the interviewer is unable to find an eligible respondent at home after repeated calls during CPS interview. Noninterviews may also result for such reasons as refusal to cooperate, sickness in the respondent household, or impediments. To compensate for losses resulting from noninterviews, CPS estimates are adjusted by imputing characteristics of interviewed households and persons.<sup>9</sup> The noninterview adjustment takes account of the residence of the noninterview within geographic subdivisions, whereas the coverage adjustment is performed only at the national level. As is true for age loss, however, there are selective biases involved in failure to interview that cannot be completely eliminated by imputation procedures.

Within the CPS schedule for an interviewed household, there may be one or more unanswered questions, referred to as *nonresponse*. If the schedule, as a whole, meets minimum requirements for a completed interview, item nonresponses are removed by imputation for the purpose of tabulations. The imputation procedures are tailored to the subject matter. For item imputation, account of any relevant information in answered questions is taken. In some instances, the imputation is logically necessary (e.g., imputation that a person marked as wife of head of household is female); the imputation is, at best, probable in nature, but subject to error. Since there is no way of assuring that

<sup>8</sup> See the section on second-stage ratio adjustment in ch. V, p. 10.

<sup>9</sup> See the section on adjustment for noninterviews in ch. V, p. 10.

TABLE VII-3 Estimates of the Percent of Net Undercount of the Population in the 1970 Census, by Sex, Race, and Broad Age Groups

Year and age	All classes	All races		White			Negro		
		Male	Female	Total	Male	Female	Total	Male	Female
1970									
All ages	25	33	18	19	25	14	77	99	55
Under 5 years old	35	36	33	21	23	20	101	104	98
5 to 9 years old	30	32	29	23	24	22	73	77	69
10 to 14 years old	13	14	12	10	11	9	32	35	28
15 to 19 years old	12	15	8	9	13	5	37	43	32
20 to 24 years old	23	33	14	18	25	11	85	121	52
25 to 34 years old	43	57	28	34	43	24	125	185	67
35 to 44 years old	31	53	9	20	36	5	107	177	40
45 to 54 years old	21	36	6	14	27	1	87	124	53
55 to 64 years old	26	29	24	11	22	19	80	92	70
65 years old and over	18	9	24	18	12	22	12	-3.1	42

\* The minus sign indicates a net overcount.

Source: [48, table 6, "Preferred Estimates"].

of item imputation will balance, even on an expected basis, item nonresponse also introduces potential bias into the CPS estimates.

**Potential bias from response errors**—Some of the recorded responses on completed CPS questionnaires are in error, errors that may come from the respondent, from the interviewer, or, in some cases, from an interaction between respondent and interviewer. A respondent may deliberately give a wrong answer, or the question may be misunderstood due to imprecise question wording, inadequate interviewing techniques, the use of the telephone in interviewing, or—especially when reporting for another member of the household—the respondent may be misinformed or in doubt as to the correct response. The interviewer, on the other hand, may elicit a wrong answer by the way the question is asked, or the interviewer may misunderstand the response and record something the respondent has not meant.

Insofar as the chance of error in the recorded response contributes to the variability of CPS data, its effect is measured in the simple response variance and the correlated component of response variance, discussed on page 78. In addition to its effect on variance, response error has a bias potential, since there is no law in probability that tends to balance out misclassification errors. The bias potential is likely to be especially serious whenever a substantial number of response errors are deliberate, since deliberate errors are likely to be overwhelmingly one-directional.

**Potential bias from rotation groups**—A household selected for the CPS sample is kept in sample for 4 consecutive months, then left out of the sample for the next 8 months, and finally put back in sample for 4 additional months. (See ch. III). In any given month, the total CPS sample is comprised of eight subsamples called rotation groups: (1) Households in sample for the first

time, (2) households in sample for the second time, . . . and (8) households in sample for the eighth and last time.

Each of the rotation groups is a probability sample drawn from the same total population. It might be supposed that if CPS estimates were made separately for each rotation group, the estimates would differ somewhat because of sampling error, but the expected values would be the same for all eight estimates. As a matter of fact, this is not true. It has been convincingly demonstrated, from accumulations of CPS data, that the expected value of the estimates varies among the rotation groups with the most pronounced differences occurring between the rotation group in the sample for the first time when compared with any of the other groups in sample.

The estimate from a particular rotation group can be compared with the estimates from all groups by the rotation group index. This index is the ratio of the estimate based on the sample units in a particular rotation group to the average estimate from all eight rotation groups combined, multiplied by 100. (Table VII-4 shows indexes for rotation groups by the number of months they have been in sample.) The indexes are averaged over a period of 3 years to improve their reliability. For unemployed, the index of 110.2 for first month households indicates that the estimate from households in sample the first month is about 11 times as great as the average over all eight rotation groups; the index of 100.1 for unemployed for second month households is about the same as the average over all rotation groups.

It should not be assumed that estimates from one of the rotation groups can be taken as a standard in the sense that it would provide unbiased estimates, while the estimates for the other seven would be biased. In view of the possible imperfections in each step in CPS operations, it is far more likely that the expected value of each rotation group is biased—but to varying degrees.

**Table VII-4. Rotation Group Bias Indexes in the CPS for Selected Labor Force Characteristics and Type A Noninterview Rate, Average: 1973-1975**

Employment status and sex	Month in sample								Estimated standard error of index
	1	2	3	4	5	6	7	8	
TOTAL POPULATION. 16 YEARS OLD AND OVER									
Civilian labor force	101 7	100 0	99 6	100 3	100 2	99 3	99 1	100 0	0 2
Employed	101 1	100 0	99 6	100 2	100 1	99 5	99 5	100 1	2
Agriculture	100 8	98 3	98 2	98 7	101 2	100 3	100 8	101 5	3 5
Nonagriculture	101 1	100 0	99 7	100 2	100 0	99 5	99 4	100 1	2
Unemployed	110 2	100 1	98 5	101 7	102 1	95 7	92 9	98 8	1 2
Looking, full time	107 4	99 9	99 1	101 3	101 1	97 3	94 4	99 6	1 7
Looking, part time	119 9	100 9	96 5	103 5	105 9	89 8	87 8	95 7	2 3
Not in labor force	97 4	100 0	100 7	99 6	99 7	101 2	101 5	99 9	4
Keeping house	98 1	100 1	100 5	99 6	99 6	101 0	101 3	99 9	5
Going to school	93 8	99 3	100 7	98 7	100 4	102 7	103 8	100 9	1 1
Other reasons	97 5	100 2	101 2	100 1	99 6	100 9	100 9	99 5	8
MALES. 16 YEARS OLD AND OVER									
Civilian labor force	100 9	100 0	99 8	100 2	100 0	99 6	99 4	100 1	2
Employed	100 6	100 0	99 8	100 1	100 0	99 8	99 7	100 2	2
Agriculture	99 7	98 1	98 2	99 3	100 9	100 7	101 1	101 7	3 5
Nonagriculture	100 7	100 1	99 9	100 2	99 8	99 7	99 6	100 1	2
Unemployed	105 8	101 2	100 1	101 4	101 2	97 2	94 6	98 4	1 6
Looking, full time	103 8	101 2	100 9	100 7	100 6	98 1	96 2	98 6	2 3
Looking, part time	115 2	101 0	96 4	105 5	105 1	92 4	87 5	97 0	5 4
Not in labor force	96 7	99 9	100 8	99 2	100 0	101 5	102 2	99 8	9
Keeping house	85 7	100 9	102 4	102 5	96 0	102 7	103 4	106 5	5 3
Going to school	94 1	99 5	100 1	98 1	100 5	102 7	103 9	101 2	1 8
Other reasons	97 9	100 1	101 0	99 4	100 0	101 0	101 6	99 1	9
FEMALES. 16 YEARS OLD AND OVER									
Civilian labor force	102 8	99 9	99 2	100 3	100 5	98 7	98 6	100 0	5
Employed	101 8	100 8	99 4	100 1	100 3	99 1	99 2	100 1	5
Agriculture	106 1	99 3	98 3	96 1	101 6	98 5	99 5	100 5	6 5
Nonagriculture	101 8	100 0	99 4	100 3	100 3	99 1	99 2	100 1	5
Unemployed	115 2	98 9	96 7	102 0	103 0	94 1	91 0	99 1	1 5
Looking, full time	112 1	98 3	96 7	102 0	101 7	96 2	92 1	100 8	2 4
Looking, part time	123 3	100 8	96 5	102 1	106 5	88 0	88 0	94 8	3 4
Not in labor force	97 6	100 0	100 7	99 7	99 6	101 1	101 2	100 0	4
Keeping house	98 2	100 1	100 5	99 5	99 6	101 0	101 2	99 8	4
Going to school	93 4	99 0	101 4	99 2	100 3	102 6	103 5	100 3	1 8
Other reasons	96 4	100 5	101 9	102 0	98 8	100 5	99 5	100 7	2 0
Type A noninterview rate	129 3	89 9	85 2	85 3	122 6	97 0	96 2	94 2	(X)

X Not applicable.

\*These standard errors are rough estimates only. The standard error of any differences between indexes will be smaller than the square root of the sum of variances due to substantial completions among the monthly estimates.

Total CPS estimates, which use the combined data from all rotation groups, are subject to biases that are functions of the biases of the individual groups [3].

Since the expected values vary appreciably among the rotation groups, it follows that the CPS survey conditions must be different in one or more significant ways when applied separately to these subgroups. The ways in which the survey conditions tend to change with number of times in sample have been identified; the importance of each of these changes in affecting expected values has not been definitely established.

One way in which the survey conditions differ among rotation groups is reflected in the noninterview rates. The interviewers, being unfamiliar with households in sample for the first time, are likely to be less successful in calling when a responsible household respondent is available. Thus, noninterview rates generally start above average with first month families, decrease with more time in sample, go up a little for households in sample the fifth month (after the 8 months the household is not in sample), and then decrease again in the final months. (See the bottom line of table VII-4.) These noninterview changes are unlikely to be distributed proportionately among the various labor force categories; consequently, it is reasonable to assume that the representation of labor force categories will differ somewhat among rotation groups and that these differences can affect the expected values of the estimates. The suggestion has been made that the observed CPS rotation group bias could be entirely produced by such differences [78]. Although this is theoretically possible, with the many other factors discussed in the following, it seems unlikely that only one would be the sole cause.

Direct effects on rotation group bias come from different survey conditions that influence the response itself. Even if two samples—one of first-month households and another of later-month households—are identical in terms of a distribution of labor force characteristics, the expected responses from the two samples can differ because of the following:

- The attitude of later-month respondents toward the survey or toward particular questions is somewhat different as a result of earlier interviews.
- The later-month interviewer is biased in the questioning or recording by knowledge of the earlier interviews.
- A change occurs in the distribution of household respondents, for example, heads of households versus other household members.
- A change occurs in the distribution of interviews obtained in person versus those obtained by telephone.
- A change occurs in the distribution of self-respondents, for example, males versus females.
- A change occurs in the interview procedure such as the asking of additional labor force questions in selected rotation groups.

Although these different survey conditions have influenced some individual responses, they do not all have a significant effect. If, for example, knowledge of the earlier interviews actually biases the interviewer, the bias would show itself in reporting the same labor force category, when in fact circumstances had changed. The CPS procedure protects against this type of bias by requiring an independent interview—i.e., the interviewer does not have responses from previous interviews with him and asks the questions as in the prior interview—not accepting “same as

last month” as a response. It is believed that interviewers adhere to these instructions with reasonable consistency. An experimental test of possible interviewer bias indicated a small, but inconclusive, effect on the rotation group bias index for unemployed males [2, section 3B].

The strength of the independent interview comes from the protection it provides against falsely continuing a respondent in a given category when, in fact, his labor force situation has changed. The independent interview has a corresponding weakness, since it doesn't provide protection against inconsistent responses, i.e., false differences in labor force category from 1 month to the next for respondents in an essentially unaltered labor force situation.

The volume of inconsistent responses in CPS is known to be great. Gross changes for respondents reporting for 2 consecutive months are far too numerous when evaluated against realistic assessments of labor force turnover in the total population [30; pp. 80-81]. Moreover, this evaluation has been confirmed by experiments that have included field reconciliation. These experiments demonstrate that a substantial proportion of gross changes are spurious, i.e., labor force situations are reported as different when they are essentially unchanged. Since these inconsistent responses, like other nonsampling errors, do not balance out, they also contribute to the differences in expected values for the various rotation groups.

**Potential bias from processing errors**—Processing is basically an office operation, and its errors are easier to study and to control than errors occurring in the field. Processing error components are believed to be under control and relatively minor. Because of ratio estimation, it is difficult to assign an amount or direction of bias arising from processing error.

Processing losses and other types of error occur during the course of transforming CPS schedules, completed by interviewers, into sets of estimates and tabulations. (See ch. IV, p. 51. For example, because processing deadlines must be met, it may be necessary to prepare the CPS tabulations before all complete questionnaires are available; imputations are made to substitute for missing or inconsistent questionnaire entries; residual error remains after the coding and editing operations have been completed; the checks on the FOSDIC operation occasionally fail to detect reading errors—failure to read individual marks or even whole pages or picking up a few blanks as though they were marks. However, the error rate of FOSDIC reading is so slight that improvement may be impossible.

**Potential bias in estimation procedure**—As a general rule, bias is an undesirable characteristic associated with a survey design and is to be avoided to the extent that it is feasible. Nevertheless, in certain circumstances, the admission of some small bias into the design can have the corollary effect of reducing variance so that there is a net gain in accuracy. Under such circumstances, bias may be acceptable, since the most accurate estimate for a given budget is a prime objective of survey design.

Although unbiased estimation formulas are available, CPS estimates are based on ratio estimates that are technically biased. Because CPS composite estimates are based on ratio estimates, they are technically biased also. The biased estimates are preferred, because they are subject to less variance than the corresponding unbiased estimates. The extent of the bias in the composite estimate differs from the ratio estimate, because the data from the separate rotation groups are weighted differently in the two estimates [3].



TABLE VII-5 Gross Difference Rate for Selected Labor Force Characteristics, CPS Reinterview Unreconciled Sample, by Quarter: 1975

Characteristic	January-March		April-June		July-September		October-December	
	Gross difference rate	Standard error	Gross difference rate	Standard error	Gross difference rate	Standard error	Gross difference rate	Standard error
	(percent)	(percent)	(percent)	(percent)	(percent)	(percent)	(percent)	(percent)
Employed in agriculture	0.6	0.2	0.9	0.2	0.7	0.2	0.6	0.2
Employed in nonagriculture								
Working, full time	5.6	5	6.1	5	7.0	5	5.9	4
Working, part time	7.1	5	6.3	5	7.0	4	7.4	4
With a job, not at work	2.4	3	2.1	3	3.8	4	1.3	2
Unemployed	3.6	4	3.2	4	3.7	4	3.0	3
Not in the labor force	6.7	5	6.3	5	5.8	5	6.1	4

It is necessary to recognize that some sample households are selected with probability differing from others in the same sample. This is done by assigning special weight factors to the records for such units. (For example, see the section on special weights in ch. V, p. 56, and app. G, p. 140.) Occasionally, to make the overall weight unbiased for a unit, the special weight factor must be greater than 4.0 but is arbitrarily cut back to 4.0 for the tabulations. The lower weight reduces the variance of the estimates at the cost of a small element of bias.

Evaluation of the importance of the various potential biases—The accuracy of CPS estimates is decreased by bias—not necessarily by each source of bias but by the net effect of all bias sources combined. Unfortunately, there are no estimates of total net bias for CPS estimates, because it has not been feasible to simultaneously study the effects of all sources of biased error against a standard of truth. Since the overall measure is not available, some assessment is needed of how important individual sources of potential bias may be in the total picture—important to the extent that a given source is likely to determine the direction and the approximate size of the total net bias.

Response error is undoubtedly the most important potential source for bias in CPS estimates of labor force characteristics for the population as a whole.<sup>7</sup> Evidence of this comes from (1) the extreme sensitivity of response to the survey design, what might be considered a trivial difference in survey conditions can produce pronounced differences in the estimates obtained (as evidenced by the rotation group indexes in table VII-4), and (2) the larger number of response differences in the reinterview, the

larger the gross difference rate, the larger the net error (response bias) can be.

For the latter point, consider the schematic representation of the results of the original interview and the reinterview (unreconciled) (given on p. 79) for the characteristic unemployed. The ratio  $(b+c)/n$  is referred to as the gross difference rate or the proportion of persons for whom inconsistent responses were obtained in the original interview and reinterview. Since the responses are inconsistent, we conclude that one of the two responses did not provide the true labor force classification. Thus, the level of the gross difference rate is a direct indication of the total number of response errors. By this reasoning, the larger the gross difference rate, the greater the number of response errors that, in turn, increase the likelihood of a nontrivial net error. That is, the more response errors into and out of a given category, the greater the net error can be. Estimates of the gross difference rate, by quarter, for calendar year 1975, for selected labor force characteristics, are shown in table VII-5.

The primary defense for CPS, against the potential bias of nonresponse (i.e. noninterview and item nonresponse), is to keep the rates of nonresponse low. Where this has been achieved, any sizeable biases included in the imputation of characteristics to nonrespondents are likely to be relatively unimportant, since such biases would affect the estimates proportional to the rates for undercoverage, noninterview, and item nonresponse. (See table V-4, table IV-1, and the section on close out for final processing in ch. IV, p. 53.)

A special study has tested the distribution of labor force characteristics of the noninterviewed in the CPS. The study was only partially successful, since it omitted refusal cases and obtained interviews from only about half the noninterviews studied. The labor force distribution for those noninterviews converted to interviews was very similar to the imputed distribution of noninterviews—with the exception of one small subgroup of

<sup>7</sup> The importance of the various potential sources of bias in estimates of labor force characteristics for some subgroups of the population (e.g. Negro males) may differ from the biases for the population as a whole, however, no assessment of the importance of individual sources of bias for subgroups has been attempted due to our limited knowledge about the effect of these biases on estimates for specific population subgroups.

employed persons. Persons with job, not at work, were substantially overrepresented among the noninterviewed, presumably because of entire households out-of-town on vacation and inaccessible during CPS week. Theoretically, the hard core of persons noninterviewed in the CPS and not converted to interviews in the special study could have serious biases not hinted at by the results of the special study. To have an appreciable effect on total estimates, however, these biases would have to be drastic indeed, since the hard core of the noninterviewed is only some 3 percent of the total CPS sample.

Similarly, bias due to item nonresponse on labor force questions would also have to be drastic, since the overall nonresponse rate for labor force items averages less than 0.5 percent on an item basis, that is, for each labor force item, the number of blanks divided by the total number of times the item should have been reported. There are about five blanks per 100 questionnaires, a 5-percent rate on a questionnaire basis [33, exhibit 2].

For some characteristics, however, the item nonresponse rates are considerably higher, for example, in the collection of income data. In order to maximize the amount of usable information, modifications in income data collection procedures, designed to reduce the nonresponse rate, were introduced in the 1969 March CPS and have continued since that time. However, even with such improvements, about 16 percent of all persons 14 years old and older interviewed in the March 1975 CPS did not report complete income information.<sup>8</sup> Improved income edit and allocation procedures were also introduced, starting with the 1968 March CPS; the main feature of the new procedures was a more refined method for imputing missing income data to expand the use of information already known about that person [56]. On the whole, it is believed that the impact of nonresponse on the aggregate CPS income data is considerably less than the nonresponse rate suggests, because families at the higher income levels tend to report their main sources of income but omit relatively minor amounts from interest, dividends, and similar sources [28]. In terms of the high response rates achieved and maintained in the CPS survey, it seems unrealistic to assume that the potential bias of nonresponse could be as important in its effect on total estimates as the potential bias of response error.

As with nonresponse, the primary defense against the potential bias introduced by undercoverage is to keep the rate of undercoverage low. While undercoverage is low for the population as a whole and for the white population, there are several age-sex groups among the not-white population (race reported as Negro or other) where the undercoverage is quite high. For example, in table V-5 for ages 20-21 years old for the not-white group, the coverage rate for males is 0.733 and for females the rate is 0.852, indicating coverage losses of 26.7 percent and 14.8 percent, respectively. For estimates of the labor force characteristics of the population as a whole, however, even such high undercoverage cannot be viewed as a potentially large source of bias, since these age-sex-race categories represent a relatively small proportion of the whole population.

The processing of CPS data has been routinized, the opportunities for error restricted, and the occurrence of error carefully quality controlled. Under these circumstances, the volume of

processing errors is far smaller than the volume of field errors, and the potential for bias is correspondingly low. Thus, the fact that processing errors may be producing some trivial increases or decreases in the total net bias of CPS estimates can, for practical purposes, be disregarded.

The choice of technically biased estimation formulas, as the basis for published CPS estimates in place of alternative unbiased estimates, was a conscious trade-off in which small biases were accepted in order to attain smaller sampling variability for the estimates—and smaller mean square errors. In the face of larger potential biases from other sources, the fact that the estimation formulas are not strictly unbiased is of no practical consequence.

### Circumstances Affecting the Objective Measurement of Labor Force Status

Heretofore, we have assumed that each individual in the population could be assigned to a labor force category (e.g., employed, unemployed, or not in the labor force) in an unambiguous manner. This would mean that the true number of persons in each category could be measured, using the CPS definitions and concepts. This assumption also implies that the shifting of individuals between categories one month to the next reflects real change in labor force status. For the majority of individuals in the population, this assumption is justified. In most instances, establishing that a person is or is not employed is fairly straightforward. However, there are many persons whose labor force situation is uncertain in that it is not objectively clear whether they are unemployed or not in the labor force.

For the majority of persons not employed, the concept of looking for work is the key to classifying them as either unemployed or not in the labor force.<sup>9</sup> When the person ceases looking for work, he shifts from being unemployed to not in the labor force on the basis of the CPS definitions. In many instances, it is difficult to apply this concept objectively, since the transition from being unemployed to not in the labor force may not be an abrupt change. Rather, this change may occur over a period of time owing to the nature of the job seeking activities, and the point in time at which an individual considers himself to have ceased looking for work may be a rather subjective determination on the respondent's part. Thus, the responses obtained in successive months for these individuals may be inconsistent in that they show a change in labor force status while no substantive change has taken place.

Some evidence to this effect is provided by the rotation group indexes shown in table VII-4. These indexes show that changes in the survey conditions result in estimates that differ even though each estimate is supposed to be a measure of the labor force status for the same population.

The following data appear to be consistent with the hypothesis that persons having an uncertain labor force situation are most likely to change responses as a result of repeated interviewing and, thus, change the rotation group indexes. Further, these data appear consistent with the intuitive assumption that such persons are a larger group relative to unemployed than to employed or not in labor force, that they represent a greater proportion of persons available for part-time work, and, finally, are even

<sup>8</sup> For each person in the sample 14 years old and older, questions were asked on the amount of money income received in the preceding calendar year from eleven different sources. These supplemental income questions are included on the March CPS questionnaire each year.

<sup>9</sup> See ch. IV, p. 42, for the definitions of employed, unemployed, and not in the labor force.

more frequent among females who might be available for part-time work.

To illustrate, table VII-4 shows that the first-month rotation-group index for the unemployed is 110.2, while the second-month index is 100.1. This suggests that, for two groups of individuals in essentially the same labor force situation, substantially more will be reported as unemployed when interviewed for the first time as compared with their reports when interviewed for the second time. Within the unemployed, the difference between first- and second-month indexes for those persons looking only for part-time work is even greater than for the unemployed as a whole—a category where one might expect a greater proportion of individuals with ambiguous labor force situations. For this category, the first- and second-month indexes are 119.9 and 100.9, respectively. The difference is even greater for unemployed females looking for part-time work where the first- and second-month indexes are 123.3 and 100.8, respectively.

#### Effect of Bias in Monthly Levels on Estimates of Month-to-Month Change

While the estimates of monthly level have large potential bias, there is stability in CPS survey conditions that assures high correlation in the direction and magnitude of the total net bias from month-to-month and even from one year to the next. As a result, much of this bias cancels out when estimating month-to-month change. The rate of noninterview remains relatively stable from month-to-month as does the distribution of noninterviews among persons not at home, temporarily absent, or those who refuse to be interviewed. (See table IV-1). Also, each monthly estimate maintains the same proportions of the eight rotation groups, which tends to neutralize the effects of rotation group bias on estimates of change. Further, estimates of the coverage loss, associated with the CPS sample, are stable from month-to-month (see table V-4), indicating similar stability in the bias associated with this source.

Accordingly, where estimates of level for CPS statistics are subject to total net biases that are potentially large—probably even larger than the sampling error—the biases of month-to-month changes are believed to be small, relative to their sampling errors. On this assumption, estimates of change would be only slightly less accurate than they are precise.

#### THE MEAN SQUARE ERROR OF THE CPS ESTIMATES

The standard error of a survey estimate will be an unsatisfactory measure of the accuracy of the estimate whenever the bias is large relative to the standard error. This is so because the standard error measures the deviation of the possible sample estimates from their expected value which, if the estimates are biased, is different than the true value.<sup>10</sup> A measure of the accuracy of the estimate when a bias exists is the square root of the mean square error (generally referred to as the *root mean square error*). The *mean square error* is defined as the average of the squares of the deviations of the possible sample estimates from the true value. That is,

$$MSE(x) = E(x_k - x)^2$$

where  $x$  is the true value of the characteristic being estimated. The root mean square error (RMSE) corresponds, in terms of measuring accuracy, to the sampling error in measuring precision.

The mean square error of the estimate can also be expressed as the total variance of the estimate plus the square of the net bias. Thus, the mean square error of an estimated total ( $x$ ) is given by

$$MSE(x) = Var(x) + [Bias(x)]^2 \quad (VII/5)$$

The term  $Var(x)$  includes all the components of variance discussed in the section on precision of the CPS estimates, pages 77-80. Biases may result from the method of sampling, the measurement methods, or other sources of nonsampling error discussed on pages 81-88; the bias term in formula (VII/5) refers to the net of all of these biases.

As long as the bias term in formula (VII/5) is small, relative to the standard error, the standard error alone provides an adequate measure of accuracy. It understates the RMSE in this situation but not seriously, and confidence intervals formed using the standard error are not serious distortions of the distribution of estimates around the true value. If the bias is large, the standard error is misleading, since the measure of accuracy and the distortion of confidence limits are great.

The effect of the magnitude of the bias component on the relationship between the standard error and the root mean square error is illustrated by table VII-6.

When the bias of the sample estimate is relatively small, as compared to the standard error of the sample estimate, for example, less than 20 percent, table VII-6 shows the standard error

TABLE VII-6 Effect of Bias on the Ratio of the Root Mean Square Error to the Standard Error

Bias in $x$ (positive or negative)		$\frac{\sqrt{MSE(x)}}{\sqrt{Var(x)}}$
0		1.000
0.05	$\sqrt{Var(x)}$	1.001
0.10	$\sqrt{Var(x)}$	1.005
0.20	$\sqrt{Var(x)}$	1.02
0.50	$\sqrt{Var(x)}$	1.12
1.00	$\sqrt{Var(x)}$	1.41
1.50	$\sqrt{Var(x)}$	1.80
2.00	$\sqrt{Var(x)}$	2.24
3.00	$\sqrt{Var(x)}$	3.16

Source: [21, p. 58].

<sup>10</sup> The concept of true value is discussed in [22, ch. 12, sec. 2].

TABLE VII-7 Effect of Bias on the Probability of an Error Greater Than 1.96 Times the Standard Error

Bias in $\bar{x}$	Probability of Error		
	Less than $-1.96 \sqrt{\text{Var}(\bar{x})}$	More than $1.96 \sqrt{\text{Var}(\bar{x})}$	Total
0.0	0.0250	0.0250	0.0500
0.02 $\sqrt{\text{Var}(\bar{x})}$	0.0238	0.0262	0.0500
0.06 $\sqrt{\text{Var}(\bar{x})}$	0.0217	0.0287	0.0504
0.10 $\sqrt{\text{Var}(\bar{x})}$	0.0197	0.0314	0.0511
0.20 $\sqrt{\text{Var}(\bar{x})}$	0.0154	0.0392	0.0546
0.60 $\sqrt{\text{Var}(\bar{x})}$	0.0052	0.0869	0.0921
1.00 $\sqrt{\text{Var}(\bar{x})}$	0.0015	0.1685	0.1700
1.50 $\sqrt{\text{Var}(\bar{x})}$	0.0003	0.3228	0.3231

Source: [6, p. 14].

alone understates the RMSE by less than 2 percent. For estimates of change, the relationship between the root MSE and the standard error would be the same as shown in the table, when the bias is expressed in terms of the standard error of the change.

The effect of bias on the probability of a sampling deviation further from the true value than 1.96 times the standard error is shown in table VII-7. If the estimate is unbiased, under normal assumptions, we would expect a sample result to fall 1.96 times

the standard error beyond the true value only 5 percent of the time, a probability of error of 2.5 percent in either direction. The table shows the bias has little effect on the total probability of an error, provided it is less than 10 percent of the standard error. Even if the bias is 20 percent of the standard error, column 4 of the table shows the distortion of the total probability of error is modest (about 5.46 percent versus 5.0 percent). However, columns 2 and 3 show that, as the bias increases, the error becomes increasingly nonsymmetric.

## Chapter VIII.

# CALCULATION AND PRESENTATION OF SAMPLING ERRORS

### INTRODUCTION

When discussing sampling error, or the precision of a sample result, we are referring to how closely we can reproduce, from a sample, the results that would be obtained if we should take a complete count or census, using the same methods of measurement, questionnaire, interview procedures, type of interviewers, supervision, etc. The sampling error is measured by the standard error; the square of the standard error is the variance.

The estimation of variances for a multipurpose survey that uses a complex sample design and estimation procedure is a complicated undertaking. The two objectives presently considered in estimating variances of the major statistics of interest for the CPS are—

1. To measure the precision of the survey findings.
2. To evaluate the impact of each of the stages of sampling and estimation on the overall precision of the survey.

For a given method of preparing estimates for the survey, the total variance, by which the overall precision of the survey is measured, is given by the sum of the variance contributions from each of the stages of sampling in the survey and the correlated and uncorrelated components of response error. (See formula (VII/3) on p. 78.) For the combined A- and C- design samples, the total variance should include contributions from selecting one stratum out of each pair of NSR strata, PSU's within the strata, USU's within the PSU's, the choice of interviewers, and the responses given during the interviews.

To achieve the first of these objectives, the estimated variance should include the sum of all of the variance components representing the random error in the survey. For the second objective, each of the variance components relative to the total needs to be separately identified. Both of these objectives are partly frustrated, because no practical method has been found that allows the total of all of the variance components to be measured or separately identified, stays within the budget, and does not interfere with the conduct of the survey. Similarly, it is not possible to separately identify each of the variance components that are included in the variance estimators currently used. For example, the current method of estimating the total variance includes estimates for only part of the correlated component of response variance (a portion in NSR strata) and cannot separately identify it; also, although the simple response variance is included, it cannot be separately identified. Although the procedure is imperfect, the estimated components of the total variance are useful in evaluating the efficiency of the sample design. The current variance estimation procedure does a better job of reflecting the impact of the separate stages of estimation on the estimated total variance.

### EARLIER METHODS OF ESTIMATING SAMPLING ERRORS

Several methods of approximating variances of estimates based on CPS national samples have been used, each of which

has incorporated certain compromises and simplifying assumptions needed to reduce the task to manageable proportions. The different methods are usually introduced when changes in the CPS sample design are made after a decennial census. The estimation methods and the associated assumptions have been dictated primarily by the computing facilities available and the ability to adapt advances in variance estimation theory to the CPS situation.

In general, the simplifying assumptions have tended to produce estimated sampling errors larger than the sampling error that should have been obtained and, because of this, we have probably failed to accept the validity of some of the CPS results that seemed marginal.

### Sampling Errors of Unbiased Estimates

An expression of the variance of the CPS estimate should reflect the total impact of the several stages of sampling and the separate steps in the estimation process. In the early period of the CPS, it was felt that constructing an estimator for this complicated variance expression involved more computation than was feasible with available computing equipment; punchcard equipment and mechanical desk calculators were the only computing facilities at hand. To make some of the earlier variance estimates, starting with the first attempts at estimation in 1946 the CPS was treated as a single-stage ratio estimate to total population of the form  $x' = \frac{x}{y} \bar{Y}$ , where  $x$  is the simple inflator (or unbiased) estimate of total population with the characteristic  $X$ ,  $y$  is a similar estimate for a total population 14 years old and over, and  $\bar{Y}$  is an independent estimate of population 14 years old and over. For this estimate, the variance estimator is fairly simple and involves computations readily handled by a desk calculator. The procedure was used for variance estimation until 1954; at that time, the effects of ratio estimation, by broad age groups and sex, were studied.

The method—Estimates of variance under these assumptions made use of the random group technique in self-representing (SR) strata [21, ch. 10, sec. 16]; rotation groups were used as the random groups within clusters of PSU's, defined by noninterview clusters. In nonself-representing (NSR) strata, a system of collapsing strata was used [22, ch. 9, sec. 5]. Variance estimation under these conditions, as applied to the 230-area CPS design, is described in more detail in [44, p. 50].

Impact of simplifying assumptions—For a number of reasons, these variance estimates over-estimated the sampling errors of the actual CPS. The assumed form of the CPS estimate did not take into account the gains expected from the first stage of ratio estimation and ignored the impact of correlations within the separate age-sex-race categories actually used in the CPS second-stage of ratio estimation. Using the separate rotation group totals as random groups also overstated the variance in SR PSU's because of the problem of rotation group bias. (See ch. VII, p. 84.) Also, the random group variance estimator assumes simple random sampling of USU's, rather than the systematic sampling actually used, and probably caused some

overstatement of the variance. The collapsed stratum variance estimator also introduces an overestimate of the variance [22, ch. 9, sec. 5]; this problem persisted until later designs, incorporating both A- and C-design samples, made it possible to eliminate this bias from the variance estimates.

Finally, the impact of controlled selection in designating sample PSU's was ignored; at the present time, none of the routine methods of CPS variance estimation developed so far properly reflects the variance reduction arising from this source.<sup>1</sup>

### Variance Estimates by the Replication Method

The sampling error of an estimate, based on any sample design using any estimation procedure no matter how complex, may be estimated by the method of replication. This method requires that the sample selection, collection of data, and estimation procedures be independently carried through (replicated) several times. The dispersion of the resulting estimates can be used to measure the sampling errors of the full sample.

The method—Obviously, one would not consider repeating the entire Current Population Survey several times each month simply to estimate sampling errors. A practical alternative has been to draw a set of random subsamples from the full sample surveyed each month, using the same principles of selection, and to apply the regular estimation procedures to these subsamples. We refer to these subsamples as *replicates*. Variance estimates produced in this manner represent a substantial improvement in accuracy and precision over earlier procedures.

The replication method was used to estimate CPS sampling errors for the 230-area design [44, p. 57], after the first electronic computers became available, and was continued into the 330- and 357-area designs. This system was used each month, over a period of about 7 years, to provide estimated variances for the first- and second-stage ratio estimates and for the composite estimate of level. By seasonally adjusting the composite estimates from each replicate, it also became possible to approximate the variance of CPS estimates with and without seasonal adjustment.

Impact of simplifying assumptions—Because the simplifying assumptions used at this time were less restrictive than the former ones that treated the CPS as the ratio of unbiased estimates, the variance approximations more nearly reflected the impact of the actual estimator, although the variances were still somewhat overestimated. Each of the replicates was subjected to the second-stage ratio estimate in the same age-sex-race categories by essentially the same process as used for the full sample at the time. However, rather than using a separate set of first-stage ratio estimate factors for each replicate (each replication had only half of the sample NSR PSU's), the factors appropriate for all sample NSR PSU's were used. Similarly, the noninterview adjustment factors and the special weights for subsampling, developed for the full sample, were used for each replicate. These simplifications mean that the impact of the special weighting, the noninterview adjustments, and first-stage ratio estimation are not completely represented in the variance estimates. Moreover, using replications to estimate variances has the same problem of overestimation as for the unbiased estimate

in the collapsing of NSR strata and estimating the within-PSU sampling variance with rotation groups. This variance estimation procedure also fails to reflect the effect of controlled selection in designating the sample PSU's.

The size of the memory for the computer available at this time severely limited the number of replications and the items for which variances could be estimated. When the procedure was first developed for the UNIVAC I computer, it was used to produce variance estimates for 14 items, based on 40 replications. Each month, these computations required more computer running time than was needed for the basic CPS processing and tabulation. Subsequently, to reduce costs, variances for the 14 items were regularly estimated, based on 20 replications. Empirical research also showed that variances, based on 20 replications, were quite similar to those based on 40 replications. The theoretical loss in precision of variance estimates, by reducing the number of replications, is discussed by Gurney [68]. This loss in precision was considered acceptable for labor force statistics, because the standard errors published are usually based on averages over several months; the problem would be more serious for statistics collected less frequently.

The replications were random selections from the population of all possible independent half samples, defined with a given pairing of strata and USU's. The subsequent work of McCarthy [71; 72], discussed in appendix K, has shown that more precise variance estimates would be possible by an equivalent sample of replications from an orthogonal set.<sup>2</sup>

### Keyfitz Methods

In the mid '50s, Keyfitz proposed a method of variance estimation that provides consistent estimates of the variance for most stratified complex estimators, based on relatively simple quadratic functions of the observations in each stratum [26]. The method, as originally developed, applies only when two primary units are selected independently from each stratum. However, useful approximations can be obtained for other sample designs by grouping or subdividing strata, as required. This system for variance estimation was introduced during the use of the 357-area design [44, p. 56]; for the discussion, we will identify the system, used at that time, as the First Keyfitz Method. The method is based on the fact that the variance of the sum of two independent and uncorrelated sample estimates can be expressed as the square of the difference between the estimates. With this method, estimates of the variance could be produced with greater precision (see [4]) and substantially less effort than with the Replication Method, using the computer facilities available at that time. Variances for the combined A- and C-design samples are now estimated by extending the method and applying the procedure to the 461-area design (the modified procedure is called the Current Method); it continues to apply the essential principle of Keyfitz variance estimation.

The Current Method of variance estimation is described in the next section; it differs from the First Keyfitz Method in two minor respects. Although both of the differences undoubtedly improve

<sup>1</sup>The impact of controlled selection of PSU's on the variance component arising from sampling PSU's is discussed in [5].

<sup>2</sup>For a supplemental inquiry in March 1973 (the Occupational Changes in a Generation Survey), a set of 243 orthogonal replications was defined for the 461-area CPS sample and used for variance estimation. The procedure for defining the replications in that study is based on [17].

the variance of the variance estimates, it is not clear how the typical variance over-estimates are affected.

For the First Keyfitz Method, the variances were estimated in SR strata using random groups composed of selected combinations of rotation group totals within noninterview clusters. There were four sets of such replicated half samples, each a combination of rotation groups subjectively selected to minimize the effect of the rotation group bias on the variance estimate. The variance in SR strata, under the Current Method, is estimated by using a set of more rigorously defined replicated half samples that ignore rotation groups (see app. K, p. 150); the method should result in a significant increase in the reliability of the variance estimates, since it more closely reflects the sample design.

The following term appears in the derivation of the variance estimator in appendix K, formula (K/28):

$$\gamma_a = \hat{Y}_a / \bar{Y}_a$$

where  $\hat{Y}_a$  is the current independent population control figure for the a-th age-sex-race category and  $\bar{Y}_a$  is the expected value of the unbiased estimate of the current population, based on the CPS sample. (See app. K, p. 155.) In the First Keyfitz Method, all  $\gamma_a = 1$ , while the derivation for the Current Method assigns a value to each, based on current information for the terms in the formula for  $\gamma_a$ .

