# CHAPTER I. THE ACCURACY OF CENSUS TOTALS

#### A. Introduction

The degree of accuracy in a count of the total number of people in a country is a function of the accuracy with which the entire census operation was conducted. The head count may be either more or less accurate than the enumeration of some of the parts, such as age or marital-status groups, but if all the census procedures are of poor quality and the characteristics of the population have not been accurately determined there is little likelihood that the head count will be correct. Indeed, one of the ways of appraising the quality of the head count consists of analysing the accuracy of data on various characteristics of the population. This analysis may not only reveal evidence of inaccurate classification of the individuals enumerated, which is likely to be associated with a faulty head count, but also may reveal a tendency to omit certain categories of the population, which of course has a direct effect on the accuracy of the total count.<sup>1</sup>

Proper evaluation of the head count requires consideration not only of the census figures for components of the population and other information bearing on the quality of the work, but also of any collateral vital statistics, migration statistics, and other relevant data, and the reliability of these collateral data. Ideally all the materials should be examined, and evaluated simultaneously, but in practice this is impossible. In this manual the ways of analysing the different items are taken up separately.

# FACTORS AFFECTING THE COMPLETENESS OF A CENSUS COUNT

The factors which contribute to the completeness of a census count will be but briefly mentioned here. These factors have already been considered in greater detail in other publications dealing with census procedures.<sup>2</sup> The main function served in reviewing them here is to help direct the analyses of the accuracy of census details. For example, social customs sometimes operate against reporting certain segments of the population in certain countries; accordingly, if it is known that among a certain population group the custom of not talking in public about baby boys is prevalent, then special attention can be directed toward evaluating the completeness of census enumeration of male children among that group.

It is assumed that administrative and technical procedures at the central census office have been arranged as well as possible so that clerical errors or other mistakes in handling the census returns are reduced to a minimum. The factors affecting the completeness of the census count can then be listed as follows:

1. Obtaining full public co-operation. Sometimes the general public may not fully comprehend the significance of a census because of ignorance resulting from illiteracy, lack of communication facilities et cetera. In the past in some areas censuses have been connected with taxation, military service, rationing, or other benefits or obligations; where the public believe that such a connexion exists it is difficult to obtain their full and unbiased co-operation. Superstition and various social customs sometimes are important; if it is believed that "bad luck" will follow if one gives out personal information about the members of his family, there will be trouble in obtaining such information. In some cases enumerators may have to be recruited from an educated class which does not enjoy the full confidence of less educated segments of the population.

In general, but not always, this type of problem is most acute in the case of poor and illiterate populations, among those groups sometimes characterized as "underdeveloped".

2. Problems of geography. In an area which is difficult of access because of geographical barriers such as mountains, swamps, deserts, bodies of water et cetera, and where the road and transportation system is inadequate, considerable trouble will be had in locating the population. Often, where such geographical conditions are found, there is also an absence of adequate maps, which complicates the problem.

3. House identification. The intent of every census is to relate the population to specific areas; this involves being able to identify the dwelling or other place where each person is to be counted. Furthermore, it is necessary to identify dwellings so that coverage can be assured; the enumerator must know that he has or has not as yet enumerated a given person. Among some populations, such identification is difficult for various reasons: there are houses which lack exact addresses, persons who dwell on boats or barges, nomads or seminomads who have mobile dwellings, or persons who have no fixed place of residence.

4. Obtaining good enumerators. If there are few educated and literate people in a given area, it may be difficult to obtain enough persons qualified to do a good job of enumeration. Sometimes also, even when there are enough qualified people in an area, it may be difficult to recruit them for census work. There may also not have been enough opportunity to provide enumerators with the necessary training to enable them to interpret the census instructions correctly.

Other problems affecting the accuracy of census enumerations may be mentioned. Some such problems are peculiar to one country only. It is useful, before appraising the accuracy of the census results, to review all the problems which may have been encountered and to consider what defects in the enumeration may have resulted. Special efforts should be made to appraise the

<sup>&</sup>lt;sup>1</sup>For the analysis of age and sex classifications as clues to the accuracy of the head count, see chapter III.

<sup>&</sup>lt;sup>2</sup> See: United Nations, Population Census Methods, Population Studies, No. 4; and Handbook of Population Census Methods.

completeness of the census counts in those areas or among those population groups which are known to be subject to conditions unfavourable for census taking.

