THE WORK OF THE UNITED STATES BUREAU OF THE CENSUS WITH EMPHASIS ON SAMPLE DESIGNS AND CONTROL OF ERRORS IN CENSUSES AND SURVEYS

By M. N. MURTHY
Indian Statistical Institute

SUMMARY. In this paper an attempt is made to present the work of the United States Bureau of the Census with emphasis on the methodological aspects of sample design and control of errors in censuses and surveys. Though I claim no originality of its contents, I alone am responsible for the selection, presentation and accuracy of the material included in this paper. By no means the coverage in this paper is to be considered complete. At best it attempts to give an idea of the types of methods being used at the Bureau of the Census in the field of data collection and compilation without attempting to give much of details to be precise. It may be mentioned that more attention is given in this paper to population and agricultural fields than to economic fields. This paper is based on a number of published and unpublished papers of the United States Bureau of the Census.

1. United States statistical system

1.1. *Federal statistical system.* In the United States, a decentralised statistical system has come into existence because of historical considerations and the nature of the governmental set-up. The need for coordination of the activities of the different statistical organisations in a decentralised system was recognised as early as 1933 when the Central Statistical Board was set up for this purpose. Since then, the responsibilities of coordination in the field of statistics have been transferred to the Office of Statistical Standards in the Bureau of the Budget of the Executive Office of the President.

The State Governments are, in general, independent of the Federal Government in carrying out statistical programmes. The development of statistical systems in the States has been uneven. In practice, formal and informal cooperative arrangements exist by which the statistical activities of the States and the Federal Government are coordinated in a number of subject matter fields such as vital statistics and agriculture.

Even a brief account of the statistical system in the United States would be incomplete without a mention of the activities of Universities and other private organisations in the field of statistics. In general, the coverage by individual private organisations is restricted and they usually specialise in particular fields with a view to develop the methodology of data collection, compilation and interpretation. There is some amount of cooperation in certain fields, between the statistical activities of the Federal and State Governments on the one hand and the Universities on the other.

The Federal Statistical System may be considered to consist of four types of statistical organisations differentiated on the basis of their functions and the use to which the data collected by them are put to : (i) General Coordinating Agency; (ii) General-purpose Statistical

Agencies; (iii) Analytical and Research Agencies; (iv) Administrative and Regulatory Agencies.

1.2. *Office of Statistical Standards.* The Office of Statistical Standards in the Bureau of the Budget of the Executive Office of the President is primarily a coordinating agency and its main functions are : (i) coordination of the activities of the various statistical agencies with a view to avoid duplication and omission; (ii) evolution of statistical standards in concepts and definitions, and in presentation and publication of data; (iii) review of the questionnaires and forms prepared by the various statistical agencies; (iv) statistical programming and budgeting; and (v) promotion of international cooperation in the field of statistics.

1.3. *General-purpose statistical agencies.* These agencies are primarily responsible for regular collection, compilation and publication of data in specified fields for general use. As a group, they account for a large proportion of the statistical activities of the Federal Government. The following are mainly general-purpose agencies : (i) Agricultural Marketing Service (Department of Agriculture); (ii) Bureau of the Census (Department of Commerce); (iii) Bureau of Labour Statistics (Department of Labour); (iv) National Office of Vital Statistics (Department of Health Education and Welfare).

1.4. *Analytical and research agencies.* The following are agencies which use statistics collected by other agencies for future projections, policy making, and construction of composite measures of social and economic conditions : (i) Council of Economic Advisors (Executive Office of the President) ; (ii) Division of Agricultural Economics (Department of Agriculture) ; (iii) Farm Economics Research Division (Department of Agriculture); (iv) Bureau of Foreign Commerce (Department of Commerce) ; (v) Business and Defence Service Administration (Department of Commerce); (vi) Office of Business Economics (Department of Commerce); (vii) Bureau of Mines (Department of Interior); (viii) Division of Research and Statistics (Board of Governors of the Federal Reserve System).

1.5. *Administrative and regulatory agencies.* These agencies collect statistics as a by-product of their routine duties. Usually the statistics collected by them are of specific rather than of general interest. But some of these agencies, such as the Social Security Administration and Internal Revenue Service, collect data of general interest as a by-product of their administrative functions.

1.6. *Advisory committees.* The Advisory Committee on Statistical Policy, set up by the American Statistical Association in 1931, advises the Bureau of the Budget on broad matters of public policy affecting the Federal Statistical System. The Advisory Council on Federal Reports, which was established in 1942 and which is financed by business organisations, advises the Budget Bureau on the reduction of reporting burdens, cost of Government reports, and on increasing their usefulness. The Labour Advisory Committee on Statistics, set up in 1945, advises the Government on the interests of organised labour in Federal Statistical programmes. Besides these, there are other advisory bodies to advise the Government on matters relating to statistics in certain specified fields.

1.7. *Principal federal statistical programmes.* As an important part of statistical programming and budgeting, the Office of Statistical Standards prepares each year a consolidated statement for major statistical programmes. A general idea regarding the range and scope of principal federal statistical programmes is given by Table 1 showing the budget estimates for the financial year ending June 30, 1961.

TABLE 1. BUDGET ESTIMATES FOR PRINCIPAL CURRENT STATISTICAL PROGRAMMES BY BROAD SUBJECT AREAS

sl. no.	statistical programme	1961 budget estimates (in million dollars)
1.	labour statistics	8.5
2.	demographic statistics	6.7
3.	prices and price indexes	4.0
4.	production and distribution statistics	15.7
5.	construction and housing statistics	1.3
6.	national income and business financial accounts	5.8
7.	total	42.0

Source : Special analysis of principal federal statistical programmes in the 1961 budget.

2. Organization and functions of Bureau of the Census

2.1. *Functions.* In 1902 the United States Congress set up a permanent office to take census and to collect other statistics. Originally this office was in the Department of Interior, but was transferred to the Department of Commerce and Labour in 1903. When this Department was split in 1913, the Bureau of the Census was placed in the Department of Commerce. The Bureau of the Census collects, compiles, analyses and distributes statistical data relating to population, housing, agriculture, construction, governments, manufactures, mineral industries, business, foreign trade and related subjects under the broad technical and administrative authority from the Secretary of the Department of Commerce. This job is done in such a way as best to carry out the mandates of the constitution of the U.S.A. and other authorizing and mandatory legislative enactments and to meet the needs of the business and other users of statistical information. The Bureau also organizes training programmes of different types in the field of statistics to train the participant trainees from the United States and other countries.

2.2. *Census and surveys.* The Bureau conducts the Census of Population and Housing every ten years (years ending in '0'), and every five years it conducts the Census of Agriculture (years ending in 4 and 9), the Census of Manufacture, Mining Industry and Business (years ending in 3 and 8) and the Census of Governments (years ending in 2 and 7). Besides conducting these censuses, the Bureau conducts a number of current sample surveys such as the monthly Current Population Survey, the National Health Survey, the Monthly Retail Trade Survey, the Annual Survey of Manufacture, the Monthly Wholesale Trade Survey, and the Construction Survey. Some of these surveys are undertaken by the Census Bureau in its capacity as data collecting agent for other government agencies.

2.3. *Organization.* The Director and the Deputy Director who are responsible for the work of the Bureau of the Census are assisted by five Assistant Directors and the Chiefs of the International Statistical Programmes Office and Public Information Office. The Assistant Directors are responsible for the work in the fields of (i) Statistical Standards; (ii) Demographic Fields; (iii) Economic Fields; (iv) Operations and (v) Administration.

TABLE 2. ORGANIZATION OF THE U.S. BUREAU OF THE CENSUS AND THE DISTRIBUTION OF EMPLOYEES AS ON JANUARY 1, 1959

sl. no.	department/division/section	number of employees
1.	Office of the Director including International Statistical Programme Office and Public Information Office	129
	Statistical Standards	
2.	Statistical Research Division	29
3.	Statistical Reports Division	25
4.	Electronic System Division	84
	Demographic Fields	
5.	Agricultural Division	63
6.	Population Division	252
7.	Housing Division	54
8.	Demographic Surveys Division	*
9.	Statistical Methods Division	*
10.	Decennial Operations Division	16
	Economic Statistical Fields	
11.	Business Division	84
12.	Industry Division	132
13.	Foreign Trade	61
14.	Government Division	84
15.	Transportation Division	14
16.	Economic Operations Division	488
17.	Construction Statistics Division	*
	Operations	
18.	Field Division	1755
19.	Geography Division	100
20.	Machine Tabulation Division	357
	Administration	
21.	Administrative Service Division	341
22.	Budget and Management Division	77
23.	Personnel Division	61
24.	Census Operations Office at Jeffersonville	427
25.	total (Bureau of the Census)	4033

*Not available.

Source: Federal Executive Departments and Agencies—prepared by the States Senate Committee on Government Operations, January, 1959.

2.4. *International Statistical Programmes Office.* This office conducts the Bureau's foreign consultation and training programmes and represents it in international statistical activities. In this capacity it provides consultation on statistical matters to foreign countries, arranges programmes for foreign visitors and trains foreign technicians in census and other statistical methods. This office also helps other agencies to recruit and train technicians for work abroad. It conducts the research needed for its training and consultation programmes, assembles foreign statistics through exchange of publications and provides statistical information to foreign governments and international organisations.

2.5. *Public Information Office.* This office plans and directs the Bureau's public information programme which is designed to ensure public cooperation in current and forthcoming censuses and surveys.

2.6. *Statistical Standards.* This office consists of three Divisions and their activities are the following :

(i) The Statistical Research Division develops and promotes effective use of mathematical, statistical and psychological methods and techniques. This Division conducts research on all phases of statistical methodology and provides guidance to other Divisions.

(ii) The Statistical Reports Division advises on publication policy, standard terminology, statistical presentation and the like. It prepares the Statistical Abstract of the United States and its supplements.

(iii) The Electronic System Division conducts research on all phases of electronic high speed digital computer systems and makes available such systems to the other Divisions of the Census Bureau and to other groups for processing of mass data.

2.7. *Demographic Fields.* This office consists of 6 Divisions and their functions are given below :

(i) Agriculture Division designs and conducts current survey and census programmes covering agriculture, agricultural activities, agricultural products, irrigation and drainage.

(ii) Population Division designs and conducts programmes covering the demographic, social and economic characteristics of the people.

(iii) Housing Division collects and compiles the data on housing characteristics through census and survey programmes.

(iv) Demographic Surveys Division conducts current surveys and special censuses in the demographic fields.

(v) Decennial Operations Division programmes and processes data collected in the 1960 Censuses of Population and Housing.

(vi) Statistical Methods Office provides guidance on sampling, quality control and related matters for the demographic subject and operations divisions. It may be noted that the Statistical Research Division has this responsibility for the Bureau of the Census as a whole.

2.8. *Economic Fields.* This office consists of 7 Divisions and their functions are given below :

(i) Business Division designs and conducts surveys and census programmes covering retail, wholesale and service trades.

(ii) Industry Division designs and conducts current survey and census programmes covering manufacturing, mineral and related industries.

(iii) Foreign Trade Division designs and conducts survey programmes covering the export and import trade of the United States and foreign trade and shipping statistics.

(iv) Government Division designs and conducts current survey and census programmes covering governmental structure, activities finances and employment in State and local governments

(v) Transportation Division designs and conducts census programmes covering transportation industry and various segments thereof.

(vi) Economic Operations Division programmes and processes data collected in censuses of Business, Manufactures, Transportation and Mineral Industries and current surveys in the economic fields.

(vii) Construction Statistics Division designs and conducts survey programmes covering all phases of construction statistics.

2.9. *Operations.* This office consists of 3 Divisions and their functions are given below :

(i) Field Division which has 17 Regional offices throughout the United States plans and collects data for current surveys and census programmes.

(ii) Geography Division conducts geographic research and establishes geographical areas for the collection and presentation of statistical data.

(iii) Machine Tabulation Division administers the programme concerning the transfer of processed information to punch cards and the tabulation of statistical data through the conventional tabulating machines.

2.10. *Administration.* This office consists of 3 Divisions and their functions are given below.

(i) Administrative Service Division provides procurement record management, reference services, guidance on printing matters and other central services within the Bureau. It also conducts searches of Census records for age documentations.

(ii) Budget and Management Division administers the budget fiscal services, administrative internal audit and work measurement programmes and promotes a Bureau—wide management programme.

(iii) Personnel Division administers the personnel management programme including organization, planning, classification, pay administration, personnel recruitment and utilization, employee relations and training.

2.11. *Advisory committees.* The Bureau of the Census makes use of the advice given by a number of advisory committees. A number of Census Advisory Committees representing specialized business and professional associations advise the Census Bureau on its programmes. A committee set up by the American Statistical Association also advises the Census Bureau on all programmes and on matters of general policy. Within the Census Bureau, the executive staff composed of the Assistant Directors, Deputy Director and Director meet almost every week to review and coordinate the work of the Bureau.

3. Current population survey

3.1. *Introduction.* The need for data on employment and unemployment was keenly felt in the United States of America during the depression of 1930's and since then there has been a growing demand for this information. In 1940, the Work Projects Administration started issuing the monthly report on labour force on the basis of a sample survey initiated by its Research Division during 1939. Since August 1942, the Bureau of the Census has taken up this work.

In October 1943, the Census Bureau designed a new sample for the monthly labour force surveys. Every month the new sample consisted of about 21,000 interviewed households spread over 68 sample primary stage units (PSU) which were counties or groups of counties. In February 1954 the sample was redesigned to make the sample households spread over 230 sample PSU's which included the orginal 68 PSU's and in 1956 the sample was further expanded to cover 35000 interviewed households spread over 330 PSU's.