### CURRENT METHOD FOR ESTIMATING SAMPLING ERRORS

Exactly the same theoretical basis underlies variance estimation by the Current Method and the First Keyfitz Method; minor differences lie in the application of the theory. Both of these variance estimates use the Taylor approximation by dropping terms with derivatives higher than the first. However, by taking account of the A- and C-design samples, the bias in the variance estimate from collapsing NSR strata has been eliminated. At present, there is no routine method of computing variances for seasonally adjusted data; an earlier study has shown that the seasonal adjustment of CPS estimates has relatively little impact on the variances.

The theory is discussed and variance estimators are derived in appendix K, using illustrations oriented to the Current Method. The impact of the simplifying assumptions used are also assessed. The method may be summarized as follows: A linear form is derived from the expression for a CPS estimator (it may be any complicated estimator which is differentiable) by using the Taylor approximation and dropping terms which involve second- and higher-order derivatives. The variance for this linear form is shown to be the same as for the complicated estimator under a set of assumptions consistent with the CPS situation. The variance of the linear form is estimated by applying the Keyfitz method. The work of Tepping [32], Woodruff [79], Woodruff and Causey [80], Frankel [13], McCarthy [72], and Keyfitz [26] are used in the derivations.

The computer program for this system can routinely produce estimates of variance for up to 100 specific statistics each month. The total variances are shown (with some of the variance

components) for several forms of the CPS estimate: e.g., the unbiased estimate with and without first-stage estimation and combined with the second stage of ratio estimation, the composite estimates of level and month-to-month change, and other estimates. Estimates of sampling errors produced by this system are shown for selected items in tables VIII-1 through VIII-4 Statistics, other than the routine 100, can be specified so that the same computer programs can provide variance estimates for supplemental items which are occasionally included in the survey.

### VARIANCES FOR STATE AND LOCAL AREA ESTIMATES

For statistics estimated at the national level, variances are estimated from the sample data by the methods previously given. For local areas which are coextensive with one or more sample PSU's, a variance estimator can be derived, using the methods of variance estimation used in the SR portion of the national sample; this can be done for areas made up of sample PSU's from any of the A-, C-, or D-designs. However, estimates for areas which have contributions from NSR strata, such as a State, have variance estimation problems that are much more difficult.

#### Complicating Factors

The following factors complicate the estimation of variances for States:

1. Estimates are typically required as averages over several CPS months, e.g., annual averages.
2. The NSR CPS national sample PSU's falling in a State have actually been designated by controlled selection. (See app. A.) It is relatively safe to ignore this when approximating the variance of estimates at the national level; however, for State variance estimates, this aspect of the sample design is much more important. Ignoring it will usually imply substantial overestimates of the component of variance from sampling of PSU's [5].
3. Estimating variances in the NSR part of a State, by collapsing strata similar to the system used in the national sample, would result in low precision in the variance estimate, because, typically, there would be few collapsed pairs of strata defined for a single State. Also, the component of variance from sampling PSU's can be more important than for national estimates, because, in some States, the proportion of population in NSR strata is larger than the national average. A further problem arises because the NSR strata usually are made up of PSU's from more than one State, and estimates are based only on sample PSU's from within the State.
4. Those State estimates having contributions from supplemental CETA PSU's have the complicating factors mentioned and an additional one: Although the supplemental PSU's have been selected from strata restructured to be wholly within the State, the definitions of the restructured strata depend on the particular set of CPS PSU's already in sample, and this random definition of the strata can add an unknown component of variance to the estimates.

### Imputed Variances for States

For the reasons enumerated in the previous section, variances for estimated annual averages for selected States are not based entirely on computations using sample data from these areas. For the States that do not have CETA supplemental PSU's, approximate variances have been derived that take advantage of the results obtained in computing variances for national estimates, and in averaging variances for similar characteristics over time. (See the following section.) Some of the variance estimates, based on census data for the strata defined within the State, have also been incorporated. (See the discussion of iterative procedures conducted within the States in ch. VI, p. 69.) A number of different methods have been used for years prior to and including calendar years 1975 and earlier; these methods are currently under review.

In general, we believe the standard errors provided for the States and SMSA's are conservative, i.e., they are probably over-estimates.

### GENERALIZING ESTIMATED SAMPLING ERRORS

With some exceptions, the sampling errors provided with sample data in Census Bureau reports are generalizations. They are usually derived by combining variances computed for the statistic in several surveys over time or are based on variances for a set of statistics known to have similar variance behavior.

#### Why Generalized Sampling Errors Are Used

With a variance estimation process developed for an electronic computer (such as the current method for CPS national tabulations, for example), it would be possible conceivably, to compute and show an estimate of the sampling error, based on the survey data, for each figure in the results tables of a report. There are a number of reasons why this is not done: It is not possible to show a separate sampling error for each figure that will be needed in analyzing the results of a study; even if each published sample estimate were accompanied by its sampling error, one cannot predict the combinations of results (e.g., ratios, differences, etc.) that may be of interest to the reader. Also, the presentation of individual sampling errors would double the size of the results tables, and the cost of computing the large number of variances to be estimated would be excessive.

Sampling errors, when estimated from sample data, have sampling errors of their own. The estimated variance for a statistic for a particular month generally has a larger relative sampling error than the value of the estimated statistic. This means that the estimates of sampling error may vary considerably from one time period to another or for related characteristics (that might be expected to have nearly the same relative sampling error). Therefore, some method of stabilizing the sampling error estimates, for example, by generalization or by averaging over time, is needed to improve their usefulness.

Experience has shown that groups of statistics, which usually can be determined in advance, do tend to have similar relvariances<sup>3</sup> when estimated from the CPS. The relvariances for

statistics in these groups should have improved reliability if an averaging or generalization process is used to take advantage of this experience.

#### Method of Generalizing the Sampling Errors

Estimated sampling errors are averaged by use of a relvariance function of the form

$$V_x^2 = a + b/x \quad (\text{VIII/1})$$

where  $V_x^2$  is the relvariance of the estimate  $x$ , and  $a$  and  $b$  are two parameters, fitted by a least-squares process, to a set of observed related estimates and their computed relvariances.

The particular relvariance function has been used since 1947 for the CPS and its supplements and was chosen from among several that were considered [21, ch. 12, sec. B-15]. This function is used to average estimated sampling errors of means or totals based on estimated counts of the population having a characteristic; variances of estimates based on values reported for the sample unit (e.g., aggregate expenditures, amount of income, etc.) do not lend themselves to averaging in this way.

As a rationale for using this function, consider the relvariance of a sample estimate as the product of the relvariance from simple random sampling and a design effect. Defining  $P = 1 - Q = X/N$  as the proportion of the population having the characteristic  $X$  where  $N$  is the population size, and writing the design effect as  $F$ , the relvariance of the estimated total  $x$  based on a sample of  $n$  members from the population is approximately

$$V_x^2 \doteq FQ/(np)$$

The expression can also be written

$$V_x^2 \doteq - (F/n) + F/(nx)$$

which has the form of expression (VIII/1). One determines the values of  $a$  and  $b$  for a set of observed estimates and relvariances by minimizing the least-squares residuals of the observed relvariance relative to the relvariance from the function. In using this least-squares method for the CPS, the expressions for  $a$  and  $b$  are not explicit; they are computed by an iterative process first introduced in 1949. Prior to 1949, they were computed by a number of other methods that proved less satisfactory.

Choice of sampling errors to generalize—To apply the method, one must compute, by usual methods, relvariances for a set of statistics that meet certain specific requirements. In practice, the set should comprise at least 20 observed estimates with their relvariances, although occasionally fewer observations than this are used.

The statistics in the set should have a similar relationship between each estimate and its relvariance, i.e., about the same value of the design effect. In practice, this means the statistics should be similar regarding clustering by PSU, by USU, and among persons within housing unit. For example, relvariances for estimates of total persons classified by a characteristic of the housing unit or of the household, such as the total urban population, number of recent migrants, or persons of Spanish origin,

<sup>3</sup> Relvariance refers to the square of the relative sampling error, or the square of the coefficient of variation.



would tend to have fairly large relvariances, because these characteristics usually appear among all persons in the sample household (and often among all households in the USU as well). On the other hand, lower relvariances would result (for equivalent size of estimate) for estimated population by personal income, education, marital status, or detailed age categories of the sample person.

Computed relvariances are required for estimates over a wide range so that observations are available to insure a good fit of the curve at high, low, and intermediate levels of the estimates. One needs to remember also that, because the CPS is a ratio estimate, statistics made up of very large proportions of second-stage ratio estimate cells may have variances very near zero, for example, males 30 to 34 years old in the labor force.

It must be remembered that, by reading the relvariance of an estimate from a curve fitted in this way, we introduce some error. Fitting a curve to the set of estimates and relvariances may substantially modify some extreme relvariance values that really are extreme.

**How the relvariance function is used**—After the parameters  $a$  and  $b$  of expression (VIII/1) are determined, it is a simple matter to construct a table of sampling errors of estimates of level for publication with a report. In practice, such tables show the sampling errors that are appropriate for specific estimates of level, and the user is instructed to interpolate sampling errors for figures not explicitly shown in the table. (See, for example, table L-1 in app L.)

A table of sampling errors for proportions is also shown in appendix L. It has been derived by making use of the approximation

$$\frac{v_x^2}{y} \approx v_x^2 - v_y^2$$

where the relvariances of  $x$  and  $y$  are read from the relvariance function (VIII/1). This formula was introduced in 1948 after empirical research showed that it produced useful approximations. The approximation is appropriate when the correlation between the ratio  $x/y$  and the denominator  $y$  is close to 0; the approximation is an overestimate if the correlation is positive. For most situations, the correlation is reasonably close to zero [21, ch. 12, sec. B-15].

## STATISTICAL REVIEW OF PUBLICATIONS

It is the policy of the Census Bureau to discuss the reliability of data in a "Source and Reliability Statement," published as part of each Census Bureau report. The principles and methods observed in presenting survey errors are discussed in detail in [14] which is a revision of Technical Paper 32 [65]. Such principles and methods have been observed since 1946 in the reports of results from the CPS.

Most publications presenting CPS results include text which describes and analyzes the data, pointing out trends and other relationships that may be of special interest to users of the data. For reports published by the Census Bureau, the textual material undergoes a statistical review, prior to publication, to insure that

all statements in the text take into account the sampling variability and other information on survey error. At the Bureau of Labor Statistics, the analysis and statistical review of the labor force publications regularly take into account the measures of survey error provided by the Bureau of the Census.

The analysis of survey findings depends, in part, on the use of confidence intervals to evaluate the statistical significance of estimates prepared from sample data. The sample estimate and an estimate of its standard error can be used to construct interval estimates with prescribed confidence that the interval includes the average result of all possible samples. To illustrate, if all possible samples were selected, each of these surveyed under essentially the same conditions, and an estimate and its estimated standard error calculated from each sample, then approximately 95 percent of the intervals, from two standard errors below the estimate to two standard errors above the estimate would include the average value of all possible samples. The interval, from two standard errors below the estimate to two standard errors above the estimate, is called a *95-percent confidence interval* [14, p. 6]. Approximately 90 percent of the intervals, from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate, would include the average value of all possible samples; these intervals are called *90-percent confidence intervals*.

It is a Census Bureau policy to consider the difference of two estimates statistically significant without indicating a qualification if the difference exceeds twice the standard error of the difference, i.e., to use 95-percent confidence intervals. If a report comments on the difference, the appropriate statistical test performed on the two estimates to see that this criterion is met. Differences with ratios to their standard errors in the range 1.6 to 2.0 may be mentioned if the appropriate qualifying language is used. Thus, for an estimated difference in this category, we may indicate there is "some evidence" of an increase or decrease. The ratios 1.6 to 2.0 imply the use of confidence intervals in the range of 90 to 95 percent.<sup>4</sup>

The first responsibility for including the impact of the survey errors in the analysis of the results in a Census Bureau report has been placed on the author of each report. The analyses of survey findings are usually written by specialists in the subject matter on the Census Bureau staff. The Chief of the Statistical Methods Division is responsible for insuring that Census Bureau standards discussed in [14; 65], are applied and must attest that the analysis meets the quality standards before the report is published.

The author of a report submits a copy of the proposed analysis of the results of a study and provides evidence that the required statistical tests have been performed. The Statistical Methods Division then checks that all tests have been applied and applied correctly. The division assists the author in performing the tests in several ways. A program of training is provided to authors in recognizing the need for the review; the methods of applying the appropriate tests are illustrated by lectures; work with a specially written analysis of a hypothetical "The Book and Dexterity Survey." A computer program is available which can be directed to execute a number of tests; data provided to it and will perform much of the tedious arithmetic involved in some of the tests. Counseling on the more difficult and unusual tests is also provided.

<sup>4</sup> The sample sizes are large enough that we may safely assume the normality of all but some very small CPS estimates to be approximately normal.

# VARIANCE ESTIMATES TO DETERMINE OPTIMUM SURVEY DESIGN

The sample design and the estimation procedures used in the CPS have been modified on several occasions during the history of the survey. These changes result from a continuing program of research and development having the objective of optimizing the use of the techniques and resources which are available at a given time; changes also occur when reliability conditions or tabulation areas are revised. Knowledge of the components of variance attributable to each of the several stages of sampling

and the effect of the separate steps in the estimation procedure on the variance of the estimate are needed to develop an efficient sample design.

Formulas for the estimation of the components of variance that result from the stages of sampling and estimation are discussed with formula (K/37) in appendix K, p. 157. The discussions given there assume the CPS tabulations are derived using the estimator given by formula (V/1), on p. 63, which does not include composite estimation. The formulas for estimating the variance components and the impact of the estimation procedure on the total variance for composite estimates of level and month-to-month change are outlined in [36: 53] and are the basis of

TABLE VIII-1 Components of Estimated Sampling Variance for CPS Composite Estimates of U.S. Level, Monthly Averages: 1975

Population 16 years old and over	Average estimate of level <sup>1</sup> (x 10 <sup>3</sup> )	Average standard error of level <sup>2</sup> (x 10 <sup>3</sup> )	Distribution of variance <sup>3</sup>		
			Within PSU (percent)	Between PSU (percent)	Between stratum (percent)
(1)	(2)	(3)	(4)	(5)	(6)
<i>Not-in-labor-force, total</i>	58.655	228.01	83.7	14.7	1.6
Not white	7.239	85.02	95.0	5.1	-
<i>Civilian labor force, total</i>	92.612	228.01	83.7	14.7	1.6
Not white	10.529	85.02	95.0	5.1	-
Part time	8.197	112.68	96.4	2.6	1.0
Teenagers, 16-19 years old	8.799	82.53	93.9	6.0	1
<i>Employed in agriculture, total</i>	3.381	103.54	87.0	9.7	3.3
Males	2.801	85.43	91.4	5.1	3.5
Not white	284	31.15	78.2	23.6	-1.8
Teenagers, 16-19 years old	453	31.42	87.4	12.7	1
<i>Employed in nonagriculture, total</i>	81.402	243.39	90.3	8.0	1.7
Males	48.429	146.17	96.9	1.5	1.6
Not white	8.787	86.24	98.0	1.8	1
Farm residence	1.948	80.92	85.6	15.4	-1.1
With a job, not at work	5.007	107.41	97.3	2.9	-2
Self-employed	5.626	96.85	88.7	11.0	3
Teenagers, 16-19 years old	6.593	80.27	96.2	3.9	1
<i>Unemployed, total</i>	7.830	123.75	93.1	5.4	1.5
Not white	1.459	55.17	95.5	5.1	5
Teenagers, 16-19 years old	1.752	55.80	96.0	3.9	2
<i>In SMSA's, total</i>	103.355	446.32	93.8	4.4	1.7
Central city	44.956	367.27	102.8	-4.6	1.8
Rural nonfarm residence	36.919	454.82	93.1	4.9	2.0
Rural farm residence	6.520	185.44	106.9	-5.8	-1.2
Household heads	72.422	165.80	99.9	-6	8

- Entry represents zero.

<sup>1</sup> The arithmetic mean of the 12 monthly estimated levels for the year

<sup>2</sup> The square root of the arithmetic mean of the variances of the 12 monthly estimated levels for the year

<sup>3</sup> The percent of the estimated total variance attributed to the three stages of sampling (See p. 96.)

the estimated variances given in the tables discussed in the following sections. Annual averages have been used to improve the stability of the estimated variances.

### Variance Components Due to Stages of Sampling

Tables VIII-1 and VIII-2 indicate, for the composite estimates given by formula (V/5), p. 64, how the several stages of sampling affect the total variance of level and month-to-month change, respectively.

Column (4) in each of these tables shows that, for all of the characteristics, the portion of sampling error which includes the variance from sampling within the PSU is the largest. Characteristics related to agricultural employment, total civilian labor force, and persons not in the labor force have somewhat smaller contributions than other characteristics. Because total civilian labor force, plus total not in labor force, must add to the independent second-stage ratio estimate control figures (which are assumed to have no variance), the standard errors for these totals are the same for the two characteristics; a similar statement

TABLE VIII-2. Components of Estimated Sampling Variance for CPS Composite Estimates of U.S. Month-to-Month Change, Monthly Averages: 1975

Population 16 years old and over	Average estimate of change <sup>1</sup> (x 10 <sup>3</sup> )	Average standard error of change <sup>2</sup> (x 10 <sup>3</sup> )	Distribution of variance <sup>3</sup>		
			Within PSU (percent)	Between PSU (percent)	Between stratum (percent)
(1)	(2)	(3)	(4)	(5)	(6)
<i>Not in labor force, total</i>	657	167.01	98.9	0.7	0.4
Not-white	115	68.49	99.0	1.3	3
<i>Civilian labor force, total</i>	624	167.01	98.9	7	4
Not-white	110	68.49	99.0	1.3	3
Part time	295	109.27	100.7	-1.3	5
Teenagers, 16-19 years old	533	81.44	97.2	3.2	4
<i>Employed in agriculture, total</i>	209	69.36	87.4	12.6	-1
Males	144	54.35	95.2	4.5	3
Not-white	29	25.41	86.6	15.8	-2.4
Teenagers, 16-19 years old	95	29.23	89.1	11.0	-
<i>Employed in nonagriculture, total</i>	656	170.36	100.2	-8	6
Males	457	112.57	101.9	-2.4	5
Not-white	101	64.88	100.8	-8	-
Farm residence	49	53.67	97.4	3.2	6
With a job, not at work	1,532	131.15	97.2	2.7	-
Self-employed	134	77.01	97.3	2.6	2
Teenagers, 16-19 years old	393	74.33	99.2	1.3	5
<i>Unemployed, total</i>	442	123.61	100.0	-7	7
Not-white	87	57.54	99.3	9	-2
Teenagers, 16-19 years old	192	63.27	99.1	1.0	-
<i>In SMSA's, total</i>	147	148.00	95.0	6.5	-1.4
Central city	73	133.13	100.1	-4	3
Rural nonfarm residence	155	160.01	93.6	7.2	8
Rural farm residence	80	98.51	88.1	13.3	-1.3
Household heads	150	89.85	100.3	-	-3

- Entry represents zero.

<sup>1</sup>The arithmetic mean of the absolute values of the 12 month to month changes for the year

<sup>2</sup>The square root of the arithmetic mean of the variances of the 12 month to month changes for the year

<sup>3</sup>The percent of the estimated total variance attributed to the three stages of sampling

applies to estimates for the not-white categories for these characteristics as well.

Negative figures in columns 5 and 6 imply negative estimates of the associated variance components; however, because the variances between PSU's and between strata are relatively small, and the variance estimators have a variance of their own, some negative estimates are to be expected.

When estimating month-to-month change, the component which includes the within-PSU portion of the estimated total

variance tends to be even larger than for level (compare columns (4) for tables VIII-1 and VIII-2). Each month-to-month difference is based on information from the same set of sample PSU's in the two months.

In table VIII-2, the ratio of column 2 to column 3 indicates the degree of confidence in estimates of month-to-month change, if one assumes the change and its standard error is constant by month during 1975. For example, the ratio for not in the labor force is about 3.9 ( $657/167.01 \approx 3.9$ ) and, because it is

Table VIII-3. Variance of Composite Estimates of Monthly U.S. Level and Variance Factors for Selected Estimators, Monthly Averages: 1975

Population 16 years old and over	Variance of composite estimate of level <sup>1</sup> (x 10 <sup>9</sup> )	Variance factor <sup>2</sup>			
		Unbiased estimator	Unbiased estimate with		
			First stage	Second stage	First and second stage
(1)	(2)	(3)	(4)	(5)	(6)
<i>Not in labor force, total</i> .....	51.991	4.0698	4.0132	1.1543	1.1046
Not-white .....	7.228	4.9031	4.0965	1.1022	1.0979
<i>Civilian labor force, total</i> .....	51.991	8.3244	8.2473	1.1543	1.1406
Not-white .....	7.228	8.3173	7.3137	1.1022	1.0979
Part time .....	12.696	1.2010	1.2017	1.0859	1.0832
Teenagers, 16-19 years old .....	6.811	3.2240	3.2128	1.0359	1.0306
<i>Employed in agriculture, total</i> .....	10.720	1.4785	1.0603	1.5300	1.1076
Males .....	7.299	1.5001	1.0460	1.5649	1.1060
Not-white .....	.970	1.1583	1.0150	1.2478	1.1424
Teenagers, 16-19 years old .....	.987	1.1842	1.0792	1.1633	1.0670
<i>Employed in nonagriculture, total</i> .....	59.237	6.0689	6.0057	1.2151	1.1666
Males .....	21.365	6.7699	6.6832	1.2189	1.1569
Not-white .....	7.437	5.6909	5.1624	1.2178	1.1762
Farm residence .....	6.547	1.2052	1.1749	1.2660	1.2306
With a job, not at work .....	11.537	.8752	.8710	.8496	.8473
Self-employed .....	9.380	1.2662	1.2622	1.1835	1.1844
Teenagers, 16-19 years old .....	6.444	2.3790	2.3715	1.0841	1.0699
<i>Unemployed, total</i> .....	15.313	1.1609	1.1578	1.0101	1.0007
Not-white .....	3.044	1.2898	1.2074	.9965	.9758
Teenagers, 16-19 years old .....	3.114	1.1254	1.1202	.9669	.9582
<i>In SMSA's, total</i> .....	199.206	2.6815	2.6155	1.1950	1.1318
Central city .....	134.885	1.5732	1.3271	1.4023	1.2137
Rural nonfarm residence .....	206.862	2.0887	1.6124	1.5997	1.1358
Rural farm residence .....	34.389	1.8227	1.2250	1.8465	1.2361
Household heads .....	27.490	6.6847	6.5833	1.2884	1.2706

<sup>1</sup>The arithmetic mean of the 12 variances of monthly level for the year.

<sup>2</sup>The ratio of the variance of the indicated estimator to the variance of the composite estimate of level. Factors are based on monthly variances averaged for the year.

greater than 2.0, suggests the change can be measured with better than 95-percent confidence. Under these same assumptions, the month-to-month change can be detected with 95-percent confidence or better for about half of the characteristics shown.

### Total Variance as Affected by Estimation

Table VIII-3 shows how the separate estimation steps affect the variance of estimated levels. The figures for civilian labor

force show, for example, that the variance of the composite estimate of level is  $51.991 \times 10^9$  (equal to the square of the standard error of level for this characteristic given in column 3 of table VIII-1). The variance of the simple unbiased estimate for this characteristic would be 8.3244 times as large (given by the factor in column 3 of table VIII-3); however, if the first stage of ratio estimation is also included with the unbiased estimator, the variance would be 8.2473 times as large as the composite estimate of level. The unbiased estimate with the second stage of ratio estimation (but without the first stage) would be only 1.1543 times as large. If the first and second stages of ratio

Table VIII-4. Variance of Composite Estimates of U.S. Month-to-Month Change and Variance Factors for Selected Estimators, Monthly Averages: 1975

Population 16 years old and over	Variance of composite estimate of change <sup>1</sup> (x 10 <sup>9</sup> )	Variance factor <sup>2</sup>
		Unbiased estimate with first and second stage
(1)	(2)	(3)
<i>Not in labor force, total</i>	27.892	1.4066
Not-white	4.691	1.3220
<i>Civilian labor force, total</i>	27.892	1.4066
Not-white	4.691	1.3220
Part time	11.940	1.1695
Teenagers, 16-19 years old	6.633	1.1328
<i>Employed in agriculture, total</i>	4.810	1.3903
Males	2.954	1.4442
Not-white	645	1.2170
Teenagers, 16-19 years old	854	1.1846
<i>Employed in nonagriculture, total</i>	29.024	1.4746
Males	12.673	1.3530
Not-white	4.209	1.4634
Farm residence	2.881	1.4648
With a job, not at work	17.201	.9590
Self-employed	5.931	1.4152
Teenagers, 16-19 years old	5.524	1.1988
<i>Unemployed, total</i>	15.281	1.0926
Not-white	3.311	1.0778
Teenagers, 16-19 years old	4.003	1.0302
<i>In SMSA's, total</i>	21.905	2.2049
Central city	17.722	1.9659
<i>Rural nonfarm residence</i>	25.604	1.9913
<i>Rural farm residence</i>	9.704	1.7474
<i>Household heads</i>	8.072	2.0252

<sup>1</sup>The arithmetic mean of the 12 variances of month-to-month change for the year

<sup>2</sup>The ratio of the variance of the unbiased estimator, after the first and second stage of ratio estimation, to the variance of the composite estimate for month-to-month change. Factors are based on monthly variances averaged for the year

estimation are included, i.e., the estimator given by expression (V/1), p. 63, its variance would be 1.1406 times the variance of the composite estimate.

The gain in variance for level for each characteristic due to the composite estimate is given by the factor shown in column (6) of table VIII-3 and for month-to-month change by the factor in

column 3 of table VIII-4. These factors show that the composite estimate improves the variance of all items shown for level and change, except for some unemployment items and for "with a job, not at work" in nonagriculture. These latter characteristics tend to have low month-to-month correlations. (See ch. V, p. 62.)

## Chapter IX.

# CURRENT POPULATION SURVEY USED FOR SUPPLEMENTAL INQUIRIES

### INTRODUCTION

#### How The CPS Is Used For Supplemental Inquiries

In addition to meeting the needs for information on the labor force status of the population each month, the CPS is used to collect data for a variety of studies of the general population and for specific subsets of the population. Such studies have been conducted in the past for the use of Federal and State agencies, private foundations, and other organizations.

The studies make use of the CPS facilities through one of the following approaches—

- By including a limited number of supplemental questions as part of the CPS interview.
- By identifying a desired subgroup of sample households or sample persons during the basic CPS interview and leaving a separate questionnaire to be completed either for the household or for a designated member and returned by mail.
- By identifying a desired subgroup after the basic CPS interview and subsequently conducting a mail survey to obtain the necessary information.
- By utilizing all or part of the sample for interviews conducted between the regular CPS interviews.

In addition, for data collection efforts which cannot be accommodated through these approaches, the sampling frames, developed for the CPS, are used to select samples for surveys conducted independently of the CPS. Examples include the Health Interview Survey, a weekly, national survey designed to collect information on the morbidity and health-related characteristics of the population; the National Crime Surveys, which are conducted monthly and provide national data quarterly on the victims of crimes in the population; and the Annual Housing Survey, which collects data on the characteristics of the housing inventory for selected SMSA's and on a national basis. In these surveys, the sample households are selected within the PSU's to be independent of those used in the CPS, although the same sample PSU's are used. (See app. E.)

Independent surveys can also be based on old CPS sample households that have completed their regular 8 months of CPS interviews.<sup>1</sup> For such surveys, the sample selection costs are minimized and travel costs are reduced since units have been located previously. In addition, since some household data are available on CPS records for stratification purposes, samples may be more efficiently selected to cover specific population subgroups.

<sup>1</sup>The conditioning of responses to a proposed study by earlier CPS interviews must be considered in all methods of using CPS sample households.

### CRITERIA FOR ACCEPTABLE SUPPLEMENTAL INQUIRIES

As a basis for undertaking special surveys for Federal agencies or other sponsors, a number of criteria have been developed which determine the acceptability of such work. The present criteria were developed by the Bureau of the Census and reviewed and elaborated upon by a committee of the American Statistical Association [1]. Although not listed in the following as a separate criterion, the confidentiality requirements of the Bureau's enabling legislation [34] apply to all data collected by the Census Bureau. In no case does the the Bureau provide information which will permit the identification of an individual or a household, such as name or address.

The staff of the Census Bureau, working with the sponsor agency, develops the survey design, including the methodology, questionnaire wording, followup program, and the size of sample. Although these features are subject to review by sponsor, the Census Bureau conforms to the same standard: accuracy and reliability that apply to regular work.

The following criteria are reviewed to determine the acceptability of a program that proposes to use CPS facilities:

1. The subject matter of the inquiry must be in the public interest. Some illustrations are given by the supplemental inquiries conducted during calendar years 1975 and 1976 and listed in appendix I.
2. The inquiry must not have an adverse effect on the other Census Bureau programs. The questions must not cause respondents to question the importance of the survey and result in losses of response or quality. It is essential that the perception of the Bureau, to the public and Congress as the objective factfinder for the nation, is not damaged; other important functions of the Bureau, such as the decennial censuses or the economic censuses, must not be affected in terms of equal response rates or in congressional acceptance and approval of these programs.
3. The subject matter must be compatible with the basic survey. This is to avoid introducing a concept that affects the accuracy of responses to the basic CPS interview; for example, a series of questions incorporating a revised labor force concept that could inadvertently affect responses to the standard labor force items would not be included.
4. The inquiry must not slow down the work of the basic survey. In general, the supplemental inquiry must not require more than 20 minutes of interview time per household, nor create conflicts or competition for the use of Census staff resources that arise in dealing with the supplemental inquiry. The Census Bureau will not jeopardize the schedule for completing the CPS or other Census Bureau work by completing a supplemental inquiry within some specified time frame.

5. The subject matter must not be overly sensitive. This criterion is imprecise, and its interpretation has changed over time. For example, the subject of birth expectations, once considered sensitive, has been included as a CPS supplemental inquiry. Also, in appendix L, the question on the birth history of never-married women, asked in June 1975, would not have been considered suitable a few years ago.
6. It must be possible to meet the objectives of the inquiry through the survey method. That is, it must be possible to translate the survey objectives into meaningful questions, and the respondent must be able to supply the information required to answer the questions.
7. If the supplemental information is to be collected during the CPS interview, the inquiry must be suitable for the personal and/or telephone procedures used in the CPS. Additional flexibility in the interview process is possible if the information is to be collected by a supplemental questionnaire dropped off at the survey household or by an independent interview.
8. The nature of the inquiry must be such that the Census Bureau is uniquely qualified to conduct the survey. If the

proposed survey can be conducted by a competent public or private survey organization, the Census Bureau will not undertake the project. The Census Bureau does not engage in competitive bidding with other public or private survey organizations, nor does it solicit or cause requests to be initiated for the use of the Bureau's statistical services.

#### CLEARANCE BY THE OFFICE OF MANAGEMENT AND BUDGET

Even though the proposed inquiry is compatible with the criteria given in the preceding section, the Census Bureau does not make the final decision regarding the appropriateness or desirability of the supplemental survey. The Office of Management and Budget, through its Statistical Policy Division, reviews the proposal to make certain it meets Government-wide standards regarding the need for the data and appropriateness of the design and insures that the survey instruments, strategy, and respondent burden are acceptable.



## REFERENCES

1. American Statistical Association. Census Advisory Committee. "Report of the Subcommittee on Criteria for Surveys for Other Federal Government Agencies." *American Statistician*, February 1969, pp. 17-19.
2. Bailar, Barbara A. "The Effect of Repeated Interviewing." Paper presented at the Conference on Econometrics and Mathematical Economics, Seminar on the Analysis of Panel Microdata, Ann Arbor, Mich., Apr. 5-6, 1973.
3. ———. "The Effects of Rotation Group Bias on Estimates From Panel Surveys." *Journal of the American Statistical Association* 349 (March 1975): 23-30.
4. Banks, Martha J., and Shapiro, Gary M. "Variances of the Current Population Survey, Including Within- and Between-PSU Components and the Effect of the Different Stages of Estimation." *Proceedings of the Social Statistics Section, Annual Meeting of the American Statistical Association, Fort Collins, Colo., Aug. 23-26, 1971*. Washington, D.C.: American Statistical Association, 1972, pp. 40-49.
5. Brooks, Camilla. "The Effect of Controlled Selection on the Between-PSU Variance." *Proceedings of the Social Statistics Section, Annual Meeting of the American Statistical Association, Atlanta, Ga., Aug. 25-28, 1975*. Washington, D.C.: American Statistical Association, 1976, pp. 336-341.
6. Cochran, William. *Sampling Techniques*. New York: John Wiley and Sons, 1953.
7. Dippo, Cathryn. "Expansion of CPS to Provide Reliable State Estimates of Unemployment." *Proceedings of the Social Statistics Section, Annual Meeting of the American Statistical Association, Atlanta, Ga., Aug. 25-28, 1975*. Washington, D.C.: American Statistical Association, 1976, pp. 387-394.
8. Eckler, A. Ross. *The Bureau of the Census*. New York: Praeger, 1972.
9. Executive Office of the President, Office of Management and Budget, Statistical Policy Division. *Federal Statistics: Coordination, Standards, Guidelines: 1976*. Washington, D.C.: Government Printing Office, 1976.
10. ———. *The Statistical Reporter*. Washington, D.C.: Government Printing Office (monthly).
11. Fellegi, I. P. "Response Variance and Its Estimation." *Journal of the American Statistical Association* 59 (December 1964): 1016-1041.
12. Frankel, Lester R., and Stock, J. Stevens. "On the Sample Survey of Unemployment." *Journal of the American Statistical Association* 37 (March 1942): 77-80.
13. Frankel, Martin R. *Inference From Survey Samples*. Ann Arbor: University of Michigan, Institute for Social Research, 1971.
14. Gonzalez, Maria, Ogus, Jack L., Shapiro, Gary, and Tepping, Benjamin J. "Standards for Discussion and Presentation of Errors in Survey and Census Data." *Journal of the American Statistical Association* 70 (September 1975): 1-23.
15. Goodman, Roe, and Kish, Leslie. "Controlled Selection: A Technique in Probability Sampling." *Journal of the American Statistical Association* 45 (September 1950): pp. 350-372.
16. Gurney, Margaret, and Daly, Joseph F. "A Multivariate Approach to Estimation in Periodic Sample Surveys." *Proceedings of the Social Statistics Section, Annual Meeting of the American Statistical Association, Philadelphia, Pa., Sept. 8-11, 1965*. Washington, D.C.: American Statistical Association, 1966, pp. 242-257.
17. ———, and Jewett, Robert S. "Constructing Orthogonal Replications for Variance Estimation." *Journal of the American Statistical Association* 70 (December 1975): 819-8
18. Hagood, Margaret Jarman. "Development of a 1940 Farm Level of Living Index for Counties." *Rural Sociology* June 1943, pp. 171-180.
19. ———. "Rural Level of Living Indexes, 1940." In *New Rural Sociology*, September 1943, pp. 292-293.
20. Hansen, Morris H., and Hurwitz, William N. "On the Theory of Sampling From Finite Populations." *Annals of Mathematical Statistics* XIV (1943): 333-362.
21. ———, ———, and Madow, William G. *Sample Surveys: Methods and Theory. Vol. I. Methods and Applications*. New York: John Wiley and Sons, 1953.
22. ———, ———, and ———. *Sample Survey Methods and Theory. Vol. II. Theory*. New York: John Wiley and Sons, 1953.
23. ———, ———, Nisselson, Harold, and Steinberg, ———. "The Redesign of the Census Current Population Survey." *Journal of the American Statistical Association* 50 (September 1955): 701-719.
24. ———, ———, and Pritzker, Leon. "The Estimation and Interpretation of Gross Differences and the Simple Ratio Variance." *Contributions to Statistics Presented to P. P. C. Mahalanobis on the Occasion of His 70th Birthday*. Calcutta: Pergamon Press, 1964.
25. Jabine, T. B., and Tepping, Benjamin J. "Control of Quality of Occupation and Industry Data." *Proceedings of the 39th Session of the International Statistical Institute, August 1973, Vol. XLV, bk. 3*. Vienna: Österreichisches statistisches Zentralamt, 1974, pp. 360-392.
26. Keyfitz, Nathan. "Estimates of Sampling Variance When Two Units Are Selected From Each Stratum." *Journal of the American Statistical Association* 52 (December 1957): 510.
27. ———. "Sampling with Probabilities Proportional to Size: Adjustment for Changes in Probabilities." *Journal of the American Statistical Association* 46 (March 1951): 109.
28. Ono, Mitsuo, and Miller, Herman P. "Income Nonresponse in the Current Population Survey." *Proceedings of the Social Statistics Section, Annual Meeting of the American Statistical Association, New York, Aug. 19-22, 1969*. Washington, D.C.: American Statistical Association, 1969, pp. 73-80.
29. Palmer, Susan. "On the Character and Influence of Nonresponse in the Current Population Survey." *Proceedings of the Social Statistics Section, Annual Meeting of the American Statistical Association, Washington, D.C., Dec. 27-31, 1973*. Washington, D.C.: American Statistical Association, 1974, pp. 73-80.
30. President's Committee to Appraise Employment Statistics. *Measuring Employment and Unemployment*. Washington, D.C.: Government Printing Office, 1967.
31. Sudman, Seymour. *Reducing the Cost of Surveys*. Aldine Publishing Co., 1967.