# B. Appraising the head count of a single census

The problem considered in this section of the chapter is that of evaluating the accuracy of a single census without reference to demographic data other than those of the census itself. This step should be taken in the appraisal of any census. If the data of a previous census, records on births, deaths and migration, or detailed statistics on population characteristics also exist, this step is only a part of a more extensive analysis. If the results of this one census, without detailed breakdown by population characteristics, are the only demographic data available, methods of appraisal are of necessity confined to those described in this section, unless direct checks, considered in section E of this chapter, can be undertaken.

Even in the case of those censuses which yield no other statistics than the head count, the results are invariably available for various political subdivisions of the country. In many cases, but not invariably, information is also available on the numbers of households (or families) enumerated. The enumeration procedures are generally conducted on a household basis (that is to say the information for all persons living in one household is grouped together on the census schedule and the household is separately identified). Procedures for using the figures for subdivisions of the country and the information on households, if available, for helping to appraise the total head count, are accordingly introduced here.

In the circumstances stated above, only a direct check such as described in section E below, can give a definite measure of the extent of error in the head count. However, without a direct check, the following four tests may show whether or not an error of considerable importance is probably present, and in some cases may yield an estimate of the amount of errors.

(1) Review of the enumeration procedures and information on the quality of performance.

(2) Comparing the census figures with any available data from non-demographic sources which relate to the numbers of the population or parts thereof.

(3) Relating the population distribution as revealed by the census findings to known characteristics of the subdivisions.

(4) Comparing the head and household counts.

Additional tests, some of which are described subsequently in the present section, are also recommended, depending upon the circumstances of the case.

#### REVIEW OF CENSUS PROCEDURES AND PERFORMANCE

With respect to the first test, there is little that can be said other than that all the factors involved in the census undertaking which may bear on the accuracy of results should be examined. Such an examination may very well be regarded as an integral part of the census work rather than as a separate testing procedure. The enumerators' maps should be checked to see whether the entire country was covered, and the returns should be checked to see whether every known city, town, village and hamlet was included. Available reports of the census supervisors for different areas should be searched for observation on the quality of the personnel engaged and of their performance. The completed schodules for each area should be scanned for an impression of the degree of care taken in filling them, and of the extent to which the instructions appear to have been understood and followed.

The prescribed census procedures themselves should be studied with attention to the possibilities of error which may be inherent in them. For example, if it is found that in some or all areas the enumeration was made by convoking the population of each village, the members of each tribe et cetera, at an appointed time and place instead of visiting their dwellings individually, the possibility of a substantial error may be envisaged. If the enumeration was conducted over a very long period of time, errors may be expected as a result of population flux during the interval.

An example of the use of some of the methods described above, together with others, to verify the probable extent of omissions is found in the Brazilian census of 1940. Persons connected with government and business organizations were required to report any known case of census omission on a special questionnaire; in one area, postmasters were required to identify persons in their respective districts who had not been reported at the census; students were offered rewards for information on persons not enumerated; local judicial authorities were asked to express their opinion on the probable efficiency with which enumeration was carried out in their respective areas. Direct checks were also employed. By comparing the information obtained from these several sources, it was estimated that about 1.7 per cent of the population of the country had not been enumerated, this percentage varying greatly from area to area, depending on whether population was dense or scattered and whether transportation was easy or difficult.8

#### Relating the census count to non-census counts

Often, some estimate of population can be obtained from such sources as police records, church records, tax lists, counts of persons eligible for military service, school attendance et cetera. Information relevant to population, or to households, farms et cetera, may also be obtained from censuses of housing and agriculture. In some cases such information may purport to be total counts of the population living in each area; in other instances the information may relate only to selected population segments or other relevant data which can be converted into estimates of the total population.4 These figures are then compared with the census counts for the whole country and each region. Where there is relatively "close" agreement the presumption is that both counts are reasonably correct; at least such agreement does not prove that the census count is incorrect. If there is considerable disagreement, it is necessary to check further in an effort to appraise the two counts.

<sup>&</sup>lt;sup>8</sup> International Statistical Conference, Washington, D.C., 1947. Note on the Completeness of the 1940 Brazilian Population Census, by José Carneiro Felippe and Octavio Alexander de Moraes.

<sup>&</sup>lt;sup>4</sup> Methods of deriving population estimates from non-censal statistics have been described in the preceding manual in this series, Manual I: *Methods of Estimating Total Population for Current Dates*, Population Studies, No. 10, especially chapter III.

Of course, even when both counts are in agreement, the evidence is not conclusive that the census count is correct. Both counts may be incorrect by the same amount, as for example, by the omission of the population living in some remote area. Further checks, including the conducting of sample re-enumerations, are necessary in order to provide positive proof of the degree of inclusiveness in the census. If the non-census counts and estimates are believed to be nearly accurate, however, the comparison yields at least an approximate estimate of the error in the census figure for the area concerned. Also, if there is reason to believe, for example, that the non-census figure is below the true number, and if the census figure is still lower, the inference can be drawn that the census figure is too low by at least the amount of the difference.