The concepts and definitions used in this survey have remained substantially the same since its inception. The data are collected for the calendar week containing the 12th of the month and the monthly report is published within three weeks of the completion of data collection. In July 1959, the Bureau of Labour Statistics of the Department of Labour was assigned the work of analysis and publication of the monthly report on labour force. The Bureau of the Census continues to do the work relating to sample design and collection and tabulation of data.

3.2. *Scope of the survey.* This survey with its monthly supplements provides fairly detailed information on the economic status and activities of both the farm and nonfarm population of the United States. It provides data on the personal characteristics of the total labour force and total population such as employed and unemployed, age, sex, race, marital and family status, veteran status, educational background etc. The distributions of workers, by the number of hours worked, by major occupation and industry groups are also made available from this survey. This survey provides information not only on the current labour force but also on the labour reserve.

3.3. *Concepts and definitions.* Though the CPS covers the whole of the population of the United States, the labour force survey is confined to persons of age 14 years and over. The division of the population is done as follows :

(1) total labour force : (a) armed forces; (b) civilian labour force : (i) at work (ii) with a job but not at work; (iii) looking for work; (iv) inactive employment;

(2) not in labour force : (i) keeping house; (ii) going to school; (iii) unable to work; (iv) retired and others.

The priority in the classification is given to work. Even if a person does an hour's work during the reference week for pay or on own account he is taken as employed. In case of unpaid employment, a person is said to be employed only if he works for at least 15 hours during the reference week. Persons in the labour force belonging to categories (a), b(i) and b(ii) are considered to be "employed" and those in categories b(iii) and b(iv) are considered to be "unemployed". The category b(iv) includes persons who were not at work and did not look for work as they thought there was no scope for getting a job. Since 1957, persons who have been laid off and those who are waiting to take up new jobs have been considered as unemployed and persons attending school who had new jobs to which they were to report have been included in the category "not in the labour force."

3.4. *Available information.* Fairly detailed information on a number of characteristics is available at the county level which is an important administrative division in the United States. Each of the 48 States is divided into a number of counties and there are about 3,000 counties in the whole of the country. Some of the data available at county level are illustrated by the following : population, area, rate of population increase, percent of population classified as rural farm, percent of non-white population, Hagood Index which is a rural level of living index, average value of product of all farms, principal industries etc. Within each county fairly detailed maps and the population for each of the Census Enumeration Districts (ED) are available.

3.5. *Original sampling design.* The original design was evolved in 1943 under rather severe administrative restrictions. The design had to conform to the then existing field organization of about 60 supervisors each of whom supervised the work of 5 to 15 part-time enumerators. The main objectives of the survey were to provide estimates of labour force statistics for the country as a whole on a fairly tight time schedule and to serve as a source for collecting data on topics of special current interest from time to time.

The sampling design adopted was a stratified multi-stage design where one primary stage unit (PSU), which was a county or a group of counties, was selected with probability proportional to population (ppp) from each stratum and from each selected PSU a sample of segments of about 6 households was selected directly or in stages so as to make the design self-weighting. A segment is a cluster of nearby houses or flats consisting of approximately 6 households.

3.6. *Primary stage unit.* The country is divided into 4 regions on the basis of geographic and administrative considerations. Within each region the contiguous counties were grouped to form the PSU such that the PSU is heterogeneous within itself especially with respect to farm-non-farm composition and that the total area of a PSU did not exceed 1500 square miles in the East and 2000 square miles in the West. By this procedure, 3000 and odd counties are grouped into about 2000 PSU'S.

3.7. *Stratification.* The PSU's are grouped into 68 strata which are homogeneous within themselves with respect to social and economic characteristics and that the strata populations are approximately the same. For this purpose the PSU's are first classified into four categories : (i) 12 largest cities and Washington D.C. including the surrounding metropolitan area; (ii) all other PSU's with cities of 50,000 or more population in 1930; (iii) all other PSU's with less than 23% of the population residing on farms in 1940 and those units with high in-migration rate between 1940 and 1943; and (iv) the rest of the PSU's.

The 13 areas in category (i) are treated as individual strata. The stratification in categories (ii) and (iii) is done taking into consideration the rate of migration, proportion of labour force engaged in manufacture, principal industries of the area etc. The PSU's in category (iv) are stratified on the basis of type of farming and major crops of the area.

3.8. *Selection of the PSU.* The PSU's which are treated as individual strata are said to be self-representing strata. From each of the non-self representing strata one PSU is selected with probability proportional to 1940 population.

3.9. *Selection of households.* In each of the strata, the over-all sampling fraction was taken as 1 in 2050. In case of non-self-representing strata, the sampling fraction in the selected PSU is taken as the product of the over-all sampling fraction and the inverse of probability of selection of that PSU so that the weights to be used in estimation remains

2050 for all sample units. In the sample PSU's the 1940 Enumeration Districts (ED) are arranged according to whether they are in urban, rural non-farm or rural farm area. The average household size of the county in which the ED is situated is applied to the population of that ED to get an approximate idea of the number of households, since the 1940 Census tabulation did not provide the figure for number of households at ED level. Dividing the approximate number of households by 6, the number of clusters or segments of 6 households into which the ED was to be divided was obtained. Using this as the size for the ED's in the selected PSU's, the required number of segments were selected systematically. The number of segments to be selected in an ED was the same as the number of times the ED occurs in the sample which was almost always one.

The sample segments in the selected ED's are obtained after dividing that ED into the number of segments allotted to it either in the laboratory by the geography division or in the field by direct observation. In cases where detailed maps are available, the segments are selected directly. If a sample segment is found to contain more than 20 households at the time of enquiry, subsampling is resorted to and the multiplier is adjusted.

3.10. *Estimation procedure.* Till recently the final estimate was obtained by the procedure of double ratio estimation. The estimation procedure consisted in first obtaining a ratio estimate for each of the 56 age (14)–sex (2)–colour (2) classes using the 1950 census figures for the 24 colour (2)–residence (3) –region (4) classes and then applying a ratio estimate to these figures using the current figures for age-sex-colour classes obtained from independent sources. The estimate is of the form

$$x'' = \sum_{a} \frac{x'_a}{y'_a} Y_a$$

where $$x'_a = \sum_{c} \frac{x_{ac}}{f} \frac{Z_c}{z'_c} \text{ and } y'_a = \sum \frac{y_{ac}}{f} \frac{Z_c}{z'_a};$$

x_{ac} and y_{ac} are the sample value of some characteristic and the sample value of number of persons in a particular combination of (a) age-sex-colour class and (c) colour-residence-region class; Z_c and z'_c are respectively the actual and estimated census populations in a particular colour–residence–region class; and f is the over-all sampling fraction.

3.11. *Redesign of the survey.* In the original design, only 68 PSU's were selected because it was thought necessary to have a full-time supervisor for each PSU especially at the initial stages of the survey to reduce the response and measurement errors. If this restriction was not there, it would have been possible to select a larger number of PSU's thereby decreasing the sampling variability. In 1954, it was felt that the supervisory staff was not well utilized and that more effective supervision could be achieved by strengthening the regional supervisory staff and by introducing certain control procedures without the need for a supervisor in each PSU. A redesign was thought of mainly with a view to increase the effectiveness of supervision through stronger staffing and use of better control methods and by reducing regional offices to 17. The money saved by this reduction in the number of regional offices was also utilized to increase the number of sample PSU's to 230 from the original 68 keeping the over-all sampling fraction the same as before.

In re-designing the survey, more recent data are used in stratification and selection of PSU's. One PSU is selected from each stratum with probability proportional to the 1950 census population, keeping as many of the originally selected 68 PSU's as possible in the sample. In the selection of the PSU's, the method of controlled selection was used to increase the geographic spread of the sample.

3.12. *Sample expansion.* In 1956, the over-all sample size was increased from about 21000 interviewed households per month to 35000 interviewed households. The expanded sample is spread over 330 sample PSU's consisting of 638 counties as compared to the spread of the previous sample over 230 sample PSU's comprising 453 counties. The main objective of this sample expansion was to increase the reliability of the survey results. With the increased sample size the sampling variability would be about 80% of that in the previous 230 PSU sample.

3.13. *Rotation of the sample.* The sample rotation procedure adopted in 1953 ensures 75% common sample segments between any two consecutive months and 50% common sample segments between any two successive years for the same month. This is achieved by dividing the sample of segments into 8 sub-samples and allowing each sub-sample to be canvassed during two 4-month periods separated by 8 months. After a sub-sample has been canvassed in 8 months, that sub-sample is replaced by a fresh sample. With this scheme one-eighth of the sample segments are replaced every month by a fresh sample and another one-eighth of the sample is replaced by a sub-sample canvassed 8 months earlier. It may be noted that this procedure is a compromise between complete repetition and complete replacement. The above rotation scheme together with the estimation procedure adopted aims at providing fairly good estimates of level and also of change from month to month, which is quite important, by retaining 75% common sample from month to month and at the same time does not over-burden the respondent since the respondent is given a break for 8 months during the 16 months he is in the sample.

TABLE 3. RELATIVE EFFICIENCIES OF COMPOSITE ESTIMATES AND DOUBLE RATIO ESTIMATES

sl. no.	item	estimated relvariance* composite estimate	estimated relvariance* double ratio estimate	col. (2) / col. (3)
	(1)	(2)	(3)	(4)
1.	total labour force	.000020	.000024	0.83
2.	total employed in non-agriculture	.000036	.000044	0.82
3.	total employed in agriculture	.001517	.001617	0.94
4.	males employed in agriculture	.001434	.001604	0.89
5.	males employed	.000020	.000020	1.00
6.	non-white employed	.000216	.000233	0.93
7.	females in labour force	.000159	.000196	0.81
8.	at work in non-agriculture for full time	.000066	.000073	0.90
9.	unemployed	.001958	.001725	1.14

*Square of coefficient of variation (These estimates are likely to be subject to large sampling errors).

Source : Current Population Survey (Draft) by Steinberg, Table 15.

3.14. *Composite estimate.* The rotation scheme explained above helps in getting good estimates of both the level and the change. It may be noted that the following two estimates can be obtained from the data collected : (i) the double ratio estimate mentioned in paragraph 3.10 and (ii) the estimate obtained by adding the estimate of change from the previous month calculated from the 75% common sample to the previous month's estimate. The composite estimate is obtained as an average of these two estimates. Though the optimum weights to be used may not be equal, equal weighting in this case is found to give rise to some increase in reliability for almost all items as compared with the double ratio estimate.

3.15. *Estimation of variance.* The estimation of sampling variability has to take into account the sampling design and the estimation procedure. The calculation of sampling variability in the CPS becomes difficult because (i) only one PSU is selected from each stratum; (ii) systematic sampling is used in the selected PSU's and (iii) the variance of the composite estimate used is complicated. The first difficulty may be overcome for all practical purposes with the use of collapsed strata technique which is likely to provide an over-estimate of the sampling variation. The second difficulty presents no problem at all in estimating the total variance of the estimate. The problem then is in obtaining within-PSU variance, which requires some simplifying assumptions. The third difficulty is overcome by using the following method based on the technique of interpenetrating sub-samples.

The method of calculating the sampling variability for the small number of major summary items is to obtain 20 interpenetrating half sub-samples (made up by selecting one PSU from each collapsed stratum), and prepare the composite estimate for these items for each half sub-sample. The estimated variance of an item for the total sample is then derived from the 20 half sub-sample estimates. More replications can be obtained, if desired, to increase the reliability of the variance estimate. This variance estimate has an upward bias contributed by introducing the variance between collapsed strata, but empirical studies have indicated that this upward bias is relatively small in the CPS sample.

An approximate way of calculation of the sampling variability based on more degrees of freedom is to get an estimate of variance for each of the collapsed strata and then add up the variance estimates. For instance, if x_{s1} and x_{s2} are estimates of the s-th collapsed stratum from the two PSU's, then an estimate of the variance of the estimate

$$\sum_{s} \frac{x_{s1}+x_{s2}}{2}$$

is given by

$$\sum_{s} \frac{(x_{s1}-x_{s2})^2}{4}.$$

3.16. *Approximate variance estimator.* Variances are estimated separately each month for a few of the more important individual items, but to simplify the computation of approximate variances for a large number of additional items, the relationship between the sampling variability and the value of the estimate has been studied. A satisfactory relationship for a group of similar items is believed to be given by

$$v_x^2 = a + \frac{b}{x}$$

where v^2 is the relvariance, x is the value of the estimate and a and b are constants. The utility of such an approximation depends much on the degree of its stability over time for the same group of items. It seems that this relationship is not conformed to by large values of x.

3.17. *Non-response.* The non-response rate in this survey is usually of the order of 3 to 5 percent including the contribution to non-response or non-interview due to refusals which is less than 1%. The other sources of non-interview are "not at home", "temporarily absent", "difficult communication", "on account of bad weather" etc. Attempts are made to reduce non-response by call-back and the use of telephones.