32. Tepping, Benjamin J. "Variance Estimation in Complex Surveys." *Proceedings of the Social Statistics Section, Annual Meeting of the American Statistical Association*, Pittsburgh, Pa., Aug. 20-23, 1968. Washington, D.C.: American Statistical Association, 1969, pp. 11-18.
33. Thompson, Marvin M., and Shapiro, Gary. "The Current Population Survey: An Overview." *Annals of Economic and Social Measurement* 2 (April 1973).
34. U.S. Code, Title 13: Census.
35. U.S. Congress. *Comprehensive Employment and Training Act of 1973*. Pub. L. 93-203.
36. U.S. Department of Commerce, Bureau of the Census. "Calculating the Composite Estimate and the Variances for the Composite Estimate and Month-to-Month Change." Memorandum from P. Bounpane to D. Diskin, Nov. 14, 1974.
37. ———. *County and City Data Book: 1972 (A Statistical Abstract Supplement)*. Washington, D.C.: Government Printing Office, 1973.
38. ———. "The Effect of Special Procedures to Improve Coverage in the 1970 Census." U.S. Census of Population and Housing: 1970 Evaluation and Research Program. Ser. PHC (E)-6. Washington, D.C.: Government Printing Office, 1974.
39. ———. *Current Population Survey: Interviewers Reference Manual*. Ser. CPS 250. Washington, D.C.: Government Printing Office, 1971.
40. ———. *Current Population Survey (CPS) Office Manual*. Ser. CPS-HVS 256. Washington, D.C.: Government Printing Office, 1973.
41. ———. "1970 CPS Redesign: Illustration of Computation of PSU Probabilities Within a Given 1970 Stratum" Memorandum from W.M. Perkins to J. Waksberg, Feb. 19, 1971.
42. ———. "1970 CPS Redesign: Proposed Method for Deriving Sample PSU Selection Probabilities Within 1970 NSR Strata." Memorandum from W.M. Perkins to J. Waksberg, Aug. 5, 1970.
43. ———. *The Current Population Survey Reinterview Program* January 1961 through December 1966, by Ralph Bailey, Irwin Schreiner, Leonard Baer, and Linda Kerstetter. Technical Paper 19. Washington, D.C.: Government Printing Office, 1968.
44. ———. *The Current Population Survey: A Report on Methodology*, by Joseph Steinberg, Thomas B. Jabine, and Leon Pritzker. Technical Paper 7. Washington, D.C.: Government Printing Office, 1963.
45. ———. "Data Collection Coverage and Current Program PSU's: January 1973." Form 11-1A (map). Washington, D.C.: Government Printing Office, 1973.
46. ———. "1970 Data-Collection Forms and Procedures." *U.S. Census of Population and Housing: 1970*. Ser. PHC(R)-2. Washington, D.C.: Government Printing Office, 1971.
47. ———. "Educational Attainment in the United States: March 1975." *Current Population Reports*. Ser. P-20, No. 295. Washington, D.C.: Government Printing Office, 1976.
48. ———. "Estimates of Coverage of Population, by Sex, Race, and Age: Demographic Analysis." *U.S. Census of Population and Housing: 1970 Evaluation and Research Program*. Ser. PHC(E)-4. Washington, D.C.: Government Printing Office, 1974.
49. ———. "Estimates of the Population of the United States, by Age, Sex, and Race: 1970 to 1975." *Current Population Reports*. Ser. P-25, No. 614. Washington, D.C.: Government Printing Office, 1975.
50. ———. "Estimates of the Population of the United States by Age, Sex and Race: July 1, 1974 to 1976." *Current Population Reports*. Ser. P-25, No. 643. Washington, D.C.: Government Printing Office, 1976.
51. ———. "General Housing Characteristics for the United States and Regions." pt. A. *Current Housing Reports*. Ser. H-150 73A. *Annual Housing Survey: 1973*. Washington, D.C.: Government Printing Office, 1975.
52. ———. "General Housing Characteristics for the United States and Regions." pt. A. *Current Housing Reports*. Final Report. Ser. H-150-74. *Annual Housing Survey: 1974*. Washington, D.C.: Government Printing Office, 1976.
53. ———. "General Outline of the Steps in the Keyfitz-Estimat and Variance Programs." Memorandum from P. Bounpane to D. Diskin, Nov. 6, 1974.
54. ———. "Housing Authorized by Building Permits and Public Contracts: 1975." *Construction Reports*. Ser. C40-00-1; and Ser. C40-00-13. Washington, D.C.: Government Printing Office (monthly reports and annual summaries).
55. ———. "Vacant Housing Units in the United States." *Housing Vacancies: Current Housing Reports*. Ser. H-111. Washington, D.C.: Government Printing Office (quarterly).
56. ———. "Money Income in 1974 of Families and Persons in the United States." *Current Population Reports*. Ser. P-60 No. 101. Washington, D.C.: Government Printing Office 1976.
57. ———. "Number of Inhabitants." pt. A. *U.S. Census of Population: 1970*. Vol. 1. *Characteristics of the Population*. Washington, D.C.: Government Printing Office, 1972.
58. ———. "Optimum Size of Segment: CPS Redesign." Memorandums from J. Waksberg to J. Daly, Nov. 18, 1970; July 17, 1971; and Apr. 23, 1974.
59. ———. "PSU's Included in Current Survey Samples." Form 11-4. Washington, D.C.: Government Printing Office, September 1976.
60. ———. *Reinterview Manual*. Form 11-56. Washington, D.C.: Government Printing Office, 1975.
61. ———. *Response Variance in the Current Population Survey*, by Benjamin J. Tepping and Katherine L. Boland. Working Paper 36. Washington, D.C.: Government Printing Office, 1972.
62. ———. "Rotation of PSU's for A- and C-Design Samples." Memorandum for the Record from C. Brooks and R. Hanson, Mar. 26, 1975.
63. ———. "Sample PSU's, 461-Area Current Population Survey National Sample: March 1973." Form 11-5 (map). Washington, D.C.: Government Printing Office, 1976.
64. ———. "Sample PSU's, 461-Area Sample and State Supplements: August 1976." Form 11-6 (map). Washington, D.C.: Government Printing Office, 1976.
65. ———. *Standards for Discussion and Presentation of Errors in Data*, by Maria Gonzalez, Jack L. Ogus, Gary M. Shapiro, and Benjamin J. Tepping. Technical Paper 32. Washington, D.C.: Government Printing Office, 1974.
66. ———. *State Economic Areas*, by Donald J. Bogue. Washington, D.C.: Government Printing Office, 1951.
67. ———. *U.S. Census of Population and Housing: 1970: Procedural History*. Ser. PHC(R)-1. Washington, D.C.: Government Printing Office, 1976, pp. 14-1 to 14-18.

68. ———. *The Variance of the Replication Method for Estimating Variances from Grouped Data for the CPS Sample Design*, by Margaret Gurney. Technical Notes 3. Washington, D.C.: Government Printing Office, 1970. pp. 7-12.
69. ———. *The X-11 Variant of the Census Method II Seasonal Adjustment Program*, by Julius Shiskin, Allan H. Young, and J. C. Musgrave. Technical Paper 15. Washington, D.C.: Government Printing Office, 1965.
70. ——— and U.S. Department of Labor, Bureau of Labor Statistics. "Concepts and Methods Used in Labor Force Statistics Derived From the Current Population Survey." *Current Population Reports. Special Studies Ser. P-23*, No. 62. Washington, D.C.: Government Printing Office, 1976.
71. U.S. Department of Health, Education, and Welfare, National Center for Health Statistics. *Pseudoreplication, Further Evaluation and Application of the Balanced Half-Sample Technique*, by Philip J. McCarthy. Ser. 2, No. 31. Washington, D.C.: Government Printing Office, 1969.
72. ———. *Replication, an Approach to the Analysis of Data from Complex Surveys*, by Philip J. McCarthy. Ser. 2, No. 14. Washington, D.C.: Government Printing Office, 1966.
73. U.S. Department of Labor, Bureau of Labor Statistics. *Geographic Profile of Employment and Unemployment: 1973*. Report No. 431. Washington, D.C.: Government Printing Office, 1974.
74. ———. *Geographic Profile of Employment and Unemployment: 1974*. Report No. 452. Washington, D.C.: Government Printing Office, 1976.
75. ———. *Employment and Earnings*. Washington, D.C.: Government Printing Office (monthly).
76. ———. *Employment and Earnings*. Vol. 17, No. 8. Washington, D.C.: Government Printing Office, 1971.
77. ———. "Gross Changes in the Labor Force: A Problem in Statistical Measurement," by Robert B. Pearl in *Employment and Earnings*. Vol. 9, No. 10. Washington, D.C.: Government Printing Office, 1963.
78. Williams, W. H., and Mallows, C. L. "Systematic Biases in Panel Surveys Due to Differential Nonresponse." *Journal of the American Statistical Association* 65 (September 1970) 1338-1349.
79. Woodruff, Ralph S. "A Simple Method for Approximating the Variance of a Complicated Estimate." *Journal of the American Statistical Association* 66 (June 1971): 411-414.
80. ——— and Causey, Beverly D. "Computerized Method for Approximating the Variance of a Complicated Estimate." *Journal of the American Statistical Association* 71 (June 1976): 315-321.

## Appendix A.

**CONTROLLED SELECTION OF SAMPLE PSU'S**

## INTRODUCTION

Controlled selection may be thought of as a device for achieving a more thorough stratification in sample selection than is ordinarily feasible with conventional random-stratified methods [15]. Typically, as the number of criteria used in determining strata is increased, the problem of classifying PSU's becomes rapidly more complex; PSU's that clearly belong in one stratum, according to one or two criteria, also clearly belong in different strata when other criteria are considered. Conventional stratification procedures can increase the number of strata to the point that one sample PSU is to be selected from each stratum; beyond this point, the claims of conflicting criteria can only be disregarded or compromised.

In controlled selection, the stratifying criteria are divided into two groups. The first group of criteria is used to form strata of PSU's in the conventional manner, with the final sample of PSU's defined in terms of the number of PSU's (usually one) to be selected from each stratum. The second group of criteria deals with characteristics of the various selections of PSU's possible by choosing one from each stratum and are used to derive controls or conditions that define admissible patterns of sample PSU's. *Admissible patterns* are selections of PSU's that satisfy the conditions of the second as well as the first group of criteria. Consequently, the final sample of PSU's must be stratified in relation to both the first and second group of criteria.

If only one of the admissible patterns of PSU's is worked out and designated as the final sample, the result would be a purposive selection, not a probability sample. To insure that the final sample of PSU's has a probability base, controlled selection develops a set of admissible patterns with very special properties: Each pattern in the set is assigned a probability, and the sum of the probabilities of the patterns in which a given PSU occurs is exactly equal to the probability assigned that PSU. Consequently, a probability selection of one pattern out of the set is, at the same time, a probability selection of each of the PSU's. In the following explanation, the principal

In the following explanation, the principles and procedures for accomplishing the purpose of controlled selection are discussed under four topics: (1) The controlled selection problem as it applies to the CPS, (2) the derivation of controls defining admissible patterns of sample PSU's, (3) the development of a complete set of admissible patterns and the assignment of a probability to each pattern, and (4) the selection of a sample admissible pattern to determine the sample PSU's. Some of the examples included in the discussion are based on the specific procedures used in making a controlled selection of PSU's for the CPS A-design sample.

**THE CONTROLLED SELECTION PROBLEM  
AS IT APPLIES TO THE CPS**

Three operations precede the controlled selection sampling procedure: The division of all PSU's into strata in accordance with the first group of criteria, the determination of the second group of criteria used to control the PSU sample selection, and

the assignment of a probability for each PSU of being selected in the final sample.

The method for developing strata, and the criteria used, are discussed in chapter II. Two additional criteria taken as controls are a distribution of PSU's by geographic areas and a two-way division of PSU's between those in and those not in the sample under the old design. For the most part, the geographic areas used are States, but, in some instances, a State's territory is divided into two or more subdivisions. Within each stratum, PSU probabilities are computed proportionate to 1970 census populations and then modified by the Keyfitz technique to maximize retention in sample of those PSU's in the old design sample of PSU's. The computation of these probabilities is discussed in [27: 41: 42].

Theoretically, all strata and all PSU's within each stratum could have been put through the controlled selection process as a single operation. In practice, however, the work is simplified by reducing the listing of PSU's and strata in the following ways:

1. All SR strata are omitted. These PSU's would be in any sample to be designated, and nothing is to be gained by putting them through the selection procedure.
2. In some instances in NSR strata, the Keyfitz technique results in retaining the old design sample PSU with certainty. These strata are omitted.
3. In many NSR strata, although no PSU is certain to be selected, there are individual PSU's given a zero probability of selection by the Keyfitz technique. These PSU's are omitted from the listing.
4. A fair number of NSR strata include a cluster of two or more PSU's that, from the standpoint of controlled selection, have identical characteristics—i.e., all members of the given cluster are from the same geographic area, and all are either in the sample under the old design or not in sample under the old design. Each such cluster is treated as a single PSU in the simplified listing, and its attached probability is equal to the sum of the probabilities for the PSU's making up the cluster. After controlled selection is finished, those clusters falling in sample are independently subsampled to determine which specific PSU is to be in sample. The following description of steps in the controlled selection operation does not further distinguish these PSU's as being clusters or individual PSU's.
5. Since the NSR strata do not cross the boundaries of the four census regions, it is convenient to carry out controlled selection separately, region by region.
6. With respect to strata and PSU probabilities, CPS controlled selection for the A-design and for the C-design is treated as two independent problems, each with its own set of inputs.

**DERIVE THE CONTROLS THAT DEFINE  
ADMISSIBLE PATTERNS**

The first step in controlled selection is to label each PSU in the simplified listing according to two control criteria: The geographic area of the PSU and whether the PSU was in sample.

under the old design. Then, the expected number of PSU's from each given control category to be selected in the sample is determined by summing the probabilities for all PSU's within the specified category.

If sample PSU's are selected by conventional random-stratified procedures, any pattern of PSU's might be selected, provided only that it includes only one PSU from each CPS stratum. Controlled selection rules out many of these patterns of PSU's as inadmissible and gives them no chance of being selected. An admissible pattern, as defined for the CPS, is required to have the expected number of sample PSU's from each control category. In actual practice, these expected numbers are not whole numbers but fall between integers. Therefore, the number of sample PSU's in a given control category is admissible if it is either of the two integers closest to the expected number. Determining an admissible pattern is a trial and error process, and the number of sample PSU's in a given control category must be tested at each step.

The process of deriving control statements is illustrated in tables A-1 and A-2; the numbers are based on actual CPS A-design data from the West region. The rows in table A-1 represent CPS strata; each column represents a single geographic area (i.e., a control category). The sum of the PSU selection probabilities in the geographic area, given as column totals at the bottom of the table, are the expected numbers of PSU's to be selected from the specified geographic areas. In the body of the table, the probabilities with an asterisk indicate PSU's in the old design sample. The subtotal of these probabilities for the West region is 14.5214\*; this is the expected number of old design sample PSU's to be in each admissible pattern.

Table A-2 records, in column 2, the expected numbers from table A-1 and translates them into control statements. Column 3 of table A-2 shows the numbers of sample PSU's that may be found in an admissible pattern for each control category. The final column of table A-2 provides the probability for each admissible number of sample PSU's. This probability does not apply to the individual patterns but to the whole set of patterns from which the final selection is made.

Taking the control statement for Arizona, as an example, table A-2 shows that each pattern of sample PSU's developed as a potential candidate for the A-design sample must include either one or two PSU's from Arizona (from the strata in the simplified listing). Moreover, when a complete set of potential patterns has been formed, the sum of probabilities for patterns including one PSU from Arizona must equal 0.7890, and the sum of probabilities for patterns including two PSU's from Arizona must equal 0.2110. Similar control statements for each State (or subdivisions of States) must also be satisfied. A further control statement, derived from the total over all States of expected values for old design sample PSU's, requires each admissible pattern to include a total of 14 PSU's (with a probability of 0.4786) or 15 PSU's (with a probability of 0.5214) that were also in sample in the old design.

## DEVELOP AND ASSIGN PROBABILITIES TO THE ADMISSIBLE PATTERNS IN A COMPLETE SET

A worksheet for constructing a set of admissible patterns starts from a simplified listing of the appropriate universe, by

stratum and PSU, showing the probability and control categories for each PSU, and a list of control statements for each category showing the options and probabilities for each option. An illustration of this input to the development of a set of admissible patterns is provided in table A-3; the first three columns of sections I and II. For the sake of simplicity, this illustration is based on artificial data and involves only 5 strata, 13 PSU's, 3 geographic control categories (U, V, and W), and a control for designating PSU's in the old design sample (\*). An actual example would require a lengthy worksheet and many more admissible patterns.

Note that the control statements in section II are derived from section I. For example, if the probabilities for the five PSU's identified as U or U\* in section I are added, the sum is 1.6. Accordingly, the control statement in section II provides that each admissible pattern must contain one or two U's, and, of the set of admissible patterns, two U's must have a probability of 0.67 and one U must have a probability of 0.33.

The designation of the first admissible pattern is not predetermined, since any pattern that satisfies the control statements may be taken as a starting point. As one possible approach, for example, note that the control statement options have the lowest probabilities for two U's, one V, two W's, and two \*'s. A pattern with these options may be easier to find; accordingly, check five PSU's—one in each stratum—aiming to get the specified mix. The first trial may achieve the specified mix or, failing that, may achieve another type of admissible pattern, taking account of alternate control options. In either case, the first trial can be finalized as admissible pattern 1. If, however, the first trial checked off is not an admissible pattern, note the control statement has not been met and try switching PSU's to meet that condition without violating the other controls. With only five strata, it may be just as quick to reject the first trial selection completely and start afresh, but, in the practical situation involving many strata, it is faster to modify the inadmissible pattern.<sup>1</sup>

As soon as the first admissible pattern is finalized, an appropriate probability must be assigned. The initial probabilities are examined for each PSU in the pattern and also the initial probabilities for each control option appropriate to the pattern. The lowest of all of these probabilities becomes the probability for admissible pattern 1. Once this probability has been assigned, a new column is placed beside the initial probabilities, showing probabilities that remain after admissible pattern 1. In this column, for each PSU and each control option not involved in the first pattern, the initial probability is copied, unchanged, into the new column.

In the example provided in table A-3, admissible pattern 1 (see section III) does, in fact, include two U's, one V, two W's, and two (\*'s). By checking probabilities for each PSU in the pattern and each appropriate control option (in sections I and II), the lowest initial probability is seen to be 0.19; thus, this becomes the pattern probability. When the new column of residual probabilities is computed, at least one of the reduced probabilities becomes zero (in this example, the probability for PSU in stratum 5, indicated in table A-3 as U). This fact makes it possible for admissible pattern No. 1 to be duplicated later.

<sup>1</sup> This explanation assumes the set of admissible patterns is being developed by hand. Actually, the computer provides a tool for doing the job. In true selections, switching PSU's, if necessary, to satisfy the control statements and proceeding to find successive admissible patterns until the set

no further pattern would be admissible if it included PSU U in stratum 5.

The procedure for developing admissible pattern 2 (and all succeeding patterns) basically repeats the method of pattern 1—the only difference is that the new column of residual probabilities is used rather than the initial probabilities. Again, there is

not a predetermined selection, since any pattern satisfying the revised control statements is admissible. However, the range of choices is necessarily less wide, since at least one PSU and/or at least one control option is no longer an admissible choice. Thus, as each successive pattern is finalized, the scope of choice becomes narrower until the final admissible pattern—given the

TABLE A-1. PSU Selection Probabilities, by Stratum and by Control Criteria, for Selected States of the West Region: 1970 CPS A-Design NSR Strata

Stratum identification (1970 A-design)	Selected states in the West region					Total, West region
	Arizona	Colorado	New Mexico	Utah	Remaining States	
(1)	(2)	(3)	(4)	(5)	(6)	(7)
803.....			0.0557	*0.0559	*0.4943 .3941	*0.5502 .4498
807.....		0.0414			*.8766 .0820	*.8766 .1234
809.....	*0.2079				*.7921	*1.0000
818.....		*.9033 .0710	.0069		.0188	*.9033 .0967
825.....		.0049		*.3573	*.6053 .0325	*.9626 .0374
830.....				.0130	*.9870	*.9870 .0130
832.....	.4200	.1891	.0185	.0079	.3645	1.0000
839.....	.1971				.8029	1.0000
841.....				.0672	.9328	1.0000
842.....	.0970	.1368	*.0976 .0256		*.2542 .3888	*.3518 .6482
844.....	.0560		.0759	.0176	*.4847 .3658	*.4847 .5153
847.....	*.2330	.0488	.3270	.0703	.3209	*.2330 .7670
Other strata in the West region...					*8.1722 3.8278	*8.1722 3.8278
Subtotal.....	*.4409 .7701	*1.9033 .4920	*1.0976 .5096	*.4132 .1760	*12.6664 7.5309	*14.5214 9.4786
Total.....	1.2110	1.3953	1.6072	.5892	20.1973	24.0000

\*PSU's in sample in the old A-design.

\*Probability sums shown are reduced by 1.0000 to reflect simplification in listing. (See text, app. A, p. 105.)

TABLE A-2. Expected Values and Control Statements for Geographic Areas and for PSU's  
In Old A-Design Sample, NSR Portion of the West Region: 1970 CPS A-Design

Geographic areas	Sum of selection probabilities for all NSR PSU's	Control statements	
		PSU's selected in admissible pattern	Probability
(1)	(2)	(3)	(4)
Arizona .....	1.2110	1 2	0.7890 .2110
California			4638
Subdivision A .....	2.5362	2 3	5362
Subdivision B .....	2.7436	2 3	2564 7436
Colorado .....	1.3953	1 2	6047 3953
Hawaii .....	.5671	— 1	4329 5671
Idaho .....	2.8086	2 3	1914 8086
Montana .....	2.0598	2 3	9402 0598
New Mexico .....	.6072	— 1	3928 6072
Nevada .....	.2061	— 1	7939 2061
Oregon			3583
Subdivision A .....	1.6417	1 2	6417
Subdivision B .....	1.3457	1 2	6543 3457
Utah .....	.5892	— 1	4108 5892
Washington			8527
Subdivision A .....	2.1473	2 3	1473
Subdivision B .....	2.0084	2 3	9916 0084
Wyoming .....	1.1328	1 2	8672 1328
Alaska <sup>2</sup> .....	1.0000	(X)	(X)
<b>Total for all NSR PSU's .....</b>	<b>24.0000</b>	<b>24</b>	<b>1.0000</b>
<b>PSU's in old A-sample<sup>3</sup> .....</b>	<b>14.5214</b>	<b>14</b> <b>15</b>	<b>4786</b> <b>5214</b>

\* For PSU's in sample in the old A-design.

— Entry represents zero.

X Not applicable.

<sup>1</sup> Probability shown reduced by 1.0000 to reflect simplification is listing. (See app. A, p. 105.)

<sup>2</sup> Alaska not involved in controlled selection.

<sup>3</sup> See table A-1, p. 107.

TABLE A-3. A Hypothetical Example of the Controlled Selection of PSU's  
(Worksheet Illustrating Development of a Complete Set of Admissible Patterns)

## I. PSU's. by Stratum

Stratum	Control category of PSU	Initial PSU probability	Probability remaining after admissible pattern number							
			1	2	3	4	5	6	7	8
1.....	U	0.27	0.27	0.27	0.22	0.22	0.22	0	—	—
	V	.19	.19	.19	.19	.19	.19	.19	0.04	0
	W	.54	.35	.19	.19	.01	0	—	—	—
2.....	U	.63	.63	.47	.42	.42	.41	.19	.04	0
	V	.37	.18	.18	.18	0	—	—	—	—
3.....	U	.36	.17	.01	.01	.01	0	—	—	—
	W	.40	.40	.40	.40	.22	.22	0	—	—
	V	.24	.24	.24	.19	.19	.19	.19	.04	0
4.....	U	.22	.22	.22	.22	.04	.04	.04	.04	0
	V	.28	.28	.28	.23	.23	.22	0	—	—
	W	.50	.31	.15	.15	.15	.15	.15	0	—
5.....	U	.19	0	—	—	—	—	—	—	—
	W	.81	.81	.65	.60	.42	.41	.19	.04	0

## II. Control Statements

Control category	Number of sample PSUs	Initial probability	Probability remaining after admissible pattern number							
			1	2	3	4	5	6	7	8
U.....	1	.033	.033	.033	.033	.015	.015	.015	0	—
	2	.67	.48	.32	.27	.27	.26	.04	.04	0
V.....	0	.16	.16	0	—	—	—	—	—	—
	1	.84	.65	.65	.60	.42	.41	.19	.04	0
W.....	2	.51	.32	.32	.27	.27	.26	.04	.04	0
	3	.49	.49	.33	.33	.15	.15	.15	0	—
	2	.95	.76	.60	.60	.42	.41	.19	.04	0
	3	.05	.05	.05	0	—	—	—	—	—

## III. Admissible Patterns

Stratum	Admissible patterns							
	1	2	3	4	5	6	7	8
1.....	W	W	U	W	W	U	V	V
2.....	V	U	U	V	U	U	U	U
3.....	U	U	W	W	U	W	W	W
4.....	W	W	V	U	V	V	W	U
5.....	U	W	W	W	W	W	W	W
Probability assigned to pattern	.019	.016	.005	.018	.001	.022	.015	.004
Cumulative probabilities	.19	.35	.40	.58	.59	.81	.96	1.00

• PSU's in sample in the old design

— Entry represents zero.

0 Selection that determines probability of the admissible pattern.



previous selections—is absolutely determined. Note, for example, in table A-3, section II, that after admissible pattern 7 is finalized, each of the five strata has only one PSU with any remaining probability, each of the control statements has only one option with any remaining probability, and the probabilities are all the same (i.e., 0.04). Admissible pattern 8, therefore, can be identified in only one way.

### SELECTION OF THE PSU'S IN SAMPLE

Once a complete set of admissible patterns is available, the final steps in controlled selection require the selection of the specific admissible pattern to be used for the sample.

The probabilities assigned to the set of admissible patterns for the CPS are first cumulated to 1.0000. A random number,  $R$ , is then chosen from a table of random numbers where  $0 < R \leq 1.0000$ . The first admissible pattern in the array whose cumulant is greater than or equal to  $R$  is selected as the sample pattern.

The corresponding cumulative pattern probabilities for the illustration are shown in the last row of section III of table A-3. In this example, it would be appropriate to choose a 2-digit random number. If  $0.40 < R \leq 0.58$ , then admissible pattern 4 becomes the sample pattern. Similarly, admissible pattern 5 is selected only if the random number is 0.59.<sup>2</sup>

The simplified example provided in table A-3 illustrates how the choice of a single random number selects the admissible pattern to use for the sample, gives the appropriate probability to each PSU in the listing, and also gives the appropriate probability to each of the control options in the control statements.

Observe PSU V in stratum 4, for example. This PSU is included in admissible patterns 3, 5, and 6 among those listed in section III. These patterns have probabilities of 0.05, 0.01, and 0.22, respectively. The chance that one of these patterns would be selected is 0.28 (i.e.,  $0.05 + 0.01 + 0.22$ ), and this is precisely the initial probability that PSU V in stratum 4 has in the section I listing.

Similarly, only patterns 4 and 7 include just one PSU from category U. The chance one of these two patterns would be selected is 0.33 (i.e.,  $0.18 + 0.15$ ), and this is precisely the initial probability the control statement in section II assigns for category U to be represented by just one sample PSU.

<sup>2</sup> In the simplified example, all probabilities are given to two decimal places. In the CPS design, probabilities were computed to four decimal places.

Since the CPS controlled selection is applied to a simplified listing of strata and PSU's (see the list of simplifications given on p. 105 of this appendix), the designation of a sample admissible pattern does not in itself provide a complete listing of the sample PSU's.

First, some of the PSU's in the sample pattern are actual clusters of PSU's. From each of these clusters, one PSU is selected (with appropriate probability based on size) to be the sample, representing the stratum from which the cluster comes.

Second, some of the strata are omitted, either because the stratum consists of one self-representing PSU or because the computed Keyfitz probability makes one PSU within the stratum a certainty selection. All these PSU's must be added to the list of PSU's provided by the controlled-selection sample pattern.

### SOME OBSERVATIONS

The following observations arise out of the experience in applying the controlled selection procedure.

It is possible to generate a set of control statements that is difficult, in some cases impossible, to meet for all admissible patterns in a set. When this happens, the control on the number of PSU's also in the old design sample has been relaxed. This problem did not occur in defining the set of admissible patterns for the A-design in the West but did occur in other regions.

The selection probabilities for NSR strata in the West, given in column (2) of table A-2, sum to 24.0000. However, there are 16 NSR strata in the West; the discrepancy is accounted for by the steps taken to simplify the listing of PSU's. For similar reasons, the probability in column (2) of this table for PSU's in the old A-sample, when adjusted, would be 16.5214\* rather than the figure shown.

The CPS experience in the controlled selection sample in the West was to define a set consisting of 76 admissible patterns. From the set, a pattern was chosen which had 16 sample PSU's that were also in the old A-design sample and also had the desired distribution of 1970 sample PSU's by geographic area.

Because PSU's in Alaska were defined using Census Division in 1970 and election districts in the old design sample, it was possible to define an unambiguous selection probability for the PSU's in Alaska. For this reason, Alaska PSU's were selected independently, without controlled selection probabilities. The 1960 sample PSU identified as Anchorage was, by chance, placed in the new design by a 1970 PSU, also identified as Anchorage. This coincidence is not counted among the 16 sample PSU's also in the old design sample.

## Appendix B.

# SAMPLING WITHIN PSU'S: SELECTING THE SAMPLE OF 1970 ED'S

## INTRODUCTION

The selection of the sample of ED's within sample PSU's is accomplished by arranging ED's in a specific geographic sequence, assigning a measure of size to each ED, and selecting a systematic sample of ED's with probability proportionate to the measures of size. In a census region, the selection of the ED's is performed independently for each of the four ED categories, C, B, U, and R,<sup>1</sup> within two groups of PSU's (SMSA and non-SMSA). In each of these eight categories within a region, the selection is continued across all PSU's with the random start used in sampling an ED category in a given PSU derived from sampling in the same ED category in the previous PSU.

This appendix illustrates the process by describing the selection of ED's in the A-design sample PSU 412, comprised of LaPorte and Starke Counties in northern Indiana. Since this PSU is not an SMSA and is not within an urbanized area, the PSU has only ED categories U and R. The sampling rate which makes the sample self-weighting in this PSU is 1 in 677.68. (See line (8) in table II-4.)

The selection of the sample uses summary records for each ED in the sample PSU's. In addition to the geographic codes used in sorting, the records provide the number of housing units and group quarters population (excluding inmates of institutions) used to assign a measure of size to the ED in determining the probability of selection. The measures of size are based on the 1970 census (first count) results for the ED and reflect those edits and coverage adjustments which were a part of the census process.

As explained in chapter III, the CPS is a continuing sample program, and new samples of USU's are continually being introduced. A set of sample USU's, designated with overall probability of 1 in 1,968 for the A-design, is identified by an alpha-numeric code for control; the illustration in this appendix demonstrates the selection of ED's associated with Sample A29. Sample A29 includes USU's originally intended for interview at the initial introduction of the 461-PSU design in December 1971. For various administrative reasons, A30, rather than A29, was used, but the selection of Sample A29 is chosen for the illustration, because it forms the starting point for the sample rotation system described in chapter III; the USU's in Sample A30 are easily identified once the A29 USU's are known.<sup>2</sup>

Although additional samples of USU's are regularly required for the CPS, the sampling of ED's, described in this appendix, is performed only once for the initial Sample selected for the design.

## SEQUENCE ED RECORDS

The first step in designating the sample within the PSU is to arrange ED summary records into a geographic sequence to maximize the advantage of systematic sampling.

<sup>1</sup>These ED categories are discussed ch. II, p. 18.

<sup>2</sup>See how additional samples are generated in ch. III, p. 24, and the system for allocating USU's in app. E p. 127.

The ED summary records are first assembled by PSU in a sequence of PSU's established separately, by census region. The following specifications do not identify a unique order of PSU's; however, the arrangement, once established, has been maintained and is identified as the *PSU selection order*.

1. Self-representing (SR) PSU's appear first and are arranged by SMSA and non-SMSA and, within these groups, by a more or less arbitrary order within State.
2. Nonself-representing (NSR) PSU's are first ordered in three groups, by sample design: A-design only, A- and C-design, and C-design only, and within these groups further by SMSA and non-SMSA. PSU's within these sort groups are arranged rather arbitrarily, but some consideration is given to ordering by the within PSU sampling intervals and by State.
3. PSU's in more than one State have their separate State parts treated as separate PSU's. If the States involved are in the same census region, the State parts of the PSU occupy adjacent locations in the PSU selection order.

Within the PSU selection order, the ED summary records are sequenced by the following sort priority, using identification codes on the records:

1. C, B, U, and R.
2. Special place recode—The intention of the recode is to relegate those ED's having the majority of their group quarters (noninstitutional) population residing in special places to the end of the list of ED's (within a C, B, U, or R category) in the PSU.
3. Place size—To assemble ED's from the largest place size first, followed by those from smaller places, in descending order.
4. County.
5. ED number.

As the ED number was originally assigned by serializing the mapped ED's sequentially in a serpentine order within county, a geographic order is established within the CBUR-special place-size categories. Systematically selecting ED's when listed in this sort has approximately the same effect as a secondary stratification of ED's, by city size in urban areas and by geographic location in all areas.

The first five ED's in PSU 412, within category U, are shown in table B-1 along with an assortment of other ED's listed in the described sort.

## COMPUTE A MEASURE OF SIZE FOR EACH ED

The *measure of size*  $M_e$  is the number of USU's assigned to the ED and is used in determining the probability of selecting the  $e$ -th ED. The measure of size is computed by the formula<sup>3</sup>

TABLE B-1. Selection of A-Design Sample ED's Within PSU 412: LaPorte-Starke Counties, Ind.

(Sampling interval: 1 in 677.68)

ED category	Place size code	Identification codes		Total housing units $M_0$	GO population $P_0$	Total USUs		Sample designation numbers	USU in sample A29	Hit number	List or area ED	Segment number
		County	ED			$M_0 +$	$\Sigma M_0$					
						(5) + (6) / 3 4	Cumulate (7)					
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
U	10	091 (LaPorte)	512	535	5	149	149	*222 14	60	0001	List	5014
	10		5128	43		11	160					
	10		513	8		2	162					
	10		514	665		166	328					
	10		5148	101		25	353					
	10	091 (LaPorte)	520	742	28	188	934	899 82	154	0002	List	6016
	10		526	646		162	1738					
	10		5308	177		44	2286					
	10		5378	600		150	3076					
	10		545	679		170	3746					
	9	091 (LaPorte)	5488	211	-	53	4321	4288 22	20	0007	List	301
	9		562	288		17	5023					
	9		563	173		43	5068					
	9		009	294		74	5662					
	9		*128	72		18	5910					
R	4	091 (LaPorte)	540	690	-	173	5910	*538 89	61	0001	List	E
	3		576	346		87	565					
	0		506	680		170	1339					
	0		554	525		131	1822					
	0		567	285		72	2593					
	0	149 (Stark)	002	445	-	111	3332	3249 61	29	0005	Area	
	0		017	614		104	3968					
	0		*028	214		54	4566					
	0											
	0											

- Represents zero.

\* PSU does not have ED's of category C or B or any with major proportion of population in GO.

\* Determined by successively adding sampling interval to random start.

\* Random start for ED category U.

\* Random start for ED category R.

\* Last ED in category in PSU.

$$M_e = [H_e + P_e/3] / 4$$

and rounded to the nearest integer (every ED is assigned a value  $M_e \geq 1$ ), where  $H_e$  is the total number of housing units, occupied and vacant, in the ED, and  $P_e$  is the number of group quarters (GQ) persons, excluding inmates of institutions. In table B-1, the measures  $M_e$ , computed from  $H_e$  and  $P_e$  in columns (5) and (6), are shown in column (7).

### SELECT THE SAMPLE OF ED's

The sample of ED's is designated by the computer. As the file of ED records is passed through the computer, eight independent sample selection operations are carried out, one for each of the four ED categories C, B, U, and R within the two kinds of PSU's (SMSA and non-SMSA). The results are consistent with the following steps (displayed in table B-1):

1. For each ED, establish a *cumulant*,  $\Sigma M_e$  (as in column 8) equal to the sum of the measure  $M_e$  for the e-th ED and all other ED's appearing before it in the PSU in the same ED category (C, B, U, or R).
2. Establish a *random start*, RS, for the sampling operation. As the sampling interval in the illustration is 1 in 677.68, a random start is needed within the range

$$.01 \leq RS \leq 677.68$$

A separate random start is used for each of the eight independent sampling operations; the two random starts in the illustration are shown in column (9) as 222.14 and 538.89 for categories U and R, respectively.

The computation of the random starts is discussed later in this appendix.

3. The random start becomes the first sample designation number for the ED category.
4. The sample designation number determines the ED in sample within the ED category—

- a. Round the sample designation number to the nearest integer.
- b. The ED having the smallest cumulant, in column (8), that equals or exceeds the rounded sample designation number is the sample ED.

5. Add the sampling interval to the sample designation number to determine the next sample designation number.
6. Repeat the steps indicated in items (4) and (5) until the ED sampling within the PSU is complete for the category. If the sampling interval is less than the measure of size for an ED, the sampling process may require more than one USU to be selected from the ED.

7. Sample ED's are designated similarly within each category. Table B-1 shows these steps designate nine category U- and six category R-sample ED's in PSU 412 for Sample A29.

### IDENTIFY THE SAMPLE USU IN THE SAMPLE ED

In addition to designating the sample ED's within the PSU, the process also identifies the specific sample USU to be interviewed within the sample ED. Appendixes C and D illustrate how the sample designation number is used to identify the sample USU within area and list ED's, respectively.

The sample designation number in column 9 of table B-1 for the first ED selected directs that the 222d USU of category U in the PSU is in sample. This is equivalent to the 60th USU in the 4th ED listed (i.e., the first sample ED) in the sort used for sampling (see column 10).

The entry in column 10 of table B-1 for the sample ED is derived as follows:

1. Round the sample designation number, column 9, to the nearest integer.
2. Add the measure of size  $M_e$ , column 7.
3. Subtract the cumulant, column 8, e.g.,  $222 + 166 - 328 = 60$ .

### DETERMINE THE RANDOM STARTS FOR THE SAMPLING OPERATION

In practice, the computer designates the sample ED's for a large number of PSU's within a census region in one operation. A set of random starts is required for PSU's within each census region for C, B, U, and R ED's separately within SMSA, non-SMSA PSU's. A procedure to yield a balanced set of the needed starts is automatically specified with the final choice determined by one selection from a random number table. These random starts are used as the first sample designation numbers for ED's in the first PSU in the selection order. The random starts employed for second and subsequent PSU's are computed using the outcome of the selections from the previous PSU.

Random starts must allow any combination of digits to be selected within the range of the sampling interval. If a sampling interval of, for example, 543.21 is used, the random start must be in the range 0.01 to 543.21 inclusive.

For the second and subsequent PSU's, the random start for each of the ED categories is calculated by the following procedure. This summary assumes that the random start is known for PSU 1, and sampling has been completed for a given ED category; it is desired to determine the random start in the same ED category for PSU 2, the next PSU in the selection order. Let

$RS_1$  = random start used in the ED category in PSU 1.

$W_1$  = sampling interval used in PSU 1; i.e., sampling is performed at the rate of 1 in  $W_1$ .

$m_1$  = number of sample USU's designated in the category in PSU 1.

$D_1 = RS_1 + (m_1 - 1) W_1$  the sample designation number for the last ED selected in the category in PSU 1.

$M_1 = \sum_e M_{e1}$  = total of measures of all ED's in the category in PSU 1.

Then the random start to be used in the same category in PSU 2 is given by

$$RS_2 = (D_1 + W_1 - M_1) W_2 / W_1 \quad (B/1)$$

where  $W_2$  is the sampling interval in PSU 2.

To illustrate, the random start for category U in PSU 374, which immediately follows PSU 412 in the PSU selection order, is given by substituting the following values found in table B-1 into formula (B/1):

$$D_1 = 5.643.58$$

$$M_1 = 5.910$$

$$W_1 = 677.68$$

For  $W_2 = 642.06$ , the sampling interval in PSU 374,  $RS_2 = 389.64$ .

### DETERMINE AREA OR LIST PROCEDURES FOR THE SAMPLE ED

After the sample ED's are designated within the PSU, the address registers for the ED's are screened to determine whether list- or area-sampling procedures are to be used. (See ch. II, p. 16.) Column 12 of table B-1 shows the outcome of the screening performed for the sample ED's in the illustration to determine whether each is a list or an area ED.

The procedures for sampling within the ED are illustrated in appendixes C and D for area and list ED's, respectively. These steps use the figure given in column 10 of table B-1 to determine the A29-sample USU.

### IDENTIFY USU's FOR SUBSEQUENT (ROTATING) SAMPLES

The A30-sample USU is determined from the A29 USU by adding 1 to the entry in column 10 of table B-1. Samples A31 through A48 are designated by extending this procedure.<sup>4</sup>

Incrementing the column 10 entry to designate further Samples may produce a result greater than the ED measure of size in column 7. See, for example, ED 562 in county 091 where measures 15, 16, and 17 will be used for Samples A29, A30, and A31, respectively, but Sample A32 cannot be accommodated, because the ED has only 17 measures. In this situation, ED 563, next in sort in the PSU, comes into sample with

<sup>4</sup> The practice of determining subsequent samples by adding 1 to the entry in column 10 is in some conflict with the method of determining random starts described earlier in this appendix; this problem is examined in app. E, p. 130.