The application of this test, that is to say, comparing census results with non-census estimates-is illustrated with data for Haiti, Northern Rhodesia and Southern Rhodesia in table 1 below. A census was taken in Haiti in 1918/19; its results are compared with the estimates derived from church membership data for the same period. The census count of 1,631,000 population is 30 per cent below the non-census estimate of 2,121,000. Among the several regions of the country the differences between the two counts are quite variable. In the Artibonite the two figures are very close; in the North, however, the non-census estimate is about 71 per cent greater than the census count. In order to evaluate these data each set of counts has to be explored at considerable length and other evidence introduced. This problem is illustrated at further length in appendix A.

The data for Northern Rhodesia show much closer agreement than those for Haiti. Here the census count of 1,837,000 is 9 per cent greater than the non-census count of 1,674,000 derived from tax registers. In four of

#### Table 1

#### COMPARISON OF CENSUS COUNTS AND NON-CENSAL POPULATION ESTIMATES, FOR SELECTED AREAS (numbers in thousands)

			Diffe	rence
Area and date	Census count (a)	Non-censal estimate (b)	Number (a)—(b)	Per cent of census count
Haiti, 1918/19 West North Artibonite South Northwest	1,631 671 256 241 385 78	2,121 780 437 255 541 108	$-490 \\ -109 \\ -181 \\ -14 \\ -156 \\ -30$	$ \begin{array}{r}30 \\16 \\71 \\6 \\41 \\39 \end{array} $
Northern Rhodesia, 1950 Barotse Central Eastern Northern Southern Western	1,837 278 176 376 410 212 385	1,674 268 175 322 338 204 367	+163 + 10 + 1 + 54 + 72 + 8 + 18	+ 9 + 4 +14 +18 + 4 + 5
Southern Rhodesia, 1948 Salisbury Gwelo Victoria Bulawayo Umtali	1,619 408 366 284 313 248	1,631 436 334 298 335 228	-12 -28 +32 -14 -22 +20	

Difference negligible.

the six regions the differences are 5 per cent or less, and may be considered almost negligible for many practical purposes. In two of the regions, the Eastern and Northern, the differences are quite substantial. Part of the difference between the census and non-census counts may be due to the fact that the census count is based on a sample and therefore has some sampling variance; sampling variance, however, does not account for all of the differences between the two sets of counts.

Detailed pursuance of this problem is not within the scope of this manual. It can be pointed out, however, that the procedures for obtaining non-census counts such as those of Northern Rhodesia can lead to some error. Generally, these procedures consist in obtaining figures on the numbers of taxable heads of households, collected from every village by visiting officials, with subsidiary estimates and adjustments to make estimates of total population. The figures on taxpayers may be incomplete or out of date and the estimates of their dependants, mostly women and children, may be quite inaccurate. Actually, in the present instance, the two counts for adult males are quite close, the totals differing by less than 2 per cent. For women and children, however, the noncensus counts are about 12 per cent below the census counts; it is quite possible that most of this difference is due to error in the non-census figures. Furthermore, non-census figures for some villages were out of date, having been kept without change since the time of the last visit.<sup>5</sup> In short, there are sufficient discrepancies within the non-census count to disqualify it as a positive check on the census count. It can be concluded only that in four of the regions the agreement is quite good, and in two regions there is much less agreement. It must be concluded, however, that the results of the two counts in Northern Rhodesia substantiate each other much more closely than they do in the case of Haiti.

In Southern Rhodesia the agreement between the census and non-census counts appears somewhat better than in Northern Rhodesia. In none of the regions is there exact agreement. The maximum difference, however, is only 9 per cent; in some regions the census count is higher and in others the non-census count is higher. As in the case of Northern Rhodesia, the census count is based on a sample and hence has some sampling variance; the non-census counts also are in error to some degree.<sup>6</sup>

Practically speaking, the agreement between the census and non-census counts is probably about as close as can be expected.

The observations on these three countries can be summarized as follows:

1. For Southern Rhodesia, the comparison with the non-censal estimates gives support to an inference that the census count is not subject to great error.

2. For Northern Rhodesia, the result of the test is less conclusive, but it does not suggest a large error in the census figures.

<sup>&</sup>lt;sup>5</sup> Report on the 1950 Demographic Sample Survey of the African Population of Northern Rhodesia, Central African Statistical Office, Salisbury, Southern Rhodesia, April 1952, pp. 10 and 11.