TABLE 3. DISTRIBUTION OF NON-RESPONSE RATE BY REASONS FOR THE PERIOD JANUARY-DECEMBER 1959

sl. no.	month	non-response rate				
		total	no one at home	temporarily absent	refusal	others
(0)	(1)	(2)	(3)	(4)	(5)	(6)
1.	January	4.0	1.3	1.2	.9	.6
2.	February	3.9	1.3	1.2	1.0	.4
3.	March	4.4	1.6	1.3	1.0	.5
4.	April	3.7	1.4	1.0	1.1	.2
5.	May	3.1	1.2	.9	.8	.2
6.	June	3.9	1.1	1.9	.7	.2
7.	July	5.9	1.6	3.3	.7	.3
8.	August	5.7	1.6	3.1	.7	.3
9.	September	4.0	1.5	1.4	.8	.3
10.	October	4.3	1.8	1.2	1.0	.3
11.	November	3.7	1.6	.9	.9	.3
12.	December	3.9	1.8	.9	.9	.3

Source : Summary Report of CPS Reinterview—December 1959, Table M1.

The weights or multipliers of all the interviewed households are adjusted for the non-response from some of the occupied households in the sample because of refusal, not at home, temporarily absent or other reasons. This adjustment is made separately for groups of PSU'S and within these for each of six colour-residence groups of households.

3.18. *Tabulation.* Once the data recorded on the mark-sensing cards reach the Bureau of the Census, the coding of the industry-occupation is undertaken. After this, the completed mark-sensing cards are passed through the IBM 707 document reader which transfers the data from the mark-sensing cards to the punch cards. These punched cards are passed through the 101 machine to get some consistency checks. Then the information on the cards is transferred by a card-to-tape converter to the magnetic tape. The electronic computer is used to edit and tabulate the data. It is possible to programme such that the tabulated data come out of the high speed printer in the form in which the tables are to be published. It may be mentioned that a number of multipliers are being used at the tabulation stage due to (i) response adjustment (ii) subsampling in case of heavy work-load and (iii) ratio method of estimation.

3.19. *Field organization.* The field organization consists of 17 regional offices. Each regional office has one Supervisor, and a few Programme Assistants. During the enumeration week each month, a staff of about 550 permanently hired part-time enumerators are employed to collect data for this survey from a sample of about 35000 interviewed households every month. The ratio of supervisory staff to the enumerators is approximately 1 in 8.

3.20. *Quality control programme.* To keep the quality of the data collected at a high standard, a Quality Control Programme is in operation. This programme consists of intensive initial training of enumerators, organization of refresher training courses, direct physical observation of the work of enumerators by the supervisors twice a year, reinterview of a sub-sample of the work of each interviewer once in four months by a supervisor and review of the filled-in questionnaires.

3.21. *Cost of the survey.* The CPS costs about $ 1.8 million yearly of which $ 1 million is for field operations, $ 0.5 million for planning, sample designing, office control work and tabulation and the rest for analysis and publication.

3.22. *Supplements to the survey.* Since the CPS sample has been chosen so as to represent a cross section of the population of the United States, this sample is used to obtain data on many other special aspects of socio-economic activity of the population. This is done by supplementing the labour force survey by additional enquiries the scope of which may change from month to month. Some of the topics covered by this programme are the following :

(i) labour force trend among married women and the family characteristics of the workers;

(ii) educational level of workers and the extent and type of employment of those currently in school;

(iii) annual personal and family income classified by numerous personal and economic characteristics;

(iv) recreational activities and expenditures of the population;

(v) expenditure for maintenance and repair of house; and

(vi) consumer purchases of durable consumer goods.

The sample primary stage units and field organisation of the CPS are also used for other surveys as National Health Survey and Retail Stores Survey which are described later.

4. National health survey

4.1. *Objectives.* A sub-committee of the U.S. National Committee on Vital and Health Statistics, set up in 1951, recommended that a continuing national morbidity survey be conducted on a probability sample basis to obtain data on the prevalence and incidence of disease, injuries, and impairments, on the nature and duration of the resulting disability and on the amount and type of medical care received. At the instance of the Department of Health, Education and Welfare, the Congress passed a law in 1956 authorizing the Surgeon

General of the Public Health Service to make continuing surveys and special studies of the population of the United States to determine the extent of illness and disability and related information such as the number, age, sex, occupation of persons afflicted with chronic or other disease or injury; the type of disease or injury; the amount and types of services received because of such conditions.

The National Health Survey consists of three components: (i) collection of data on health through a continuing Health Household-Interview Survey; (ii) collection of data on health through procedures other than household interview; and (iii) evaluation of the procedures of data collection and evolution of improved techniques of measurement. Here a brief description of only the first component of the survey is being given.

4.2. *Household-interview survey.* The Household-Interview Survey is being conducted by the Census Bureau for the Public Health Service of the Department of Health, Education and Welfare. Though there are limitations to the accuracy of diagnostic and other information collected in household interviews, this procedure of data collection is being used, for (i) some of the data concerning the circumstances and consequences of illness or injury and the remedial action taken or sought by the individual can be obtained more accurately from household members than from any other source; and (ii) this method of data collection greatly facilitates the comparison of healthy and sick populations.

This is a continuing survey. Every week a cross section of the population represented by a sample of about 720 households are interviewed. This would mean that in a year a sample of about 36,000 households would be covered by this survey. For this purpose, there are 120 permanent part-time interviewers throughout the country trained, directed, and guided by 17 supervisors located in the Regional Offices of the Census Bureau.

4.3. *Questionnaire.* The questionnaire consists of 40 items covering identification and socio-economic description of respondents, 12 general questions on the presence or absence of illness, accidents, or impairments for each member of the household, and 54 detailed questions for each person, for whom they are applicable, covering details of illness, accident and impairment and on medical, dental, and hospital care. For most questions, the reference period is the previous 2 weeks, but for some rare items, for which memory can be relied upon, the reference period is the previous year.

4.4. *Sampling design.* The sampling design for this survey is similar in many respects to that of the Current Population Survey (CPS) described in the previous section. It does not, however, use the sample rotation and composite estimate of the CPS. Since separate estimates were needed for 41 prespecified regions, called Tab Areas, it was necessary to modify the CPS stratification and allocation of the sample to strata, which modification gave rise to 372 strata being used in the Health Survey. The sample of 36,000 households being covered in this survey is not the same as the CPS sample but is purposely made different to reduce the burden on the sample households by taking the systematic sample of households within the selected PSU's in a direction opposite to that used in the CPS.

4.5. *Estimation procedure.* The estimation procedure is similar to the double ratio estimate used in the CPS and consists of the use of ratio method of estimation at two steps. First, the ratio estimate is applied at PSU level using the 1950 population for 12 colour-residence classes. At the second stage, the current population figures for 76 age-sex-colour classes obtained from other sources are used to obtain a ratio estimate.

4.6. *Reinterview procedure.* The Supervisor recontacts about one-sixth of the sample households interviewed by each interviewer and reinterviews one prespecified person in the household. This reinterview programme is mainly intended for three purposes: (i) training and control of quality of interviewers, (ii) measurement of interviewer variability, and (iii) detection of interviewer bias. Besides the reinterview programme there are a number of editing and processing control operations to evaluate and control the quality of the data.

4.7. *Non-response.* It may be noted that the rate of non-response is about 6% of which 1% is due to refusal and the rest due to 'not at home,' 'vacant', etc.

5. Retail stores survey

5.1. *Objectives.* Since 1951, the Bureau of the Census has been conducting a monthly Retail Stores Survey to provide estimates of the dollar volume of sales of retail stores in the United States by kind of business and to measure the trends in the volume of sales.

5.2. *Sampling design.* The present sample of retail stores consists of two parts: (a) all organizations reported as operating 11 or more retail stores in the 1954 Census and all retail stores with 1954 volume of sales more than 5 million dollars; and (b) a stratified multistage sample with the same area primary stage units as for the CPS, but with a combination of list and area samples of retail stores within the primary stage units. The sample areas are located in 230 selected primary stage units. The within-PSU sample consists of all the identified large establishments and area or segment sample of the remaining establishments. There are 1,900 segments of about 4 retail stores each in the sample for each month. The within-PSU. sample can be considered to consist of three parts: (i) all retail stores in sample PSU's with 1954 volume of sales greater than a prespecified quantity which varied according to the kind of business and all departmental stores with volume of sales less than 5 million dollars; (ii) all retail stores in the selected segments with 1954 volume of sales greater than a certain amount ($ 150,000, $ 225.000 or $ 30,000) depending on the kind of business; and (iii) all other retail stores in the sample segments. The overall sampling rate for group (i) varies from PSU to PSU; for group (ii) at is 6% each month; and for group (iii) at is ½% each month making 6% sample stores when aggregated for 12 months.

The retail stores in categories (a), b(i) and b(ii) are surveyed every month. The retail stores in category b(iii) are divided into 12 groups and every month one of these groups are surveyed to get information for preceding two calendar months. In case of the former, the data are collected by mail whereas the latter are requested in person by field representatives to report their sales figures.

5.3. *Estimation procedure.* It is of interest to note that the technique of collecting information about volume of sales for the preceding two months in case of retail stores in category b(iii), can be utilized in improving the efficiency of the estimate. With the scheme explained, it is possible to get two estimates for a given month.

(i) The usual estimate using the volume of sales reported for the month from the current sample (say X_i).

(ii) The ratio estimate using the ratio of estimates of volume of sales for this month (X'_i) to that for the previous month (X''_{i-1}) from the same current sample and the estimate of volume of sales for the previous month from the previous month's samples (X'_{i-1}), namely

$$\frac{X'_i}{X''_{i-1}}\,X'_{i-1}.$$

A composite estimate using the two estimates would be

$$X''_i = (1-W)X'_i + W\,\frac{X'_i}{X''_{i-1}}\,X''_{i-1},$$

where W is to be so chosen as to minimize the variance of the composite estimate. It may be noted that X''_{i-1} is used instead of X'_{i-1} since in a continuing survey it would be possible to use a composite estimate for the $(i-1)$-th month also.

TABLE 4. EFFICIENCY OF THE PRELIMINARY AND THE FINAL ESTIMATES AND COMPARISON OF VARIANCES OBTAINED WITH SPECIFIED CONSTANTS WITH THOSE OBTAINED WITH OPTIMUM CONSTANTS

correlation	optimum constants		relvariances of composite estimates of[1] level for a month				ratio of variance obtained with specified constants to variance obtained with optimum constants	
			with optimum constants		with specified[2] constants			
	W	K	preliminary	final	preliminary	final	preliminary	final
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
.99	.868	.876	.141	.124	.156	.136	1.11	1.10
.98	.817	.834	.199	.166	.200	.167	1.01	1.01
.95	.724	.702	.312	.238	.333	.258	1.07	1.08

[1] Expressed as multiples of relvariance for the usual estimate X'_i.

[2] $W = 0.8$ and $K = 0.83$.

Source : Woodruff, R. S. (1959): The use of rotating samples in the Census Bureau's monthly surveys, *Proc. Amer. Stat. Assn.*, Social Statistics Section, 130–138

TABLE 5. RELVARIANCES[1] OF MONTH-TO-MONTH RATIOS, MONTH-TO-YEAR AGO RATIOS AND ANNUAL TOTALS

month to month correlation	month-to-month ratios		month-to-year ago ratios		year to year correlation	annual totals	
	preliminary to final[2]	final to final[3]	preliminary to final[2]	final to final[3]		sum of final estimates	sum of simple estimates
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
.99	.020	.047	.126	.082	.90	.093	.083
.98	.040	.062	.210	.159	.85	.105	.083
.95	.098	.107	.464	.359	.75	.143	.083

[1] Relvariances expressed as multiples of relvariance for the corresponding simple estimates X'_i.

[2] Relvariance of ratio of preliminary estimate for a given month to the final estimate of the preceding month or the month year ago.

[3] Relvariance of ratio of final estimate of a given month to that of the preceding month or the month year ago.

Source : Same as for Table 4.

It may be noted that in the next month, it would be possible to get another estimate of this month's total volume of sales from the next month's sample (x_i'). This may be used in revising the composite estimate mentioned above as follows :

$$X_i'^{*} = K\,X_i'^{*} + (1-K)x_i^{*},$$

where K is chosen so as to minimize the revised composite estimate. The former estimate may be considered as 'preliminary' and the latter as 'final'. The optimum values of W and K would be

$$W = \frac{\sqrt{1-\rho^2}-(1-\rho)}{\rho},$$

where ρ is the correlation coefficient between estimates for two successive months from the same sample and

$$K = \frac{1-\sqrt{1-\rho^2}}{\rho^2},$$

if optimum value of W is used in $X_i'^{*}$. The constants used and proposed to be used are $W = 0.8$ and $K = 0.83$.

Tables 4 and 5 demonstrate that there is considerable gain in using the preliminary composite estimate and that the use of final estimate is advisable from an over-all point of view. This procedure of obtaining composite estimator is promising and should be explored wherever there is scope for it.

6. Wholesale trade survey

6.1. *Objectives and scope.* The main objectives of this monthly survey are to estimate the volume of sales by kinds of business of wholesalers and to study their trend over a period of time. This survey is confined to merchant wholesalers who constitute a major portion of the broad field of wholesale trade. Merchant wholesalers are defined as 'establishments normally known as wholesalers, merchant wholesalers or jobbers, primarily engaged in buying, taking title to and where customary, physically storing and handling goods and selling the goods at wholesale principally to retailers and industrial distributors and commercial users.' This group includes industrial distributors, exporters and importers, cash-and-carry wholesalers, drop shippers, wagon distributors etc.

6.2. *Sampling design.* The firm, and not the establishment, is considered the unit of sampling. The sources used for selecting the sample are : (i) list of wholesalers with paid employees obtained in the 1958 Census of Business; and (ii) Bureau of Old-age and Survivors Insurance list of wholesalers (with paid employees) entering business since 1958. All firms with sales above a certain value are definitely included in the sample and they number 1,400. From the population of wholesalers excluding these 1400 firms, a sample of 15,600 firms is selected using a stratified sampling scheme.