USU's 1 through 17 in that ED assigned to Samples A32 through A48, respectively.

### ASSIGN HIT NUMBERS

For the purpose of control, it is convenient to assign a number to the USU's in sequence as they are designated for the sample. The numbering system is called the *hit number* and is incorporated in the specifications of certain processing operations.

For USU's selected from 1970 census materials within list area ED's and separately for each of the four ED categories (B, U, and R), the sample USU's are numbered, beginning at 000 in each PSU, in the order selected. Hit numbers for the examples discussed in this appendix are given in column (11), table B-1.

For USU's selected from building permits, a 4-digit hit number is constructed from the 2 digits for the month the permit was issued (the one's and ten's digits) and the year of issuance (the thousand's and hundred's digits). Thus, a permit USU comprised from a building permit issued in July 1973 would be signed the hit number 7307. Hit numbers are not assigned Cen-Sup USU's.

### SEGMENT NUMBER

The segment number is a 4-digit number assigned to correlate the sampling materials and the information collected for USU's in sample. Each USU in sample is associated with a segment number in the manner discussed in chapter II, p. 16. Segment numbers, assigned in the illustration used in this appendix, are given in column (13), table B-1. The four digits of segment number are derived as follows:

1. The first (most significant) digit is the rotation group number, assigned as described in chapter III in the section on the operating systems, p. 22.
2. The second and third digits are used to provide a serial number, beginning with 01 for each rotation group within the PSU; the USU's are serialized when in the sequence, given in this appendix as the first step in designating the sample within the PSU.
3. The fourth digit is a check digit computed from the six digits given by the 3-digit PSU code and the first three digits of the Segment number. This digit is included as part of the PSU and segment code to identify all materials used by the interviewer for each segment and permits an internal check on the accuracy of recording the identification code.

## Appendix C.

# SAMPLING WITHIN PSU'S: SELECTING SAMPLE USU'S FROM AREA ED'S

### INTRODUCTION

This appendix illustrates the operational steps which define the USU to be interviewed for Sample A29 (or C14) in an area ED. Four operational steps are discussed: Locate each address recorded in the address register on a map of the ED; use this information to subdivide the ED into small land areas called chunks or blocks; assign a measure of size (number of USU's) to each chunk; and, finally, list all USU's in the ED in order, by chunk, and select the sample USU. The chunk which contains the sample USU is called the sample area segment.

Area samples for the current surveys appear only among 1970 census conventional or prelist ED's and make use of the census maps for these ED's. The area sampling process, in practice, is applied using census maps generated from a variety of sources. For illustration, this appendix uses maps that, although identified as for a specific ED, do not describe an actual 1970 Census Enumeration District.

The illustration assumes that ED 091-567 has been selected by the steps illustrated in appendix B. (See table B-1.) We assume the review of the address register for this ED has shown the quality of addresses require that area sampling must be used to define the sample USU.

### THE PROCEDURE

#### Locate Each Address Recorded in the Address Register on a Map of the ED

Figure C-1 is a reproduction of the ED map carried by the 1970 census enumerator when canvassing ED 091-567; the ED boundaries are shown in orange. During the 1970 census, each unit enumerated was assigned a census serial number which, along with its address, was recorded in the address register for the ED. If the unit did not have a street name and house number address, the enumerator was to show the unit location by a dot on the ED map along with its housing unit serial number. In the illustration, the enumerator has indicated some of these units and has also added two unnamed roads (A and B) on the map.

The information from the ED map and the address register is used to locate each address on a second map of the ED called a work copy. The best available map of the area is used; this may be a clean copy of the enumerator's original map showing map corrections the enumerator has made.

A location for each address is required to be sufficiently precise to place it within the geographic features indicated on the work copy map. If the location of as many as one-fourth of the addresses in the ED (or 50 units, whichever is less) cannot be determined, the field staff will visit the ED to determine the location of all current housing units in the ED before proceeding to the next step.

#### Use the Addresses and Geographic Features To Map the ED Into Chunks

The chunks, sometimes also referred to as blocks, are defined as having recognizable geographic features as boundaries and containing about 35 or less housing units. For reasons given later in this appendix, the most desirable chunk size is in the range of about 7 to 20 housing units (equivalent to about 2 to 5 USU's of 4 housing units).

The work copy map for the ED is given in figure C-2 to illustrate the results of the chunking operation. The chunk boundaries and numbers to identify the chunk are identified in red; the total housing units in each chunk is circled in green. Figures associated with street sectors are counts of address register units and have been entered in the process of locating units on the map; they appear only in those areas that could be chunked in alternative ways.

#### Allocate USU's to the Chunks

The probability of selecting the ED is proportionate to the total number of USU's in the ED (given in table B-1, col. 7) derived from 1970 census first counts. However, chunks are defined on the ED work copy map from the counts of housing units and GQ population, on the ED address register.

The accumulation worksheet, form BC-59 (fig. C-3), is used to allocate the USU's to the chunks within the ED.

In column (b), the form lists the number of housing units and equivalent GQ population in each of the chunks defined on the work copy map. The sum of column (b) is recorded in item 7 and represents a measure of size as given by the housing units and GQ population recorded in the address register.

Item 6 shows total USU's based on first counts (from col. 7 of table B-1) so that the average units per USU in ED 091-567 is given by the ratio of these two or  $279/72 = 3.875$ . The number of USU's per chunk is given by dividing the column (b) entries by this ratio (the equivalent operation is to multiply by the inverse of the ratio, 0.258, which is recorded in item 8). The result of the multiplication is recorded in column (c) for each block on the list.

The integral portions of column (c) entries are entered in column (d) and are a first allocation of the 72 USU's to the blocks. The sum of these crude allocations differs from 72: the difference ( $72 - 62 = 10$ ) is the number of chunks for which the allocation must be increased (rounded up) by 1. The final allocation of the 72 USU's to the 18 blocks is recorded in column e.

#### Select the Sample USU

The ED measure check list, form BC-2505 (fig. C-4), is used for selecting the sample USU. The selection procedure also identifies the sample block. The objective is to select a sample USU that is representative of the population of the ED.

**FIGURE C-1. Census ED Map  
ED 091-567  
LaPorte County, Ind.**

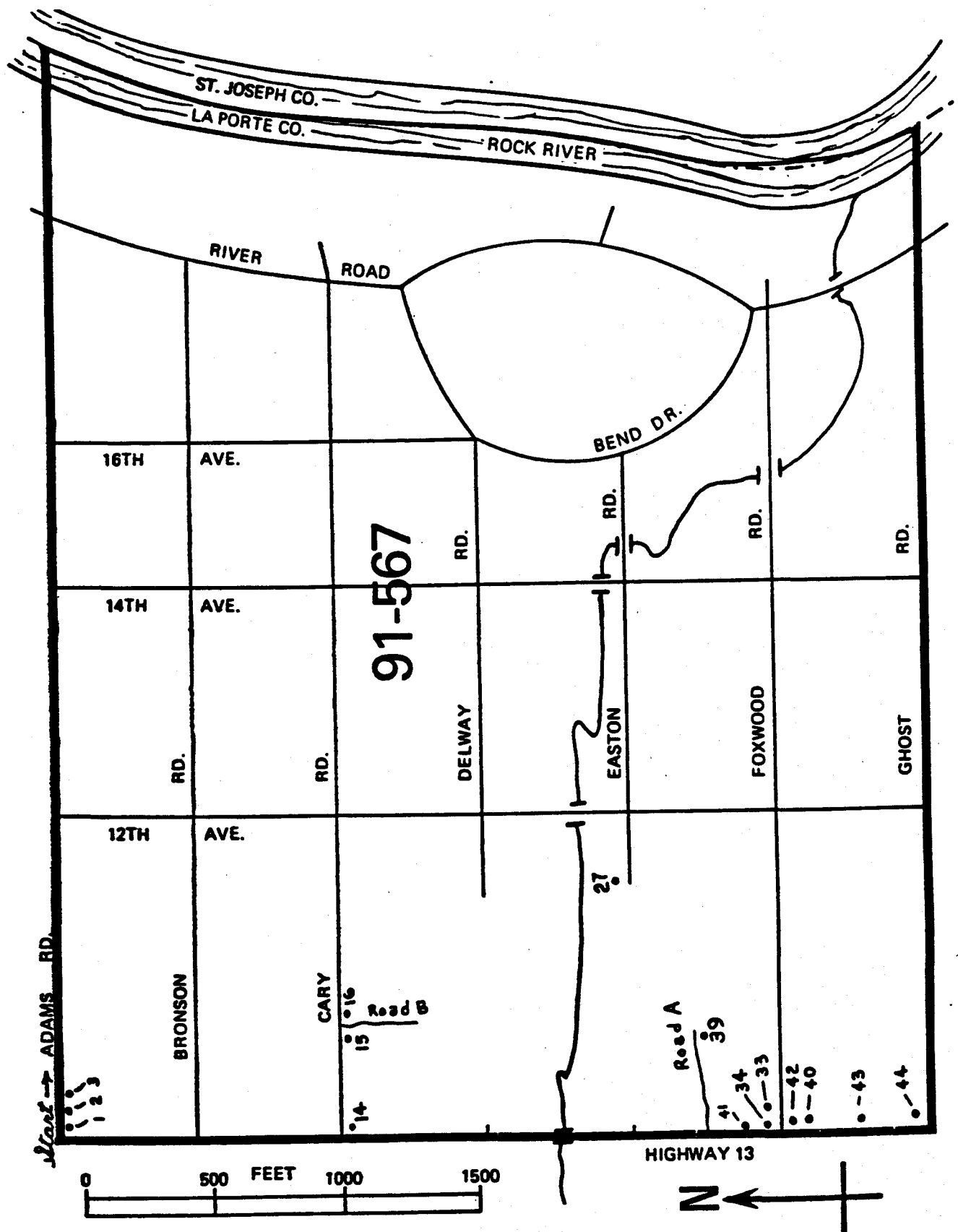
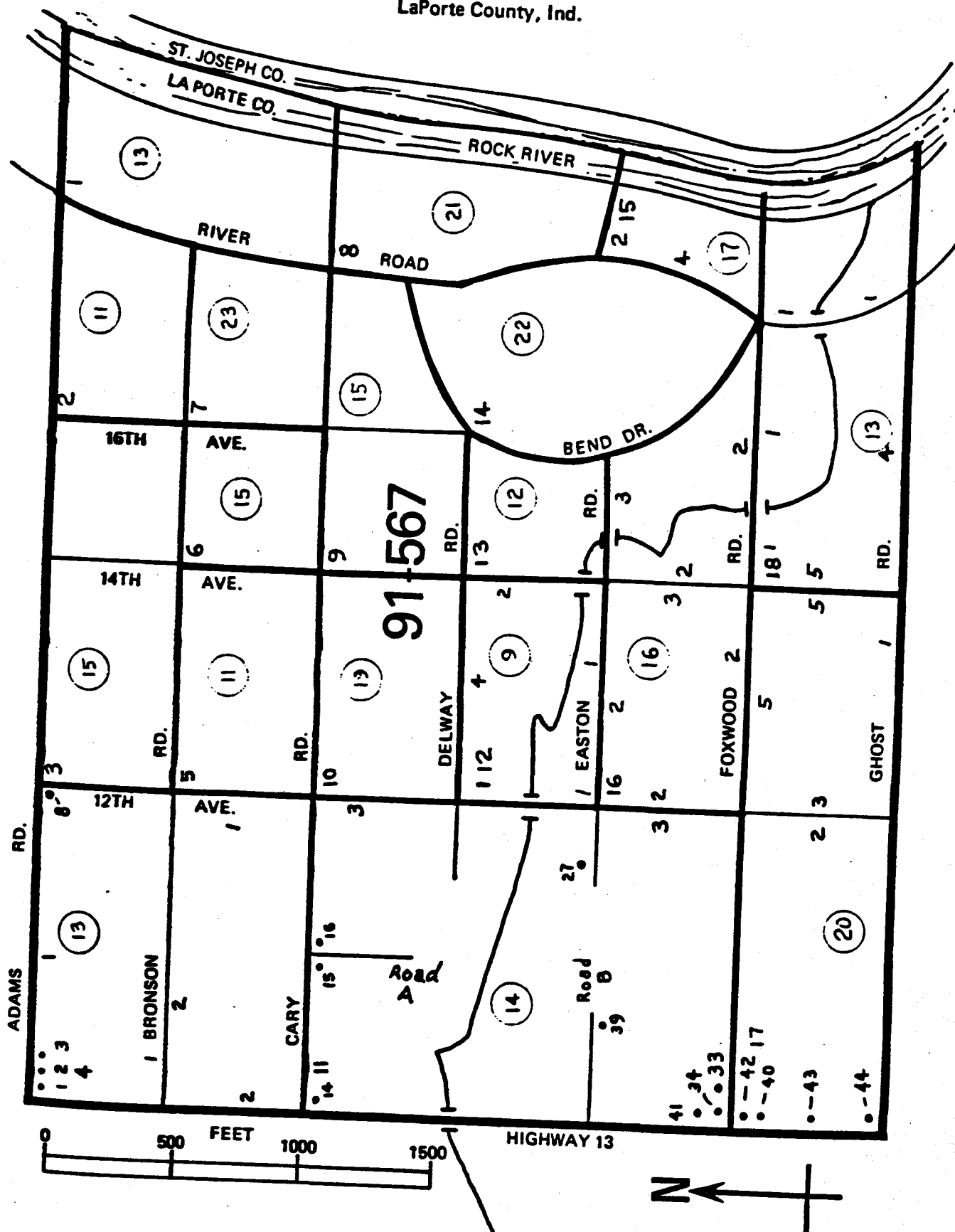


FIGURE C-2. Work Copy Map  
ED 091-567  
LaPorte County, Ind.





**FIGURE C-3. Area ED Accumulation Worksheet  
(Form BC-59)**

U.S. DEPARTMENT OF COMMERCE SOCIAL AND ECONOMIC STATISTICS ADMIN. BUREAU OF THE CENSUS						1. Survey	2. State	3. County	4. PSU		
AREA ED ACCUMULATION WORKSHEET						5. ED	6. Tot. meas. ( $M_F$ )	7. Units	8. Constant		
Block No.	Units	Measure				Block No.	Units	Measure			
(a)	(b)	Exact (c)	Initial (d)	Final (e)	Cum. (f)	(a)	(b)	Exact (c)	Initial (d)	Final (e)	Cum. (f)
1	13	3.354	3	3							
2	11	2.838	X	3							
3	15	3.870	X	4							
4	13	3.354	3	3							
5	11	2.838	X	3							
6	15	3.870	X	4							
7	23	5.934	8	6							
8	21	5.418	8	6							
9	15	3.870	X	4							
10	19	4.902	X	5							
11	14	3.612	X	4							
12	9	2.322	2	2							
13	12	3.096	3	3							
14	22	5.676	8	6							
15	17	4.386	4	4							
16	16	4.128	4	4							
17	20	5.160	5	5							
18	13	3.354	3	3							
	279		62	72							

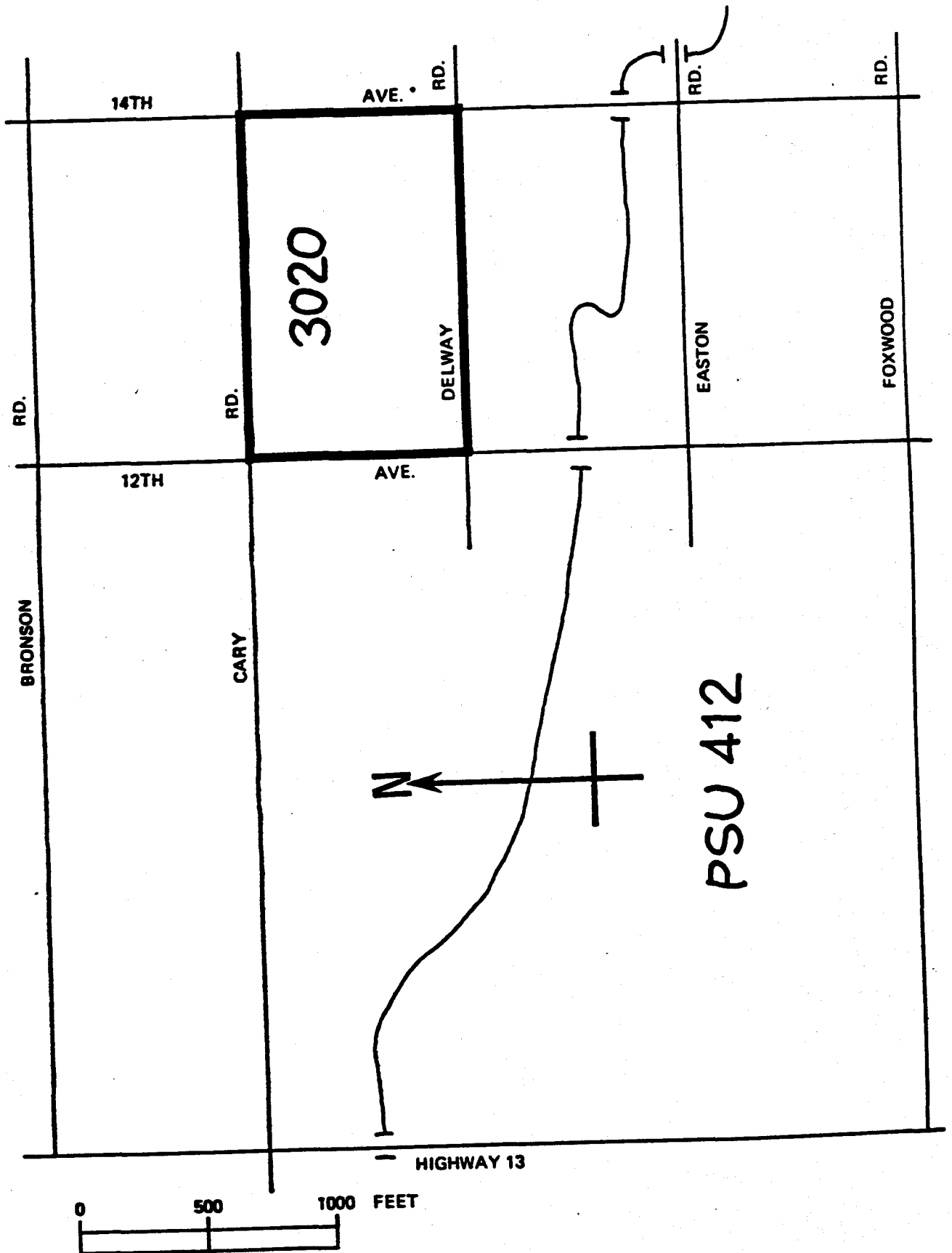
Notes

USCOMM-

FIGURE C-4. ED Measure Check List,  
Current Surveys  
(Form BC 2505)

Form BC-2505 (110-20-72)						U.S. DEPARTMENT OF COMMERCE SOCIAL AND ECON. STAT. ADMIN. BUREAU OF THE CENSUS						1. County <b>091</b>		2. ED <b>567</b>		3. PSU <b>412</b>	
ED MEASURE CHECK LIST CURRENT SURVEYS						4. Tot. meas. <b>72</b>		5. Hit meas. <b>51</b>		6. Random number <b>60</b>							
7. PSU name <b>La Porte - Starks, Indiana</b>						8. Place name											
Line (a)	Block (b)	Block fraction (c)	Sample measure (d)	Sample (e)	Segment number (f)	Line (a)	Block (b)	Block fraction (c)	Sample measure (d)	Sample (e)	Segment number (f)	Line (a)	Block (b)	Block fraction (c)	Sample measure (d)	Sample (e)	Segment number (f)
1	1	1/3	14			51	14	1/6	A42	3020D	101						
2		2/3				52		2/6	A43		102						
3		2/3				53		3/6	A44		103						
4	2	1/3				54		4/6	A45		104						
5		2/3				55		5/6	A46		105						
6		2/3				56		6/6	A47		106						
7	3	1/4				57	15	1/4	A48	3020E	107						
8		2/4				58		2/4			108						
9		2/4				59		3/4			109						
10		4/4				60		4/4	1		110						
11	4	1/3	24			61	16	1/4	2		111						
12		2/3				62		2/4	3		112						
13		2/3				63		3/4	4		113						
14	5	1/3				64		4/4	5		114						
15		2/3				65	17	1/5	6		115						
16		2/3				66		2/5	7		116						
17	6	1/4				67		3/5	8		117						
18		2/4				68		4/5	9		118						
19		3/4				69		5/5	10		119						
20		4/4				70	18	1/3	11		120						
21	7	1/6	34			71		2/3	12		121						
22		2/6				72		3/3	13		122						
23		3/6				73					123						
24		4/6				74					124						
25		5/6				75					125						
26		6/6				76					126						
27	8	1/6				77					127						
28		2/6				78					128						
29		2/6				79					129						
30		4/6				80					130						
31		5/6	44			81					131						
32		6/6				82					132						
33	9	1/4				83					133						
34		2/4				84					134						
35		2/4				85					135						
36		4/4				86					136						
37	10	1/5				87					137						
38		2/5	51	A29	3020	88					138						
39		3/5	52	A30		89					139						
40		4/5	53	A31		90					140						
41		5/5	54	A32		91					141						
42	11	1/4		A33	3020A	92					142						
43		2/4		A34		93					143						
44		2/4		A35		94					144						
45		4/4		A36		95					145						
46	12	1/2		A37	3020B	96					146						
47		2/2		A38		97					147						
48	13	1/3		A39	3020C	98					148						
49		2/3		A40		99					149						
50		3/3		A41		100					150						

FIGURE C-5. Segment Map



column 10 of table B-1; this is recorded as the hit measure in item 5, figure C-4.

One line of this check list is used to list each USU within each of the chunks in the ED using columns (b) and (c).

Before determining the USU corresponding to number 51, the 72 USU's in the ED are renumbered using a random start (in the range of 1 through 72) to specify the first USU in the renumbering. As the chunking and numbering operations are done by the same technicians, the numbering in column (d), with a random start, is to remove any possible bias in designating the sample USU.

The random start selected, 60 (recorded in item 6), means that USU number 1 appears on line 60; number 2, on line 61; number 13, on line 72; number 14, on line 1; and USU number 51 is on line 38, identifying the second USU in block number 10. This USU would be designated for interview for Sample A29.

With the USU on line 38 assigned to A29, the CPS will use USU's A30 through A48, listed on lines 39 through 57 of figure C-4 for 1972-1982.

### Identify the Sample Segment

The chunk containing the sample USU is defined as the sample segment. The interviewer, when sent out to collect CPS information for Sample A30, will have a copy of the segment map resembling figure C-5 (an enlarged map of chunk number 10 from fig. C-2). Before the segment is scheduled to enter the sample, the interviewer will visit and list all current housing units in the segment. An office clerk applies the within-segment sampling instructions, summarized in column c of fig. C-4, to designate the third unit listed and every fifth unit thereafter for interview in Sample A30.

The segment is assigned a segment number in column f of figure C-4; the USU associated with this sample selection will

always be carried under the segment number 3020 in this illustration.<sup>1</sup> Note, however, the USU for A33 is in a new chunk, and will be recorded in the office records as 3020A. The suffixes A, B, etc., indicate the geography chunk for office purposes. When a new block enters the sample, it will be listed and processed in the same way, and the designated sample USU's will continue to be identified by the same segment number.

### DEFINITION OF CHUNKS WITHIN AREA ED's

Blocks comprising more than five USU's are undesirable, since each chunk, when in sample, must be listed several times (to designate each USU and keep the listing up to date<sup>2</sup>). Larger chunks, therefore, have a disproportionately larger listing cost. For this reason, the larger blocks (with more than an expected 35 housing units) are visited to further subdivide them into smaller chunks; this special operation is called *subsegmenting*. Permitting the larger chunks would mean improved control on the variance, but the increased cost is considered unacceptable.

Chunks having only one USU are also not desired except for zero or very low population ED's where this becomes necessary. Smaller land areas are subject to larger relative differences between the current number of housing units and population and the numbers attributed to the area on the basis of the last census. These large differences, which can have serious impact on the sampling error, are reduced by avoiding the use of chunks with measures of only one USU. The occasional USU found at the time of interview to have very large size is handled by the rare events universe procedure, described in appendix G.

<sup>1</sup>The first digit of the segment number indicates that the segment is a member of rotation group 3 (See app. B, p. 114.)

<sup>2</sup>See ch. IV, p. 30.

## Appendix D.

# SAMPLING WITHIN PSU'S: SELECTING SAMPLE USU'S FROM LIST ED'S

### INTRODUCTION

About two-thirds of all ED's involved in an A- or C-design sample are of the list type, that is, members of quadrant 1 in figure II-1. The 1970 mail census areas largely coincide with ED's of this type and planning of the CPS redesign was influenced by the likelihood of success in developing a computer system for current survey sampling in these areas. The system, as developed, assembles addresses of 1970 census housing units into USU's within the ED by an operation called segmenting, designates the sample USU, and prepares a list of addresses in the USU required in conducting the interview. This computer-based and essentially error-free operation prepares most of the within-ED sampling materials needed for about two-thirds of the USU's in national samples used over the current decade.

It was recognized that the large one-time capital investment in planning and preparing data for such an operation could not be justified by the CPS requirements alone. The total needs of the Census Bureau over the current decade were estimated at 62 different national samples in the PSU's designated for the CPS sample design; during this period, the CPS was expected to use only 20 of these samples. The balance represented the predicted requirements of all other national samples in these PSU's for the decade.

The cost of preparing data for computer list sampling is a function of the number of sample ED's involved. To accommodate the CPS requirements over the decade, preparatory processing is required in about 11,200 list ED's. However, because the method of assigning USU's in a CPS ED also makes USU's available for any other survey in the same ED, the requirements of the remaining 42 samples bring the total up to only about 26,000 list ED's. Requirements for additional surveys the Census Bureau has been asked to undertake have since added to these figures.

The following sections of this appendix summarize the one-time activities which determine the 1970 census addresses making up USU's in a CPS list ED; the term *segmenting* is given to these activities. A separate section deals with the designation of the sample in special places for which the computer is not used. An illustration of a portion of a computer segmented ED is discussed in the final section.

### THE PROCEDURE

#### Prepare Corrected Tape Records of Addresses of 1970 Housing Units in List Sample ED's

Most of the one-time cost has been incurred in the preparation of a corrected file of addresses of the housing units recorded in the census address register for the sample ED's. About 10 percent of the 4,500,000 census housing units in sample TAR ED's<sup>1</sup> required an addition, a deletion or a correction. About 6,500,000 records were prepared in non-TAR sample ED's.

This step takes place after the enumerator's copy of the E: address register has been given the technical review, discussed in chapter II, p. 19, which has determined the following

1. At least 90 percent of the census HU's have a complete address (street name and house number), and the units failing this test have been identified on the address register so they can be ignored in all subsequent sampling steps.
2. Housing units in special places have been identified, and their number, as well as the number of group quarters (GQ) persons in these places, have been noted. The inmate institutional population is ignored except the special situation mentioned on page 123 for ED's made up entirely of such population.
3. The ED is within a permit issuing jurisdiction; this is not a necessary condition for TAR ED's.

In TAR ED's, correction of the existing census tape address register, to incorporate changes and additions noted by the enumerator, provides the required tape record of addresses. In other list ED's, a tape copy is prepared from the handwritten address register.

The corrected tape copy of census address registers for list ED's is produced by punching records or their changes directly from the address register. The process is repeated independently, and the two results are then matched to expose errors. The quality of surveys for the decade depends on the accuracy of these steps; the process has been given special supervisory attention. An evaluation of the quality of this operation shows the computer sampling operation is now based on records having an estimated 0.3 percent which are defectives ( $\sigma \pm = 0.25$  percent).

Each record in a list ED, as corrected, shows the identification and category of ED (C, B, U, or R) and, for the address of living quarters, the street name, house number, apartment number or description, block number, post office name, zip code, SMSA code, if any. This information is required in the subsequent steps which define the USU's. The block number for address refers to areas such as the typical city block for urban areas. Apartment numbers or descriptions, useful information for the interviewer and included in the address register, are not considered sufficiently reliable to warrant a correction for purpose alone; for this reason, a record was not corrected to include a new or revised apartment number reported by the census enumerator.

#### Sequence the 1970 Census Housing Unit Records Within the ED

The purposes of the sort operation are to reflect the order of addresses on the ground rather than the order in the census address register so that the gain in reliability from systematic sampling may be realized and to enable the computer to

interviewer coverage of current units is improved and interviewer travel between units is minimized.

In some respects, these objectives conflict, because USU's made up of adjacent units, i.e., compact, will typically be more homogeneous (and, therefore, less efficient) as sampling units than noncompact USU's. However, some research has shown that biases introduced by errors in coverage in the interview process can be better controlled by compact USU's. When USU's defined on the basis of 1970 units within a multiunit address are interviewed later in the decade, the living space involved is sometimes found to be allocated to housing units in a different way. This kind of problem often arises when the structure is remodeled; in some cases, this may involve nothing more than locking an interior door to create two housing units out of space formerly occupied by one.

When a USU is made up of addresses that had four or fewer housing units at the time of the census, the interviewer can be instructed to include the occupants of all HU's existing at the sample address at the time of interview, irrespective of the number of HU's actually present. At such USU's, made up of take all (TA) addresses, the interviewing process does not require the additional steps of listing and sampling the units for interview. The additional steps still must be taken at larger multiunit addresses containing more than one USU, i.e., at nontake-all (NTA) addresses (ch. IV, p. 30.)

Improved control of coverage could be further extended by defining larger addresses (for example, up to eight units) as TA USU's; however, this would result in an increase in variance. Studies of the cost and variance of alternative cluster sizes have led to the choice of four units as the optimum cluster size for estimating unemployment [58]. Prior studies had suggested that balancing optimum cluster sizes among results for several major statistics when noncompact clusters<sup>2</sup> were considered, led to the choice of six units as the optimum. Thus, somewhat different levels of unit costs and objectives, over time, have led to different interpretations of what is optimal.

To achieve the foregoing objectives, the HU records within the ED are sorted before the segmenting operation in the following order:

1. Block number
2. Street name
3. House number
4. Apartment number

For ED's from census TAR areas, this address information was included on the HU records, prepared in advance of the census. In addition, a census housing unit serial number was assigned in the census processing after the HU records were put in this sort order but before enumeration. Additional units found during enumeration were assigned the next available serial number at the end of the ED. The serial number and the address information were entered on the address register for the additions.

For prelist and conventional census areas, the address register information (including a housing unit serial number) was recorded by the census lister or enumerator but may be incomplete for some HU's. Also, in some ED's, block numbers may not

have been used. The desired basic order is approximated by first sequencing records within the ED by street name, house number, and census serial number to assemble all HU records for the same street address. Next, the lowest census serial number is entered as a recode on each HU record at an address. Finally, the recoded records are sequenced by block number (where available), recoded serial number, and apartment number.

The effect of this sort generally is to reproduce the order used by the census enumerator (in conventional areas) or the lister (in prelist areas) to achieve systematic and ordered coverage of the ED.

### Determine the Number and Size of USU's in the ED

It is necessary to determine  $N(S_e)$ , the integral number of USU's of size  $S_e$ , and  $N(S_e + 1)$ , the number of size  $S_e + 1$ , so the total USU's constructed will equal the measure for the  $e$ -th ED assigned when selecting the ED sample. Here the size is in terms of 1970 census housing units at addresses with complete street name and house number recorded in the ED address register.

The number of units listed in the ED address register reflects the result of the field data collection process, whereas the number of housing units used in computing the measure of size in the ED sampling procedure includes the effect of a number of subsequent 1970 census processing procedures that produced the first counts for the census. These two HU counts may differ because of edits of census information, census coverage improvement procedures [38; 46, pp. 6-7] and other census processing as well as the deliberate omission of units with incomplete address from consideration in constructing USU's. During the screening of the ED's designated for sampling (ch. II, p. 19), counts are determined for housing units with usable addresses and housing units and population in special places recorded in the ED address register.

The notation used for the counts in the  $e$ -th ED is summarized at the top of page 124.

The following procedure is used to compute  $S_e$  and  $S_e + 1$  and determine  $N(S_e)$  and  $N(S_e + 1)$ :

First, determine  $\hat{M}_{se}$  (the number of USU's to be assigned to special place housing units and population in the ED). Compute  $\hat{M}_{se}$  by

$$\hat{M}_{se} = M_e[\hat{P}_e/3 + \hat{H}_{se}] / [\hat{P}_e/3 + \hat{H}_e] \quad (D/1)$$

If either  $\hat{H}_{se}$  or  $\hat{P}_e$  is not zero in formula (D/1), then  $\hat{M}_{se} \geq 1$ . If  $\hat{H}_{se} = \hat{P}_e = 0$ , then  $\hat{M}_{se} = 0$  ( $\hat{M}_{se}$  is assigned the value 1 if there is institutional population in this situation).

The segmentation of the  $\hat{M}_{se}$  special place USU's is performed by manual procedures. Special place USU's are assigned to the end of the ED, and the steps defining the specific locations to be visited by the interviewers are done when the USU is scheduled to come into sample. (See p. 125.)

Second, the number of USU's to be assigned to the balance (the regular housing unit portion) of the ED is then given by

$$M_{re} = M_e - \hat{M}_{se} \quad (D/2)$$

<sup>2</sup> The noncompact cluster system, used from 1963 until the current design was adopted, defined USU's by subsampling listings of current housing units in an address segment at one in three

Item	Definition	Source of 1970 census information for the e-th ED	
		ED summary record (first count)	Address register <sup>1</sup>
1	Total HU's in ED .....	$H_e$	$\hat{H}_e$
2	Group quarters (GQ) population (institutional population is not included) .....	$P_e$	$\hat{P}_e$
3	HU's in special places <sup>2</sup> .....	(X)	$\hat{H}_{se}$
4	HU's not in special places $\hat{H}_{re} = \hat{H}_e - \hat{H}_{se}$	(X)	$\hat{H}_{re}$
5	Measure of size used in deter- mining probability of selecting the ED .....	$M_e$	(X)
	$M_e = [P_e/3 + H_e]$ $M_e > 1$		

X Not applicable.

<sup>1</sup> Based on counts of 1970 census HU's in the address register having complete street name and house number.  
<sup>2</sup> All HU's in special places whether with complete or incomplete entries in the address register.

Then compute

$$\hat{H}_{se}/M_{re} = S_e + R_e/M_{re} \quad (D/3)$$

The terms in the right side of (D/3), when expressed as integers, define  $S_e$  and  $R_e$ .  $R_e$  defines  $N(S_e + 1)$ , the number of USU's of size  $S_e + 1$ . From this, it follows that the number of USU's of size  $S_e$  is given by

$$N(S_e) = M_{re} - R_e \quad (D/4)$$

Typically, because the formula for the measure of size  $M_e$  is based on one-fourth the number of total HU's in the ED, given by the first count, and some of the regular housing units in the address register have been omitted from sampling, the value of  $S_e$  will be 3, with  $R_e/M_{re}$  a fraction near 1. Those ED's with values of  $\hat{H}_e$  and  $H_e$  that differ by more than 10 percent are reviewed to insure the counts and computations are correct.

As an illustration, see table B-1 which shows that ED 091-520 was selected for the sample in PSU 412. Column (12) of the table shows that the address register is adequate for list sampling, and since the ED is within a permit issuing jurisdiction, the ED is a member of quadrant 1 of figure II-1.

From table B-1, we have

$$\begin{aligned} H_e &= 742 \\ P_e &= 28 \\ M_e &= 188 \end{aligned}$$

Assume the review of the address register units during the screening operation shows that

$$\begin{aligned} \hat{H}_e &= 733 \\ \hat{P}_e &= 22 \\ \hat{H}_{se} &= 4 \text{ so that } \hat{H}_{re} = 729 \end{aligned}$$

Formula (D/1) shows the number of USU's to be assigned special place housing and population in the ED is, when rounded,  $\hat{M}_{se} = 3$  and the number of USU's to be allocated to the regular housing units in the ED is  $M_{re} = 185$  (from formula D/2). Formula (D/3) shows that  $S_e = 3$  and  $R_e = 174$ ; thus, the segmentation process must produce three USU's assigned to special place housing and GQ population and, for the balance of the 174 USU's of size 4 and 11 USU's of size 3.

### Assign the Addresses of 1970 Census Housing Units to USU's

The assignment of 1970 housing units to the requisite number of USU's of size  $S_e$  and  $S_e + 1$  is performed by the computer. The description in this section summarizes the relatively complex steps performed in this operation.

Three objectives are observed in assigning 1970 census housing units to the USU's. In priority order they are to—

- Keep units from addresses with four or fewer units in same USU. This is to maximize the number of USU's in up of take-all (TA) addresses to increase the coverage benefit of address sampling. (This is the most important of the three objectives.) An NTA address will have size greater than  $S_e + 1$ , must be allocated to two or more USU's will require listing and sampling steps to define the units to be interviewed.

- Minimize the total distance between addresses within the USU. As the sequence of the HU records is presumed to reflect location on the ground, this means that USU's comprised of addresses nearest in the sort are defined where alternatives are possible.
- Minimize the number of USU's that have both TA and NTA addresses. These mixed USU's introduce complications in the interviewer's task.

The following summarizes the steps in defining USU's:

First define USU's of size  $S_0 + 1$  made up of whole addresses each having  $S_0 + 1$  units—USU's of this size are assigned to whole addresses beginning at the end of the sort sequence and working toward the first address. Counts of the USU's defined are kept during the process, and up to  $N(S_0 + 1)$  USU's are defined by this procedure. The remaining steps are to define the  $N(S_0)$  and the remainder, if any, of the  $N(S_0 + 1)$  USU's in the ED.

Identify a string of usable addresses and form a set of proposed USU's of size  $S_0$ —The *string of usable addresses* is a list of housing units beginning with the first unit in the sort not yet assigned to a USU and consisting of all unassigned units at each of the sequentially available addresses until the string includes at least  $S_0$  units. Because the string is made up of all units at the addresses, it will often have more than  $S_0$  units, and several different combinations of  $S_0$  units could be identified.

1. Within the string, and mindful of the priority considerations previously enumerated, form the various possible USU's of size  $S_0$ , each of which includes the first unit in the string.
2. For each possible USU, compute the distance between the addresses in the USU. The *distance* is the count of all addresses encompassed by the units in the USU including addresses that may already have been assigned to some other USU.
3. Determine  $D(S_0)$ , the minimum distance of any of the possible USU's in the string, and identify the USU having this minimum distance.

Repeat these steps for proposed USU's, of size  $S_0 + 1$  and identify the USU with minimum distance  $D(S_0 + 1)$ —The string used for this purpose needs to have at least  $S_0 + 1$  units.

Choose one of the two minimum distance USU's—In general, USU's that would minimize the total within USU distance over the whole ED are desired. That is, if

$$1 + D(S_0) < D(S_0 + 1) \quad (D/5)$$

choose the USU of size  $S_0$ , otherwise choose the other. In making the choice, however, other conditions are imposed to allow flexibility in the choice of either size of USU as long as possible during the ED segmenting. For example, if formula (D/5) is the only condition used, it would be possible to exhaust the requirement for  $N(S_0)$  USU's in segmenting the first units in the ED and force all remaining units into USU's of size  $S_0 + 1$ .

An additional condition is, therefore, introduced that involves the difference between  $N(S_0)$  and  $N(S_0 + 1)$ :

If the difference is 20 or greater, define USU's which reduce the difference by relaxing the requirement expressed in formula (D/5) even if larger within-USU distances must be

accepted. If the difference is even greater (by multiples of 20), then even larger within-USU distances are accepted. If the difference is less than 20, the condition of formula (D/5) alone applies.