<sup>&</sup>lt;sup>6</sup> Report on the Demographic Survey of the African Population of Southern Rhodesia, Central African Statistical Office, Salisbury, Southern Rhodesia.

3. For Haiti, the magnitude of the difference calls for detailed investigation of sources of error in both sets of figures; pending the result of such an investigation the possibility of a major error in the census count must be admitted.

# Relating the census counts for various areas to known characteristics of the areas

Generally the statistical agencies of every country have intimate knowledge about the geographical, social, and economic characteristics of the country's subdivisions. Such knowledge is sometimes quantitative and perhaps very often of a qualitative nature. In either event such details provide clues as to the possible population of the various parts of a country. Such knowledge does not in itself prove that the census count is correct or deficient, but it serves as a background against which to appraise the census count. For example, many countries have detailed land use maps and/or maps showing the agricultural areas or comprehensive aerial photographs. Such maps or photographs can never prove how many people should have been enumerated in each segment of the country, but they can suggest whether there may be many more, or many fewer people than the census counted. In this connexion use of density figures (population per square kilometre, for example) according to land use in the various subdivisions of the country may be the most illuminating procedure. For example, an area devoted to coffee production is likely to have a lower population density than one devoted to sugar cane or tobacco, since coffee requires perhaps only one-third as much labour per unit of land; an area devoted to cattle raising should be more sparsely settled than one devoted to dairying.

This type of appraisal can be illustrated with data for Puerto Rico. A group of municipalities were selected in which sugar-cane raising was very important; they were all located on coastal plains. The average density shown by the census in these municipalities was 470 persons per square mile; the range in density was from 341 to 706 persons. Another group of municipalities located in the mountains, in which coffee raising was very important, had an average density of 332 persons per square mile. The range in density among these municipalities was from 200 to 483 persons per square mile. Clearly, the two types of regions have different typical population densities, but for some of the sugar areas the census shows no higher density than for certain coffee areas. If it were desired to check the census counts, this comparison would suggest paying special attention to those sugar municipalities which reported low population densities, and perhaps those coffee municipalities which reported very high densities.

Other information can be brought to bear in accordance with the known conditions of the country. Perhaps a given part of a country contains extensive public lands which are largely closed to settlement; few people should be enumerated as living there. The results of censuses of agriculture, industry, or housing, if available, are likely to be especially useful. All such types of information cannot be listed here; the statisticians in each country can seek the information appropriate to their purpose.

# COMPARING THE HEAD AND HOUSEHOLD COUNTS

Comparison of these two counts is done most easily by computing the numbers of persons per household. Several such measures are available: (a) the arithmetic average number of persons per household; (b) the median number of persons per household; (c) households classified by number of members, such as oneperson households, two-person households et cetera. For the sake of simplicity the tests illustrated below are based on the arithmetic average number of persons per household (or family).

# COMPARISON OF URBAN AND RURAL AREAS

This test involves calculating the average number of persons per household for urban and rural areas, not necessarily by political subdivisions. In most countries the censuses classify the population as urban and rural; hence it is possible to tabulate the census results so as to obtain counts of the numbers of persons living in each type of area, and the numbers of households in each type. Such data without reference to political subdivisions will be used in these illustrations.

As a general rule, the average size of household in rural areas should be larger than in urban areas. The absence of such a relationship constitutes a warning that the census count may be deficient and that an investigation of the reasons for the observed differences is required. This type of test can be illustrated with the following data for Puerto Rico and Costa Rica. (The choice of these illustrations, of course, does not imply that the census figures for these countries are thought to be particularly unreliable.)

#### Persons per household

	•
Puerto Rico (1950)	
Total country	5.07
Metropolitan areas (containing central cities	
of 50,000 population and over)	4.69
Central cities	4.55
Population living outside central cities	5.16
Cities of 25,000 to 49,000 population	4.40
Cities of 10,000 to 24,000 population	4.56
Small cities and rural areas	5.35
Costa Rica (1950)	
Total country	5.59
San José metropolitan area	5.40
San José city	5.31
Population living outside San José city.	5.72
Other cities	5.26
Rural areas	5.74

It will be noted that in both of these countries the number of persons per household shown by the census is larger in rural areas than in the cities. Within the urban areas of these countries there seems to be little, if any, relationship between city size and number of persons per household. In some countries the average size of household may become progressively smaller as the size of the city increases; this condition, however, is not as commonly found as the urban-rural difference mentioned above. As regards the figures for Puerto Rico and Costa Rica, the conclusion from this test is that it has produced no evidence of a possible error which would require investigation.