The firms are first stratified into kinds of business and then the firms within each kind of business are stratified on the basis of the sales figures. The number of firms selected from each kind of business varies from 150 to 650 depending on : (i) the number of firms in the kind of business; (ii) the distribution of firms by their sales; and (iii) whether geographic division trends are to be studied. Within the kind of business, the allocation to the strata is made in an optimum way with respect to the sales characteristics.

Out of the sample of 17,000, 1,400 large firms report their sales every month and the remaining 15,600 sample firms are divided into four equal panels each of which report every fourth month their sales for the preceding two months. Thus every month reports are received from a sample of 5,300 firms.

6.3. *Estimation procedure.* The preliminary and final composite estimates described in connection with the Retail Trade Survey are applied in this survey also because the firms in the rotating panels report for the preceeding two months. The coefficients used for w and k are 0.7 and 0.72 respectively.

Sampling errors are computed by kinds of business for each of the monthly estimates of rates, for the ratio of current month to previous month sales and for the ratio of the current month to year ago sales.

TABLE 6. COEFFICIENTS OF VARIATION FOR MONTHLY ESTIMATES OF SALE BY KINDS OF BUSINESS

sl. no.	kinds of business	coefficient of variation for July 1958–June 1959					
		monthly sales		ratio of current month sales to			
				previous month sales		year-ago sales	
		range	median	range	median	range	median
(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1.	automotive	4–11	7	2–6	3	4–11	8
2.	electrical and electronics	2–4	3	1–2	1	2–3	2
3.	furniture, home furnishings	2–5	3	2–5	3	3–5	3
4.	hardware, plumbing, heating goods	2–3	2	1–2	1	2–2	2
5.	lumber, construction materials	2–4	2	2–4	2	2–3	2
6.	machinery, equipment, supplies	2–3	3	2–2	2	2–3	3
7.	metals, metalwork (except scrap)	3–5	4	2–3	2	3–4	3
8.	scrap, waste material	3–5	4	2–3	2	4–4	4
9.	durable goods total	1.3–2.0	1.6	0.6–1.1	0.9	1.2–2.0	1.5
10.	grocery, confectionary, meat	2–2	2	1–1	1	1–2	2
11.	farm products (edible)	3–5	4	1–2	2	2–4	3
12.	beer, wine, distilled spirits	2–4	3	1–3	2	2–3	2
13.	drugs, chemicals, allied products	2–3	3	1–2	2	2–2	2
14.	tobacco	2–4	3	2–2	2	2–2	2
15.	dry goods, apparel	2–4	3	2–3	2	3–3	3
16.	paper, allied products	2–3	3	1–2	2	2–2	2
17.	coal	3–5	4	1–2	2	2–4	2
18.	farm supplies	6–8	7	2–5	4	5–8	6
19.	non-durable goods total	1.2–1.8	1.6	0.7–1.1	0.8	1.0–1.3	1.2
20.	merchant wholesalers total	0.8–1.5	1.0	0.6–0.8	0.6	0.6–1.1	0.8

Source : *Monthly Wholesale Trade Report Bureau of the Census*, BW-60-2, February, 1960.

Since estimates of sampling errors obtained from the sample are themselves subject to error, they are to be taken to give a general rather than an exact idea of the sampling variability.

Approximately 10% of the total sales are imputed because of non-response. The sampling variability does not reflect the bias that may be introduced due to this imputation of sales in case of non-responding firms. The estimate of volume of sales is not published separately in a few kinds of business where the non-response rate is considered too high to provide reliable estimates.

7. Survey of Manufacturing Industries

7.1. *Objectives.* Since 1949, The Census Bureau is conducting an annual survey of manufacturing industries on a sample basis to provide inter-censal estimates of shipments by product classes, and of total employment, total wages and salaries, man-hours worked and total volume of shipments and cost of materials on an establishment basis classified by industry. From these estimates the estimate of value added can be derived. Besides these, estimates are provided for metals consumed, inventories, plant and equipment expenditures, fuels and electrical energy consumed, and other items.

7.2. *Sampling design.* The sampling frame used is a list of establishments as obtained in the census supplemented by the establishments which have come into existence since the census from the Bureau of Old-Age Survivors Insurance lists of manufacturing establishments. The survey covers only establishments with one or more paid employees, as does the Census of Manufactures.

A sample of about 50,000 establishments is chosen from a population of about 300,000 establishments using a stratified systematic sample with approximations to optimum allocations to size strata. All multi-unit companies with at least one establishment with 250 or more employees and all establishments with 1,000 or more employees are definitely included in the sample. Within each of the 150 industry groups, the establishments are stratified according to the number of employees into 7 size classes. The allocation to these strata is done in an optimum manner. Seven different sampling fractions are used depending on the size class. These sampling fractions are modified in certain cases so as to ensure approximately the same coefficient of variation for estimates of each industry group total. The establishments falling in each class are arranged by industry group and a systematic sample with the assigned sampling fraction is selected.

7.3. *Estimation procedure.* A difference estimate of the current total is obtained by adding the difference between estimates of the totals for the current period and the recent census period from the same sample to the population total of the census period. This estimate is being used instead of a ratio estimate mainly because of its simplicity. Further some empirical studies conducted in the Census Bureau have shown that the ratio and the difference estimates substantially provide estimates of equal precision when the sample size is fairly large.

7.4. *Coverage check.* Coverage checks using area sampling have shown that about 90% of the establishments and about 99% of employment and wages are covered by the sampling frame used in the survey.

7.5. *Redesign of the survey.* Since 1949 when the original design was evolved, the survey has been redesigned twice, once in 1955 and then in 1959. The main objective of the redesign has been to obtain fairly reliable estimates for product classes. It may be noted that an establishment may have one or more products. The most important feature of the revised design is the construction of a suitable measure of size which when used for selection of the sample would provide estimates of change from year to year with prespecified precision in an optimum manner.

For each industry the regression between the absolute change from year to year and the base period value is studied and an empirical relationship of the form

$$|d_{ijc}| = a_i + b_i \, y_{ijc}$$

is derived, where d_{ijc} is the change in the value of the character under consideration from the last year and y_{ijc} is the last year's figure for the c-th product class for the j-th establishment in i-th industry and a_i and b_i are regression constants for that industry. The measure of size to be attached to a particular establishment is given by

$$\sum_{c} (a_i + b_i \, y_{ijc}).$$

It has been shown* that selection of units with probability proportional to this size gives rise to an optimum design. The coefficient of variation which can be tolerated for a product class is specified in the form $\frac{K}{\sqrt{X_c}}$, where K is a constant and X_c is the value of the characteristic for the product class 'c'. The sample size required is a function of K, the distribution of the value of the characteristics and the sampling design. Using the value of K and the supplementary information available for the base period, the optimum cut-off point, above which size the units are completely enumerated, is derived. From the remaining units, the required number of sample units is drawn with probability proportional to the measure of size mentioned earlier.

A point which needs attention is the selection of the sample by randomization of each establishment, i.e., if the proportion of the measure of size for an establishment to the total size is P_{ijc} then that establishment will be selected with probability P_{ijc} and rejected with probability $1-P_{ijc}$. Though this method is being used in the Census Bureau because of some operational considerations, it is reorganized that substantial gains in efficiency can be achieved if the sample is selected with probability proportional to size systematically after suitable arrangement instead of by the procedure of individual randomization.

8. Foreign trade statistics

8.1. *Source of data.* The Foreign Trade Statistics work relates to both the export and the import statistics. In this section only the export statistics part of the work is explained. It may be mentioned that similar procedures are proposed to be used for import statistics also.

* By Jack Ogus of the Bureau of the Census in an unpublished article.

The source of data for export statistics is the Collector of Customs to whom each person sending any merchandise worth $ 25 or more in case of commercial shipment and $ 100 or more in case of non-commercial shipments is required by law to file a Shipper's Export Declaration. In the Export Declaration, a description of each of the commodities shipped with a six-digit commodity classification code, the country of destination, the quantity, value and shipping weight of the merchandise are available. It is the responsibility of the Customs authorities to ensure that these forms are filled in properly.

Every month, the Bureau of the Census receives about 750,000 of these filled-in forms from the 300 ports in the country. This mass of data is edited and compiled using the electronic computer. The data on the value and the quantity are published for the 3000 × 145 possible commodity by country-wise classifications.

8.2. *Use of sampling.* Though the number of documents with value less than $500 is about two-thirds of the total number of documents, their contribution to the total values is less than 10%. To reduce the work of compilation and tabulation, all the documents with values less than $ 100 are completely ignored and a 10% systematic sample of documents with value between $ 100 and $ 500 is taken in case of all countries of destination.

8.3. *Method of presentation.* The commodity totals are the estimated figures which include the total value of all shipments with value $ 500 or more and the estimated total value for all shipments with value between $ 100 and $ 500. For each commodity the total figures for shipments with value $ 500 or more are given by country of destination with a sub-total for all the countries taken together. In case of shipments with value between $ 100 and $ 500, the estimates are presented for those countries of destination where the number of documents in the sample is 3 or more. For countries of destination with the sample size less than 3, the estimates are pooled up and presented as for "others". A sub-total of these estimated figures is presented for all the countries taken together.

8.4. *Confidence limits.* The confidence chart at 60% level for the binomial distribution has been used to get approximate confidence limits for the totals being estimated. If the number of sample documents for a particular class is 'n' the confidence limits for the total number of units say 'N' in that class can be obtained by referring the binomial confidence chart (say N_1, N_2). The approximate confidence limits for the total value is obtained by multiplying N_1 and N_2 by

$$\bar{X}\sqrt{1+V_x^2},$$

where $\bar{X}$ and V_x^2 are the average and relvariance of the values of all the shipments between $100 and $ 500 obtained from past data.

In the report, a table giving the deviations of the two limits from the estimate, are given for different values of n. For getting the confidence limits at 60% for a particular estimate x based on sample size n, one has to refer to the table and subtract from the estimate the value specified in one of the columns and add to the estimate the value specified in another column. The above procedure is applied as such in case of the commodity totals. In case of the commodity × country totals, as has been mentioned earlier, the estimates are published only if the sample size is 3 or more. So in this case the table giving the deviations to be used for arriving at the confidence limits have been worked out using the conditional binomial confidence chart. It is of interest to note that the decision to take the number 3 as the sample size below which estimates are not to be given has been arrived at on the consideration of mean square error using the distribution of the tabulation cells by the number of documents.

8.5. *Increase in sampling fraction.* Recently the sampling fraction for the documents with value $ 100 to $ 500 has been increased from 1 in 10 to 1 in 2 in case of all countries of destination except Canada in which case the original sampling fraction of 1 in 10 is retained. This increase has been effected to provide more reliable estimates.

9. ECONOMIC CENSUSES

9.1. *Scope.* The economic censuses covering the manufacturing, business and mineral industry establishments are conducted once in every five years. In these censuses all establishments with at least one paid employee are included. An almost complete list of such establishments is maintained by the Bureau of Old-Age and Survivors Insurance since all such establishments are required to file tax returns. Fairly detailed information such as number of persons employed, number of man-hours worked, wages and salaries paid, inventories, value added, etc. are collected from each establishment. The enquiry is carried out by mail.

9.2. *Census of manufacture.* The last Census of Manufacture was conducted in 1958. The questionnaires were mailed to the 300,000 and odd establishments early in 1959 with instructions to fill in and mail back within 30 days. It may be mentioned that out of these 300,000 establishments only about 95000 employ 20 or more employees. About 50 percent of the questionnaires were returned filled out within the specified period of 30 days. Follow-up work was taken up and by the end of August about 98 percent of the establishments had sent in their returns. In case of small non-responding units the figures were imputed on the basis of the figures for similar establishments in the responding population in respect of number of workers.

In case of establishments with 20 or more workers some broad scrutiny was done manually and for the smaller establishments a routine mechanical scrutiny was carried out. The four digited industrial classification and seven digited product classification are adopted. The industry classification of an establishment is determined by the machine on the basis of the products and the operations. In this Census about 240 different questionnaires were used. By December 1959, the preliminary report was ready and the detailed tables were expected to be ready by the middle of 1960.

9.3. *Census of business.* This Census is conducted on the same lines as the Census of Manufacture except that establishments with no employees are also covered. There are about 1,800,000 retail stores of which about 800,000 have no paid employees, 270,000 wholesale trading establishments and about 900,000 service trade establishments, of which about half have paid employees. The questionnaires are mailed only to the establishments with at least one paid employee found in the records of the Bureau of Old-Age and Survivors Insurance. For non-pay roll cases the information is obtained from the income tax returns. The data are usually published for 4 regions, 9 geographic divisions, States and the District of Columbia and for counties and cities over 2500 in population.

10. Census of population and housing

10.1. *Introduction.* The decennial Census of Population is required by the United States Constitution. The first census was taken in 1790 and since then censuses of population have been taken once in every 10 years. Many innovations in the collection and compilation of data have been introduced in the recent censuses. Attempts have been and are being made to make effective use of sampling, quality control procedures and high speed data processing equipment in carrying out the census.