**Assign USU serial numbers**—The USU's are given serial numbers by a process equivalent to sequencing USU's by the lowest census HU serial number each contains and numbering USU's in this order, 1 to  $M_{re}$ . USU serial numbers  $M_{re} + 1$  to  $M_0$  are assigned to the special-place USU's.

### Identify USU's for Subsequent (Rotating) Samples

The USU serial numbers assigned are used to identify the specific USU for the various Census Bureau surveys in the ED. (See app. B, p. 114. See also col. 10, table B-1 for the specific illustration for ED 091-514 which shows that USU serial number 60 would be assigned to CPS Sample A29, No. 61 in that ED to Sample A30, etc.)

### Determine the Sample in the Special Place Portion of the ED

The numbers assigned to USU's representing special place population and housing units have been assigned to the last measures in the ED. When the sample falls in one of these USU's, the sample is selected by the following steps:

1. Record the housing units and GQ population in each of the separate special places within the ED.
2. Compute the exact measure of size for each special place in the ED

$$\hat{M}_{sei} = (\hat{H}_{sei} + \hat{P}_{sei}) / 3/4 \quad (D/6)$$

where  $\hat{H}_{sei}$  and  $\hat{P}_{sei}$  are housing units and noninstitutional GQ population in the  $i$ -th special place in the ED as determined from census address register records.

3. The computation of  $\hat{M}_{sei}$  is carried to 3 decimal places, rounded to an integer and assigned as the measure of each special place by a procedure similar to the allocation of USU's to chunks when sampling within area ED's. (See app C, p. 115.) The sum of  $\hat{M}_{sei}$  over all special places must equal  $M_{se}$ , determined in (D/1).

### ILLUSTRATION

Table D-1 illustrates the result of computer segmenting for some of the 1970 housing units in list ED 091-514.

The entries in columns 3 through 6 represent HU's with complete addresses in a list ED. This information represents the input to the segmenting operation, described in assigning the addresses of census housing units to USU's on page 115.

Columns 4, 5, and 6 show address information recorded for each housing unit in list sample ED's that is pertinent to the segmenting operation. The number of housing units at each address is shown for each HU in column 3.

The entries in columns 1 and 2 show the outcome of the segmenting operation that assigns each HU in the ED to a USU.

The USU serial number, column 1, appears for each HU in the USU. Thus, the first four HU's shown in this selected portion



of the ED comprise the 58th USU when serialized. (See p. 125.) USU number 59 is comprised of two two-unit addresses separated by a single unit address at 513 Dustin Ave. This last address is part of USU 60.

Addresses having all units in a single USU are defined as *TA addresses*; all other addresses are defined as *NTA addresses*. At a TA address, the interviewer is to include all currently existing units. For an NTA address, a subsample must be selected from an up-to-date listing of the current units.

USU 63 lies wholly within the seven-unit address at 1041 Hall Ave. The units to be interviewed when this USU comes into sample will comprise an expected four out of every seven units existing at the address at the time of interview. The instructions

for listing and (where appropriate) the subsampling of units are more completely covered in chapter IV, p. 30.

USU's comprised entirely of TA addresses are known as *TA USU's*. USU 63 illustrates an *NTA USU*. Because USU 62 has two addresses, one NTA and one TA, the USU is referred to as mixed.

Column 7 of table D-1 displays the assignment of USU's to specific Samples. Table B-1, column 10, shows USU 60 of this ED is to be assigned to Sample A29. Chapter III and appendix E show how USU 61, 62, . . . are allocated to CPS Samples A30, A31, . . . and how USU's 59, 58, . . . are allocated to Samples Y73, Y74, . . . of a different survey called HIS, the Health Interview Survey.

TABLE D-1 1970 Census Housing Units Assigned to USU's by the Computer for a Selected Portion of List ED 091-514 in PSU 412, LaPorte-Starke Counties, Ind.

USU serial number	TA or NTA	Units at address				Allocation of USU to sample
		Number of units	House number	Street name <sup>1</sup>	Apartment or description	
(1)	(2)	(3)	(4)	(5)	(6)	(7)
58	TA	1	60	Frank Ave		Y7
58	TA	1	62	Frank Ave		Y7
58	TA	1	64	Frank Ave		Y7
58	TA	1	66	Frank Ave		Y7
58	TA	1	509	Dustin Ave	1	Y
59	TA	2	509	Dustin Ave	2	Y
59	TA	1	513	Dustin Ave		A
60	TA	2	409	E Fifth St		Y
59	TA	2	409	E Fifth St		A
59	TA	1	411	E Fifth St		A
60	TA	1	413	E Fifth St		/
60	TA	1	417	E Fifth St		.
60	TA	1	1021	Hall Ave		.
61	TA	3	1023	Hall Ave		
61	TA	3	1023	Hall Ave		
61	TA	3	1023	Hall Ave		
61	TA	1	1029	Hall Ave		
62	TA	7	1041	Hall Ave		
62	NTA	7	1041	Hall Ave		
62	NTA	7	1041	Hall Ave		
62	NTA	7	1041	Hall Ave		
63	NTA	7	1041	Hall Ave		
63	NTA	7	1041	Hall Ave		
63	NTA	7	1041	Hall Ave		
63	NTA	7	1041	Hall Ave		
63	NTA	1	482	E Sixth St		
64	TA	1	484	E Sixth St		
64	TA	2	11	Galt Pl		
64	TA	2	11	Galt Pl		

<sup>1</sup>Arbitrary addresses substituted to avoid disclosure of census information

## Appendix E.

# SAMPLING WITHIN PSU'S: RESERVATION OF USU'S FOR CURRENT SURVEY SAMPLES

### INTRODUCTION

This appendix describes how the USU's are assigned to the various current national samples within the sample PSU's. The method of assignment has been developed primarily to serve the requirements of the CPS, but the system has been extended, where possible, to encompass all surveys conducted within the CPS PSU's that use the current survey sampling frames discussed in chapter II.

The method of assigning USU's to surveys is illustrated by describing the allocations for a number of Census Bureau programs having USU's selected from all or a subset of the PSU's involved in the A- and the C-design CPS national samples. Each of these programs has some aspect of gradual sample rotation incorporated in its design. A similar assignment and gradual rotation procedure applies for USU's in the D-design sample (for CETA), but none of these assignments are given in the illustrations. The illustrations deal with the sample surveys listed at the bottom of this page. Some background on the USU assignments for these programs is also necessary.

Samples of USU's were initially assigned to the first three of these surveys at the time the CPS sample was first selected. The initial assignments are indicated in column 2 of figures E-1 through E-4. The USU assignments are identified by an alphanumeric abbreviation, as indicated in the preceding list, where A and C denote the CPS; Y, the HIS; and T, the QHS. An M abbreviation is given to USU's assigned to a survey that was never implemented and now considered with USU's labeled R for Reserve. (See the discussion of midpoint on p. 136.)

After the initial assignment of USU's, new surveys were undertaken by the Census Bureau, and some of the original assignments have been changed. The more current assignments are shown by the abbreviations appearing in column 3 of figures E-1 through E-4.

Since each of the surveys has some aspect of sample rotation (ch. III, p. 21), several separate samples of USU's are required for each survey over the life of the sample design. Thus, as shown in figures E-1 through E-4, a block of 19 adjacent USU's is assigned for each selection of the CPS (A30 through A48 and C14 through C32), a block of 12 adjacent USU's for each selection of the HIS (Y73 through Y84), etc.

### SYSTEM FOR ALLOCATING USU's

The system of reserving sample USU's for Census Bureau surveys is summarized in the following sections.

#### Allocations, Based on Initial CPS Selections

The USU's selected for the first CPS samples (A29 and C13) are the starting points for the system. Some illustrations of the procedure for first selections for the A-design sample A29 are provided in appendixes C and D. USU's adjacent to each of these first selections represent a sample selected with the same specifications as the initial CPS selections; the samples differ only because the random starts in the systematic selection process are advanced (or reduced) by 1 in each PSU. These adjacent samples are called *matches* of the CPS samples.

As an example, the initial sample of USU's for the HIS are matches of Sample A29. As shown in figures E-1 through E-3, each selection for A29 has adjacent to it (above it) a matching USU that is coded Y73, indicating an initial HIS selection.

#### Adjustment For Different Survey Selection Probabilities

Assigning USU's to other census surveys as matches of CPS samples implies designating samples with the same overall selection probability as the A-design (1 in 1,968) or the C-design (1 in  $2 \times 1,968$ ). Surveys with fractional multiples of the A- or C-design selection probabilities involve subsamples of matches of the CPS selections.

For example, the initial HIS Sample has an overall probability of 1 in 1,509.7446 and is selected from A-design sample PSU's. This represents a sample  $1,968/1,509.7446 = 1.3035$  times as large as the CPS A-design sample. Thus, for a systematic subsample of 0.3035 of the A29 selections, there are two USU's designated for the HIS (one next to the A29 selection, the other midway to the following A29 selection) and one HIS USU designated for each of the remaining proportion ( $1.0000 - 0.3035 = 0.6965$ ) of the A29 selections.

This system of subsampling the A29 matches for the HIS means that the HIS USU's are not evenly spaced selections

Survey	Survey abbreviation	Sample designs involved	USU's needed for rotation
Current Population Survey (CPS)	A or C	A and C	19
Health Interview Survey (HIS)	Y	A	12
Quarterly Household Survey (QHS)	T	C	8
National Crime Survey (NCS)	J	A	4
Annual Housing Survey (AHS)	F	A and C	1
Unassigned reserve USU's	R	A and C	(X)

X Not applicable

FIGURE E-1. Reservation of USU's for the 461-Area Sample Design SR PSU's

Type of hit	USU assignments		Number of USU's allocated			U: U's available between hits
	Initial assignment	Current assignment	Total for survey	Between midpoints	Between hits <sup>1</sup>	
(1)	(2)	(3)	(4)	(5)	(6)	(7)
A12						
	Y84 ↓ Y73	HIS	12			
	A29	NCS-J01	1			
	A30 ↓ A48	CPS A	19			
	R11	AHS-F1/F2	1			
	R12 ↓ R14	NCS J03	3			
	R15 ↓ R18	Reserve	4			
	M11 M12	Reserve	2			
	R21	AHS-F1/F2	1			
	R22 ↓ R28	Reserve	7			
	Y84 ↓ Y73	HIS	12			
	A29	NCS-J01	1			
	A30 ↓ A48	CPSA	19			
A22	R31	AHS-F1/F2	1			
	R32 ↓ R34	NCS J03	3			

28

64 = 2 x 32

13

<sup>3</sup>Midpoint

22

36

See footnotes at end of figure.

FIGURE E-1. Reservation of USU's for the 461-Area Sample Design SR PSU's—Continued

Type of hit	USU assignment		Number of USU's allocated			USU's available between hits
	Initial assignment	Current assignments	Total for survey	Between midpoints	Between hits <sup>1</sup>	
(1)	(2)	(3)	(4)	(5)	(6)	(7)
A2 <sup>2</sup> —Con.	R35 ↓ R38	Reserve	4	<sup>3</sup> Midpoint  22	64 = 2 x 32	1312
	T13 ↓ T6	QHS	8			
	M11 M12	Reserve	2			
	R41	AHS-F1/F2	1			
	R42 ↓ R45	NCS J02  NCS J08	4			
	R46 ↓ R48	Reserve	3			
	Y84 ↓ Y73	HIS	12			
C <sup>2</sup>	C13	Reserve	1	28	64 = 2 x 32	1312
	C14 ↓ C32	CPS C	19			
	R51	AHS-F1/F2	1			
	R52 ↓ R55	NCS J02  NCS J08	4			
	R56 ↓ R58	Reserve	3			
	M21 M22	Reserve	2			
	R61	AHS-F1/F2	1			
				<sup>3</sup> Midpoint		

See footnotes at end of figure.

FIGURE E-1. Reservation of USU's for the 461-Area Sample Design SR PSU's—Continued

Type of hit	USU assignment		Number of USU's allocated			USU's available between hits
	Initial assignment	Current assignments	Total for survey	Between midpoints	Between hits <sup>1</sup>	
(1)	(2)	(3)	(4)	(5)	(6)	(7)
C <sup>2</sup> —Con.	R62	QHS	2	22		
	R63					
	R64	Reserve	5			
	R68					
	Y84					
	Y73	HIS	12			
A1	A29	NCS-J01	1			
	A30	CPSA	19			

<sup>1</sup> The computation of this figure is given in app. E.

<sup>2</sup> Hits: 1 312 USU's apart are allocated A A C A A C etc.

<sup>3</sup> Midpoints located  $10 - 1312/2 = 666$  USU's beyond A-and C-hits

from the sampling frame. The slight increase in sampling error expected because of this deviation from a strictly applied systematic sample is considered acceptable in return for the simplification of the sampling process that is possible.

### Conflict In Designating Random Starts

The designation of a sample of adjacent USU's, in practice, represents a systematic sample with the initial random start

increased (or decreased) by 1.00 in every PSU. This practice is some conflict with the procedure given in appendix B, p. 11; where formula (B/1) shows the random starts are somewhat different.

To illustrate this, consider two hypothetical PSU's in the design: the first PSU in the selection order is assumed SR with the within-PSU sample selected at 1 in 1,968, the second PSU in that selection order is assumed NSR with subsampling at 1 196.8. Assume that the parameters used in formula (B-1) these PSU's are—

Parameters	Symbol	PSU No. 1	PSU No. 2
Within PSU sampling intervals	W	1,968	196.8
Initial random starts consisted with formula (C-1)	RS	1.00	1.00
Number of USU's designated	m <sub>1</sub>	30	(X)
Total measures in the ED category in the PSU	M <sub>1</sub>	59,031	(X)
Sample designation number of last selected ED	D <sub>1</sub>	57,073	(X)

X Not applicable.

FIGURE E-2. Reservation of USU's for the 461-Area Sample Design NSR PSU's in A- and C-Design Samples

Type of hit	USU assignment		Number of USU's allocated			USU's available between hits
	Initial assignment	Current assignments	Total for survey	Between midpoints	Between hits <sup>1</sup>	
(1)	(2)	(3)	(4)	(5)	(6)	(7)
A <sup>2</sup>						
	Y84 ↓ Y73	HIS	12			
	A29	NCS J01	1			
	A30 ↓ A48	CPS A	19			
	R11	AHS-F1/F2	1			
	R12 ↓ R14	NCS J03	3			
	R15 ↓ R18	NCS J07	4			
	T13 ↓ T6	QHS	8			
	M11 M12	Reserve	2	<sup>3</sup> Midpoint		
	R21	AHS-F1/F2	1			
	R22 ↓ R28	Reserve	7			
	Y84 ↓ Y73	HIS	12			
	C13	Reserve	1			
	C14 ↓ C32	CPSC	19			
	R51	AHS F1/F2	1			

36

64 = 2 x 32

 $P_{hi}(\frac{1968}{2})$ 

22

C<sup>2</sup>

See footnotes at end of figure.

**FIGURE E-2. Reservation of USU's for the 461-Area Sample Design NSR PSU's in  
A- and C-Design Samples—Continued**

Type of hit	USU assignment		Number of USU's allocated			USU's available between hits
	Initial assignment	Current assignments	Total for survey	Between midpoints	Between hits <sup>1</sup>	
(1)	(2)	(3)	(4)	(5)	(6)	(7)
C <sup>2</sup> —Con.	R52	NCS J02	4	28	64=2x32	$P_{hi}(\frac{1968}{2})$
	R55	NCS J08				
	R56 R58	Reserve	3			
	M21 M22	Reserve	2	Midpoint  22		
	R61	AHS-F1/F2	1			
	R62 R63	QHS	2			
	R64 R68	Reserve	5			
	Y84 Y73	HIS	12			
A <sup>2</sup>	A29	NCS-J01	1			
	A30 i	CPS A	19			

<sup>1</sup> The computation of this figure is given in app. E.

<sup>2</sup> Hits are assigned alternately to A and C-samples

<sup>3</sup> Midpoints located  $10 + (1/2)P_{hi}$  (1968/2) USU's beyond A-or C-hits

A random start of 1.0 in PSU No. 1 leads to the assignment of USU numbers to the Samples given in columns 1 and 2 of table E-1. In the procedure actually used, formula (B/1) determines the first (A29) hit in PSU No. 2, but the other sample hits for PSU No. 2 are assigned consecutively from the A29 hit, as indicated in column 4 of table E-1. If formula (B/1) is applied to determine PSU No. 2 starts for each of the 1,968 different Samples it is possible to select from PSU No. 1, the values given in column 3 rather than column 4 would result. However, the column 3 values imply that some USU's would be used more than once, and this result is unacceptable. For example, by

rounding the noninteger starts to the nearest integer, the second USU in PSU No. 2 would be used for all Samples corresponding to the column 3 computed starts of 1.5, 1.6, ..., 2.4; i.e. Samples A34 through A43

### RESERVATIONS IN SAMPLE PSU's, BY TYPE

The specific pattern of reservations of USU's for a survey depends on the sample design and whether the PSU is SR or N

**FIGURE E-3. Reservation of USU's for the 461-Area Sample Design NSR PSU's in A-Design Sample**

[illegible]

See footnotes at end of figure.



FIGURE E-3. Reservation of USU's for the 461-Area Sample Design NSR PSU's in A-Design Sample—Continue

Type of hit	USU assignment		Number of USU's allocated			USU's available between hits
	Initial assignment	Current assignments	Total for survey	Between midpoints	Between hits <sup>1</sup>	
(1)	(2)	(3)	(4)	(5)	(6)	(7)
A	A29	NCS-J01	1			
	A30	CPSA	19			

<sup>1</sup> The computation of this figure is given in app. E.<sup>2</sup> Midpoints located  $(1/2) P_{hi}$  (1968) USU's beyond A-hit

TABLE E-1. Random Starts to Assign Matching Samples in Two Hypothetical PSU's

CPS Samples	PSU No. 1	PSU No. 2	
		Calculated from formula (B/1)	Used in practice
(1)	(2)	(3)	(4)
A29	1	10	1
A30	2	1.1	2
A31	3	1.2	3
A32	4	1.2	4
A33	5	1.4	5
A34	6	1.5	6
A43	15	2.4	15
A48	20	2.9	20
(X)			
	196	20.5	196
	197	20.6	(X)
	1.959	196.8	(X)
	1.960	1	(X)
	1.961	2	(X)
	1.968	.9	(X)

X Not applicable.

The following features of the A- and C-designs used in the influence this problem:

- The total sample USU's in all PSU's in a C-design sample 1/2 the number in the A design.
- The same PSU's are SR in both the A- and C-designs.
- The number of NSR PSU's in the C-design is one-half the number in the A-design.
- Some PSU's are NSR and in sample for both A- and C-designs; within each of these PSU's, the probability of selecting the USU's is the same for the A- and for C-design samples.

#### Reservations in SR PSU's

The foregoing features of the A- and C-designs mean that out of every three selections in SR PSU's (also called *hits*) be allocated to the A-design sample; the third hit is used in the C-sample. (See fig. E-1.) This pattern of hits is shown in column (1) where the two A-selections are identified as the A1 and A2 hits; the third selection, the C-hit.

The remaining columns of figure E-1 are explained by examining the entries appearing for the A1 hit. Columns (2) and (3) show that USU's identified as Sample A29 and originally assigned to the CPS have been reassigned to the National Survey and are identified as Sample J01.<sup>1</sup> This same reassignment appears for the A29-A2 hit. The first USU for the labeled "C13" in column 2 and originally intended as sample associated with the original assignment for A29, has been reassigned to the reserve (the National Crime Survey use design sample only). An entry of "reserve" in column 3 shows that the USU is not currently assigned and is available for future survey. Other entries show that Samples A30 through A34 are reserved for the use of the CPS for the A1 and A2 selections and a similar set, C14 through C32, following the C-selection.

<sup>1</sup> This reassignment was made before any of the sample USU's were viewed.

FIGURE E-4. Reservation of USU's for the 461-Area Sample Design NSR PSU's in C-Design Sample

Type of hit	USU assignments		Number of USU's allocated			USU's available between hits
	Initial assignment	Current assignments	Total for survey	Between midpoints	Between hits <sup>1</sup>	
(1)	(2)	(3)	(4)	(5)	(6)	(7)
C	T13 ↓ T6	QHS	8			
	C13	Reserve	1			
	C14 ↓ C32	CPS C	19			
	R51	AHS	1	28		
	R52 ↓ R58	Reserve	7		56 = 2 x 28	$P_{hi}(1968)$
	M21 M22	Reserve	2	<sup>2</sup> Midpoint		
	R61	AHS	1			
	R62 R63	QHS	2			
	R64 ↓ R68	Reserve	5		18	
	T13 ↑ T6	QHS	8			
	C13	Reserve	1			
	C14 ↓					

<sup>1</sup> The computation of this figure is given in app. E.

<sup>2</sup> Midpoints located 10 -  $P_{hi}(1968)$  2) USU's beyond C-hit

The 19 USU's following the A29 hit are followed by USU R11, now assigned to the Annual Housing Survey. Samples F1 and F2. Similarly, USU's R12, R13, and R14 have been assigned to the National Crime Survey and are followed by four USU's (R15-R18), not currently assigned to any survey.

All of the five sample surveys in this illustration were originally designed to incorporate some form of routine rotation of sample

USU's. For the CPS, the samples start with the lower numbered USU's and are replaced by larger numbers (e.g., A30 is replaced by A31, etc.). The direction of replacement is indicated by the numbering of Samples for the surveys and is also shown by the arrow in column 2.

Column 7 records the total number of USU's between the hits; this number is a constant (1,312) for all SR PSU's. Column

6 records the number of different sample USU's between each hit in which interviews are conducted during the life of the sample design for the five surveys listed at the beginning of this appendix on p. 127. (The computation of the required number of USU's between hits is discussed on this page.

Column 5 records the number of USU's reserved between each hit and a point following the hit called the midpoint. The midpoint is involved in determining the entry for column 6. For SR PSU's, midpoints are located halfway to the following hit plus 10 USU's; e.g., the midpoint is 666 USU's ( $10 + 1,312/2$ ) after the A1 hit and  $1,312 - 666 = 646$  USU's before the A2 hit.

### Reservations In NSR PSU's In Both A- and C-Design Samples

The reservation of USU's for surveys for the group of PSU's in the A- and C-design samples simultaneously is recorded in figure E-2 in the same way as for self-representing PSU's in figure E-1. However, the following distinctions must be recognized.

The A- and C-samples have been designed so that an NSR PSU in both the A- and C-samples simultaneously will have the same within-PSU sampling interval for both designs. For this reason, the sampling interval (the number of USU's between the hits) is  $2 P_{hi}$  ( $1,968/2$ ), where  $P_{hi}$  is the probability of selecting the  $hi$ -th PSU in stratum  $h$ . (See table 11-4.) Column 1 of figure E-2 shows that the hits are assigned alternately to the A- and C-designs. There is equal spacing between the A- and C-hits.

Midpoints for this set of PSU's are located half way to the following hits plus 10 USU's; e.g., the midpoints after the A- or C-hits are

$$10 + (1/2) P_{hi} (1,968/2) = 10 + P_{hi} (492)$$

USU's and

$$P_{hi} (1,968/2) - 10 - P_{hi} (492) = P_{hi} (492) - 10$$

USU's before the C- or A-hits.

As shown in column 6, the spacing between the A- and C-hits must be at least 64 USU's to provide for all samples over the decade. (See the computation of the required number of USU's between hits on this page. Consequently, the sampling interval between the two A hits must be at least 128 USU's (2 times 64). If the PSU is so small relative to its stratum population that the sampling interval within the PSU is less than 1 in 128, the PSU will be exhausted before the decade. (See the replacement of PSU's in ch. III on p. 25.)

### Reservations In NSR PSU's In The A-Design Sample

The explanations given for figures E-1 and E-2 apply to figure E-3 as well. The location of the midpoint for this class of PSU's has a definition differing from all other PSU's; it is half way to the next hit.

The discussion on the computation of the required number of USU's between hits shows that a minimum of 80 USU's between the two A-hits is required to avoid exhausting all USU's in the sample PSU before the decade.

### Reservations In NSR PSU's in the C-Design Sample

The explanations given for figures E-1 and E-2 apply to figure E-4 as well. The midpoint is located  $10 + P_{hi} (1,968/2)$  USU's beyond the C-hits.

The following discussion shows that PSU's of this category must have a within-PSU sampling interval of at least 56 to avoid exhausting the PSU before the decade.

### COMPUTATION OF THE REQUIRED NUMBER OF USU's BETWEEN HITS

It is necessary to determine the number of USU's available in each PSU for assignment to the surveys planned. PSU's with an insufficient number of USU's are subject to replacement. (See the replacement of PSU's in ch. III, p. 25.) As will be seen, the number of USU's available depends on whether the PSU is SR or NSR and on its being in sample in the A-design, the C-design or in both designs.

The concept of the midpoint is involved in these computation and developed in the following manner. Early in the planning of the A- and C-design samples, a method of providing a double CPS sample for interview in March of each year was under consideration. For this purpose, it would appear reasonable to assign the two needed USU's<sup>2</sup> exactly half way between each A- and C-sample hit; thus, at the beginning of the decade, there would be equal spacing between the CPS USU's and the additional USU's interviewed in March. However, because the additional USU's would not rotate like regular CPS USU's, the hits would be irregularly spaced by the end of the decade (when the CP would be using Samples A48 and C32). Because of this, the USU's for doubling the sample were relocated to reduce the impact of the spacing problem. These locations are called the *midpoints* and have separate definitions that depend on the kind of PSU. Although the survey that would double the sample each March has not been implemented, it has left its mark on the system of allocating USU's to the samples.

The required number of USU's is computed by a three-step process for all PSU's in each of the four types of PSU, given figures E-1 through E-4.

- First: Determine the half point that divides the interval between hits into two equal halves (note that the location of the midpoint does not necessarily halve the interval).
- Second: Determine the number of reservations in the half interval with the largest number of reservations; call this number  $R$ .
- Third: Use  $R$  to determine the required number of USU's between hits.

Since figure E-3 represents the simplest case, it is described first.

### NSR PSU's In The A-Design Sample

For these PSU's, the midpoint is defined so that it exactly halves the interval between hits. The total USU's reserved

<sup>2</sup> These are the USU's abbreviated as M and discussed on p. 127.

each of the half-intervals is recorded in column 5 of figure E-3. The larger of the two entries is 40 reserved USU's. A complete interval, therefore, requires 80 USU's, the entry in column 6. If  $1.968 (P_{hi}) < 80$ , the PSU will be involved in replacement.

### Self-Representing PSU's

The midpoint for these PSU's is defined 10 units beyond half the interval. Column 5 of figure E-1 shows that there are 28 USU's assigned between the A1 hit and the first midpoint, and, therefore,  $28 - 10 = 18$  assignments in the first half-interval after A1. The complete set of reservations for the six half-intervals from A1 to the next A1 selection, assuming all USU's in figure E-1 are assigned to a survey, is—

A1 to half point .....	18
Half point to A2 .....	32
A2 to half point .....	28
Half point to C .....	32
C to half point .....	18
Half point to A1 .....	32

The largest of these half-interval reservations is 32; therefore, the interval between the A1, A2, and C-hits can be no less than

64. As all SR PSU's will have intervals of 1,312, there is clearly no danger of exhausting all USU's in SR PSU's.

### NSR PSU's In The A- And C-Design Samples

The number of reservations in the four half-intervals from the A to the next A-hit (fig. E-2) are—

A to half point .....	26
Half point to C .....	32
C to half point .....	18
Half point to A .....	32

The largest of the half-interval reservations is 32; thus, the AC interval is 64, and the interval between A-hits is 128. If  $P_{hi} (1.968) < 128$ , the PSU will be involved in replacement.

### NSR PSU's In The C-Design Sample

The numbers of reservations in the two half-intervals are 18 and 28; thus, the interval between C-hits must be 56. If  $P_{hi} (1.968) < 56$ , the PSU will be involved in replacement.

## Appendix F.

# MAINTAINING THE DESIRED SAMPLE SIZE

## INTRODUCTION

Each of the separate, continuing operations involved in selecting the sample, conducting interviews, and preparing estimates from the CPS has been reviewed and made as efficient as possible over the course of the nearly 40 years the CPS has been in existence. One reaction to change in survey conditions, such as a further cost restriction, is to modify the sample size within the existing design. Two common examples of this are the need periodically to reduce the overall probability of selection so that a constant sample size is maintained with the continued growth in total population and to occasionally reduce the sample size to meet a revised budget or to remain within a fixed budget when CPS costs increase elsewhere. The CPS sample size is reduced by deleting one or more subsamples of USU's that have been designated for this purpose as a part of the original CPS design.

The problem of incorporating a substantial increase in the CPS sample size is handled by other means.

## SAMPLE REDUCTION

### National Sample

The original sample of USU's for Samples A29 and C13 can be identified as the sum of 101 subsamples. These subsamples are called *cut groups* and are used for the systematic reduction of the total CPS national sample. The number of cut groups chosen (i.e., 101) is rather arbitrary except that it needs to be a

number prime to 8, the number of rotation groups, because cut groups were assigned without regard to rotation group.

All USU's (except Cen-Sup) have a cut group identification number. Except for permit USU's, the number was assigned to the USU when Samples A29 and C13 were selected. The number, once assigned to a USU, is retained for all matching USU's in subsequent samples.<sup>1</sup> Permit USU's are given cut group numbers as they are selected. The treatment of Cen-Sup USU's is discussed on page 139.

The cut group number was assigned by first sorting USU's in a specific sequence and then numbering them 1 through 101 repeatedly through all USU's. The sort sequence is—

1. Census region
2. SR and NSR
3. SMSA/non-SMSA
4. C, B, U, and R
5. PSU order of selection<sup>2</sup>
6. Hit number<sup>2</sup>

A reduction in the national sample can be easily accomplished by deleting all USU's in one or more cut groups. Thus, if there are  $k$  cut groups remaining in sample at some time, the sample may be reduced by 1 in  $k$  by deleting one of the  $k$  cut groups. Cut group numbers are chosen for deletion in a specific sequence designed to keep the deletions spaced out as evenly as possible

<sup>1</sup> A USU made up of housing units adjacent to units in a second USU is called a match. (See app. E, p. 127.)

<sup>2</sup> See app. B.

TABLE F-1. Basic Weights for A-, C-, and Combined A- and C-Design National Samples

National samples	First month in sample	Cut groups remaining	Basic weights		
			A-design sample	C-design sample	Combined A- and C-design sample
(1)	(2)	(3)	(4)	(5)	(6)
A29/C13 .....	(X)	101	1968 0000	3936 0000	1312.00
A30/C14 .....	Dec 1971	96	2070 5000	4141 0000	1380.33
A31/C15 .....	Aug 1972	95	2092 2947	4184 5895	1394.86
A32/C16 .....	Apr 1973	93	2137 2903	4274 5806	1424.86
A33/C17 .....	Dec 1973	91	2184 2637	4368 5274	1456.17
A34/C18 .....	Aug 1974	89	2233 3483	4466 6966	1488.89
A35/C19 .....	Apr 1975	89	2233 3483	4466 6966	1488.89
A36/C20 .....	Dec 1975	89	2233 3483	4466 6966	1488.89
A37/C21 .....	Aug 1976	88	2258 7273	4517 4545	1505.81
A38/C22 .....	Apr 1977	88	2258 7273	4517 4545	1505.81
A39/C23 .....	Dec 1977	85	2338 4470	4676 8940	1558.96

X Not applicable

groups have been used to modify for the CPS samples. The tableabilities (of 1 in 1.968 for Sample C13) were modified for Sample 101 cut groups. The selection A30) is given by

6) = 1 in 2.070.5

5 cut groups remaining, a 1 in 96 samples A31/C15 were introduced, of the remaining 96 groups. rotation groups of a Sample (e.g., e basic overall selection probability weight. However, in a given several different Samples involved, chart, fig. III-1, where, for February is comprised of four rotation from Sample A32, and three from s that each of these Samples has

The system of deleting USU's from the sample and adjusting the basic weights for the surviving USU's is followed for all segments except the Cen-Sup. For this group of segments, USU's are never deleted in the reduction process. The reduction is handled by modifying the weight adjustment factors that are routinely applied to each Cen-Sup USU during estimation.

### Supplemental Samples for the States

The same system for identifying USU's in the D-design sample in terms of cut groups was incorporated in the sampling procedure. These cut groups were intended for use in reducing the State supplemental samples. However, it was realized that the sample requirements and rates of population change would differ by State; therefore, it would be futile to specify a common cut group number to be deleted over all State D-design samples. Thus, although the system is available for reducing a State supplemental sample, it has not yet been used.

Since State supplemental samples were selected for use while national Samples A33 and C17 were involved, the State samples have been assigned to 91 cut groups, a number consistent with the surviving groups in Samples A33/C17.

A-, C-, and D-design sample treatment for these USU's

s as defined for the system (EU); the term is actually inverse but is made up of sample. There are eight rotation group, and each is to contain 100 or more quarters population) a r of the REU for the rotation; a member for eight CP: red Sample as the first; e, it will have units sub balance of the period; a total number of units i ample, a USU in rotation ave 100 or more units; views, becomes a men USU will have subsan n rotation group 5 of th

y the REU is defined th its in such USU's in th is are discussed later

### s Universe

ted for interview in ea practice, for each lar the units is assigned: roup to which the lar ple is selected for int: succeeding (matchin hth subsample, which assigned at some ti amples and interview e way as a normal U: es all of the large US subsample of units fr unbiased sample of EU. No special weic are.

### IS PROCEDURE

tributed by variatio it can introduce bir ions. The biases ca

CPS estimates to be understated; however, in light of the relatively few USU's involved, the biases are not serious.

### Failure to Include All Large USU's in the REU

It is only during the period originally assigned for interview that USU's discovered to be in excess of 100 units are put into the REU. The rotation chart (fig. III-1) shows that this covers eight interviews assigned over a period of 16 months. Because of the definition, USU's that grow to exceed the 100 unit limit after being retired from the CPS in the regular rotation are not included in the REU.

If the rotation chart is extended, it will be seen that the eighth month of interview of the eighth subsample from a large USU in the REU will occur in the 72d month, counting from the month the USU first entered the sample in regular rotation. During the 56-month period (given by the difference of  $72 - 16 = 56$ ), other sample USU's in the original rotation group could have become large but will not be in the REU and, therefore, will not be represented by a subsample of units for interview.

This bias could be eliminated by periodically, during the 56-month period, canvassing USU's no longer in sample and, for those USU's found to be large, applying the rare events procedure for the balance of the 72-month period. This is not being

done primarily because of the additional cost and the inconvenience of administering the additional visits.

### Introduction of New Redesign

As indicated in the definition, we need to have the REU comprise the accumulation of large USU's over eight CPS Samples to have subsamples of units from these USU's included in the sample without special weights. But, while a new sample design is being introduced and before eight Samples have been used, special weights are needed until the necessary accumulation is completed. However, some of the weights required for this reason are unacceptably large.

A general rule limits any single adjustment to the basic CPS weight<sup>1</sup> for a USU to a factor no greater than 4.0. The objective is to control the variance, if necessary, at the expense of admitting some small bias. The extent of bias introduced in the REU procedure is not large and may be inferred from table G-1. The table shows the theoretical and the actual special weight factors assigned to subsamples of units from USU's that are large enough to become members of an REU and are interviewed in the first eight samples.

<sup>1</sup> CPS basic weights are discussed in app. F

TABLE G-1. Special Weights for Subsamples of Units From the Rare Events Universe, by Sample

National samples <sup>1</sup>		State supplemental samples	Special weights	
Phase I areas	Phase II areas		Theoretically required	Actually used
A30/C14	A30/C14	D34	27.5	4.0
	<sup>2</sup> A31/C15		27.5	4.0
A31/C15	<sup>4</sup> A31/C15	D35	3.75	3.75
A32/C16	A32/C16	D36	2.5	2.5
A33/C17	A33/C17	D37	1.875	1.875
A34/C18	A34/C18	D38	1.5	1.5
A35/C19	A35/C19	D39	1.25	1.25
A36/C20	A36/C20	D40	1.0714	1.0714
A37/C21	A37/C21	D41	1.0	1.0

<sup>1</sup> See ch III p. 22, for discussion of areas in phase I and phase II

<sup>2</sup> Maximum weight assigned is 4.0

<sup>3</sup> For interviews in phase II in August 1972-June 1973

<sup>4</sup> For interviews in phase II in August 1973-June 1974

## Appendix H.

# RULES FOR USUAL PLACE OF RESIDENCE

### INTRODUCTION

The following rules and examples that define and illustrate the usual place of residence of a person in a sample unit have been adapted from the *CPS Interviewer's Reference Manual*, CPS-250 [39, ch. 3].

The usual place of residence is ordinarily the place where a person usually lives and sleeps. A usual place of residence must be specific living quarters held by the person to which he is free to return at any time. Living quarters that a person rents or lends to someone else are not considered his usual place of residence during the time these quarters are occupied by someone else. Likewise, vacant living quarters that a person offers for rent or sale during his absence are not considered his usual place of residence while he is away. The following rules illustrate how this concept is applied; the table on page 143 summarizes these rules.

### RULES

#### Families With Two or More Homes

Some families have two or more homes and may spend part of the time in each. For such cases, the usual residence is the place in which the person spends the largest part of the calendar year. Only one unit can be the usual residence. For example, the Browns own the home in the city and live there most of the year. They spend their summer vacation at their beach cottage. Neither house is rented in their absence. Assuming the Browns are occupants of a housing unit in sample, and—

- The Browns are at the beach cottage, report the beach cottage as "occupied by persons with usual residence elsewhere"; if the city home was in sample, the unit would be reported as "temporarily absent."
- The Browns are at the city home, interview the Browns at their city home; if the beach cottage was in sample, it would be reported as "vacant seasonal."

#### Seamen

Consider the usual place of residence for crew members of a vessel to be their homes, rather than on the vessel, regardless of the length of their tours and regardless of whether they are at home or on the vessel at the time of the visit (assuming that they have no usual place of residence elsewhere).

#### Other Occupants

A mail address alone does not constitute a usual place of residence. Usual residents not only include members of the family, but may also include—

- Lodgers
- Servants

- Farm hands
- Other employees who live in the unit and consider it their usual place of residence

In addition, usual residents include persons who live in sample unit but are temporarily absent, such as—

- Unmarried children away at school
- Persons traveling on business
- Railroad men
- Bus drivers
- Persons visiting, on vacation, or temporarily in hospital

#### Persons With No Usual Place of Residence Elsewhere

Persons with no usual place of residence elsewhere are interviewed at the sample unit. Such persons include migrants, persons trying to find permanent living quarters, persons staying temporarily in the unit who have no other of their own.

Entire households with no usual place of residence elsewhere include persons who have rented or lent their usual living quarters to others and are traveling or temporarily staying in another unit. Often, such people offer their usual living quarters for sale, sometimes as furnished units, during their absence.

#### Citizens of Foreign Countries Temporarily in the United States

Determine whether to interview citizens of foreign countries temporarily in the United States according to the following

1. Living on premises of an embassy. Do not interview citizens of foreign countries and other persons who are living on premises of an embassy, ministry, legation, chancery, consulate.
2. Not living on premises of an embassy. Interview citizens of foreign countries and members of their families temporarily in the United States and not living on the premises of an embassy, etc., only if they are students or are employed here and have no usual place of residence elsewhere in the United States.

Interview citizens of foreign countries, permanently in the United States, but not on the premises of an embassy, etc., if they have no usual place of residence in the United States. However, do not list citizens who are merely visiting or traveling in the United States.

#### Members of the Armed Forces

Consider the usual place of residence for members of the Armed Forces (either men or women) as the sample household if they are stationed in the locality and usually sleep in the unit.