The 1960 Census of Population and Housing represented a major advance as compared with the 1950 and earlier censuses. The research and evaluation work on the 1950 Census, the increased use of sampling developments and the application of electronic computer in large scale statistical operations were major factors influencing the new developments introduced in the 1960 Census. Following the general research and development phases that set the frame work for the new methods to be adopted, specific planning work for the 1960 Census of Population and Housing which is the 18th Decennial Census was initiated in January 1956. In the Spring of 1956, a Budget Review Committee was formed. By Fall of 1956 a number of Advisory Committees were brought into existence which included the Federal Agency Population Council and the Council of Population and Housing Census Users representing a number of federal agencies and the public. This shows that planning for the 1960 Census had begun well in advance.

Considerable pre-planning was necessary because major changes in the methods of collection and compilation of data were being contemplated for this Census. These changes were aimed at reducing the work, increasing the speed of publication of results and improving the accuracy of the census data.

10.2. *Complete and sample censuses.* In the 1960 Census of Population and Housing data on age, sex, colour, race and marital status and on some housing items were collected on the basis of complete enumeration whereas the data on employment status, occupation, industry, class of workers, education, income and place of work and other population items and a substantial number of housing items were collected only for a 25% sample of households.

The use of sampling in census work would help in (i) reducing the time interval between data collection and publication of results; (ii) reducing the over-all cost of operations; (iii) utilization of at least part of the cost reduction in improving the quality of census results and in inclusion of additional items or tabulations in the census. The experience of the 1950 Census in which data on migration, education, and income and some other items were collected from a 20% sample of persons has been satisfactory in that the use of sampling has not seriously limited the widespread and successful use of the results.

Though use of sampling introduces sampling variability in the results, the non-sampling variability which arises due to defective collection and processing of data is likely to be considerably less in the sample census than in the complete census especially for characteristics which involve difficult concepts and or which are difficult to remember. So in comparing the sample census and complete census for such items, one should consider the total variability which is usually known as mean square error and is the total of sampling and non-sampling variabilities.

TABLE 7. EMPLOYMENT STATUS BY SEX FOR 20% SAMPLE AND FOR COMPLETE CENSUS IN 1950—MASSACHUSETTS, SOUTH CAROLINA AND NEVADA

(in 000)

sl. no.	employment status and sex	Massachusetts		South Carolina		Nevada	
		20% sample	complete census	20% sample	complete census	20% sample	complete census
(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1.	males 16 years and over	1710	1733	682	688	64	65
2.	labour force	1315	1331	550	553	52	53
3.	civilian labour force	1293	1308	534	537	50	50
4.	employed	1208	1226	516	520	46	47
5.	unemployed	85	84	17	17	3	3
6.	not in labour force	395	403	133	136	11	12
7.	females 16 years and over	1906	1906	736	733	56	56
8.	labour force	632	631	247	246	18	18
9.	civilian labour force	631	630	246	245	18	18
10.	employed	603	603	236	236	17	16
11.	unemployed	28	28	10	10	1	1
12.	not in labour force	1274	1275	489	488	38	38

Source : 1950 Census of Population, Vol. II, Tables 25 and 60.

TABLE 8. EXPECTED RELATIVE MEAN SQUARE ERROR OF ESTIMATED CELL FREQUENCIES FOR INDIVIDUAL ITEMS BASED ON A COMPLETE CENSUS AND ON A 25% SAMPLE OF HOUSEHOLDS

population of area	2500		10000			50000		
	RMSE with		cell frequency	RMSE with		cell frequency	RMSE with	
cell frequency	complete census	25% sample		complete census	25% sample		complete census	25% sample
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
12	7	10	50	14	20	250	34	46
50	14	19	200	30	40	1000	85	105
125	22	31	500	52	67	2500	180	200
500	49	62	2000	160	160	10000	620	650
1250	89	102	5000	320	330	25000	1520	1530

Note : Computations assume a relative response bias of 6% and response variance equal to the sampling variance for a 25% sample of households.

Source : Proposal for extension of sampling and related changes to accelerate completion and reduce costs in the 1960 censuses of population and housing, Table B, June 1957.

10.3. *Data collection.* The work of data collection for the 1960 Census of Population and Housing was generally done in two stages. In the first stage the enumerators collected the data on 100% items and during the first stage enumeration they left sample questionnaires in every fourth household with the request to fill-in and return by mail. During the second stage, the enumerators contacted those sample households which either did not return the filled-in questionnaire or sent in incorrectly filled-in questionnaires. In areas which are sparsely populated, the work of both the stages was done in one operation. Such areas accounted for about 15 to 20 percent of the whole country.

In this census, the technique of self-enumeration was adopted supplemented by enumerative follow-up where necessary. First, during the last week of March 1960, the U.S. Post Office delivered to all the addresses in the United States a questionnaire, called Advance Census Report (ACR), containing questions relating to relation to head, sex, colour (or race), date of birth and marital status and some housing items. The head of the household was requested to fill in the questionnaire and hold it for the census enumerator. During the first stage, the enumerators visited all households in the country and in cases where the ACR was already filled in, reviewed and transcribed the answers to a special form, called FOSDIC (Film Optical Sensing Device for Input to Computers) schedule. In cases where the ACR was not filled out, the census enumerator obtained the information directly on the FOSDIC schedule.

In the second stage, the census enumerator transcribed the answers in the filled-in sample questionnaires to the FOSDIC schedule and in cases where the questionnaire was either not returned or returned incorrectly filled-in, the enumerator obtained the required information in the FOSDIC schedule by personal interview or by telephone.

The census enumeration was done with reference to the census date 1st April 1960. Besides enumerating all persons living in the household on the census date, the temporary absentees and visitors were also enumerated. An effort was made to enumerate all the persons in transient places such as hotels, motels etc., on the night of 31st March with a view to ensure that all people get enumerated. For visitors and the transient guests in hotels, motels etc., Individual Census Reports (ICR) were filled-in which were sent to their places of usual residence to ensure their enumeration once and only once in the census.

10.4. *Field organization and training.* The over-all supervision of the field work of the census was done by the 17 regional offices of the Census Bureau. The field organization for the census consisted of about 400 district offices each in the charge of a District Census Supervisor. There were about 10,000 Crew-Leaders who trained and supervised the work of about 160,000 Enumerators. Each District Census Supervisor was assisted by a Technical Officer in technical matters regarding training, data collection and field review. Similarly each Crew-Leader was assisted by a Field Reviewer in reviewing the work of the enumerators.

The whole of the area of the United States was divided into small areas, called Enumeration Districts (ED), which could be well identified on the field. One or more ED's were grouped to form an Enumeration Assignment (EA) to be assigned to one enumerator in the census. On an average each EA consisted of about 300 to 400 households.

The Bureau of the Census prepared detailed instructions for the field work at different levels. The Census Bureau staff trained about 40 Chief Instructors who in their turn trained the 400 and odd Technical Officers. The Technical Officers imparted training to the 10000

Crew Leaders who in turn trained the 160000 Enumerators for the stage 1 work. After the completion of stage 1 work, about one-third of the enumerators were trained for stage 2 work.

10.5. *Tabulation.* The data collected in the FOSDIC schedules during the stage 1 enumeration is subjected to a field review. Then the filled-in FOSDIC schedules are microfilmed. The data in the microfilms are automatically transferred to the magnetic tape on the FOSDIC machine. A mechanical edit is carried out on the census data using the electronic computer to search for inconsistencies in the data. The computer is instructed not only to find any inconsistency present in the data but is also instructed to suitably modify or substitute the figures wherever necessary. A complete record of control totals and of editing and imputations for each small area is left by the computer and significant discrepancies in control totals and excessive amount of editing are investigated. After the edit, the computer tabulates all the information collected in stage 1 and the high speed printer is instructed to print out the data in the form suitable for publication.

Use of the FOSDIC schedule has eliminated the punching operation which is time consuming and which is liable to introduce errors. Further the exclusion of industry-occupation and any other items requiring manual coding from the 100% questionnaire eliminates the need for coding operation for stage 1 data thereby speeding up the tabulation. The use of electronic equipment for data processing has made it possible to furnish the population counts not only for the country as a whole, but also for all counties, cities and towns and other small political units by the end of October 1960.

The same procedure of FOSDIC, microfilm, magnetic tape and use of computers is proposed to be used to tabulate the stage 2 data also. It may be noted that stage 2 data have to pass through a coding operation.

10.6. *Quality control procedures.* Quality Control (QC) procedures are being widely used in the 1960 Census to control the errors at different levels. Quality Control procedures have been instituted in : (i) printing and cutting of FOSDIC schedules, (ii) collection, assembly and binding of schedules into enumerator books, (iii) preparation of enumerator kits, (iv) enumeration, (v) coding (vi) microfilming, (vii) FOSDIC reading, and (viii) computer tabulation.

Quality Control procedures are being used to improve the outgoing quality of the data by finding and eliminating defects from unacceptable lots and by removing unacceptable producers (by termination or retraining). Details of some of the quality control procedures are given in Section 13.

10.7. *Evaluation programme.* A fairly comprehensive evaluation programme has been evolved for the 1960 Census of Population and Housing to assess the coverage and the content errors. By coverage error in the population census is meant error in counting people or households. The content error is the error in the data collected for the population covered in the census arising from wrong reporting, recording, transcribing, coding, and tabulation. The gross error in the results which is the sum of the erroneous inclusions and omissions can under certain circumstances be taken as an indication of the response variance in the results and the net error can be taken to reflect the bias of the techniques used in data collection and tabulation. This matter is discussed in greater detail in Section 12.

It may be mentioned that a Post Enumeration Survey (PES) was conducted after the 1950 Census of Population to assess the errors in the census. The best available evidence shows that the PES under-estimated the under-enumeration by about 50% because the coverage of the PES was especially deficient in population classes where the risk of under-enumeration in the 1950 Census was the highest. Some evaluation programmes are devised in the 1960 Census to take care of this deficiency.

10.8. *Reverse record check.* In this programme, it is proposed to select a sample of persons from the 1950 Census list supplemented by the list of persons who entered this country after April 1950 and who are registered in January 1960, children born since April 1950 and persons omitted from the 1950 Census but detected by the 1950 PES. An attempt would be made to obtain the addresses of the sample persons and to determine whether they are enumerated in the 1960 census or not. This method is likely to overcome the defect of the PES, namely of the PES tending to miss the same category of persons in the Census. A pilot study revealed that it would be possible to contact at least 80% of the sample persons.

10.9. *Reverse record check of special sample.* This is essentially similar to the study mentioned above but is confined to three special population groups, viz, (i) aged social security beneficiaries; (ii) selective service registrants; and (iii) students enrolled in colleges and universities in February 1960. From each of these groups a sample would be selected and attempts would be made to determine whether the sample persons are enumerated or not in the 1960 census.

10.10. *Re-enumerative studies for coverage error.* The objective here is to obtain estimates of the gross and the net errors in coverage in the census. It is proposed to re-enumerate the population in an area sample using intensively trained enumerators. An estimate of under-enumeration is proposed to be obtained by a *de-facto* enumeration of the population and by verifying whether they have been enumerated or not in the census.

At the time of re-enumeration of the sample areas, the enumerators would be required to locate the housing unit which immediately precede and follow each of the enumerated units. This would help in estimating the under-enumeration of housing units in the census. Provision has also been made to study the over-enumeration by the method of re-enumeration.

It was proposed to use the post office to get an idea by asking them to identify households not enumerated in the census. The proposal was that the post office would be provided with one card for each address enumerated in the census and that the post office would identify the omitted housing units in the process of delivering the census cards. But this project was dropped due to lack of funds. However, a small scale project was undertaken to study the feasibility and effectiveness of the use of post office personnel to improve coverage. In this study a sample of the census cards were withheld as a check over the work of the post office personnel.

10.11 *Measurement of content errors.* Intensive reinterviews were conducted for a sample of 5000 households included in the 25% sample by specially trained enumerators. Some of these enumerators were not furnished with the census data already collected and the data obtained at the reinterview were to be matched with the data collected in the census. Cases of discrepancies would be referred back to the field for reconciliation.

Another study of the content error consists in comparing the data collected in the CPS with that obtained in the census. The CPS-census match would be confined to only the 25% sample households included in the CPS sample which would be about 8000 households.

10.12. *Processing-error studies.* In the first stage of the census, the enumerator is required to transcribe the information from ACR to the stage 1 FOSDIC schedule. In the second stage, the enumerator is required to transcribe the information from the sample questionnaire to the stage 2 FOSDIC schedule. A review would be made of a sample of this work to determine the extent of gross and net errors arising from the transcription operation.

Two coding operations are involved in stage 2 of the census. First is a general coding of detailed relationships, place of birth, place of work and income and the second is a specialized coding of industry and occupation, both are done by specially trained coders. It is proposed to estimate the gross and the net errors in coding by independently recording the data obtained in the census for a sample of households and comparing the records with the original codes. Discrepancies would be investigated and reconciled.

11. Census of agriculture

11.1. *Objective and scope.* Since 1840, agricultural censuses have been taken every 10 years until 1920 from which year the agricultural census is made a quinquennial census instead of a decennial one. At present the agricultural census is taken in the years ending in 4 and 9. The objective of the census is to obtain data on the number of farms, utilization of land, amount of farm products produced and sold, livestock number, farm implements, farm wages and expenditure by farmers. The last census was taken in 1959 and a brief description of this is given here.

All dwelling units and places with agricultural operations are covered by the census. For the purpose of the census, a place is said to have agricultural operation if (i) one or more hogs, cattle, sheep, goats, horses or mules are kept; (ii) a combined total of 20 or more chickens, turkeys, and ducks are kept; (iii) any grain, hay, tobacco or other field crops are grown; (iv) a combined total of 20 or more fruit trees, grape vines; or (v) any vegetables, berries, or nursery or greenhouse products are grown on the place for sale. Though this is the definition followed for getting information in the census by the enumerator, the definition of a farm adopted at the processing stage is the following :

(i) places with 10 or more acres, with at least $ 50 sales of agricultural products (not more than half of this value can be in forest products); or

(ii) place with less than 10 acres with at least $ 250 sales of agricultural products (not more than half of this value can be in forest products).

About 90% of the questionnaires filled-up in the field meet this definition and get tabulated.

11.2. *Method of enquiry and field staff.* The data in the census are collected by enumerators by interviewing the farm operators. In 1959, the data were collected by a crew of 30,000 enumerators under the supervision and guidance of about 2,100 crew leaders who in their turn were trained and supervised by about 75 technical instructors. Each enumerator worked in an area, called Enumeration Assignment. The enumerators were paid on piece-rate basis, but in some areas they were paid by the hour.

11.3 *The questionnaire.* The questions included in the census were selected from requests and suggestions received from many sources, such as the Department of Agriculture, State Agricultural Colleges, farm organizations etc., in consultation with a special advisory committee which included representatives from a number of agricultural organizations such as American Farm Bureau, Federal Farm Equipment Institute, Census Advisory Committee of the American Statistical Association etc.

The total questionnaire consists of 14 effective sections and 310 questions. There are 38 different versions of this questionnaire in which some of the questions have been omitted taking into consideration the region in which the questionnaire was to be used. The topics covered by the different sections are (i) ownership, (ii) crops harvested, (iii) land used, (iv) irrigation, (v) race, age, residence, off-farm work and other income, (vi) forest products, (vii) poultry and livestock and livestock production, (viii) dairy products, (ix) animals sold, (x) fertilizers and lime, (xi) selected farm expenditure, (xii) farm labour, (xiii) equipment and facilities and (xiv) rental agreement, farm values and mortgage debt.

It may be mentioned that the items (viii) to (xiv) are collected only for a 20% sample of farms. The farms with serial numbers 2 and 7 are included in the sample. Besides this questionnaire, there is another questionnaire which is filled for each landlord who has any land worked on shares by one or more tenants, renters or sharecroppers.

11.4. *Reference period.* The reference period for the data collected in this census is 1959. The census is conducted in the fall of that year. Though for many of the items it may be possible to give the figure for the year in the fall, there are items for which an estimate has to be made by the farmer as the exact figure may not be available at the time of the census.

11.5. *Self-enumeration.* Self-enumeration was attempted for the first time in 1950 by sending the questionnaires to all farmers except in the Southern States and has been continued since then. The farmers were required to fill out the questionnaire. The enumerators were instructed to collect them and to collect the information in cases where the questionnaires were either not filled out or filled out incorrectly. About 50% of the farmers, to whom the questionnaires had been sent, filled out the questionnaires by the time of the census.

11.6. *Evaluation programmes.* Evaluation Studies have been conducted in each census since 1945 to assess the coverage and content errors. The coverage study conducted on a limited scale in 1945 indicated an under-enumeration of the farms by 10%, which under-enumeration was more pronounced in the class of relatively smaller farms and less among larger farms. A brief description of the evaluation studies conducted since 1945 is given below.

11.7. 1950 *Post-Enumeration Survey.* In the 1950 Census of Agriculture which was conducted along with the Census of Population in the Spring, a Post-Enumeration Survey (PES) was conducted in about 4000 sample segments drawn from about 300 sample primary stage units. The rural part of the sample consisted of about 1200 segments selected at the rate of 1 in 950 and the sample segments had about 5500 farms.

The rural enumerators were supplied with identification particulars and original census data for an associated sample of farms obtained from the census list. The enumerators were instructed to (i) search for farms and households in the sample segment but not in the associated sample; and (ii) search for content errors in the questionnaires filled out in the cousus for the farms in the associated sample. Though the original census data were provided, the enumerators were instructed not to look at them until they have re-enumerated the farm and then to reconcile the two responses. The re-interview was confined to data on land utilization and production and did not cover livestock data.

Some main results of the PES are given in Table 9. It may be mentioned that gross error before reconciliation was greater for farms where no census data were provided than for farms where census data were made available at the time of the PES and that the gross under-enumeration among farms under 10 acres was the largest with 16.5% with a standard error of 4.6%. 58% of the over all under-enumeration was due to overlooking of the farm at the time of the census enumeration. 39%was due to exclusion at the tabulation stage and the remaining 3% was due to enumeration as part of another farm.

TABLE 9. ESTIMATE OF PERCENT OF UNDER-ENUMERATION AND THE CORRESPONDING STANDARD ERROR FOR SOME IMPORTANT CHARACTERISTICS

sl. no.	characteristics	percent of under-enumeration	standard error
(0)	(1)	(2)	(3)
1.	number of farms	5.1	0.7
2.	land area	2.0	0.5
3.	crop area	2.1	0.6
4.	corn harvested area	7.9	3.4
5.	wheat harvested area	1.6	1.8
6.	cotton harvested area	7.9	3.4

Source : Evaluation of United States censuses of agriculture 1945-1950, (Draft).

11.8. 1954 *Evaluation programme for agriculture.* After the 1954 Census of Agriculture which was conducted in the late Fall, evaluation studies were conducted in a sample of about 3500 farms in about 760 rural segments selected with an over-all sampling fraction of 1 in 1500 from 260 sample primary stage units. Associated samples were provided for the size classes less than 1000 acres, 1000 to 10000 acres and greater than 10000 acres. The third class was represented by a separate sample of 365 farms drawn in clusters of 5 from 73 counties having large farms with a sampling fraction of 1 in 20. The rural area was represented in the evaluation programme by a sample of about 3000 CPS segments.

In this study the enumerators were not provided with any census data. After the re-enumeration, the data collected were matched with the census data and in case of discrepancies, the enumerators revisited the concerned farms to determine the reasons for the discrepancies. About 75 experienced enumerators worked in the first phase of this study and about 20 of these 75 enumerators worked in the second phase. The main results of this study are given in Table 10. The reinterview did not include crop production and livestock.

TABLE 10. ESTIMATE OF PERCENT OF UNDER-ENUMERATION AND THE CORRESPONDING STANDARD ERROR FOR SOME IMPORTANT CHARACTERISTICS

sl. no.	characteristics	percent of under-enumeration	standard error
(0)	(1)	(2)	(3)
1.	number of farms	8.1	0.9
2.	land area	5.4	1.9
3.	crop area	4.0	1.1
4.	corn harvested	3.4	1.2
5.	wheat harvested area	5.3	4.1
6.	cotton harvested area	0.9	1.5

Source : Same as for Table 9.

11.9. 1959 *Evaluation programme for agriculture.* The evaluation study for the 1959 Census of Agriculture consisted of pre-census enumeration in half the sample and post-census enumeration in the other half. This study was conducted in a sample of 772 rural segments. The farms with area greater than 5000 acres were represented by a separate list sample drawn from census records of large farms in selected counties.

12. NON-SAMPLING ERROR

12.1. *Introduction.* In recent years increasing attention is being paid to the assessment and control of non-sampling errors in general and of response error in particular in conducting large scale censuses and surveys. The realisation that non-sampling error would be considerable and difficult to control in large scale operations as the complete censuses is exerting a strong influence on the decision to have a well-controlled sample survey or a complete census to obtain specific items of information.

It is of interest to note that as long back as 1871 General Francis A. Walker, Director of the 1870 Census, had realised the importance of non-sampling errors when he wrote the following.*

"In such a state of things it would seem to be the duty of those charged with the publication of these statistics to indicate in respect to each class the degree to which the figures may be relied upon, and as nearly as may be practicable, the proportion of omission or error. It is undoubtedly true that many will by such a course become advised of these deficiencies who never would have discovered them. It is probably true also that many persons will, when candidly advised of the necessary limitations of such statistics proceed to the conclusion that they are worthless, and thus reject the whole. It is unquestionable, that the results of the Census would obtain more credit if put forth without any admissions or exceptions; but I have not deemed such a course fair to the public."

12.2. *Concepts and definitions.* Non-sampling error occurs at almost all the stages of planning and execution of large scale data collection and compilation work. Non-sampling error may be considered to consist of non-sampling bias and non-sampling variation. The former is a relatively old concept compared to the latter and a number of studies have been conducted in the Census Bureau and elsewhere in the last two decades or so to estimate this type of bias. The need for assessment of non-sampling variation is increasingly being felt in the recent years.

Non-sampling bias is defined as the difference between the average value over all possible attempts at data collection and compilation under essentially the same conditions, which include the sampling design and the estimation procedure in case of sample surveys, and the value proposed to be estimated. As compared to this, sampling bias is defined as the average value over all possible samples. Non-sampling bias for any particular situation can be estimated by comparing the value obtained in one operation with that obtained by repeating this operation with more care and better control. The difficulty with this procedure in some situations arises due to the possibility of the second improved repetition of the operation being conditioned by the original operation.

* Quoted by Eckler (1953).

The total variation in a survey result is defined as the variation between the results based on all possible attempts at data collection and compilation under essentially same conditions which include the sample design and the estimation procedure in case of sample surveys. This total variation can be analysed into three components (i) sampling variation, (ii) non-sampling variation and (iii) interaction between the sample and the data collection and compilation process. If necessary the non-sampling variation may further be broken into the variations contributed by different stages of operation, such as, data collection, coding, punching, tabulation etc., and possible interactions thereof.

The bias and variation introduced at the data collection stage are termed response bias and response variation respectively. In this chapter brief descriptions of some of the studies conducted and being conducted to assess the response errors are given and in the next chapter some methods of controlling these errors would be given. Some studies for evaluating the response bias have already been mentioned in Sections 3.10 and 3. 11.

12.3. *Post-Enumeration Survey* (1950 *Census*). The 1950 Census Post-Enumeration Survey (PES) consisted of reinterviewing a sample of about 30000 persons in 470 of the 3000 and odd counties in this country by about 250 experienced and well trained interviewers. The gross and net errors of classification in the 1950 census based on the PES are given in Table 11 for some items to indicate the extent to which response bias may affect the results of the Census.

TABLE 11. GROSS AND NET ERRORS IN THE 1950 CENSUS OF POPULATION BASED ON THE POST-ENUMERATION SURVEY

sl. no.	classification	number of classes	number of cases[1]	percent mis-classified	typical class		
					number in class	net error in class total	
						amount %	standard error %
(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1.	age (5 year class)	16	26,950	6	1,686	0.8	0.8
2.	occupation group[2]	10	9,502	15	950	1.8	1.8
3.	industry group[2]	14	9,464	11	676	1.9	1.8
4.	individual income[3]	15	9,012	31	601	10.8	3.2
5.	number of rooms[4]	9	19,300	18	2,144	2.5	1.3

[1] Approximate number of cases in PES sample for which both census and PES data were available for each classification.

[2] Employed workers 14 years of age and over.

[3] Persons 16 years of age and over (sampling fraction one-half that used for first 3 items).

[4] Occupied dwelling units (sampling fraction roughly twice that used for first 3 times).

Source : Eckler, A. R. and Hurwitz, W.N. (1957) : Response variance and biases in censuses and surveys, Table 4.

12.4. *Comparison of* 1950 *Census with the CPS*. Since the results for labour force obtained from the CPS can be expected to be more reliable than those obtained in the census, the figures obtained in the 1950 census and the CPS for labour force characteristics were compared to get an idea of the response bias in the census. The results of this comparison are given in Table 12.

TABLE 12. RELATIONSHIP BETWEEN THE FIGURES FOR LABOUR FORCE CHARACTERISTICS FROM THE 1950 CENSUS AND THE CPS (APRIL 1950)

sl. no.	labour force status	net error in census as percent of CPS	
		amount	standard error
(0)	(1)	(2)	(3)
1.	civilian non-institutional population 14 years and over	+ 0.7	(a)
2.	in labour force	− 5.0	0.5
3.	employed	− 4.1	0.5
4.	unemployed	−19.4	3.7
5.	not in labour force	+ 8.2	0.7

(a) The CPS estimates are adjusted by a ratio estimation procedure, to an independent estimate of the civilian non-institutional population 14 years of age and over and hence this total is not subject to sampling variance, though it is, of course, subject to errors of other sorts.

Source : Eckler, A. R. and Hurwitz, W.N. (1957), Response variance and biases in censuses and surveys, Table 5.

12.5. *Enumerator variability study.* A study using the method of interpenetrating sub-samples (IPS) was conducted during the 1950 Census in 24 counties in Ohio and Michigan to determine the effect of enumerator variability on the results. More than 700 enumerators participated in this study. A total of 125 geographical areas with an average population of about 6500 were taken from the 24 counties. Within each geographical area a number of enumeration districts (ED) were formed which were paired and assigned to the enumerators in that area at random.

The enumerator variance was calculated separately for each of the 125 areas and the average of this variance over the areas was compared with the variance between the ED's completed by the same enumerators. This study was conducted for about 100 items and for almost all the items the *F*-test indicated statistical significance. It has been shown* that the total variation including both sampling and response variations in the case of estimating a population proportion P is given by $\frac{PQ}{n}$, where n is the sample size and that the response variation is given by

$$\sigma\frac{2}{d}=\frac{\sigma_d^2}{n}\,[1+\rho(n-1)],$$

where ρ is the intraclass correlation among the response deviations in a survey and $\sigma_d^2 = PQ-\sigma_s^2$, σ_s^2 being the sampling variance for a sample of one unit. Approximations to σ_d^2 and PQ have been worked out for a number of characteristics covered in the 1950 census from the PES. It is found that σ_d^2 is of the order of 5 to 10 percent of PQ for characteristics which could be measured with high reliability and is of the order of 50 percent of PQ for characteristics which could be measured with low reliability.