TABLE H-1. Summary Table for Determining Usual Place of Residence

Persons associated with the sample unit	Include as member of the sample unit	
	Yes	No
<b>PERSONS STAYING IN SAMPLE UNIT AT TIME OF INTERVIEW</b>		
Any person in unit, including members of family lodgers servants visitors etc Ordinarily stay here all the time (sleep here) Here temporarily—no living quarters held for persons elsewhere Here temporarily—living quarters held for persons elsewhere	Yes Yes	No
In the Armed Forces Stationed in this locality usually sleep here Temporarily here on leave—stationed elsewhere	Yes	No
Students—Here temporarily attending school—living quarters held for person elsewhere Not married nor accompanied by own family Married and accompanied by own family	Yes	No
<b>ABSENT PERSONS WHO USUALLY LIVE HERE</b>		
Inmates of specified institutions—absent because inmate in a specified institution regardless of whether or not living quarters held for person here		No
Persons temporarily absent on vacation, in general hospital, etc. (including veterans' facilities that are general hospitals)—living quarters held here for person	Yes	
Absent in connection with job Living quarters held here for person—temporarily absent while on the road in connection with job (e.g., traveling salesmen, railroad men, bus drivers) Living quarters held here and elsewhere for person but comes here infrequently (e.g., construction engineers) Living quarters held here at home for unmarried college student working away from home during summer school vacation	Yes Yes Yes	No
In the Armed Forces—were members of this household at time of induction but currently stationed elsewhere		No
In school—away temporarily attending school, living quarters held for person here Not married nor accompanied by own family Married and accompanied by own family Attending school overseas	Yes	No No No
<b>EXCEPTIONS AND DOUBTFUL CASES</b>		
Persons with two concurrent residences Regularly sleep greater part of week in another locality Regularly sleep greater part of week here	Yes	No
Citizens of foreign countries, temporarily in the United States Living on premises of an embassy, ministry, legation, chancellery, or consulate Not living on premises of an embassy, ministry, etc. If living and studying here and no usual place of residence elsewhere in the United States If merely visiting or traveling in the United States	Yes	No No No
Student nurses living at school		No

**Students and Student Nurses**

Unmarried students temporarily absent, at school, college, or trade or commercial school in another locality, are interviewed at their usual place of residence and should not be interviewed in the locality where they are attending school. A student who has left home permanently should be interviewed where he is living.

Interview a married student living with his wife and family in the locality where he is attending school and not at his parent's home.

Do not consider any student who is attending school overseas, at the time of your visits, as a household member. However, if he is in the household at the time of interview (e.g., home on summer vacation) and has no other place of residence in the United States, count him as a household member.

Since student nurses are not students in the usual sense because they generally live year round at their place of training, the rules applicable to other students do not apply. Interview a student nurse as a resident of the hospital, nurses' home, or other place in which she lives while receiving her training, even though she may have a home elsewhere that she occasionally visits. Similarly, if she were living in a nurses' home, she would be interviewed as a resident there. If she is living at home while in training, interview her at home.

#### **Persons With Two Concurrent Residences**

If a person sleeps part of the week in one place and part in another, enumerate him in the unit in which he sleeps the greater part of the week.

#### **Persons in Vacation Homes, Tourist Cabins, or Trailers**

Interview persons living in vacation homes, tourist cabins, trailers if they usually live there or if they have no usual residence elsewhere. Do not interview them if they usually live elsewhere.

#### **Inmates of Specified Institutions**

Persons who are inmates of certain types of institutions at time of interview are not considered members of the sample. They are usual residents at the institution.

These institutions include the following types:

- Correctional institutions
- Homes of the aged, infirm or needy
- Mental institutions
- Nursing, convalescent, and rest homes
- Other hospitals and homes providing specialized care

# Appendix I.

## SUPPLEMENTAL SURVEYS USING THE CPS SAMPLE IN 1975 and 1976

### INTRODUCTION

During calendar years 1975 and 1976, the CPS interviews included supplementary questions on the following topics. The supplementary questions were directed at sample households in all eight rotation groups in the A- and C-design samples, unless otherwise specified. Interviews for the basic CPS began in D-design sample households for selected States in September, 1975; after this time, the supplemental questions were asked at households in the A-, C-, and D-design samples, unless indicated otherwise. The publications for these topics are not always based on the full sample.

### 1975 SUPPLEMENTAL SURVEYS

January	No supplement.	June	Marital and Birth History and Birth Expectations (DHEW, National Institute of Child Health and Human Development, NICHD)—Study includes dates of first and most recent marriage, number of children ever born, and dates of birth and future birth expectations. Birth history is asked of never-married women in the one outgoing rotation group only.
February	(Department of Health, Education, and Welfare, DHEW, National Center for Education Statistics, NCES)—A set of screening questions are added to CPS to identify a sample of preschool, child care institutions. The schools were later contacted by a mail questionnaire, independent of the CPS.	July	Survey of Languages (DHEW, NCES)—Questions concerning place of birth, mother tongue, usual and other languages spoken, and difficulty of speaking or understanding English for persons 4 years old and over.
March	Annual Demographic Supplement—Sponsored annually by the Bureau of the Census and the Bureau of Labor Statistics (BLS), to obtain data on family characteristics, household composition, relationship to head, marital status, migration of the population since March 1, 1970, income from all sources during the previous calendar year, information on weeks worked (whether worked full time or part time), time spent looking for work or on layoff from a job, and the occupation and industry classification of the job held longest during the year. The sample is supplemented by about 2000 units at which households of Spanish origin were residing in the preceding November.	August	Food Stamp Recipients (DHEW)—Survey of receipt of food stamps during the previous 12 months, cost and value of stamps received in the most recent month, receipt of other transfer payments, monthly income, and housing expenses. This is a repeat of an earlier survey conducted in August 1974 and is more detailed than the April or December 1975 surveys.
April	Food Stamp Recipients (DHEW)—The cost and value of food stamps received by any household member during the first three months of the calendar year is asked and the number of months in the previous calendar year when a household member purchased food stamps.	September	National Immunization Survey (DHEW, United States Public Health Service)—To obtain current data measuring the extent of protection against polio, influenza, diphtheria, whooping cough, tetanus, measles, mumps, smallpox, and chicken pox. Also measured are the incidence of diabetes and chronic heart and lung conditions. Based on questions in six out of the eight rotation groups.
May	1. Survey of Multiple Jobholding and Premium Pay—Sponsored by the BLS to obtain information relating to incidence and characteristics of persons working at more than one job during survey week, to determine whether wage and salary workers reporting more than 40 hours of work during the reference week received premium pay for their overtime hours, and to measure the extent of usual overtime work. In addition, data on the beginning and ending hour of work, usual number of days and hours worked during a week, the hourly wage for hourly workers, and labor union membership for wage and salary workers are collected.	October	School Enrollment Survey (Census, BLS, NCES)—To obtain information on school enrollment (including junior college enrollment), date last attended school, date of high school graduation, and future educational plans of high school seniors. Basic enrollment data are provided for persons 3 years old and over for households in the A-, C-, and D-design samples.
			2. Adult Education (DHEW, NCES)—For persons 17 years old and over who have participated in some form of adult education in the past 12 months; type of activity, reasons for participation, sponsorship, method and place of instruction, length of participation, source of payment, and type of credit received are asked.

**November** Survey of Earnings (BLS)—Study of current earnings of wage and salary workers using more specific questions than used in the May survey. Respondents are to report specific wage rates rather than estimates of weekly earnings. For four rotation groups (in the returning portion) of the A-design sample the regular household respondent is used, for the C-design sample, the interviewer interviewed the respondent for himself. Interviews of the supplementary questions were omitted for the other half of the A-design sample and for all D-design sample units.

**December** For all households in the A-, C-, and D-design samples:

1. Farm Wage Workers Survey (U.S. Department of Agriculture)—An annual survey of total farm days and wages, total nonfarm days and wages, chief activity during the year, and migratory status of hired farm workers.
2. Food Stamp recipients (DHEW)—One question concerning food stamps purchased by household members during November 1975.

#### 1976 SUPPLEMENTAL SURVEYS

**January** No supplement.

**February** No supplement.

**March** Annual Demographic Supplement—(See March 1975 listing, except the reference date for migration questions is March 1975.)

**April** Food Stamp Recipients—(See April 1975 listing.)

- May**
1. Survey of Multiple Jobholding and Premium Pay—(See May 1975 listing.)
  2. Survey of Jobsearch Activities of the Employed (BLS)—Persons who have been employed at their present job for 4 weeks or more are asked if they have looked for another job during the past 4 weeks, and, if so, the methods they have used and the main reason they are looking.
  3. Survey of Jobsearch Activities of the Unemployed (BLS)—To obtain information on the

kinds of jobs unemployed persons are seeing and the methods and intensity with which such persons are trying to find work. The information is collected directly from the unemployed persons in the sample by a self-enumeration/mailback questionnaire.

**June** Survey of Children Ever Born and Birth Expectations—A shorter version of the June 1975 supplement, omitting the questions on marital history and birth date of each child.

**July** Survey of Work History and Job Search Activities of Persons Not in the Labor Force (BLS)—To obtain data on the work experience and labor force attachment of persons currently not in the labor force but who indicate they would like a job now or intend to look for work within the next months. All persons of this category in households in their 4th and 8th months in sample July and August are in the study. Each of these persons is asked to complete a self-enumeration/mailback questionnaire.

- August**
1. Survey of Work History and Job Search Activities of Persons Not in the Labor Force—(See July 1976 listing.)
  2. Food Stamp Recipients—(See August 1976 listing.)

**September** National Immunization Survey—(See September 1975 listing.)

**October** School Enrollment Survey—Similar to the survey in October 1975, however, new questions are included concerning types of vocational training and relationship of such training to a job. Questions concerning future education plans of school seniors, included in 1975, are deleted.

**November** National Voting Survey—This is a biennial survey to obtain information on voter participation, registration by demographic characteristics, reasons for not voting or registering are obtained, as is information on voting activities in previous elections.

**December** Survey of Farm Wage Workers—Similar to the survey in December 1975. New questions are added to obtain information on when workers migrated and the States in which they work.

## Appendix J.

# ORGANIZATION AND TRAINING OF THE DATA COLLECTION STAFF

### INTRODUCTION

The data collection staff for all Census Bureau programs is administered through 12 regional offices by the Chief of the Field Division with headquarters in Washington.

### REGIONAL OFFICES

#### Organization

The staffs of the regional offices carry out the Bureau's field data collection programs including sample surveys and censuses. Currently, the regional offices supervise a total staff of about 2,800 part-time interviewers for continuing current programs; approximately 1,350 of these interviewers work on the Current Population Survey national and State supplemental samples. When a census is being taken, the staff becomes much larger. The location of the regional offices and the boundaries of their areas of responsibility are displayed in figure J-1 and on the maps discussed in appendix M. Regional office areas were originally defined to equalize the office workloads for all programs. Figure J-2 shows the average number of CPS national and State supplemental sample units assigned for interview per month in each regional office.

A regional director is in charge of each regional office and program coordinators report to him through an assistant regional director. The CPS is the responsibility of the demographic program coordinator who has on his staff the CPS program supervisor. The latter has a staff of three to five office clerks working essentially full time on the CPS. All of them are full-time civil service employees who work out of the census office in the regional office city.

The typical regional office has 75 to 100 interviewers available for the CPS national sample; this represents a ratio of roughly 1 in 55 of the households assigned for interview in the CPS national sample shown in table J-1. A similar ratio applies for interviewers in the state supplementation. The office usually has 12 or more senior interviewers who assist the CPS program supervisor in on-the-job training and observation of the interviewer staff. Like other interviewers, the senior interviewer is a part-time employee who works out of his or her home.

Despite the geographic dispersion of the sample areas, there is a considerable amount of personal contact between the supervisory staff and the interviewers. This is accomplished mainly through the training programs and various aspects of the quality control program. For some of the outlying PSU's, it is necessary to use the telephone and written communication to keep in continual touch with all interviewers. In addition to communications relating to the work content, there is a regular system for reporting progress and costs.

A substantial portion of the budget for field activities is allocated to controlling the quality of the interviewer's work; i.e., interviewer group training and home study, observation, and

reinterview. During a recent 12-month period (ending September 1976), about 28 percent of the budget was allocated to quality control; the remaining 72 percent was for interviewer salaries and travel, clerical processing in the regional offices, recruitment, and the supervision of these activities. The 28 percent does not include the cost of the quality control function in the edit of the interviewer's work in the regional office each month. (See ch. IV, p. 45.)

#### Recruitment of Interviewers

Approximately one-fourth of the CPS interviewers leave the staff each year. As a result, the recruitment and training of new interviewers is a continuing task in each regional office.

To be accepted as a CPS interviewer, a candidate must pass the Field Employee Selection Aid test on reading, arithmetic, and map reading. The interviewer is required to live in the PSU in which the work is to be performed and have a residence telephone and an automobile.

As part-time employees, the interviewer works 40 hours or less per week. In most cases, new interviewers are paid at grade GS-3 and are eligible for payment at the GS-4 scale after 1 year of experience.<sup>1</sup> They are paid mileage for the use of their own cars while on the job and for classroom and home study training time, as well as time spent in interviewing and traveling.

#### Interviewer Training Procedures

Initial training for new interviewers—Each interviewer, when appointed, undergoes an initial training program.

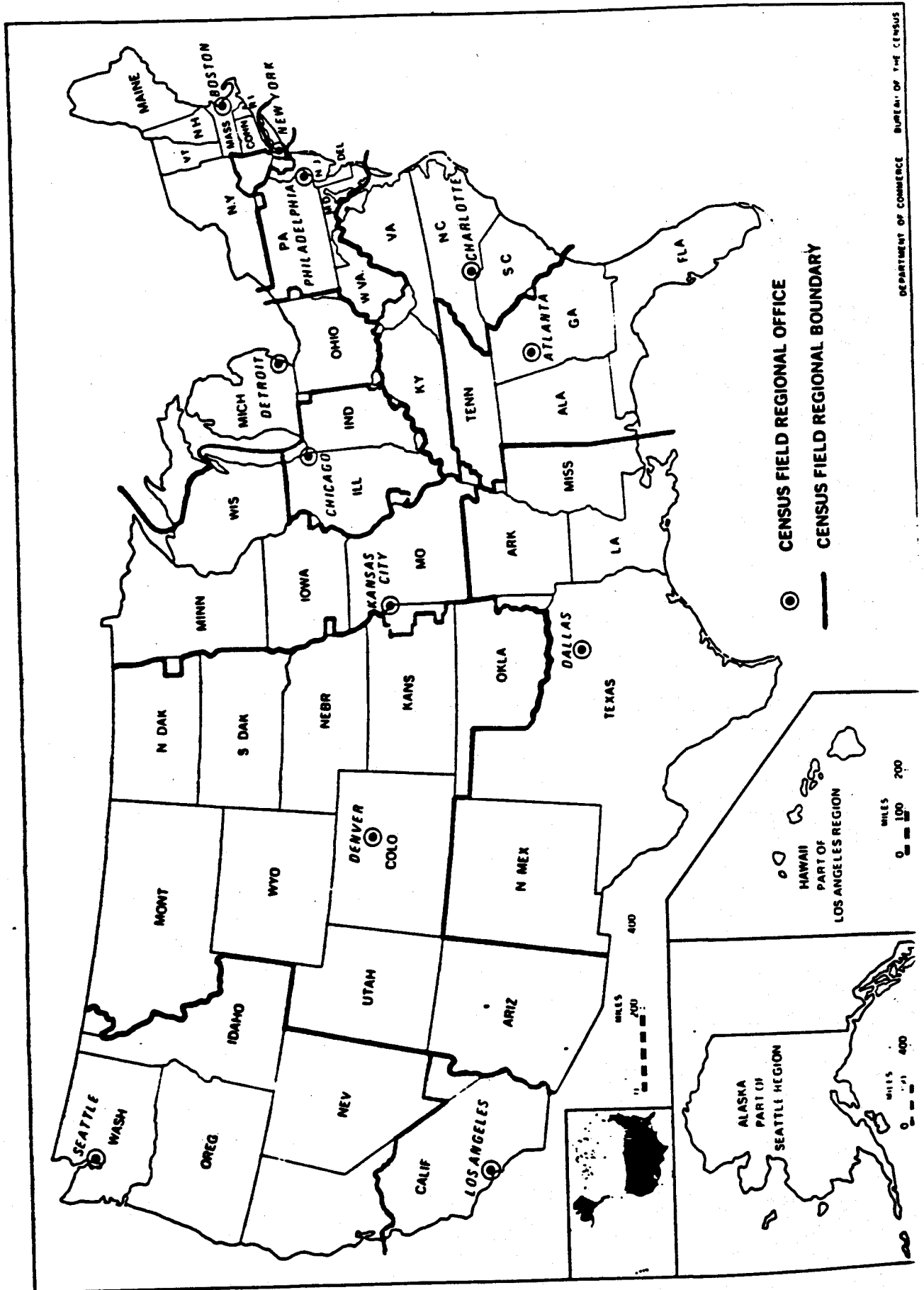
Prior to the first month's assignment, a home study exercise is completed. The interviewer is given 2-1/2 days of classroom study by the supervisor and, as a minimum, is observed by the supervisor or senior interviewer during the first 2 days of interviewing the first assignment. The classroom study includes comprehensive instruction on the completion of survey forms, with special emphasis on the CPS questionnaire, and the labor force concepts that must be understood in conducting the interviews. A large part of the classroom training is devoted to practice interviews that reinforce the correct interpretation and classification of the respondent's answers.

The interviewer completes a home study exercise before the second month's assignment and, during the second month's interview assignment, is observed for at least 1 full day by the supervisor or the senior interviewer who gives supplementary training, as needed.

The interviewer completes a home study exercise and a final review test prior to the third month's assignment. A home study assignment is completed before each of the assignments in the fourth, fifth, and sixth months. During at least 1 of these last 3

<sup>1</sup> As of October 1976, step 1 of the GS-3 and GS-4 salary levels are equivalent to \$3.56 and \$4.00 per hour, respectively.

FIGURE J-1. Census Field Regional Boundaries



months, the supervisor or the senior interviewer conducts a 1-day observation of the interviewer's work.

**Training for all interviewers**—As a part of each monthly assignment, the trained interviewer is required to complete a home-study exercise usually consisting of questions concerning labor force concepts and survey and coverage procedures. Three times a year, the interviewers are gathered in groups of about 12 for 1 day of refresher training. These sessions are conducted by the supervisor with the aid of the senior interviewers; regular CPS and supplemental survey procedures are covered.

The work of the interviewer is monitored on a continuing basis by the regular programs of reinterview, observation, and the edit, which is performed every month. (See ch. IV, p. 45.) The super-

visor schedules special visits for observation or classroom training for these interviewers shown to require it.

### Training of Regional Office Staff

The CPS staff in the regional office is trained by the CPS program supervisor. The training includes the same initial and continuing monthly training materials given to all interviewers. The CPS staff activities involved in sampling, preparing interviewer assignments, editing, and control of the survey materials are covered by the *CPS Office Manual* [40].

TABLE J-1. Monthly Sample Workloads, by Regional Office

Regional office	1970 census population of region  (thousands)	Monthly assigned interviews, occupied and vacant units <sup>1</sup>		
		National sample, A- and C-designs <sup>2</sup>	State supplements, D-design	
			Through July <sup>3</sup> 1976	September <sup>4</sup> 1976
(1)	(2)	(3)	(4)	(5)
Atlanta	17.794	5.526	—	—
Boston	17.717	4.863	1.914	1.374
Charlotte	18.071	4.856	534	391
Chicago	15.738	4.015	—	—
Dallas	20.772	5.901	1.470	1.162
Denver	12.761	3.344	5.359	4.074
Detroit	20.100	5.158	21	—
Kansas City	16.987	4.646	652	624
Los Angeles	14.371	3.992	701	523
New York	17.003	4.198	—	—
Philadelphia	18.527	4.656	755	553
Seattle	13.371	3.923	2.492	1.985
Total	203.212	55.078	13.898	10.686
Number of sample PSU's	(X)	461	*165	*155

— Entry represents zero.  
X Not applicable.

<sup>1</sup> See table IV-1, line 3

<sup>2</sup> Monthly average for 1975

<sup>3</sup> Average number of units for September-December 1975

<sup>4</sup> D-design sample reduction became fully effective in September 1976; units shown are estimates for September 1976

\* PSU's which contain D-design sample USU's only

## Appendix K.

# VARIANCE ESTIMATION: CPS NATIONAL SAMPLES

### INTRODUCTION

This appendix outlines the method of computing variances for national estimates produced from the CPS samples. The procedures are presented under three topics: The simple unbiased estimator is used to illustrate how variances are estimated for statistics that can be expressed as linear functions of sample totals; the Taylor series approximation is applied to a more complex statistic to derive its variance, under assumptions that make it equivalent to the variance of a linear combination of sample totals—the form of the linear expression is derived from the formula for the complicated statistic; and the derivation of the linear expression and the method of computing its variance are illustrated for statistics produced by the combined first- and second-stage ratio estimator for the CPS—formula (V/1).<sup>1</sup> The notation used is defined as it is introduced.

### VARIANCE OF A LINEAR FUNCTION OF SAMPLE TOTALS

The CPS unbiased estimator (ch. V) is used to illustrate the process. The methods illustrated can be applied to any linear estimate.

The CPS variance estimators presume the existence of clusters of USU's, called *Keyfitz clusters* after the author who proposed the variance method [26]. To illustrate, if  $x_1$  and  $x_2$  are the estimates from two independent samples of USU's selected within an area, we assume that

$$Ex_1 = Ex_2$$

$$Ex_1x_2 = Ex_1 Ex_2$$

From these assumptions, it follows that the variance of the sum ( $x_1 + x_2$ ) is given by

$$\text{Var}(x_1 + x_2) = E(x_1 - x_2)^2 \quad (K/1)$$

The operator  $E$  indicates the expected value of the function over all possible samples. An area with these properties is an example of a Keyfitz cluster.

We define clusters of CPS USU's so that these conditions hold for CPS samples within the clusters, and variances can then be estimated by use of expression (K/1). The definition of clusters is somewhat different in self-representing (SR) and nonself-representing (NSR) strata.

The process of selecting the CPS sample of USU's within SR strata means that any SR PSU or combination of SR PSU's will meet the conditions of a Keyfitz cluster. One possible combination would consider all 156 SR PSU's in the CPS as one Keyfitz cluster. However, a number of considerations, including the advantage of increasing the number of degrees of freedom in the

variance estimate, suggest there should be a number of smaller more geographically homogeneous clusters. For SR strata, Keyfitz clusters have been defined, each about equal in size with a population about 2.4 million. In the Northeast Census Region, there are 17 SR Keyfitz clusters: North Central, South, 14; and West, 10.

For NSR strata, the conditions required of a Keyfitz cluster met by the sample in each of the 110 pairs of A-design strata combined in the process of selecting the C-design sample.<sup>2</sup> (the section on determining the C-design sample PSU's in ch. p. 12.) Each of these pairs of strata plays the role of the Keyfitz cluster for estimating NSR variances for the combined A- and C-design samples. In this discussion, however, they will be referred to as pairs of strata to avoid confusion with another grouping of NSR strata. For this second grouping, the 110 pairs of strata have been combined into 54 *groups of strata*, about equal in size. Each group of strata is usually a combination of two, but occasionally three, pairs of strata. All PSU's in each group are in the same census region. Each of the 110 pairs of NSR strata is treated as a Keyfitz cluster when the total NSR variance is estimated for the combined A- and C-design samples or for the A-design sample alone; each of the 54 groups of NSR strata is used as a Keyfitz cluster when the NSR portion of the total within-strata variance component is estimated for the A-, C-, or combined A- and C- design samples or when the total NSR variance is estimated for the C-design sample only.

### Estimating the Variance in SR Strata

The estimate of variance among USU's within SR PSUs is based on replicated half samples within the 54 Keyfitz clusters of SR PSU's. Within each Keyfitz cluster, 8 half-sample replications are selected, at random, from the 16 subsamples of designated within each cluster for the purpose of estimating variances.

Half-sample replications—To specify the half-sample replications, the USU's within each Keyfitz cluster are placed in a sort closely resembling the order in which they were selected and assigned to rotation group (the selection order is described in app. B, p. 111). Each USU is assigned to one of 32 positions within the cluster, by using random starts and assigning random numbers in the sequence, 1 to 32. The numbering cycle is repeated for the cluster as necessary and carried over from cluster to cluster. The random starts depend on the C. B. U. R category of 1 and whether located in or not in an SMSA. This system yields panels of USU's within each cluster; each panel reflects a systematic subsample of USU's selected by a process resembling the sampling procedure employed in original sampling of USU's and rotation groups within sample PSU's.

The USU's are assigned to panels without regard to design or rotation group; i.e., the 32 panels are not assigned to USU's in rotation group 1, then in rotation group 2, separately for USU's in the A-design and then in the

<sup>1</sup> The additional variance estimation procedures for the composite estimator are covered in [36].

<sup>2</sup> This discussion ignores the impact of controlled selection of PSU's (A) which results in some lack of independence in sampling among the within the A-design and the C-design samples.



sample. For this reason, variances can be estimated using the same panel assignments for the A-, C-, or combined A- and C-design samples. However, a consequence of this method of assigning panels is to allow some variability to be introduced by the rotation group bias (ch VII, p. 83), because there may be some misrepresentation of rotation groups, by panel. The rotation group bias may affect the response variance components that appear in the variance estimates.

**Samples within Keyfitz clusters**—The original process of selecting USU's from the sample frames (as in ch. II and apps. B, C, and D) and the method of assigning the USU's to the 32 panels suggest the sample in each Keyfitz cluster could be treated as selected from 16 strata, each stratum having two panels of sample USU's. With this treatment, independent half samples could be designated, each derived by selecting one of the two panels in each of the 16 strata. A total of  $2^{16}$  independent, replicated half samples could be designated, and the variation among estimates made from each of these replicated half samples could provide the basis for an estimate of the variance of the full sample. McCarthy [72] has shown that an estimate of the variance, having equivalent precision, can be made using a subset of the  $2^{16}$  replications. The subset consists of 16 balanced (or orthogonal) replications. The number of half-sample replications in the orthogonal set is given by  $4[(n+3)/4]$ , where  $n$  is the number of strata and only the integral portion of the quotient in brackets is used. The CPS variances in SR are computed using a subsample of these orthogonal replications, a set of eight partially balanced half-sample replications.

The decision to use a sample of 8 partially balanced replications, derived from 32 panels, follows from several considerations. The precision of the variance estimates for SR would be improved by using more than 8 replications (16 replications would allow use of the orthogonal set from the 32 panels). However, this would require more computer memory and reduce the number of statistics for which variances are estimated. Also, the total variance in NSR strata is not estimated by replications so that the number of replications affects only the reliability of the variance in SR PSU's. As an alternative, the CPS sample could, in like manner, have been allocated to 16 (rather than 32) panels from which an orthogonal set of 8 replications could have been designated. However, an evaluation of the precision of the variance estimates suggested that better results would be forthcoming using a subsample of 8 partially balanced replications, selected from the orthogonal set of 16 (from 32 panels) jointly with the other half of each sample than by using the orthogonal set of 8 from 16 panels.

**Illustration of variance estimate in SR strata**—To illustrate the estimation of variance for a linear sum of sample totals, consider the unbiased estimates in SR PSU's (i.e., the sample totals inflated by the weights described in table II-4). Let  $x_s = x_{sr1} + x_{sr2}$  be the unbiased estimate in the  $s$ -th Keyfitz cluster, expressed as the sum of the estimates from two half subsamples designated for the  $r$ -th replication (i.e., for the  $r$ -th replication, a half subsample, designated as  $r_1$ , has a second half subsample,  $r_2$ , comprised of the other half); similar expressions apply for any of the replications. The USU's in the A- and C-design samples can be separately identified so that

$$\begin{aligned} x_s &= x_{As} + x_{Cs} \\ &= (x_{Asr1} + x_{Asr2}) + (x_{Csr1} + x_{Csr2}) \end{aligned}$$

where  $x_{As}$  and  $x_{Cs}$  represent the weighted sample totals of the A- and C-design USU's in the  $s$ -th Keyfitz cluster in SR strata. For the A-design (similarly, for the C-design),  $x_{Asr1}$  and  $x_{Asr2}$  are the weighted A-design sample totals of the two half samples in replication  $r$ .

Writing

$$H(x, A_s) = \frac{3}{2} (x_{Asr1} + x_{Asr2})$$

as the estimate for the characteristic in the  $s$ -th Keyfitz cluster in SR strata based on the A-design sample alone, it follows that an estimate of the total may be written as the sum over all Keyfitz clusters in SR strata:

$$H(x, A) = \sum_s^{SR} H(x, A_s) = \frac{3}{2} \sum_s^{SR} (x_{Asr1} + x_{Asr2}) \quad (K/2)$$

and based on the C-design sample alone

$$H(x, C) = \sum_s^{SR} H(x, C_s) = 3 \sum_s^{SR} (x_{Csr1} + x_{Csr2}) \quad (K/3)$$

The estimate from the combined A- and C-design samples is

$$H(x, AC) = \frac{2}{3} H(x, A) + \frac{1}{3} H(x, C)$$

and from (K/2) and (K/3) it follows that

$$H(x, AC) = \sum_s^{SR} H(x, AC_s) = \sum_s^{SR} (x_{sr1} + x_{sr2}) \quad (K/4)$$

In the  $s$ -th Keyfitz cluster, the variance can be written

$$\text{Var}(x_{sr1} + x_{sr2}) = E(x_{sr1} - x_{sr2})^2$$

subject to the assumptions for Keyfitz clusters made at (K/1). This variance can be approximated by

$$\text{Var}(x_{sr1} + x_{sr2}) \doteq \frac{1}{R} \sum_r^R (x_{sr1} - x_{sr2})^2$$

which is an average over a set of  $R$  replications in the  $s$ -th Keyfitz cluster. Because the sampling is assumed independent among clusters, it follows that an estimate of the variance in all SR strata based on the combined A- and C-design samples is approximated by

$$\begin{aligned} \text{Var}[H(x, AC)] &\doteq \frac{1}{R} \sum_s^{SR} \sum_r^R (x_{sr1} - x_{sr2})^2 \\ &\doteq \frac{1}{R} \sum_s^{SR} \sum_r^R (\Delta x_{sr})^2 \end{aligned} \quad (K/5)$$

Approximations to the variance for estimates based on the A- or on the C-design samples alone follow directly using (K/2) or (K/3).

## Estimating the Total Variance in NSR Strata

The total variance in NSR strata is estimated by using sample data from pairs and combinations of pairs of A-design strata.

The following diagram represents the  $s$ -th pair of A-design strata ( $s1$  and  $s2$ ) from which a C-design sample PSU has been selected. The A-design stratum that has both an A- and C-design sample PSU selected from it is identified as stratum  $s1$  with  $s11$  and  $s12$  identifying the A- and the C-design sample PSU's, respectively. The A-design sample PSU from stratum  $s2$  is identified as  $s21$ . The procedure is symmetrical, so that the C-design sample PSU might also have been selected from stratum  $s2$ . If this happens, the stratum designations are interchanged so that the diagram can represent all situations.

**Sample estimates**—Let  $x_{s11}$ ,  $x_{s12}$ ,  $x_{s21}$  be estimates for a characteristic given by the weighted-sample totals of all USU's in the corresponding NSR PSU's. In this illustration, the weights for the unbiased estimator for the combined A- and C-design samples are used. (See line 12 in table II-4.)

Let  $J$  be an unbiased estimate of the total for the characteristic of interest in the two strata in the  $s$ -th pair. For the A-design sample alone—

$$J(x, A_s) = \frac{3}{2} (x_{s11} + x_{s21}) \quad (K/6)$$

the estimate based on the C-design sample alone is

$$J(x, C_s) = 3 x_{s12} \quad (K/7)$$

and the estimate from the combined A- and C-design samples is

$$J(x, AC_s) = \frac{2}{3} J(x, A_s) + \frac{1}{3} J(x, C_s) \quad (K/8)$$

$$= x_{s11} + x_{s21} + x_{s12} \quad (K/9)$$

**Variance of the sample estimate**—Variances for these estimates are written by applying the definition, e.g.,

$$\text{Var } [J(x, A_s)] = E [J(x, A_s) - E J(x, A_s)]^2$$

where the operator  $E$  represents the expected value over all possible samples. Because we have assumed sampling within stratum  $s1$  is independent of sampling in stratum  $s2$ , the variances are from (K/6),

$$\text{Var } [J(x, A_s)] = \text{Var} \left( \frac{3}{2} x_{s11} \right) + \text{Var} \left( \frac{3}{2} x_{s21} \right) \quad (K/10)$$

and because of the relation of the A- and C-design samples

$$\begin{aligned} \text{Var } [J(x, C_s)] &= 2 \left[ \text{Var} \left( \frac{3}{2} x_{s11} \right) + \text{Var} \left( \frac{3}{2} x_{s21} \right) \right] \\ &\quad + \left[ E \left( \frac{3}{2} x_{s11} \right) - E \left( \frac{3}{2} x_{s21} \right) \right]^2 \end{aligned} \quad (K/11)$$

The total variance in NSR strata for the A-design sample given by (K/10), includes the variance between PSU's strata, the variance among USU's within PSU's, and the related and uncorrelated components of response error. The expression  $\text{Var} \left( \frac{3}{2} x_{s11} \right)$ , for example, represents the contribution of all of these components to the variance of the A-design sample estimate from stratum  $s1$ . The first term on the right of (K/11) represents these same variance components for estimates on the C-design sample alone and is twice the corresponding term in (K/10), because the estimate is based on only one PSU rather than the two used in the A-design sample. The second term on the right of (K/11) is the between-stratum variance component and expresses the variance arising because of the additional stage of sampling in the C-design; i.e., the probability selection of one of the two A-design strata pair of NSR strata.

Because sampling in the A-design is independent of sampling in the C-design, the variance of  $J(x, AC_s)$  follows expression (K/8) and is

$$\text{Var } [J(x, AC_s)] = \frac{4}{9} \text{Var } [J(x, A_s)] + \frac{1}{9} \text{Var } [J(x, C_s)]$$

Using expressions (K/10) and (K/11), it follows that

$$\begin{aligned} \text{Var } [J(x, AC_s)] &= \frac{3}{2} \left[ \text{Var} \left( \frac{3}{2} x_{s11} \right) + \text{Var} \left( \frac{3}{2} x_{s21} \right) \right] \\ &\quad + \frac{1}{4} \left[ E \left( \frac{3}{2} x_{s11} \right) - E \left( \frac{3}{2} x_{s21} \right) \right]^2 \end{aligned} \quad (K/12)$$

**Estimates of the variance of the sample estimate**—To estimate the variance (K/13) in the  $s$ -th pair of NSR strata formation from the sample, we make use of the expression

$$\begin{aligned} \alpha_s &= Q [(x_{s11} + x_{s12})/2 - x_{s21}]^2 \\ &\quad + T [x_{s11} - x_{s12}]^2 \end{aligned} \quad (K/13)$$

where the value of the constants  $Q$  and  $T$  are assigned 1

stratum s1	s11 (A-design sample PSU)	s12 (C-design sample PSU)
stratum s2	s21 (A-design sample PSU)	

Within a given pair of A-design strata  $s_1$  and  $s_2$ , the expected value of  $\alpha_s$  over all possible A- and C-design samples (i.e., samples of strata, PSU's, USU's, interviewers, responses, etc.), is given by

$$E\alpha_s = \left(\frac{3Q}{4} + T\right) [\text{Var}(x_{s11}) + \text{Var}(x_{s21})] + Q [E(x_{s11}) - E(x_{s21})]^2 \quad (K/15)$$

If the constants  $Q$  and  $T$  in (K/15) are assigned the values  $Q = 1/4$  and  $T = 21/16$ , then

$$E\alpha_s = \frac{3}{2} [\text{Var}(x_{s11}) + \text{Var}(x_{s21})] + \frac{1}{4} [E(x_{s11}) - E(x_{s21})]^2 \quad (K/16)$$

$$= \text{Var}[J(x, AC_s)] \quad (K/17)$$

as in (K/13). Because the sampling and estimation procedures in the  $s$ -th pair of strata are independent of the procedures in other pairs,<sup>3</sup> the estimated variance in all NSR strata may be expressed by summing the estimator (K/14) over all pairs of NSR strata. A variance estimator for the CPS unbiased estimate in NSR strata is

$$\begin{aligned} \text{Var}[J(x, AC)] &= \sum_s^{\text{NSR}} \text{Var}[J(x, AC_s)] \\ &= \sum_s^{\text{NSR}} \left\{ \frac{1}{4} [(x_{s11} + x_{s12})/2 - x_{s21}]^2 + \frac{21}{16} (x_{s11} - x_{s12})^2 \right\} \quad (K/18) \end{aligned}$$

Where the values  $Q = 1/4$  and  $T = 21/16$  have been assigned to the estimator (K/14) in each stratum pair.

The total NSR variance of the A-design sample only may be estimated by

$$\text{Var} J(x, A) = 4\left(\frac{3}{2}\right)^2 \sum_s^{\text{NSR}} \left| \frac{A_{s2}}{A_s} x_{s11} - \frac{A_{s1}}{A_s} x_{s21} \right|^2 \quad (K/19)$$

where  $A_s = A_{s1} + A_{s2}$  and  $A_{s1}$  and  $A_{s2}$  are characteristics correlated with the characteristic  $X$  in the two strata  $s_1$  and  $s_2$  in the  $s$ -th Keyfitz cluster. (Total stratum population from an earlier decennial census may be used.) Formula (K/19) is the so-called collapsed-stratum variance estimator where two strata are combined in each case [21, ch. 9, sec. 15]. This formula tends to give an overestimate of the variance [22, ch. 9, sec. 5].

An estimator of the NSR variance of the C-design sample only makes use of Keyfitz clusters, defined as groups of pairs of strata, as discussed earlier in this appendix. This is the more familiar form of the collapsed-stratum variance estimator.

<sup>3</sup> Ignoring the impact of controlled selection of PSU's.

$$\begin{aligned} \text{Var}[J(x, C)] &= \sum_g^{\text{NSR}} \text{Var}[J(x, C_g)] \\ &= \sum_g^{\text{NSR}} \frac{g L_g}{L_g - 1} \sum_s^{L_g} |x_{gs12} - x_g A_{gs} / A_g|^2 \quad (K/20) \end{aligned}$$

$L_g$  = the number of pairs of A-design strata combined into the  $g$ -th group of strata (the  $g$ -th Keyfitz cluster, usually  $L_g = 2$ ).

$A_{gs}$  =  $A_{gs1} + A_{gs2}$ , the census population of the two A-design strata in the  $s$ -th pair of strata in the  $g$ -th group.

$A_g$  = the sum of  $A_{gs}$  over the pairs of strata in the  $g$ -th group.

$3x_{gs12}$  = the estimate of characteristic  $X$  for the  $s$ -th pair of strata in the  $g$ -th group, based on the C-sample PSU.

$3x_g$  =  $\sum_s^{L_g} x_{gs12}$  the sum of the C-design estimates in the pairs of strata in the  $g$ -th group of strata.

### Variance Components for the Stages of Sampling

Formula (K/5) may also be used to estimate the variance within sample PSU's in NSR strata by using subsample panels in the pairs of NSR strata specifically constructed for this purpose. These variance estimators are given in [53].