* Hansen, Hurwitz and Bershad (1961).

12.6. *Comparison of complete census and sample survey.* Under certain specified conditions the mean square error of the sample proportion is given by

$$\text{mse}(p) = \frac{\sigma_d^2}{n}[1+(\bar{n}-1)\rho]+ \frac{N-n}{N-1}\frac{\sigma_s^2}{n}+B^2,$$

where N is the total number of persons in the area, n the number of persons in the sample ($n = N$ in the census) and $\bar{n}$ the number of persons covered by each enumerator. Table 13 presents a comparison of the results of a complete census and of a 25% sample survey assumed to be conducted under substantially the same conditions. The values taken for the parameters occurring in the above formula are given below: $\bar{n} = 1000$ (census), $\bar{n} = 250$ (sample), $P = 0.04$, $B = 0.003$, $\sigma_d^2 = 0.13$ $PQ = 0.005$, $\rho = 0.03$, $\sigma_s^2 = PQ-\sigma_d^2 = 0.033$.

TABLE 13. COMPARISON OF THE TOTAL VARIATION IN THE RESULTS OF A COMPLETE CENSUS AND OF A 25% SAMPLE SURVEY IN ESTIMATING A POPULATION PROPORTION FOR DIFFERENT POPULATION SIZES

sl. no.	population size	complete census					25% sample survey				
		n	σ_d^2	σ_s^2	B^2	$\frac{\sqrt{\text{MSE}}}{P}.100$	n	σ_d^2	σ_s^2	B^2	$\frac{\sqrt{\text{MSE}}}{P}.100$
(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
1	1000	1000	.000165	—	.000000	32	250	.000169	.000099	.000009	42
2	2500	2500	.000062	—	.000009	21	625	.000068	.000060	.000000	4 27
3	5000	5000	.000031	—	.000009	16	1250	.000036	.000020	.000009	20
4	10000	10000	.000016	—	.000009	13	2500	.000017	.000010	.000009	15
5	25000	25000	.000006	—	.000009	10	6250	.000007	.000006	.000009	11
6	100000	100000	.000002	—	.000000	8	25000	.000002	.000001	.000009	9

Source: Hanson, M. H., Hurwitz, W. N. and Bershad, M. A. (1961) : Measurement of errors in censuses and surveys. *Bull. Int. Stat. Inst.*, 38, (2), 359-374.

12.7. *Measurement of non-sampling variability in the 1960 Census.* A comprehensive experiment is being conducted in the 1960 Census to measure the variability introduced at different stages of the operation. Two methods are being used (i) replication method where the data are collected more than once from the same units and (ii) where two or more random groups or interpenetrating sub-samples of units are formed at each stage of operation and assigned to different enumerators, supervisors, etc.

Method (i) is being resorted to in two experiments. The first experiment consists in requesting a sample of 1,000 sample households, who have already returned the filled-in sample questionnaire, to report again. In the second experiment, the second-stage enumerators would visit a sample of 5,000 sample households who have returned the sample questionnaires and collect the data once again from them.

In the experiment, where method (ii) is being used, pairs of enumerators would be purposively assigned to clusters of contiguous enumeration districts (ED's) in the selected areas in the same manner as in the other areas; within each cluster the 25% sample households would be divided into two random groups one of which would be assigned to one enumerator at random and the other would be assigned to the other enumerator; in half the clusters one of the enumerators would be assigned at random to a crew leader of a neighbouring crew-leader district and the other enumerator would be supervised by the regularly assigned crew-leader. This study would include 100 crew-leaders, 1,600 enumerators and 320,000 households.

12.8. *Response error in milk and milk products.* Some studies conducted on the data on milk and milk products collected in the 1945 Census of Agriculture give an idea as to the extent of the response error in the data. Some results of the effect of the 1945 Census editing on the statistics for certain dairy items in four counties are given in Table 14. Table 15 gives an idea of the response or enumerator error in reporting average annual production per cow milked.

TABLE 14. EFFECT OF THE 1945 CENSUS EDITING ON THE DATA FOR CERTAIN DAIRY ITEMS IN FOUR COUNTIES

sl. no.	county	percent changes introduced by editing					
		milk produced		whole milk sold		cream sold	
		gross increase	gross decrease	gross increase	gross decrease	gross increase	gross decrease
(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1.	Grundy, Illinois	11.7	22.5	16.9	27.0	15.2	4.6
2.	Marshall, Kentucky	1.6	2.2	12.1	21.8	2.9	12.4
3.	Baker, Oregon	24.4	6.4	28.6	10.7	11.4	2.6
4.	Foster, N. Dakota	1.2	2.7	70.6	1.0	1.4	0.3

Source : Same as for Table 9.

TABLE 15. PERCENTAGE OF FARMS REPORTING THE SAME FIGURE FOR THE AVERAGE ANNUAL PRODUCTION PER COW

sl. no. of enumeration district	average annual milk production per cow commonly reported (gallons)	farms reporting this figure	
		number	percent of total
(0)	(1)	(2)	(3)
1.	300	3	8.1
2.	400	22	71.0
3.	400	24	92.3
4.	600	14	37.8
5.	400	22	88.0
6.	300	5	17.2
7.	450	5	20.0
8.	405	35	97.2
9.	405	25	80.6
10.	405	11	50.0
11.	400	15	46.9
12.	800	9	27.3
13.	400	28	87.5
14.	400	10	41.7
15.	405	24	85.7
16.	405	21	83.8
17.	600	18	52.0
18.	400	16	53.3
19.	500	3	60.0

Source : Same as for Table 9.

12.0. *Agriculture variance study* (1959 *Census*). In the 1959 Census of Agriculture, a study is being conducted to evaluate the enumerator contribution to the variance of an estimate. This study is being conducted in 10 counties consisting of 105 Enumerator Assignment (EA) and 18000 farms. Two enumerators are assigned to each Enumerator Assignment.

The alternate sample farms in the 20% sample are alloted to the two enumerators. The remaining 80% of the farms are grouped into groups of 4 farms and in each group two farms are assigned at random to one enumerator and the other two farms to the other enumerator. It is proposed to analyse the results for each ED in the form of an analysis of variance table.

13. Quality control in surveys

13.1. *Introduction*. Though Statistical Quality Control (SQC) techniques are being used in industry to assess and to control the quality of finished products, application of these techniques to control the quality of data collection and compilation is relatively new. Many of the quality control programmes put into operation by the Bureau of the Census have demonstrated their utility in controlling the quality at the various stages of census and survey work.

Among the various applications of these techniques to the work of the Census Bureau, special mention may be made of the use of SQC in controlling the quality of data collection and coding in the Current Population Survey and of the comprehensive quality control programme which covers different aspects of the work relating to the 1960 Census of Population and Housing, namely, (i) printing and cutting of FOSDIC enumeration schedules: (ii) collection, assembly and binding of schedules into enumerator boxes; (iii) preparation of enumerator kits; (iv) enumeration; (v) coding; (vi) microfilming the FOSDIC schedules (vii) FOSDIC reading; and (viii) computer tabulation.

SQC is applied to census and survey work with the main object of process control through identifying points at which processes go outside of tolerances that have been established and taking corrective action for increasing the outgoing quality with corrective action. For this purpose, usually SQC plans which have a built-in device to initiate corrective action are used. In the Census Bureau, more attention is given to process control through SQC than to use of acceptance plans on the finished work. For a particular situation, the best plan is defined as that which ensures the highest out-going quality for a given cost or the lowest cost for a specified outgoing quality.

No attempt would be made here to describe in detail all the various SQC plans being used by the Census Bureau. A good illustration of the applications of the SQC to survey work is provided by the use of quality control techniques in the coding operation of the Current Population Survey. A fairly detailed description of this situation and a brief description of some other selected SQC programmes are being presented here to give an idea of the type of plans used.

13.2. *Coding work in CPS*. The coding operation in the CPS consists in coding of about 51,000 person-schedules in each month during a period of one week. Once in every three months, both industry and occupation are coded whereas in the remaining months only industry is coded. The coders receive the work in lots of about 250 schedules. The standard verification operations consist of verifying the codes given by the coder by the verifier and of preparing a report for each lot giving details of discrepancies noted and corrected.

As stated earlier the sampling plan is used to control the process rather than to decide whether to accept or reject the finished work. The plan controls the work of individual coders instead of just controlling the over-all quality. In order to select a suitable plan it was necessary to specify the level of accuracy needed. After studying the past error records, it was decided to have the acceptable quality level at 0.75 errors per 100 schedules in normal months when only industry is coded and 1.50 errors per 100 schedules in quarterly months when both industry and occupation are coded.

Each coder's work is initially verified completely for a period of 3 months. If his average quality for this period is better than the acceptable quality and if he has not exceeded the acceptable error rate in more than one month during this period he is labelled as a *Q* (quality) coder. Further, the coder should have coded more than a specified number of schedules during the initial 3 month period for being considered for the qualification of *Q* coder.

In case of a *Q* coder, a 10% sample of schedules from each of the lots completed by him is verified every month. A cumulative record of the performance of each of these coders is maintained. This cumulative error rate is compared with a table providing allowable error rates and the *Q*-coder retains his status as long as the cumulative error rate is below the allowable error rate. If at any time he exceeds the allowable error rate, he is disqualified from being a *Q*-coder and his work is verified completely. A disqualified coder should code at least 2,000 schedules with acceptable quality before he can be requalified. The allowable error rates in case of *Q*-coders are worked out in such a way that the chance that a *Q*-coder whose true error rate is of acceptable quality would be disqualified is about 1 in 20. Any coder who has not been able to qualify initially or who has been repeatedly disqualified is removed from the operation.

Since it is not an acceptance sampling plan in the usual sense some lots with a high error rate may be allowed to pass through. However, since the standards for initial qualification and for remaining qualified are strict it is felt that the chance of the outgoing quality being unsatisfactory is negligibly small. An estimate of the outgoing quality with this plan is approximately given by

$$\frac{N_q}{N}\,\frac{9}{10}\,\frac{e_q}{n_q}\,100,$$

where N is the total number of schedules coded, N_q is the number of schedules coded by *Q*-coders, n_q is the number of verified schedules coded by *Q*-coders and e_q is the number of errors reported for *Q*-coders.

TABLE 16. ESTIMATES OF AVERAGE OUT-GOING QUALITY IN THE CPS CODING WORK, 1956

sl. no.	month	average outgoing quality number of errors per 100 schedules
(0)	(1)	(2)
1.	June	0.19
2.	July	0.35
3.	August	0.17

Source : Current Population Survey (Draft) by Steinberg.

13.3. *Re-interview in CPS.* The reinterview programme in the CPS consists of reinterviewing one-third of the sample households canvassed by one-fourth of the enumerators every month. In case of discrepancies between the original interview and reinterview an effort is made to reconcile the figures by probing questions and if the figures could not be reconciled the figures obtained at the reinterview by the Supervisor are taken as reliable. The reinterview includes checks for coverage of units, coverage of persons and content of schedules.

The types of possible errors are prespecified. After the reinterview, the number of errors committed by each of the enumerators whose work is being checked is compared with a table giving the acceptable number of errors. Those enumerators whose error rate is more than the acceptable error rate are retrained for a day before the next month's enumeration. If an enumerator's work remains unsatisfactory over a period of time, steps are taken to replace the enumerator.

In this case, it is to be noted that the errors in enumeration occur partly due to the enumerator and partly due to the respondent. Studies have been made in an effort to determine the contribution of the respondent to the errors detected at the reinterview. The results of such a study are given in Table 17. From the table it appears that the contribution of the respondent to the error rate is about 50%. Though it would be desirable to measure the quality of an interviewer in terms of his contribution to the total error, this has not been done because of the subjective nature of the cause for the difference between the original interview and reinterview figures.

TABLE 17. DISTRIBUTION OF THE ERRORS IN THE ORIGINAL INTERVIEW BY CAUSES, MAY 1954-FEBRUARY 1955 (BASED ON 25% ERRORS)

sl. no.	cause for error	percent
(0)	(1)	(2)
1.	respondent misinterpreted the question	12.8
2.	respondent understood the question but reported incorrectly	34.7
3.	*due to respondent*	47.5
4.	questions not asked properly	3.1
5.	enumerator misinterpreted the answer	17.1
6.	enumerator misrecorded the answer	2.3
7.	*due to enumerator*	22.5
8.	*different respondent*	27.3
9.	*others*	2.7
10.	*total*	100.0

Source : Same as for Table 16.

SAMPLE DESIGNS AND CONTROL OF ERRORS IN THE U.S. BUREAU OF CENSUS

The technicians of the Bureau of the Census are the first to acknowledge the limitations of the studies and the programmes undertaken by them in this field. They feel that the experience to date is too limited to give any assurance that the system of quality control would adequately signal trouble when it arises. It is also felt that more intensive work is necessary in the evaluation of coverage and in determining the types of error contributed by the respondents.

13.4. *Quality control projects in the* 1960 *Census. Some preliminary operations.* The plan in case of printing of schedules consisted of examination of a pair of consecutive sheets for every 5,000 impressions for prespecified defects. Six schedules were printed on one single sheet of paper and the cutting operation involved the separation of these 6 schedules. For this operation the top and the bottom sheet in each lot of 1000 sheets were inspected for defects.

In case of binding and preparation of enumerator kits continuous sampling plans were used consisting of inspecting 100% till a prespecified *m*-th defect-free unit is reached and of inspecting every *k*-th unit thereafter where *k* is a constant. The numbers *m* and *k* are so chosen as to ensure a given level of outgoing quality.

TABLE 18. OUTGOING QUALITY FOR A NUMBER OF PREPARATORY OPERATIONS IN THE 1960 CENSUS

sl. no.	operation	unit	work load thousands	percent complete	fraction rejection to date	outgoing quality
(0)	(1)	(2)	(3)	(4)	(5)	(6)
1.	printing	schedule[1]	54,025	100	.026[2] (.014)	[3]
2.	cutting	schedule[1]	54,025	100	.007	[3]
3.	collation	schedule[1]	7,520	100	.0004[4]	.003
4.	binding	book	953	100	.001	.022[5]
5.	preparation of kits	box of 2 kits	86	45	[4]	.001[5]

[1] Front and back impressions.

[2] The figure in bracket includes some rejects by the Government Printing Office.

[3] Zero systematic error, unknown random error.

[4] Corrected rather than rejected.

[5] Approximate estimate.

Source: Hanson, M. H. : Quality Control in the 1960 Censuses of Population and Housing. Memorandum to the Panel of Consultants, dated February 2, 1960.

Field review. The enumerators are instructed to bring their work for the first day to the Crew Leader or the Field Reviewer for a review of their work. During this review, the reviewer checks on specified items, such as, missing units, listing book defects etc. The number of errors committed by the enumerator for the 6 specified items of review are compared with a table showing the acceptable error-rate. On the basis of the number of times the enumerator exceeds the acceptable error-rate in this review, he is either (i) informed that no further review would be needed till he finishes his work or (ii) instructed to report for a second review or (iii) relieved of his work. There would be scope for as much as four reviews in special cases. A specimen table is given here to indicate the type of action the reviewer is permitted to take in different situations (Table 19).

TABLE 10. ACTIONS TO BE TAKEN BY THE CREW LEADER UNDER DIFFERENT CIRCUMSTANCES

field review visit number	no further review till final review	number of checks failed further review required	 enumerator's appointment terminated
(1)	(2)	(3)	(4)
1	0	1 or 2	3 or more
2	0	1	2 or more
3	0	0	1 or more

Source : Same as for Table 18.

Coding operation. Controlling of the quality of coding has received more attention than in the previous censuses. This is mainly due to the surprising results of a study undertaken to determine the quality of verification in coding operation. The rates of failure of verifiers to detect planned errors are given in Table (20).

TABLE 20. RATES OF FAILURE OF VARIFIERS TO DETECT PLANNED ERRORS

study no.	rate of failure
(1)	(2)
1	0.29
2	0.47
3	0.43
4	0.69

Source: Same as Table 18.

The plan to be used in the 1960 Census consists of coding of the census data for a sample of persons by 3 coders independently and of effecting a two by two comparison of the coded data. It is felt that this plan would help in getting a fairly good idea of the quality of the coding work of the coders.

14. Electronic data processing machines

14.1. *Introduction.* During the last decade considerable progress has taken place in using electronic computers to process the census and survey data. The Bureau of the Census has been using electronic computers since 1951. At present about 75 percent of the tabulation of the Census Bureau is done with electronic computers and for the other 25 percent of the work the punch-card tabulating machines are used.

In 1951 an electronic computer, known as the UNIVAC I, Universal Automatic Computer, was installed in the Census Bureau to process some of the data of the 1950 Census of Population. A second UNIVAC I, was purchased in 1955 to tabulate the data collected in the 1954 Census of Business, Manufactures and Mineral Industries. In 1958 two electronic computers, designated 1105 which are much faster than UNIVAC I for arithmetical operations were installed to process the 1958 Economic Census data and to cope up with the work load of the 1960 Censuses of Population and Housing.

The electronic computers are also being used to do much of the current work such as the processing of the Current Population Survey, the National Health Survey, the Retail and Wholesale Trade Surveys, the Annual Survey of Manufactures and the Foreign Trade Statistics. About 50% of the computer time is used for current work, and the computers are being used to accomplish an increasing amount of work.

14.2. *Computer characteristics.* The main feature that distinguishes electronic data processing machines (EDPM) from scientific computers is that the former need devices by which a large mass of data can be fed into and realised from the computer at a reasonably high speed consistent with the speed of the computor. An EDPM usually consists of four sections : (i) a memory in which information including instructions can be stored; (ii) a control unit which keeps track of the sequence of the instructions; (iii) an arithmetic unit which interprets and carries out the instructions given to it by the control unit; and (iv) an input-output section by means of which information can be introduced into the memory and results communicated to the user of the machine.

14.3. *UNIVAC* I. The UNIVAC I system consists of (i) a central computer which contains the electronic circuits for control, arithmetic processing, memory and checking and (ii) ten uniservos which serve to feed in and read out results from the computer. The system operates on magnetic tape to which information could be transferred from punched cards using a card to tape converter.

The system can read the information on the magnetic tape at the rate of 13000 characters per second. An 8-inch reel of tape about 1500 feet long has a capacity of about 1,200,000 characters. The UNIVAC I can accomplish 1905 additions or subtractions per second, 465 multiplications per second, 257 divisions per second and 2760 comparisons per second.

14.4. *UNIVAC* 1105. This 1105 system also consists of a central computer and 18 uniservos and it also operates on magnetic tape. The storage in 1105 has two banks of cores, which are small rings of ferromagnetic material capable of representing "1" or "0" with a total capacity of 8192 words. Any word is accessible in 8 micro-seconds (million-th of a second).

The computer works according to the sequence of instructions stored in its memory. The sequence of instructions for a particular job is called a programme. The control section functions to extract each instruction from the storage, place it temporarily in a control register, interpret it and direct its execution on the specified operands. Under the automatic control, the computer executes the instructions in the specified sequence automatically.

14.5. *Card-to-tape converter.* The Card-to-Tape converter transfers the data punched on the standard 80 column punch-cards to the magnetic tapes. During this process it is possible to select and rearrange data and to reject impossible punches. Card reading and tape recording stages are all self-checking. This assures a high degree of accuracy throughout the conversion. The maximum conversion rate is 240 cards per minute and the effective overall rate is about 120 cards per minute under normal production conditions.

14.6. *FOSDIC.* The FOSDIC (Film Optical Sensing Device for Input to Computers) automatically transfers to magnetic tape the information on micro-filmed copies of original documents without the need to punch out the information. This is a very important development in the field of large scale data processing as it eliminates entirely the time consuming punching and verification operations. To effect this it is necessary to record the data on special schedules by making marks on specified places using an ordinary pencil or pen. Such special schedules are being used in the 1960 Census of Population and Housing.

There is considerable flexibility in transforring the information from microfilms to magnetic tape in that the scanning mechanism can be made to scan the microfilm frame in almost any manner. The FOSDIC can scan the microfilm at the rate, as high as, 300 frames a minute, depending on the amount of information on the frames.

14.7. *Computer operation.* There is a great deal of flexibility in the utilisation of EDPM at the Census Bureau. This flexibility arises from being able, during periods of high work lead, (i) to operate the computers in 2 or 3 shifts; (ii) to commission the computers owned by certain universities; and (iii) to adjust the time programming between census current and special jobs

The computers are also extensively being used to edit the data. In case of current Foreign Trade work, the rejects are corrected and included in the tabulation for the next and not the current period. This procedure helps in producing the results according to schedule. In case of special and census jobs, if the rejects are not considered to be serious, they are corrected or adjusted in the computer itself and only the major defects are identified by the computer which are handled by manual intervention.

ACKNOWLEDGEMENT

I wish to thank the Indian Statistical Institute for having sponsored my study tour with the funds made available to it by the Ford Foundation, and the United States Bureau of the Census for having extended to me all necessary facilities during my stay in the United States of America (January-September 1960) and for allowing me to publish this paper.

I am grateful to Mr. M. H. Hansen Dr. C. L. Dedrick, Mr. W. N. Hurwitz and Dr. J. J. Maslowski, for the personal interest they had taken in suggesting and arranging my programme inside and outside the Bureau of the Census. I wish to express my appreciation of all the courtesies shown to me by the staff of the Statistical Research Division and of the International Statistical Programmes Office in the Bureau of the Census. I should thank Mr. M. H. Hansen, Dr. J. J. Maslowski and Mr. D. B. Lahiri for their useful comments and suggestions on the draft of this paper.

References

Brunsman, H. G. (1960): Processing and editing the data. Presented at the Annual Meeting of the Population Association of America.

Burgess, R. W. (1960): Top management of the Census Bureau. Talk delivered at the United States Bureau of the Census.

Daly, J. F. and Hansen, M. H. (1957): Data processing on electronic computers in the United States Bureau of the Census, (mimeographed).

——— (1958): Some advantages and limitations of automatic computers in processing statistical data. Presented at the Annual Meeting of the American Statistical Association.

Eckler, A. R. (1953): Extent and character of errors in the 1950 Census of Population and Housing. *Amer. Stat.*, 7, 15-23.

Fasteau, H. H. (1960): Camera and film control plan for 1960 decennial census. Office Memorandum dated 22nd April, 1960.

Hansen, M. H. and Hurwitz, W. N. (1944): A new sample of the population. *Estadistica*, 2, 483–497.

——— and Hurwitz, W. N. (1946): The problem of non-response in sample surveys. *J. Amer. Stat. Assoc.*, 41, 517–529.

——— Hurwitz, W. N., Marks, E. S. and Mauldin, W. P. (1951): Response errors in surveys. *J. Amer. Stat. Assoc.*, 46, 147–190.

——— Hurwitz, W. N. and Madow, W. G. (1953): *Sample Survey Methods and Theory*, I and II, John Wiley & Sons, New York.

——— Hurwitz, W. N., Nisselson, H. and Steinberg, J. (1955): The redesign of the current population survey. *J. Amer. Stat. Assoc.*, 50, 701–719.

Hansen, R. H. and Marks, E. S. (1958): Influence of the interviewer on the accuracy of survey results. *J. Amer. Stat. Assoc.*, 53, 635–655.

——— (1958): Procedures for the 1960 Census of Population and Housing. Presented at the Annual Meeting of the American Statistical Association.

——— Pritzker, L. and Steinberg, J. (1959): The evaluation and research programme of the 1960 censuses. Presented at the Annual Meeting of the American Statistical Association.

——— Hurwitz, W. N. and Bershad, M. A. (1960): Quality control in the 1960 Censuses of Population and Housing. Office Memorandum dated 2nd February 1960.

——— Hurwitz, W. N. and Bershad, M. A. (1961): Measurement of errors in censuses and surveys. *Bull. Int. Stat. Inst.*, 38, (2), 359-374.

Hauser, P. M. and Leonard, W. R. (1956): *Government Statistics for Business Use*, John Wiley & Sons, New York.

Marks, E. S. and Mauldin, W. P. (1950): Response errors in surveys. *J. Amer. Stat. Assoc.*, 45, 424–438.

Steinberg, J. (1955): Current Population Survey-Draft, subsequently published as Technical Paper No. 7 by United States Department of Commerce (1963).

United States Bureau of the Budget (1959): *Statistical Services of the United States Government*. U.S. Government Printing Office, Washington, D.C.

——— (1960): Special analysis of principal federal statistical programmes in the 1961 budget, Extracted from the Budget of the United States Government for the United States Government for the Financial Year ending June 30, 1961.

United States Bureau of the Census (1953) : The sample survey of retail stores. *Technical Paper* No. 1, Department of Commerce, Washington, D.C.

——— (1956) : Expansion of the Current Population Survey sample (mimeographed). Series B-23, No. 3.

——— (1956) : *Methods and Procedures of 1954 Census of Agriculture. Special Reports*, 3, (2).

——— (1957) : *Fact Finder of the Nation*. U.S. Government Printing Office, Washington, D.C.

——— (1958) : *Concepts and Methods in the Current Employment and Unemployment Statistics*. Department of Commerce, Series P-23, No. 5.

——— (1959) : *Summary Report on the Current Population Survey Reinterview*, (mimeographed), subsequently published as Technical Paper No. 6 by United States Department of Commerce, (1963).

——— (1959) : Sampling plan for the 1959 annual survey of manufactures. Office Memorandum dated August 21, 1959.

——— (1960) : *Monthly Wholesale Trade Report*, Department of Commerce, BW—60—2 February, 1960.

——— (1960) : Evaluation of the United States censuses of agriculture: 1953–1959, Office Memorandum.

United States Bureau of Labour Statistics (1959) : *Monthly Report on Labour Force*, December 1959. U.S. Government Printing Office, Washington, D.C.

United States Public Health Service (1958) : *Origin and Programme of the United States National Health Survey*. Publication No. 586-A1, U.S. Department of Health, Education and Welfare, Washington, D.C.

——— (1958): *The Statistical Design of the Health Household—Interview Survey*, Publication No. 586–A2, U.S. Department of Health, Education and Welfare, Washington, D.C.

——— (1958) : *Concepts and Definitions in the Health Household-Interview Survey*. Publication No. 586-A3, U.S. Department of Health, Education and Welfare, Washington, D.C.

Woodruff, R. S. (1959) : The use of rotating samples in Census Bureau's monthly surveys. Presented at the Annual Meeting of the American Statistical Association.

Paper received : March, 1962.