The between-stratum component of the variance for the combined A- and C-design samples is given by summing the second term on the right of expression (K/13) over all NSR clusters. An estimator for this variance may be derived from (K/15) by assigning, in each pair, the value  $Q = 1/4$  and a value for  $T$  so that  $3Q/4 + T = 0$  (i.e.,  $T = -3/16$ ). We will make use of this later when discussing estimates of variance components in the last section of this appendix.

### VARIANCE OF NONLINEAR ESTIMATORS

This section outlines the theoretical background for producing variances of statistics that are estimated from the CPS using nonlinear expressions, for example, statistics derived from ratio estimates, regression estimators, and other estimators that may involve products or compound ratios of random variables. Variances of these expressions are developed by deriving a linear function defined in terms of partial derivatives of the original (nonlinear) expression. The partial derivatives of second and higher order are ignored, but, for most statistics produced from the CPS, the variance of the linear function is a close approximation to the variance of the original expression. However, for some complicated estimators, the use of the Taylor approximation may give a poor estimate of the variance.

The variance estimation procedure for the unbiased estimator, described in the preceding section, can be applied to any linear estimator so that it can serve as a model for the linear forms derived in this section.

## Theoretical Foundation. A Summary

The variance estimators presently used are related to methods originally adapted from a paper by Keyfitz [26], later extended in papers by Tepping [32] and Woodruff [79]. The discussion in this section follows Tepping's paper.

Let  $u = \{u_1, u_2, \dots, u_k\}$  be a vector of statistics and let their expected values be given by the vector of population parameters

$$U = \{U_1, U_2, \dots, U_k\}$$

Let the population parameter of interest be a function  $F(U)$  that is to be estimated by  $F(u)$ . In formula (V/1),  $F(u) = x''$ , which is a function of a large number of terms  $x'_a$  and  $y'_a$ , each involving other terms defined in formulas (V/2) and (V/3); the vector  $u$ , in this case, comprises the statistics given by all of these sample observations. Each of these is assumed to have an expected value, a member of vector  $U$ .

To terms of the first degree in  $(u_i - U_i)$ , the Taylor series approximation for  $F(u)$ , assuming differentiability, is given by

$$F(u) \doteq F(U) + \sum_i^k (u_i - U_i) \frac{\partial F(U)}{\partial U_i} \quad (K/21)$$

where  $\frac{\partial F(U)}{\partial U_i}$  is read: "The value of the first derivative of  $F(u)$

with respect to  $u_i$  evaluated at  $U_i$  (the expected value of  $u_i$ )."

As all terms on the right side of (K/21) are constants except the  $u_i$ , the variance of the right side is the same as the variance of

$$G(u) = \sum_i^k u_i \frac{\partial F(U)}{\partial U_i} \quad (K/22)$$

The function  $G(u)$  is the sum of  $k$  variables  $u_i$ , each multiplied by a constant,  $\frac{\partial F(U)}{\partial U_i}$ , and is linear. In practice, this means that a complicated function of the random variables  $F(u)$  has a variance that can be approximated by the variance of  $G(u)$  which, because it is linear, may be estimated by a process similar to that previously illustrated for the unbiased estimate in this appendix.

### Implications of the Assumptions

The accuracy of the approximation in using the variance of  $G(u)$ , for the variance of the more complicated expression  $F(u)$ , depends on the validity of a number of assumptions that are implied in the derivation.

1. The partial derivative of  $F(u)$  with respect to  $u_i$  is assumed to exist at  $U_i$  for each  $i$ . Most of the functions that involve totals, ratios, or means of statistics estimated by the CPS are differentiable, and the partial derivatives exist when evaluated at expected values. However, special problems arise in estimating variances for medians<sup>4</sup> and other order statistics, functions involving absolute numbers, etc.

<sup>4</sup> An approximation to the sampling error of a median is illustrated in app. L.

2. It is assumed the approximation is valid when partial derivatives of  $F(u)$ , with respect to  $u_i$  of second and higher order are dropped. The validity of this approximation has been investigated by Frankel [13] and Woodruff and Causey [8] for a wide variety of complex estimators using CPS data generally, their findings show minimal impact on the accuracy of the variances for these estimators, using sample size of the CPS.
3. The  $U_i$ , expected values of the  $u_i$ , are assumed available to evaluate the partial derivatives at  $U_i$ . In some situations, proximate values of the  $U_i$  are available from census data but often the evaluation must use the only data at hand, from the sample itself, adding a component of error to estimated variance. The CPS sample is usually large enough to produce stable estimates of the  $U_i$ ; when problem situations are foreseen, they may sometimes be dealt with assuming a modified method of applying the estimator. For example, if the denominator of a ratio estimate approaches zero in a ratio estimate cell, the sample data may be combined with other ratio estimate cells to provide a more stable basis for the ratio estimate. (See the discussion of ratio estimate factors used for the A- and C-design in ch. V, p. 60.) In some instances, the  $U_i$  are evaluated using sample estimates averaged over recent months. We also take advantage of variances routinely produced each month, to derive variances averaged over time. Single-variance estimates made for characteristics that occur frequently in the population may be less satisfactory when made by this method.

## THE VARIANCE OF A NONLINEAR ESTIMATE: AN ILLUSTRATION

Application of the theory for approximating the variance of a nonlinear estimator is illustrated by the derivation of the form and its variance for the CPS combined first- and second-stage ratio estimate, formula (V/1).

Formula (V/1), as written, does not explicitly reflect two estimation steps actually performed, i.e., the special weights applied because different sampling fractions are used and the interview adjustments. However, the variance estimate reflects the variation introduced by these special weights, the intermediate second-stage ratio estimates for the population reported as Negro and other, by age and sex, and the ratio estimate of total population, by age, sex, and collapsed race for current population controls, are treated as one second-stage estimate. (See, ch. V.) In practice, of course, the weights reflect all of the steps in the second stage of ratio estimation although the variances obtained do not.

The following discussion assumes the combined A- and C-design samples are used and, therefore, are reflected in variance estimators derived here. Variances for the A-design samples alone are derived by similar procedures.

### The Nonlinear Estimator

The formula for the CPS combined first- and second-stage ratio estimates based on the combined A- and C-design samples

given by formula (V/1), is

$$x'' = \sum_a \frac{NSR \left( x_{aSR} + \sum_c \frac{x_{ac}}{z_c} Z_c \right)}{y_{aSR} + \sum_c \frac{y_{ac}}{z_c} Z_c} \hat{y}_a$$

$$= \sum_a \frac{x_a}{y_a} \hat{y}_a \quad (K/23)$$

where

$x_{aSR}$  = the weighted sample total, from USU's in SR PSU's, of population with the desired characteristic in the a-th age-sex-race category. The weights are the inverse of the probability of selecting the USU's. In practice, the weights also include special weighting and the noninterview adjustment factors.

$y_{aSR}$  = same as for  $x_{aSR}$ , but for the total population.

$x_{ac}$  = the weighted sample total, from USU's in NSR PSU's, of population with the desired characteristic in the a-th age-sex-race category in the c-th collapsed-race-residence category.

$y_{ac}$  = same as for  $x_{ac}$ , but for the total population.

$z_c$  = estimated total 1970 census NSR population in the c-th collapsed-race-residence category, based on 1970 census populations of the NSR sample PSU's, weighted and summed over all NSR strata.

$Z_c$  = the 1970 census population in NSR in the c-th collapsed-race-residence category.

$\hat{y}_a$  = independent population total for the United States in the a-th age-sex-race category for the current CPS month.

### Determine the Partial Derivatives

To apply the theory derived earlier, we observe in formula (K/23) that  $x'' = F(u)$ , where  $u$  is the vector of statistics

$$u = \{x_{aSR}, y_{aSR}, x_{ac}, y_{ac}, z_c\} \quad (K/24)$$

The first two statistics in (K/24) are defined in SR strata, and the last three are defined in NSR strata only. Each of these terms represents a number of statistics; for example, there is an  $x_{aSR}$  for each of the 68 age-sex-race categories represented by the subscript  $a$ , and there is a  $z_c$  for each of the 6 collapsed-race-residence categories within the NSR portion of a census region. Each of these statistics is a linear sum of sample totals. The vector of the expected values of these statistics is

$$U = \{E x_{aSR}, E y_{aSR}, E x_{ac}, E y_{ac}, E z_c\} \quad (K/25)$$

$$= \{x_{aSR}, y_{aSR}, x_{ac}, y_{ac}, z_c\} \quad (K/26)$$

where the terms in (K/26) are defined, respectively, as the expected values of the separate statistics in (K/25).

The partial derivative of  $x''$ , with respect to the first of the variables in  $u$ , is

$$\frac{\partial x''}{\partial x_{aSR}} = \hat{y}_a / y_{aSR} + \sum_c \frac{NSR}{z_c} \frac{y_{ac}}{z_c} Z_c$$

This term, evaluated at the expected value of the variables, is

$$\hat{y}_a / E y_{aSR} + \sum_c \frac{NSR}{E z_c} \frac{E y_{ac}}{E z_c} Z_c \quad (K/27)$$

But, because  $E z_c = Z_c$ , the denominator of (K/27) becomes

$$E y_{aSR} + \sum_c \frac{NSR}{E z_c} E y_{ac} = \hat{y}_a$$

$\hat{y}_a$  is the expected value of the unbiased estimate of U.S. population in the a-th age-sex-race category. For convenience, expression (K/27) is written as

$$\hat{y}_a / E y_{aSR} + \sum_c \frac{NSR}{E z_c} E y_{ac} = \hat{y}_a / \hat{y}_a = \gamma_a \quad (K/28)$$

In the following section, we discuss the approximate values assigned to  $\gamma_a$  in practice.

The partial derivatives of  $x''$ , with respect to each of the variables given at (K/24) and evaluated at the expected values given at (K/26), are shown in column 3 of table K-1. The entries in column 4 of the table will be discussed later.

### Derive The Linear Form of the Estimate

The linear form for which variances are to be estimated for the combined A- and C-design samples follows from formula (K/22). It is convenient to write the linear form of the estimate in two parts:

$$G(u) = H(x'', AC) + J(x'', AC) \quad (K/29)$$

corresponding to the terms in  $G(u)$  that apply to SR and NSR strata. The first two rows in column 3, table K-1, give, for SR strata,

$$H(x'', AC) = \sum_a \gamma_a \left( x_a - \frac{\bar{x}_a}{\bar{y}_a} y_a \right) \quad (K/30)$$

and for NSR strata, the remaining three rows give

$$J(x'', AC) = \sum_a \gamma_a \sum_c \left\{ \left( x_{ac} - \frac{\bar{x}_a}{\bar{y}_a} y_{ac} \right) + \frac{\bar{x}_a}{\bar{y}_a} \frac{\bar{y}_{ac}}{Z_c} z_c - \frac{\bar{x}_{ac}}{Z_c} z_c \right\} \quad (K/31)$$

TABLE K-1. Partial Derivatives of the Function  $x''$  at (K/23)  
With Respect to the Variables in (K/24)

Variable in (K/24)	Partial derivative with respect to the variable	Derivative evaluated at expected value given in (K/26)	
		Theoretical values	Values used in practice
(1)	(2)	(3)	(4)
$x_aSR$ .....	$\hat{Y}_a/Y_a$	$\delta_a$	$\delta_a$
$y_aSR$ .....	$-x_a \hat{Y}_a / (Y_a)^2$	$-\delta_a \bar{X}_a / \bar{Y}_a$	$-\delta_a X_a / Y_a$
$x_{ac}$ .....	$\hat{Y}_a Z_c / (Z_c Y_a)$	$\delta_a$	$\delta_a$
$y_{ac}$ .....	$\frac{-x_a Z_c Y_a}{(Y_a)^2 Z_c}$	$-\delta_a \bar{X}_a / \bar{Y}_a$	$\frac{-\delta_a X_a}{Y_a}$
$z_c$ .....	$\sum_a \frac{\hat{Y}_a}{Y_a} \left\{ \frac{x_a}{Y_a} y_{ac} - x_{ac} \right\} \frac{Z_c}{Z_c^2}$	$\sum_a \frac{\bar{X}_a}{\bar{Y}_a} \delta_a \frac{\bar{Y}_{ac}}{Z_c} - \frac{1}{Z_c} \sum_a \delta_a \bar{X}_{ac}$	$\sum_a \frac{X_a}{Y_a} \frac{Y_{ac}}{Z_c} - \frac{X_c}{Z_c}$

In using these forms for variance estimation, the expected value terms (see K/26) are needed for substitution into (K/30) and (K/31). The values are typically not available, and the following approximations are made. Column 4 of table K-1 shows the values to be used as a result of the assumptions described by the following:

1. The expected values, given in column 3 of table K-1, are defined theoretically as expected values of unbiased estimates. In practice, the best available estimates of these values have been used; thus, for the ratios  $\bar{X}_a/\bar{Y}_a$ , the ratios  $X_a/Y_a$  are used where  $X_a$  and  $Y_a$  are the combined first- and second-stage estimates of the characteristic and of total population in the  $a$ -th age-sex-race category, as estimated from the current CPS sample.
2. When the ratios  $\bar{X}_a/\bar{Y}_a$  are substituted for  $\bar{X}_a/\bar{Y}_a$  in the evaluation of the partial derivative of  $x''$  with respect to  $z_c$ , we also substitute the combined first- and second-stage ratio estimates for the product of unbiased estimate terms and  $\gamma_a$ ; thus, in NSR we substitute

$$Y_{ac} = \gamma_a \bar{Y}_{ac}$$

$$X_c = \sum_a \gamma_a \bar{X}_{ac}$$

3. The term  $\gamma_a$  is the ratio of the U.S. independent population control for the  $a$ -th age-sex-race category to the expected value of the unbiased estimate of population for the

same category. The ratio is defined at (K/28), and, ideally all  $\gamma_a$  should be 1.0; however, we have evidence that for many of these categories, is consistently greater than 1.0. (See, for example, table V-3 where a related<sup>5</sup> ratio is given for each of the categories; also, table V-4 shows inverse of the ratio in table V-3, averaged over all age-sex-race categories for selected months.) For increased stability we make use of averages of these related ratios (that identified as  $\gamma'_a$ ), computed, as follows, for each of the age-sex-race categories. The numerator for each ratio is average over 7 months (the current month and the 6 preceding months) of the independent population control figure; the denominator is a similar 7-month average of first-stage ratio estimate of population. The values of  $\gamma'_a$  month  $t$  are computed by the formula

$$\gamma'_{at} = \frac{\sum_{m=t-6}^t \bar{Y}_{am}}{\sum_{m=t-6}^t \bar{Y}_{am}}$$

where the subscript  $m$  denotes a month,  $\bar{Y}_{am}$  is the first stage estimate of total population in the  $a$ -th age-sex-race category for month  $m$ , and  $\bar{Y}_{am}$  is the independent population control for the corresponding month and age-sex-race category. The values of  $\gamma'_a$  are computed and applied separately, by rotation group.

<sup>5</sup> Table V-3 shows the ratio of the independent population control to first-stage estimate of population.

## The Variance Estimator For The Linear Form

The variance estimators for the linear forms in (K/30) and (K/31) follow from the estimators given earlier.

Variance estimate in SR strata—We can write the linear form for the estimate for SR strata, (K/30), as

$$H(x''AC) = \sum_s^{SR} H(x''ACs) \quad (K/32)$$

and, from (K/4), as the sum of the two half samples in the  $r$ -th replication

$$H(x''AC) = \sum_s^{SR} \sum_a \gamma_a [(x_{asr1} + x_{asr2}) - \frac{X_a}{Y_a} (y_{asr1} + y_{asr2})]$$

An estimate of the variance for (K/32), based on  $R$  replications, is, from (K/5)

$$\text{Var}[H(x''AC)] = \frac{1}{R} \sum_r \sum_s^{SR} \left| \sum_a \gamma_a (\Delta x_{asr} - \frac{X_a}{Y_a} \Delta y_{asr}) \right|^2 \quad (K/33)$$

where

$$\Delta x_{asr} = x_{asr1} - x_{asr2}$$

and

$$\Delta y_{asr} = y_{asr1} - y_{asr2}$$

Variance estimate in NSR strata—For NSR strata, we can write the linear form at (K/31) as

$$\begin{aligned} J(x''AC) &= \sum_s^{NSR} J(x''ACs) \\ &= \sum_s^{NSR} \sum_a \gamma_a \sum_c |x_{acs} - \frac{\bar{X}_a}{\bar{Y}_a} y_{acs} \\ &\quad + \frac{\bar{X}_a}{\bar{Y}_a} \frac{\bar{Y}_{ac}}{\bar{Z}_c} z_{cs} - \frac{\bar{X}_{ac}}{\bar{Z}_c} z_{cs}| \end{aligned}$$

and, using the approximations in column (4) of table K-1,

$$\begin{aligned} J(x''AC) &\doteq \sum_s^{NSR} \left\{ \sum_a \gamma_a x_{as} - \sum_a \gamma_a \frac{X_a}{Y_a} y_{as} \right. \\ &\quad \left. + \sum_a \frac{X_a}{Y_a} \sum_c \frac{Y_{ac}}{\bar{Z}_c} z_{cs} - \sum_c \frac{X_c}{\bar{Z}_c} z_{cs} \right\} \quad (K/34) \end{aligned}$$

An expression for  $J(x''AC)$ , based on three units, as in (K/9), can be written by substituting the following sums into (K/34):

$$x_{as} = x_{as11} + x_{as21} + x_{as12}$$

$$y_{as} = y_{as11} + y_{as21} + y_{as12}$$

$$z_{cs} = z_{cs11} + z_{cs21} + z_{cs12}$$

The new expression for the linear form at (K/34) has a variance estimator as in (K/18), so that

$$\begin{aligned} \text{Var}[J(x''AC)] &\doteq \frac{1}{4} \sum_s^{NSR} \sum_a \gamma_a [(x_{as11} + x_{as12})/2 - x_{as21}] \\ &\quad - \sum_a \gamma_a \frac{X_a}{Y_a} [(y_{as11} + y_{as12})/2 - y_{as21}] \\ &\quad + \sum_c \frac{1}{\bar{Z}_c} \left( \sum_a \frac{X_a}{Y_a} y_{ac} - X_c \right) [(z_{cs11} + z_{cs12})/2 \\ &\quad - z_{cs21}]^2 \\ &\quad + \frac{21}{16} \sum_s^{NSR} \left\{ \left| \sum_a \gamma_a (x_{as11} \right. \right. \\ &\quad \left. \left. - \frac{X_a}{Y_a} y_{as11}) - \sum_c (X_c - \sum_a \frac{X_a}{Y_a} y_{ac}) \frac{z_{cs11}}{\bar{Z}_c} \right| \right. \\ &\quad \left. - \left| \sum_a \gamma_a (x_{as12} - \frac{X_a}{Y_a} y_{as12}) \right. \right. \\ &\quad \left. \left. - \sum_c (X_c - \sum_a \frac{X_a}{Y_a} y_{ac}) \frac{z_{cs12}}{\bar{Z}_c} \right| \right|^2 \quad (K/35) \end{aligned}$$

$$\begin{aligned} &\doteq Q \sum_s^{NSR} |\Delta \theta_1(x'', s)|^2 \\ &\quad + T \sum_s^{NSR} |\Delta \theta_2(x'', s)|^2 \quad (K/36) \end{aligned}$$

with values of 1/4 and 21/16 assigned to  $Q$  and  $T$ , respectively.

Variance estimate for total of SR and NSR strata—The variance of the estimate ( $V/1$ ) is given, approximately, by the sum of the two variance expressions (K/33) and (K/36).

## Variances and Components of Variances of Alternative Estimates

If we combine the variance estimator for SR strata from (K/33) with (K/35) for NSR strata and insert appropriate values for  $B$ ,  $P$ ,  $Q$ ,  $T$ ,  $W$ , and  $\gamma_a$  in the resulting expression, a variance estimator can be written as

TABLE K-2. Constants for Use in Expression (K/37) To Approximate the Variance of Selected Estimators

Estimator	Values to be assigned in expression (K/37)					
	$\delta_a$	B	P	Q	T	W
Total variance						
1. Unbiased estimate	1	1	—	1/4	21/16	
2. Unbiased estimate with first stage of ratio estimation	1	1	—	1/4	21/16	
3. Unbiased estimate with second stage of ratio estimation	$\delta_a$	1	1	1/4	21/16	
4. Unbiased estimate with first and second stages of ratio estimation	$\delta_a$	1	1	1/4	21/16	
Variance between strata						
5. Unbiased estimate	1	—	—	1/4	-3/16	
6. Unbiased estimate with first stage of ratio estimation	1	—	—	1/4	-3/16	
7. Unbiased estimate with second stage of ratio estimation	$\delta_a$	—	1	1/4	-3/16	
8. Unbiased estimate with first and second stages of ratio estimation	$\delta_a$	—	1	1/4	-3/16	

— Entry represents zero.

$$\begin{aligned}
 \text{Var}(x) = & \frac{B}{R} \sum_s \sum_a \left| \sum_a \gamma_a (\Delta x_{asr} - P \frac{X_a}{Y_a} \Delta y_{asr}) \right|^2 \\
 & + Q \sum_s \left| \sum_a \gamma_a [(x_{as11} + x_{as12})/2 - x_{as21}] \right. \\
 & \quad \left. - P \sum_a \gamma_a \frac{X_a}{Y_a} [(y_{as11} + y_{as12})/2 - y_{as21}] \right. \\
 & \quad \left. - W \sum_c \frac{1}{Z_c} (X_c - P \sum_a \frac{X_a}{Y_a} Y_{ac}) [(z_{cs11} + z_{cs12})/2 \right. \\
 & \quad \left. - z_{cs21}] \right|^2 \\
 & + T \sum_s \left\{ \left| \sum_a \gamma_a (x_{as11} \right. \right. \\
 & \quad \left. \left. - P \frac{X_a}{Y_a} y_{as11}) - W \sum_c (X_c - P \sum_a \frac{X_a}{Y_a} Y_{ac}) \frac{z_{cs11}}{Z_c} \right| \right. \\
 & \quad \left. - \left| \sum_a \gamma_a (x_{as12} - P \frac{X_a}{Y_a} y_{as12}) \right. \right. \\
 & \quad \left. \left. - W \sum_c (X_c - P \sum_a \frac{X_a}{Y_a} Y_{ac}) \frac{z_{cs12}}{Z_c} \right| \right|^2 \quad (K/37)
 \end{aligned}$$

By proper choice of the constants summarized in table K-2, expression (K/37) will state approximations to the variance selected estimators for the combined A- and C-design sample.

The term on the right of (K/37) that is multiplied by B involves the summation over the 54 Keyfitz clusters in SR strata. A term of this same form provides an estimator for the total within-stratum variance from SR and NSR strata, when applied, in addition to the 54 Keyfitz clusters defined for this purpose (by combining NSR strata). As estimated, the within-PSU variance consists of the variance among USU's within PSU's, plus the simple response variance (ch. VII, p. 78); it does not represent the correlated component of response variance.

Subtracting the estimated total within-PSU variance, computed by this process, from the estimated total variance in (K/37) produces a difference consisting of the variance between strata and a residual. The residual variance<sup>8</sup> comprises the variance among PSU's within strata and a portion of the correlated component of response variance in NSR strata, reflected in (K/37).

The variance between strata comprises the variance among stratum totals within the pairs of strata (the NSR Keyfitz clusters). The constants in table K-2 show how to estimate the variance between strata.

This process can be repeated for the four alternative estimators, indicated in table K-2, to estimate the total variance, allocate it, approximately, to the within-PSU variance (separately for SR and NSR strata, if desired), the variance between strata and the residual (or between-PSU) variance.

<sup>8</sup> The residual variance is referred to as the between-PSU variance in VIII-3 and VIII-4.



## Appendix L.

# RELIABILITY STATEMENT FOR A CPS REPORT: AN ILLUSTRATION

### INTRODUCTION

The illustration of a reliability statement in this appendix is extracted from a report on educational attainment [47]. This report is based on sample data produced from the Current Population Survey in March 1975 and from decennial census data.

To conserve space, selected portions of the original statement have been omitted or modified. Tables of standard errors for Blacks and other races and an illustrative example of the standard error of a percent change have been omitted; the standard error table for estimated numbers (table L-1) has been somewhat abbreviated, and the illustrative examples have been changed to be consistent with these emendations.

In this appendix, the references to tables other than tables L-1 and L-2 are to the original Census Bureau report on educational attainment.

### SOURCE AND RELIABILITY OF THE ESTIMATES FOR EDUCATIONAL ATTAINMENT: MARCH 1975

#### Source of Data

The estimates in this report for the years 1970 and 1975 are based on data collected in March of 1970 and 1975 in the Current Population Survey (CPS) of the Bureau of the Census. Estimates for 1950 and 1960 are based on data obtained from the decennial censuses of population for those years. Brief descriptions of these sources and procedures of data collection are presented below.

**Current Population Survey (CPS)**—The following table provides a description of some aspects of the Current Population Survey Design.

The estimating procedure used for CPS data involves the inflation of weighted sample results to independent estimates of the civilian noninstitutional population of the United States, by age, race, and sex. For 1972 through 1975, these independent estimates were based on statistics from the 1970 Census of Population; statistics on births, deaths, immigration, and emigration; and statistics on the strength of the Armed Forces. For data collected in the Current Population Surveys in the years 1960 through 1971, the independent estimates used were based on statistics from the 1960 Census of Population.

**Decennial Census of Population**—Decennial census data in this report are based on complete counts or on the samples associated with the census. Descriptions of the 3-1/3-percent sample of the 1950 Census and the 5-percent sample of the 1960 Census can be found in the appropriate census publications.

#### Reliability of the Estimates

Since the CPS estimates in this report are based on a sample, they may differ somewhat from the figures that would have been obtained if a complete census had been taken using the same schedules, instructions, and enumerators. There are two types of errors possible in an estimate based on a sample survey—sampling and nonsampling. For estimates in this report, indications of the magnitude of sampling error are provided, but the extent of nonsampling error is unknown. Consequently, particular care should be exercised in the interpretation of figures based on a relatively small number of cases or on small differences between estimates.

**Nonsampling Variability**—As in any survey work, the results are subject to errors of response and nonreporting in addition to sampling variability. Nonsampling errors can be attributed to many sources, e.g., inability to obtain information about all cases in the sample, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, inability to recall information, mistakes made in collection such as in recording or coding the data, mistakes made in processing the data, and mistakes made in estimating values for missing data. To date, emphasis has been placed on identification and control of nonsampling errors and not on providing estimates of magnitude of such errors in the data.

**Sampling Variability**—The reliability of an estimate is described in terms of standard errors that are primarily measures of sampling variability, i.e., the variations that occur by chance, because a sample, rather than the whole of the population, is surveyed. As calculated for this report, the standard error also partially measures the effect of certain response and enumeration errors, but it does not measure, as such, any systematic biases in the data. The chances are about 68 out of 100 that an estimate from the sample would differ from a complete census figure by less than the standard error. The chances are about 90 out of 100 that this difference would be less than 1.6 times the

Time period	Number of sample areas	Number of counties <sup>1</sup>	Households eligible		Sample units visited, not interviewed <sup>2</sup>
			Interviewed	Not interviewed	
March 1975.....	461	923	45,000	2,000	8,000
March 1970.....	449	863	48,000	2,000	8,500

<sup>1</sup>The number of counties and independent cities included in the sample areas. These areas were chosen to provide coverage in each State and the District of Columbia.

<sup>2</sup>Sample units visited, but found to be vacant or otherwise not to be interviewed.

standard error, and the chances are about 95 out of 100 that the difference would be less than twice the standard error.

All statements of comparison involving Bureau of the Census data appearing in the text are significant at a 1.6-standard error level or better, and most are significant at a level of more than 2.0 standard errors. This means that, for most differences cited in the text, the estimated difference is greater than twice the standard error of the difference. Statements of comparison qualified in some way (e.g., by the use of the phrase "some evidence") have a level of significance between 1.6 and 2.0 standard errors.

Percent distributions are shown in the report only when the estimated base of the percentage is greater than 75,000. Because of the large standard errors involved, there is little chance that percentages would reveal useful information when computed on a smaller base. Estimated totals are shown, however, even though the relative standard errors of these totals are larger than those for the corresponding percentages. These smaller estimates are provided primarily to permit such combinations of the categories as serve each user's needs.

**Reliability of an Estimated Percentage**—The reliability of an estimated percentage, computed by using sample data for both numerator and denominator, depends upon both the size of the percentage and the size of the total upon which the percentage is based. Estimated percentages are relatively more reliable than the corresponding estimates of the numerators of the percentages, particularly if the percentages are 50 percent or more.

#### Standard Error Tables and Their Use

**Standard Errors for Data Based on the CPS**—Instead of providing individual standard error tables for each characteristic

of interest, generalized standard error tables for estimated numbers and estimated percentages, by race, are provided in tab L-1 and L-2.

The figures presented in these tables provide approximate to standard errors of various CPS estimates shown in this report. In all the standard error tables, standard errors for intermediate values not shown may be approximated by interpolation. In order to derive standard errors that would be applicable to a wide variety of items and could be prepared at a moderate cost, a number of approximations were required. In addition, where two or more items have nearly equal standard errors, such as total population and white population, one table is used to represent them. As a result, the tables of standard errors provide an indication of order of magnitude of the standard errors rather than the precise standard error for any specific item.

**Standard Errors for Data Based on the Decennial Censuses**—Sampling errors of all data from the 3-1/3- and 5-percent samples of the decennial censuses shown in this report are small enough to be disregarded. However, the standard error may be found in the appropriate census volumes.

**Illustration of the Use of Tables of Standard Errors**—Table 1 of this report shows that in March of 1975 there were 539,000 white males 20 or 21 years old who had completed one year of college. Table 1 also shows that there were 3,268,000 total white males in that age group. Table L-1 shows the standard error of the 539,000 estimate to be approximately 31,000.<sup>1</sup> The chances are 68 out of 100 that the estimate will have been a figure differing from a complete census figure by less than 31,000. The chances are 95 out of 100 that the

<sup>1</sup> Determined by interpolation and rounding to the same accuracy as the

TABLE L-1. Standard Errors for Estimated Numbers of Persons Enrolled in School, CPS Estimates for Total or White Population

(68 chances out of 100; thousands)

Estimated number of persons	Total persons in age group									
	100	250	500	1,000	2,500	5,000	10,000	25,000	50,000	100,000
10	4.4	4.6	4.6	4.6	4.6	4.7	4.7	4.7	4.7	
30	6.8	7.6	7.8	7.9	8.0	8.0	8.1	8.1	8.1	
50	7.4	9.3	9.9	10.2	10.3	10.4	10.4	10.4	10.4	
100	—	11.4	13.2	14.0	14.4	14.6	14.7	14.7	14.7	
300	(X)	—	16.1	21.3	23.9	24.7	25.1	25.4	25.4	
500	(X)	(X)	—	23.3	29.5	31.2	32.1	32.6	32.8	
1,000	(X)	(X)	(X)	—	36.1	41.7	44.2	45.6	46.1	
3,000	(X)	(X)	(X)	(X)	—	51.0	67.5	75.7	78.2	
5,000	(X)	(X)	(X)	(X)	(X)	—	73.7	93.2	98.8	
10,000	(X)	(X)	(X)	(X)	(X)	(X)	—	114.1	131.8	
30,000	(X)	(X)	(X)	(X)	(X)	(X)	(X)	—	161.4	
50,000	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	—	
100,000	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	

— Entry represents zero.  
X Not applicable.

mate would have been a figure differing from a complete census figure by less than 62,000 (twice the standard error).

Table 6 of this report shows that in March 1975 there were 4,776,000 25- to 64-year-old female service workers in the population, and table C shows that about 52.4 percent of these females had completed at least four years of high school. Table L-2 shows the standard error of 52.4 percent to be approximately 1.0.<sup>1</sup> Consequently, chances are 68 out of 100 that the estimated 52.4 percent would be within 1.0 percentage points of a complete census figure. Chances are 95 out of 100 that the estimate would be within 2.0 percentage points of a complete census figure, i.e., the 95-percent confidence interval would be from 50.4 to 54.4.

### Standard Error of a Difference

For a difference between two sample estimates, the standard error is approximately equal to the square root of the sum of the squared standard errors of the estimates; the estimates can be of numbers, percents, ratios, medians, etc. This will represent the actual standard error quite accurately for the difference between two estimates of the same characteristic in two different areas or for the difference between separate and uncorrelated characteristics in the same area. If, however, there is a high positive correlation between the two characteristics, the formula will overestimate the true standard error.

**Illustration of the Computation of the Standard Error of a Difference Between Estimated Percentages**—Table 6 of this report shows that there were 2,793,000 25- to 64-year-old male service workers in the population in March of 1975, and table C shows that 60.4 percent of these males had completed at least four years of high school. The apparent difference between the 52.4-percent figure for females of the same occupation and the percentage for males is 8.0 percent. The standard error of 52.4 percent is 1.0 percent, as shown previously. Table L-2 shows the

standard error of 60.4 percent on a base of 2,793,000 to be approximately 1.3 percent.<sup>1</sup> To get the standard error of the estimated difference use the following formula:

$$\sigma_{(x-y)} = \sqrt{\sigma_x^2 + \sigma_y^2}$$

Therefore, the standard error of the difference of 8.0 percent is about

$$1.6 = \sqrt{(1.3)^2 + (1.0)^2}$$

This means that the chances are 68 out of 100 that the estimated difference, based on the sample estimates, would vary from the difference derived using complete census figures by less than 1.6 percent. The 68-percent confidence interval around the 8.0-percent difference is from 6.4 to 9.6, i.e.,  $8.0 \pm 1.6$ . A conclusion that the average estimate of the difference derived from all possible samples of the same size and design lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. The 95-percent confidence interval is 4.8 to 11.2. Thus, we can conclude with 95-percent confidence that the percentage of service workers, 25 to 64 years old having completed at least four years of high school, is actually greater for males than for females.

### Standard Error of a Median

The standard error of an estimated median depends upon the form as well as on the size of the distribution from which the median is determined. An approximate method for measuring the reliability of a median is to determine an interval about the estimated median, such that there is a stated degree of confidence that the median, based on a complete census, lies within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

TABLE L-2 Standard Errors of Estimated Percentages, CPS Estimates for Total or White Population  
(68 chances out of 100)

Base of percentage (thousands)	Estimated Percentage				
	2 or 98	5 or 95	10 or 90	25 or 75	50
100	2.0	3.1	4.3	6.2	7.2
250	1.3	2.0	2.8	4.0	4.5
500	.9	1.4	1.9	2.8	3.2
1,000	.6	1.0	1.4	2.0	2.3
2,500	.4	.6	.9	1.2	1.4
5,000	.3	.4	.6	.9	1.0
10,000	.2	.3	.4	.6	.7
25,000	.1	.2	.3	.4	.5
50,000	.1	.1	.2	.3	.3
100,000	.1	.1	.1	.2	.2

<sup>1</sup> Determined by interpolation and rounding to the same accuracy as in the table.

1. Determine the standard error on a 50-percent characteristic, using the appropriate base.
2. Add to and subtract from 50 percent the standard error determined in step 1.
3. Using the distribution of the characteristic, calculate the confidence interval corresponding to the two points established in step 2.

A two-standard error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1.

**Illustration of the Computation of a Confidence Interval for a Median**—Table 3 shows that the median number of school years completed by white males, 65 years old and over, in the West to be 10.9 years in March of 1975. Table 3 also indicates the size, or base, of the distribution from which this median was determined is 1,286,000.

1. Table L-2 shows that the standard error of 50 percent on a base of 1,286,000 is about 2.1 percent.
2. To obtain a two-standard error confidence interval on the estimated median, initially add to and subtract from 50 percent twice the standard error found in step 1. This yields percentage limits of 45.8 and 54.2.
3. From table 3, it can be seen that 40.9 percent had completed less than 9 years of schooling and 14.2 percent had more than 9 and less than 12 years of schooling. By linear interpolation, the lower limit on the estimate is found to be about

$$9 + (12 - 9) \left( \frac{45.8 - 40.9}{14.2} \right) = 10.0$$

Similarly, the upper limit may be found by linear interpolation to be about

$$9 + (12 - 9) \left( \frac{54.2 - 40.9}{14.2} \right) = 11.8$$

Thus, the 95-percent confidence interval around the 10.9 ranges from 10.0 to 11.8.<sup>2</sup>

### Standard Error of a Ratio

Table A shows data for the percent change between The standard error of a percent change can be obtained from the standard error of a ratio  $\frac{x}{y}$ . For the data in this report, the following formula for the standard error of  $\frac{x}{y}$  may be used:

$$\sigma \left( \frac{x}{y} \right) = \left( \frac{x}{y} \right) \sqrt{\left( \frac{\sigma_x}{x} \right)^2 + \left( \frac{\sigma_y}{y} \right)^2}$$

In this formula, the ratio  $\frac{x}{y}$  can be a ratio of two estimates, a ratio of percents or a ratio of medians, where  $\sigma_x$  represent the standard errors of  $x$  and  $y$ . For the data in port, the estimates  $x$  and  $y$  are assumed to be uncorrelated; they are positively correlated, this formula will give an estimate.

**Illustration of the Computation of the Standard Error of Percent Change**—In table A, the percent change is calculated by the formula  $\frac{x-y}{y}$ , where  $x$  is the estimate for 1975 and  $y$  is the estimate for 1950. The standard error of  $\frac{x-y}{y}$  is the standard error of  $\frac{x}{y}$ . Since  $y$  represents the estimate for 1950 census data,  $\sigma_y$  can be assumed to be zero. The formula for the standard error of the percent change,  $\frac{x-y}{y}$ , simplifies to

$$\sigma \left( \frac{x-y}{y} \right) = \left( \frac{1}{y} \right) \sigma_x$$

<sup>2</sup> Although the confidence interval is symmetrical in this illustration, it is not necessarily true for medians.

## Appendix M.

# LOCATION AND DESCRIPTION OF SAMPLE AREAS

## INTRODUCTION

A number of maps and lists showing the location and geographic composition of sample PSU's associated with the CPS are available as separate publications.<sup>1</sup> This appendix describes the information shown on each of the documents pertaining to the identity of PSU's in the CPS national and CETA State supplement samples.

## SOURCES

### Two Four-Color Maps of the United States

Each of these maps (forms 11-5 and 11-6) covers the 50 States, is printed in four colors at scale 1 to 5,000,000 (approx. 30 X 40 inches), and shows sufficient detail to identify most of the individual sample PSU's. Figures M-1 and M-2 of this appendix are black and white reductions of these maps to page size; since these figures are greatly reduced in scale and do not distinguish the four colors, only the general location of the areas covered by the sample PSU's is possible.

The following information is displayed on the two four-color maps:

Form 11-5—Sample PSU's, 461-Area Current Population Survey National Sample, March 1973 [63].

1. The sample PSU's in the 461-area national sample are separately identified as members of the 376-area A-design and the 266-area C-design national samples.
2. Self-representing and nonself-representing PSU's are separately identified for the A- and C-design national samples.
3. The PSU's are identified by a PSU number. There are more than 461 numbers shown, since parts of some PSU's are identified separately for administrative purposes.
4. The location of the 12 regional offices and the regional office boundaries are shown.
5. Replacement PSU's for the national samples are identified. These PSU's enter or leave the sample at some time during the period of August 1972 through August 1982.
6. The PSU replacement schedule appears on the map and is reproduced as in table M-1 of this appendix. All PSU's shown in table M-1 are separately identified on form 11-5. The PSU replacements scheduled beyond August 1982 are not shown. (See ch. III, p. 25.)

Form 11-6—Sample PSU's, 461-area sample and State supplements, August 1976 [64].<sup>2</sup>

- 1 Sample PSU's are separately identified as members of the 461-area national sample and of the supplemental samples selected to provide State estimates.

<sup>1</sup> These materials may be ordered at cost; use the order form appearing at the end of this publication.

<sup>2</sup> Two errors in this map have come to our attention: McCurtain County, Okla., should not have a number, and PSU 357 (Carter-Shannon, Mo.) should not be indicated as a PSU involved in replacement.

2. The PSU's treated as SR and NSR in making the State estimates are separately identified. The SR and NSR classification of PSU's in the national sample is not necessarily the same when making State estimates.
3. The PSU's are identified by a PSU number. In some cases, the numbers reflect parts of PSU's separately identified for administrative purposes.
4. The location of the 12 regional offices and the regional office boundaries are shown.
5. Replacement PSU's for the State samples are identified. These PSU's will enter or leave the State samples some time during the period of August 1976 through August 1982.
6. The PSU replacement schedule appears on the map and is reproduced as in table M-2 of this appendix. All PSU's shown in table M-2 are separately identified on form 11-6. Replacements scheduled beyond August 1982 are not shown.

### One Three-Color Map Covering Selected PSU's of the Boston and New York Regional Offices

Form 11-1A—Data Collection Center Coverage and Current Program PSU's, January 1973 [45].

This three-color map covers a portion of the PSU's assigned to the Boston Regional Office and all of the PSU's assigned to the New York Regional Office. The map is needed to describe sample PSU's in the Boston office that are considered in four areas, defined as follows (these four areas, A, B, C, and D, are separately identified on both maps, forms 11-5 and 11-6):

Area A (in Connecticut)—all of the State, except Windham County.

Area B (in Massachusetts)—all of the State, except Barnstable, Dukes, and Nantucket Counties.

Area C (in New Hampshire)—Hillsborough, Merrimack, and Rockingham Counties.

Area D (in Rhode Island)—all of the State.

Form 11-1A is an older map that identifies regional offices as "Data Collection Centers," a title no longer used. The map scale of 1 inch to 10 miles makes the map 30 X 40 inches. The three colors used on form 11-1A are compatible with the color code used on form 11-5.

### List of PSU's in Current Survey Samples

Form 11-4—PSU's included in current survey samples.

This publication [59] lists the counties, independent cities (and towns in some New England States), that make up each sample PSU in any of the current survey programs of the Census Bureau including the CPS. Where the primary sampling units have been partitioned for administrative purposes, the parts are separately identified.

Separate parts of this publication list the PSU's (in order, by PSU number) by the 12 regional offices and, a second time, in

one numerically sequenced list. The PSU's are identified as members of each of the current survey samples. PSU replacement schedules for all Census Bureau current survey programs are provided.

### PSU Replacement Schedules

**Tables M-1 and M-2**—These tables appear also on forms 11-5 and 11-6, respectively, and show how sample representation is maintained when PSU's are replaced to minimize use of the same USU in more than one sample survey. (This problem is discussed in ch. III, p. 25.) The following two examples illustrate situations described in tables M-1 and M-2.

**Table M-1**—This table shows that an A-design stratum, originally represented by PSU 914 (Newton, Tex.), will have its representation shifted to PSU 554 (Grant, Ark.). This replacement occurs in December 1979 when sample A42 is first introduced; PSU 554 will remain in the national sample through August 1982.

These actions are also shown in table M-2, where the impact of the shift on the State samples is also recorded. Arkansas is one of the States for which the national sample is inadequate to provide State estimates with the reliability desired. (See. ch. VI,

p. 68.) Analysis of the national sample has shown that additional sample USU's are required for the State estimates. Arkansas, prior to December 1979, before Grant, Ark., enters the national sample. Accordingly, table M-2 shows that a supplemental sample in PSU 531 (Union, Ark.) is interviewed until December 1979 when PSU 554 (Grant, Ark.) enters the sample for the State. Because, even without PSU 914 (Newton, Tex.), the national sample in Texas is considered adequate to provide reliable estimates for that State, the sample for Texas is not otherwise affected by this shift.

**Table M-2**—This table shows a shift in the national sample between PSU's 865 (Torrance, N. Mex.) and 825 (Millard, Utah) beginning with sample A39 in December 1977. In this case, New Mexico and Utah require their national samples to be supplemented by additional PSU's for State estimates, and necessary to maintain the sample size in each of these States when the national sample PSU's are changed.

Table M-2 shows that a D sample is provided for Utah by 785 (Sanpete, Utah) until December 1977 when PSU (Millard, Utah) enters the sample. The replacement for the sample in PSU 865 (Torrance, N. Mex.), needed for New Mexico, is found also to be Torrance, N. Mex. As Torrance is not enough to supply the sample for the entire decade, it will be necessary in this situation to use some USU's in this PSU a second time for other surveys.

TABLE M-1. PSU Replacement Schedule for the CPS National Sample: August 1972-August 1982

Current PSU			Replacement PSU			
No.	County and State	Sample design	No.	County and State	Starting sample	Date of change
356	Hickory-St. Clair, Mo.	A	357	Carter-Shannon, Mo.	A36	Dec. 1
390	Roseau, Minn.	A	460	Schoolcraft, Mich.	A38	Apr. 1
394	Wells, N. Dak.	C	395	Nelson, N. Dak.	C18	Aug. 1
464	Clark-Hamlin, S. Dak.	A and C	379	Custer, S. Dak.	A41	Apr. 1
			384	Sully, S. Dak.	C25	Apr. 1
546	Baylor, Tex.	A	547	Dallam, Tex.	A41	Apr. 1
556	Gaines, Tex.	A	566	Wheeler, Tex.	A38	Apr. 1
589	Wilson, Tex.	A	629	Karnes, Tex.	A34	Aug. 1
655	Macon-Trousdale, Tenn.	A	539	DeKalb, Tenn.	A44	Apr. 1
657	McClain, Okla.	A	545	Montague, Tex.	A35	Apr. 1
687	Lampasas-Mills, Tex.	A	989	Hamilton, Tex.	A40	Aug. 1
806	Klickitat, Wash.	A	846	Morrow, Oreg.	A42	Dec. 1
851	Washington, Idaho	A	807	Wallowa, Oreg.	A37	Aug. 1
807	Wallowa, Oreg.	A	851	Washington, Idaho	A46	Aug. 1
865	Torrance, N. Mex.	A	825	Millard, Utah	A39	Dec. 1
869	Del Norte, Calif.	C	883	Wheeler, Oreg.	C25	Apr. 1
883	Wheeler, Oreg.	C	869	Del Norte, Calif.	C28	Apr. 1
873	Phillips, Mont.	C	823	Chouteau, Mont.	C18	Aug. 1
823	Chouteau, Mont.	C	843	Blaine, Mont.	C23	Dec. 1
843	Blaine, Mont.	C	873	Phillips, Mont.	C30	Aug. 1
912	Parmer, Tex.	A	588	Swisher, Tex.	A35	Apr. 1
914	Newton, Tex.	A	554	Grant, Ark.	A42	Dec. 1

TABLE M-2. PSU Replacement Schedule for the 461-Area Design and Supplemental Sample for State Estimates, CETA: August 1976-August 1982

Current PSU			Replacement PSU			
No.	County and State	Sample design	No.	County and State	Starting sample	Date of change
294	Adams. N Dak	D	293	Slope. N Dak	D46	Aug 1982
390	Roseau. Minn	A	460	Schoolcraft. Mich.	A38	Apr 1977
			346	Todd. Minn	D38	Apr 1977
464	Clark-Hamlin. S Dak	A and C	379	Custer. S Dak	A41	Apr 1979
			384	Sully. S Dak	C25	Apr 1979
531	Union. Ark	D	554	Grant. Ark.	A42	Dec 1979
914	Newton. Tex	A				
465	Tripp. S Dak	D	468	Jones. S Dak	D38	Apr 1977
468	Jones. S Dak	D	465	Tripp. S Dak	D41	Apr 1979
546	Baylor. Tex	A	547	Dallam. Tex	A41	Apr 1979
556	Gaines. Tex	A	566	Wheeler. Tex	A38	Apr 1977
655	Macon-Trousdale. Tenn	A	539	DeKalb. Tenn	A44	Apr 1981
687	Lampasas-Mills. Tex	A	989	Hamilton. Tex	A40	Aug 1978
724	Lander. Nev	D	725	Eureka. Nev	D40	Aug 1978
725	Eureka. Nev	D	724	Lander. Nev	D44	Apr 1981
785	Sanpete. Utah	D	825	Millard. Utah	A39	Dec 1977
865	Torrance. N Mex	A	865	Torrance. N Mex	D39	Dec 1977
869	Del Norte. Calif	C	883	Wheeler. Oreg	C25	Apr 1979
883	Wheeler. Oreg	C	869	Del Norte. Calif	C28	Apr 1981
806	Klickitat. Wash	A	846	Morrow. Oreg	A42	Dec 1979
823	Chouteau. Mont	C	843	Blaine. Mont.	C23	Dec 1977
843	Blaine. Mont	C	873	Phillips. Mont	C30	Aug 1982
773	Fremont. Idaho	D	851	Washington. Idaho	A46	Aug 1982
807	Wallowa. Oreg	A				
852	Judith Basin. Mont	D	763	Garfield. Mont	D38	Apr 1977
763	Garfield. Mont	D	852	Judith Basin. Mont	D44	Apr 1981

FIGURE M-1. 461-Area PSU's for the Current Population Survey National Sample  
(Form 11-5)

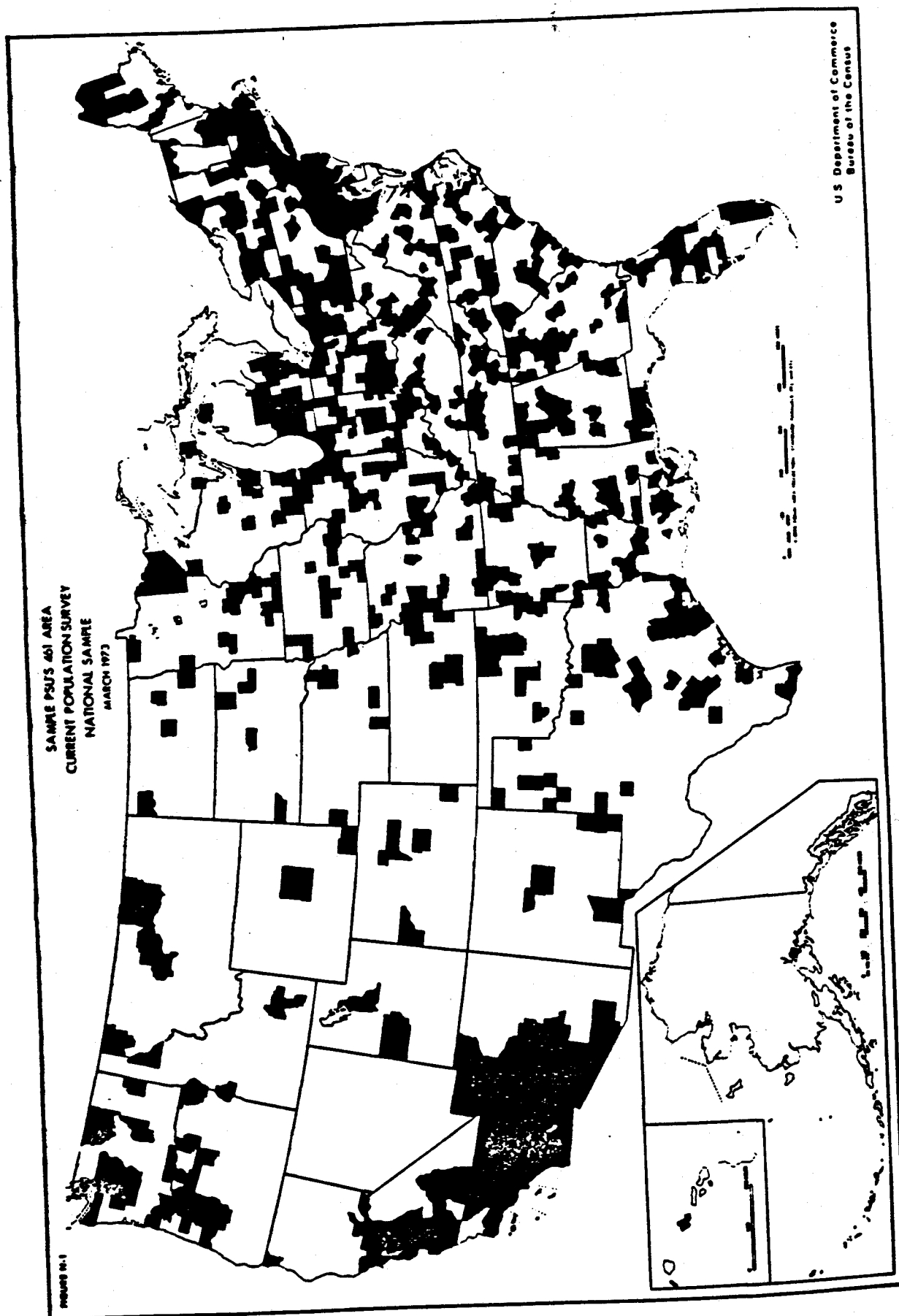
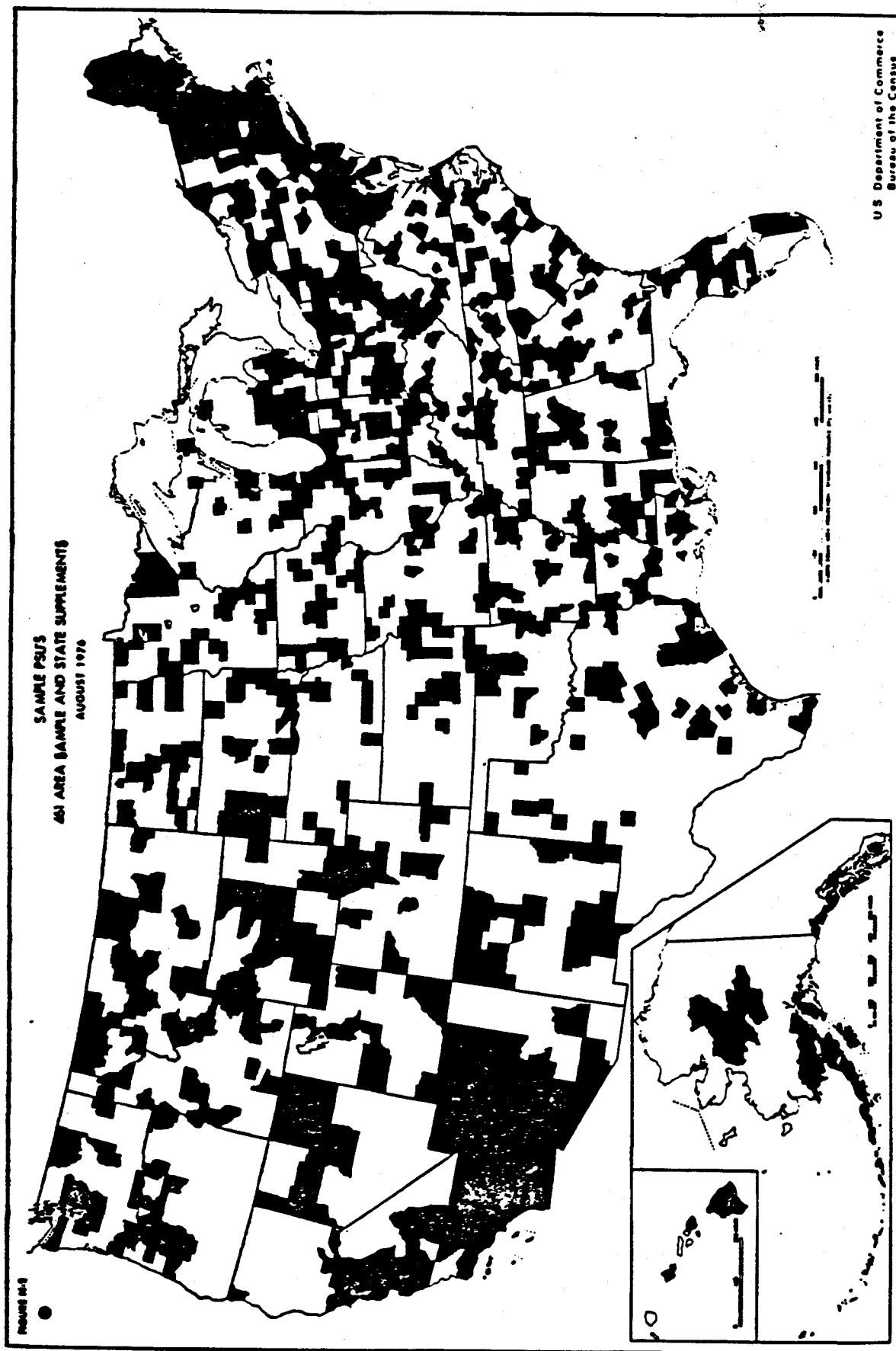




FIGURE M-2. 461 PSU's for the Area Sample and State Supplements: August 1976  
(Form 11-6)



# SUBJECT INDEX

## Note:

Definitions of terms are given on page numbers in *italics*.

## Accuracy, 77, 88

## Address(es)

- Coverage check, 46
- ED, *see* Enumeration district
- Listing sheet (form 11-211A), 33, 34, 46
- Non-take-all (NTA), 33, 126
- Register, 16, 122, 123
- Sample, 3, 16, 17
- Segment, 17, 30
- String of usable, 125
- Take-all (TA), 126
- USU, 17, 19, 125

## A-design, *see* Sample

## Admissible selection pattern, *see* Selection pattern

## Adjustment, *see* Seasonal adjustment

## American Statistical Association Census Advisory Committee, *see* Committees

## Area

- ED accumulation worksheet (form 8C-59), 118
- Sample segment, 16, 30, 115
- Segment listing sheet (form 11-2121), 31
- Urbanized, 18
- USU, 16

## Armed Forces, 33, 65, 142

## Associate Director

- Demographic Fields, 27
- Statistical Standards and Methodology, 27

## Averaging, *see* Estimates, and Variance

## Bias, 77, 81, 86

- Concept, 81
- Effect on estimated change, 88
- Mean square error, 88
- Net, 82, 85, 88
- Potential causes, 82
- Evaluation, 86
- Rare events universe (REU), 140
- Selection of USU's, 30
- See also* Ratio estimation, and Rotation group

## Blocks, 16, 115

## Book and Dexterity Survey, 94

## Building permits, *see* Permits

## Bureau of Labor Statistics, 3, 27, 54, 63, 94 Census Labor Force Committee, *see* Committees

## Causey, Beverley D., 92, 154

## C. B. U. R. categories, *see* Enumeration district

## C-design, *see* Sample

## Cen-Sup (Census supplemental)

- Sample, 17
- Segment, 18, 28
- USU, 18, 56

## Census

- Advisory Committee, American Statistical Association, *see* Committees
- Conventional (nonmail), 16
- CPS Match Study, 65
- Enumerative Check, 2
- Evaluation Program, 1970, (E6), 18
- Mail procedure, 16
- Population, 1970, 4, 16
- Prelist, 16
- Supplemental, *see* Cen-Sup
- X-11 method, *see* Seasonal adjustment

## CETA program, 68

*See also* Rotation system

## Changes

- Gross, 21, 85
- Labor force status, 87
- Month-to-month, 21, 88, 96, 98

## Chunks, *see* Blocks and Segment

## Coding

- Industry and occupation, 28, 51
- Quality control, 52

## Color, 57

## Committees

- American Statistical Association, Census Advisory Committee, 100
- Bureau of Labor Statistics-Census Labor Force Committee, 27
- CPS Policy Committee, 27
- Interagency Technical Committee on Labor Force Statistics, 27
- President's Committee to Appraise Employment and Unemployment Statistics (Gordon Committee), 3

## Components of variance, *see* Variance

## Composite estimate, *see* Estimate

## Comprehensive Employment and Training Act of 1973, *see* CETA

## Computer

- Controlled selection of PSU's, 106
- Data acceptance checks, 53
- Edit, 45
- Edit and recode program, 53
- Selection of sample, 111, 122
- Services Division, 28
- Use of, 3

## Conditioning of responses, 47, 49

## Confidence intervals, 88, 94

## Confidentiality requirements, 36, 100 *See also* Security

## Construction Statistics Division, 28

## Content

- Reinterview, 47
- See also* Error

## Control

- Card (form CPS 260), 33, 37, 38
- Data collection, 45
- Quality, 45, 51, 147
- Selection of PSU's, 12, 91, 92, 105, 150

## Cost functions, 50

## Coverage, 82

- Check, household, 47
- Error, 48
- Ratio, 61, 62, 65, 82
- See also* Error and Permit

## CPS, *see* Current Population Survey

## Crime Survey, National, 100, 127, 135

## Current Population Survey (CPS)

- Basic and reimbursable programs, 44, 145
- Census Bureau Divisions responsible for, 27
- Chronological history, 2
- Design, National sample, 6
  - 68-area, 2
  - 230-area, 3
  - 235-area, 13, 14
  - 266-area, 13, 14
  - 330-area, 3
  - 333-area, 3
  - 357-area, 12, 14
  - 376-area, 12, 13, 14
  - 449-area, 4
  - 461-area, 1, 4, 6, 14
- Introduction of new design, 22, 141
- Office Manual, 149
- Policy Committee, *see* Committees
- Program supervisor, 47, 147
- Questionnaire (form CPS-1), 33, 39, 40, 41
- Research, 4
- Specifications, 6
- See also* Variance

## Cut groups, 138

## Data

- Acceptance checks, 53
- Collection centers, 163
- Preparation Division, Jeffersonville, Ind., 28, 51
- Processing Division, 51
- Production cycle, 27, 53
- Tapes, public-use, 1

## D-design, *see* Sample

## Demographic Fields, Associate Director for

- Demographic program coordinator, 147
- Demographic supplement, annual, 65, 145, 146
- Demographic Surveys Division, 27
- Design  
Effect, 69, 70, 71, 93  
Requirements for CETA, 68  
See also Current Population Survey and Sample
- Divisions associated with CPS operations, 27
- Document sensing, 2  
See also FOSDIC
- Dwelling places, see Group quarters, Housing units
- ED, see Enumeration district
- Edit  
Computer edit and recode program, 53  
Qualified (QE) system, 46, 52  
Questionnaire, 45, 50, 149  
Regional office, 45  
See also Computer, Income, and Interviewer
- Educational Attainment, March 1975, 159
- Efficiency  
Increase, 1943-1975, 5  
Interviewer efficiency ratio, 50
- Employed persons, see Labor force concepts
- Employment and Earnings, 1, 63
- Enumeration district (ED), 16  
Accumulation worksheet (form BC-59), 118  
Address, 16, 19, 122  
Address register, 16  
Area, 16, 19  
Geographic categories, C, B, U, R, 18, 111  
List (or prelist), see Address ED  
Map, 115  
Measure checklist (form BC-2505), 119  
Measure of size, 18, 111  
Tape address register (TAR), 16
- Error(s)  
Content, 48, 83  
Coverage, 48, 61, 62, 65, 82, 88  
Net, 48, 82  
See also Bias  
Mean square, 60, 77, 88  
Misclassification, 83  
Nonsampling, 81, 88  
Processing, 85, 87, 90  
Response, 83, 86  
Sampling, methods of calculation, 90  
Standard, 77, 88, 90
- Estimate(s)  
Adjustment, noninterview  
National, 56, 91  
State, 74  
Annual averages, 64, 70
- Estimates—Continued  
Composite, 3, 21, 62, 64  
Bias, 86  
Change, 92, 96, 98  
Level, 92, 95, 97  
State, 75  
See also Variance  
Consistent, 55, 60  
Discontinuity in, 22  
Housing units, 66  
Inflation-deflation, 59  
Interval, see Confidence intervals  
Linear form of, 92, 155  
Monthly level, 88, 95, 97  
Population, independent, 4, 59, 82  
Precision of variance, 90  
Procedure for States, 64  
Annual averages  
1973, 64, 68  
1974 and 1975, 65, 68, 73  
1976, 68, 73  
Ratio, 2, 3, 4  
First and second stage, 55, 58, 59  
Intermediate second stage, 60  
National, 58  
State, 74  
Variances, 90, 98, 153  
Reliability of, 159  
SMSA, 1, 64  
State, 64, 68, 73, 92  
Unbiased, 55  
See also Bias, Confidence intervals, Housing units, and Keyfitz
- Evaluation, see Interviewer
- Field Division, 28, 147
- FOSDIC (Film Optical Sensing Device for Input to Computers), 2, 28, 52, 85
- 4-8-4 rotation system, see Rotation system
- Frankel, Martin R., 92, 154
- Gordon Committee, see Committees
- Group quarters (GQ), 15, 30
- Half point, 136
- Health Interview Survey, 127
- Hit number, 114, 134
- Household, 15  
Coverage check, 47  
Indicator weight, 66, 67  
Survey, Quarterly, 67, 127
- Housing  
Monthly estimates, 66  
Survey, Annual, 66, 100, 127, 135  
Units, 15  
Vacancy Survey (HVS), 41, 67  
Questionnaire (form HVS-1), 43
- HVS questionnaire, see Housing
- I & O, see Industry and occupation coding
- Income, 44, 87  
Edit and allocation, 45
- Index of inconsistency, 80
- Industry and occupation coding, 28, 51  
Quality control of, 52
- Inmates, institutional, 15, 33, 144
- Interagency Technical Committee, see Committees
- Interest, public, 100
- Interval estimates, see Confidence intervals
- Interview  
Independent, 85  
Personal, 44, 101  
Telephone, 44, 54, 85, 101  
Week, 33, 54  
See also Bias
- Interviewer(s)  
Assignments, 10, 47  
Error rate, 45, 49, 53  
Evaluation, 51  
Identification, 35  
Nonqualified edit, 46  
Number of, 147  
Observation, 50, 147  
Production ratio, 50  
Recruitment, 147  
Senior, 47, 147  
Techniques, 51  
Training, 46, 50, 147  
Variance, 90  
Workload, 10, 72, 149
- Jeffersonville, Ind., Processing Center, 28, 51, 54
- Keeping house, see Labor force concepts
- Keyfitz, Nathan, 12  
Clusters, 150  
Probabilities of PSU selection, 12, 105  
Variance estimation, 91, 150, 154
- Labor  
Force, concepts, 2, 42, 87, 100  
Statistics, Committee on, see Committees  
Statistics, Bureau of, 3, 27, 54, 63, 94  
Survey, Monthly, 79  
U.S. Department of, Secretary, 68
- Legal authority for CPS, 35, 100
- Letter  
Introductory ("Dear Friend," form CPS 263), 33, 35, 36  
Warning, 51
- Level, monthly, see Estimates
- Linear form, see Estimates

- List, *see* Address  
 ED, *see* Address ED  
 Sample, *see* Address sample
- Listing  
 Check, 47  
 Sheet  
   Area segment (form 11-212), 30, 31  
   Special place (form 11-213), 30, 32  
 Updating of, 33  
*See also* Segment and Unit
- Mail procedure, 16
- Management and Budget, Office of, 27, 101  
 Statistical Policy Division, 27, 101
- Mapping, 28
- March  
 Family weighting procedure, 65  
 Supplement, Annual Demographic, 145, 146
- Match samples, 25, 127
- McCarthy, Philip J., 91, 92, 151
- Mean square error, *see* Error
- Median, 154, 161
- Microdata files, 1
- Microfilm, 28
- Midpoint, 136
- Monthly Labor Survey, MLS, 79
- Noninterview(s), 33, 46, 51, 56  
 Characteristics of, 86  
 Classifications (type A, B, or C), 33  
*See also* Bias and Estimates
- Nonmail census, *see* Census
- Nonprobability sample, *see* Sample
- Nonresponse, 82, 86  
 Item, 56, 82, 87
- Non-take-all (NTA), *see* Address and Ultimate sampling unit
- Nonwhite, 57  
*See also* Color
- Observation, 50, 147  
 Special needs, 46, 50, 149
- Office of Management and Budget (OMB),  
 Executive Office of the President, 27, 101  
 Statistical Policy Division, 27, 101
- Orthogonal replications, *see* Replications
- Overall selection probability, *see* Probability
- Overcoverage, 82
- Pattern, admissible selection, 12, 105
- Perkins, Walter M., 12
- Permit (building)  
 Coverage, 17, 30, 82  
 Segment, 17, 28, 30  
 USU, 17  
 Year built, 19, 30  
*See also* Ultimate sampling unit
- Phases I and II, 22, 24
- Policy Committee, CPS, *see* Committees
- Population  
 Division, 28  
 Estimates, 4, 59, 82
- Precision, 77  
*See also* Estimates
- President's Committee to Appraise Employment and Unemployment Statistics, *see* Committees
- Press release, 54
- Primary sampling unit (PSU), 7  
 Current design, 8, 12, 163  
 Earlier designs, 8, 12  
 Heterogeneity, 8  
 Number of, 8, 10, 13, 149  
 Replacement of, 21, 25  
   Schedule, 164, 165  
 Selection order, 111  
 Selection procedures, 10, 12, 105  
   National sample, 7, 105  
   State supplemental sample (CETA), 69  
*See also* Keyfitz and Probability  
 Size, 7, 8  
 Stratification, 8  
   National sample, 8, 11  
   State supplemental sample (CETA), 69
- Principal person weight, *see* Weight
- Probability of selection, 15  
 Keyfitz, 12  
 Overall, 14, 15, 138  
 PSU, 10, 12  
 Stratum (C-design), 12  
 USU within PSU, 13  
*See also* Keyfitz and Primary sampling unit
- Probability sample, 6, 55
- Procedures  
 Rare event universe (REU), 56, 121, 140
- Processing, close-out, 53
- Production cycle, 53
- PSU's, *see* Primary sampling unit
- Public interest, 100
- Public use, *see* Data and Tapes
- Publications, statistical review of, 94
- Quality control, 45, 51, 147
- Questionnaire  
 CPS (form CPS-1), 33, 39, 40, 41  
 Edit, 45, 50, 149  
 HVS (form HVS-1), 43  
 Missing, 53
- QHS, Quarterly Household Survey, 67, 127
- Race, 56  
 Collapsed, 56
- Random group, 90, 92
- Random starts, 113, 121
- Rare event procedure, 56, 121, 140
- Rate  
 Gross difference, *see* Bias  
 Preliminary unemployment, 53  
 Refusal, 21, 42, 86  
 Vacancy, 66  
 Tenure, 67
- Ratio estimation, 2, 3, 4  
 Bias, 85  
 First- and second-stage, 58  
 March supplement, 65  
 National, 58, 63  
 State, 74  
*See also* Weighting
- Ratio-to-moving average method, *see* Seasonal adjustment
- Recode  
 Program, 53  
 Special place, 111
- Recruitment, interviewer, 147
- References, 102
- Refusals, 21, 42, 86
- Regional office(s), 163  
 Control of operations, 45  
 Interviewer training, 147  
 Organization, 27, 147  
 Processing of data, 51  
 Questionnaire shipments, 54  
 Recruitment, 147
- Reimbursable studies, 44
- Reinterview, 46, 51, 79, 149  
 Components, 47  
 Dependent reconciliation, 79  
 Independence, 47, 79  
 Telephone, 47
- Relvariance, 93  
 Proportions, 94  
 Totals, 93
- Replication, 91  
 Method, 91  
 Orthogonal, 91, 151  
*See also* Variance
- Reservations, *see* Ultimate Sampling units
- Residence, usual place of, 33, 142  
 Urban-rural, 57, 63

- Respondent burden, 101
- Responses
  - Conditioning of, 47, 79
  - Inconsistent, 85
  - See also Bias and Variance
- Rotation chart, 22, 23
- Rotation clusters, 25
- Rotation group, 21
  - Bias, 24, 83, 88, 90, 92
  - Incoming and returning, 28
  - Index, 83, 84, 88
  - Sets, 67, 74
  - Variance estimates, 90, 91, 92
- Rotation system, 21
  - CETA, 69
  - 4-8-4, 2, 21
- Sample
  - Address, 3, 16, 17
  - Areas, location of, 10, 163
  - CETA, 68, 163
  - Designs, 2, 6
    - A, 6, 10, 13
    - C, 6, 12, 13
    - D, 68
  - See also Current Population Survey
  - Efficiency, increase in, 5
  - List, see Address sample
  - Match, 25, 127
  - Nonprobability, 55
  - Overlap, 21, 22
    - Variance effect, 70, 72
  - Probability, 6, 55
  - PSU's, see Primary sampling unit
  - Reduction of, 138
  - Replacement of, see Rotation system
  - Report (form 11-200), 28, 33
  - Rotation, see Rotation system
  - Segment, see Segment
  - Self-weighting, 2, 6, 13, 55
  - Size of, constant, 56, 138
  - Supplemental State, 1, 4, 68, 149
  - See also Cut groups
- Sampling
  - Census ED's, 18, 111
  - Error, 88, 90
  - Quota, 55
  - Stages, 6
  - Variance, 5, 77, 78, 90, 95
  - Within ED's, 19, 115, 122
  - Within PSU's, 18, 111
  - Within strata, 10, 12, 105
- Seasonal adjustment, 3, 63
  - Introduction into CPS, 3
  - Variance, 91, 92
- Secretary of Labor, see Labor
- Security, 54
- Segment, 16
  - Address (or List), 17, 122
  - Area, 16, 115
  - Cen-Sup, 18, 20
  - Folder, 30
  - Listing, 30
  - Map, 120, 121
- Segment—Continued
  - Number, 16, 114
  - Permit, 17, 20, 28, 30
  - Special place, 20, 30
- Segmenting, 122
- Selection
  - Pattern, 12, 105
  - PSU's, 7, 105
    - National sample, 7, 105
    - State sample, 69
  - See also Hit number, Keyfitz, Primary sampling unit
- Shipments, 53, 54
- SMSA, see Standard metropolitan statistical area
- Source and reliability statement, 94, 159
- Special place, 2, 30, 125
  - Listing sheet (form 11-213), 30, 32
  - Recode, 111
  - Segment, 20, 30
- Standard error, see Error
- Standard metropolitan statistical area (SMSA), 1
  - Estimates, 64
- Standards
  - Accuracy and precision, 94, 100
  - Production, 48, 51
- State, see Estimates
- Statistical Methods Division, 28, 94
- Statistical Policy Division, Office of Management and Budget, 27, 101
- Statistical Research Division, 27
- Statistical Standards and Methodology, Associate Director for, 27
- Strata
  - Collapsed, 91, 92, 152, 158
  - Groups of, 150, 153
  - Homogeneity, 10
  - Number of, 8, 12, 13
  - Pairs of, 150
  - Selection, see Primary sampling unit
- Stratification
  - National sample, 8, 11
  - State sample, 69
- Subject matter
  - Compatible, 100
  - Sensitive, 101
- Subsampling, 56
- Subsegmenting, 121
- Successive samples, see Match samples
- Surveys
  - Book and Dexterity, 94
  - Construction, 17
  - Crime, National, 100, 127, 135
- Surveys—Continued
  - Current Population, *loc. cit.*
  - Demographic Supplement, Annual, 65, 145, 146
  - Health Interview, 127
  - Household, Quarterly, 67, 127
  - Housing, Annual, 66, 100, 127, 135
  - Labor, Monthly, 79
  - Supplemental, using CPS sample, 100, 145
- Survey week, 41
- Take all (TA), 33, 126
- Tape address register (TAR), see Enumeration district
- Telephone, see Interview and Reinterview
- Tepping, Benjamin J., 92, 154
- Training of data collection staff, 147
  - See also Interviewer
- Ultimate sampling units (USU's), 6, 13, 15
  - Allocation to surveys, 127
  - Mixed, 126
  - Non-take-all (NTA), 126
  - Number between hits, 136
  - Reservation of, 25, 127
  - Selection within PSU's, 113
  - Selection within segment, 30, 121
  - Size, 4, 16, 140
  - See also Address, Area, Cen-Sup, and Permit
- Unbiased estimates, see, Estimates
- Unemployed persons, see Labor force concepts
- Unemployment rate, 53
- Units
  - Erroneous inclusions and omissions, 48, 82
  - Extra, 51
  - Housing, 15
  - Listing, 30
  - See also Housing
- Update listing, 33
- Urbanized area (UA), 18
- USU's, see Ultimate sampling units
- Vacancy rates, 66
- Value, expected, 77
- Variance
  - Averaging, 93
  - Components, 64, 69, 90, 92, 95, 96, 157
  - Composite estimator, 95, 150
  - Estimation, 90, 150
    - Bias, 90, 91, 92, 94, 158
  - Generalized, 93
  - Ratio estimate, *loc. cit.*
  - Replication, 91, 150

**Variance—Continued****Response, 77****Correlated component of, 78, 158****Model, 77****Simple, 78, 79****Total, 77, 158****See also Interviewer, Keyfitz, Sampling,  
and Seasonal adjustment****Veteran status, 2****Weight, weighting****Basic, 56, 73, 138, 141****Household indicator, 66, 67****March family 54, 65****Principal person, 66****Special, 56, 73, 86, 91, 141****Woodruff, Ralph S., 92, 154****Work, looking for, 41, 87****Work Projects Administration (WPA), 2****Workers****Part-time, 3, 88****See also Labor force concepts****Workload, see Interviewer****X-11 method, see Seasonal adjustment**

*U.S. Maps and Information  
Related to*

# The Current Population Survey: Design and Methodology

(See app. M.)

**Form 11-5** Sample PSU's, 461-Area Current Population Survey National Sample: March 1973. This four-color map of the United States, scale 1 to 5,000,000, identifies the separate types of sample areas in which interviews are conducted each month for the Current Population Survey. Price \$1.00.

**Form 11-6** Sample PSU's, 461-Area Sample and State Supplements: August 1976. A four-color map of the United States, scale 1 to 5,000,000, that identifies sample areas in which interviews are conducted each month to prepare estimates for each State. Price \$1.00.

**Form 11-1A** Data Collection Coverage and Current Program PSU's: January 1973. This three-color map, scale 1 inch to 10 miles, shows CPS sample areas in the New York and a portion of the Boston Regional Office areas. Price \$0.25.

**Form 11-4** PSU's Included in Current Survey Samples: September 1976, provides a listing of the counties, independent cities, and towns (in some States) that make up each of the sample PSU's in the current survey programs of the Census Bureau. Price \$1.15.



(Please detach along this dotted line)

## ORDER FORM

Please send me the following items (indicate the number of copies desired in the space provided at the left of each title):

- \_\_\_\_\_ Form 11-5 Sample PSU's, 461-Area Current Population Survey National Sample: March 1973 at \$1.00
- \_\_\_\_\_ Form 11-6 Sample PSU's, 461-Area Sample and State Supplements: August 1976 at \$1.00
- \_\_\_\_\_ Form 11-1A Data Collection Coverage and Current Program PSU's: January 1973 at \$0.25
- \_\_\_\_\_ Form 11-4 PSU's Included in Current Survey Samples: September 1976 at \$1.15

Name \_\_\_\_\_

Organization \_\_\_\_\_

Street Address \_\_\_\_\_

City, State, and ZIP Code \_\_\_\_\_

## MAIL ORDER FORM WITH PAYMENT TO

Subscriber Services  
Section (Publications)  
Bureau of the Census  
Washington, D.C. 20233

Postage stamps are not acceptable; currency is submitted at sender's risk. Remittances from foreign countries must be by international money order or by a draft on a U.S. bank.

MAKE CHECK OR MONEY ORDER PAYABLE  
TO SUPERINTENDENT OF DOCUMENTS

TOTAL AMOUNT \$

Payment enclosed  
(Mark one)  
☐ Check  
☐ Money order  
☐ GPO coupons

OR

Charge to  
Superintendent of  
Documents Deposit  
Account Number \_\_\_\_\_