# COMPARISON OF POLITICAL SUBDIVISIONS

This test consists of calculating the number of persons per household by urban and rural areas, for the various subdivisions of a country. If the average size of household is very unusual in any one area—either very large or very small—it serves as an indication that there may be errors in the census count which ought to be investigated further. This test can be illustrated with data for Costa Rica as follows:

	Persons per household, 1950				
Province	Total	Urban	Rural		
Total country	5.59	5.33	5.74		
San José	5.58	5.40	5.80		
Alajuela	5.80	5.30	5.94		
Cartago	5.86	5.54	5.98		
Heredia	5.79	5.57	5.91		
Guanacaste	6.31	5.69	6.42		
Puntarenas	5.34	5.21	5.39		
Limón	3.98	4.08	3.94		

Examination of these data on average size of household reveals some variation among the provinces, and shows that the province of Limón has the smallest households. Furthermore, within each province the rural area has the larger average size household, with the exception of Limón. Indeed, within the urban areas there is comparatively little variation among the provinces, excluding Limón; the maximum variation is less than one-half a person (compare Guanacaste and Puntarenas). Within the rural areas the variation is somewhat higher than among the urban areas, the maximum variation, excluding Limón, being about one person (again, compare Guanacaste and Puntarenas). Limón appears so unusual by this test as to raise the possibility that there was an undercount of population there and further investigation seems called for. The possibility to be investigated is that many persons were not counted in the census although their households were enumerated. Actually, as the following analysis shows, other plausible explanations can be found for the Limón figures.

Further analysis can be based on data relating to characteristics of the population as shown by the census.

Although the evaluation of such data is to be taken up in a later chapter, it seems convenient to introduce here, for illustration, the use of figures such as those assembled in table 2, on this page. Inspection of lines 2 to 4 inclusive reveals that Limón has an unusually high percentage of one-person households, 18.3 per cent as compared with 4.8 per cent in the total country. This could have occurred if, in some households, the wives had been omitted from the census count.

Hence, the next step is to examine the sex composition. Limón has an excess of males of all ages (line 6), in comparison with the total country; only Puntarenas has a larger proportion of males. Turning to the population aged 15 and over (line 7), it is noted again that Limón has a larger proportion of males than the country as a whole.

This high percentage of men might be a further indication that many women were missed. Accordingly, marital status composition can be studied; if many wives had been missed then there ought to be a preponderance of married men in this province. Among the married population, however, there is but a very slight preponderance of men (line 9). In this respect Limón differs but very little from the entire country. Among the single population there is a very large excess of men (line 8); in Limón 63.3 per cent of the single population are men as compared with but 52.5 per cent in the total country. Only Puntarenas has a somewhat higher proportion of single men. These figures do not support the hypothesis that wives were omitted from the census in Limón in any unusually large numbers.

Might an unusually large number of children have been omitted from the census count in Limón? In line 5 is presented the number of children under 5 years of age per 100 women aged 15 to 44 years inclusive. If the great majority of the women in the province were included, but if a large number of children were omitted, this ratio ought to be relatively small in Limón. In fact, however, this ratio is exactly the same for Limón as for the total country (line 5). Hence, the presumption is that unusually large numbers of children were not omitted from the census count in this province.

L	ine No.	Entire country	San José	Alajuela	Cartago	Heredia	Guanacaste	Puntarenas	Limón
1.	Population (in thousands)	800.9	281.8	148.9	100.7	51.8	88.2	88.2	41.4
2.	Number of households (in thousands)	143.2	50.5	25.7	17.2	8.9	14.0	16.5	10.4
3.	Number of households containing only one person (in thousands)	6.9	1.9	0.8	0.6	0.3	0.5	0.9	1.9
4.	One-person households as percentage of all households	4.8	3.8	3.0	3.6	3.3	3.4	5.5	18.3
5.	Children under 5 years per 100 women aged 15-44	74	65	81	79	68	89	80	74
	Percentage of males:								
6.	In total population	49.9	47.9	49.9	50.5	48.9	51.2	53.7	53.0
7.	In population aged 15 and over	49.3	46.4	49.1	49.7	47.9	51.3	55.5	54.3
8. 0	In single population aged 15 and over	52.5	47.3	51.9	53.3	49.3	55.2	65.3	63.3
	and over <sup>a</sup>	<b>49.9</b>	49.5	49.7	49.9	49.6	50.0	50.8	50.9

Table 2

Selected	POPULATION	CHARACTERISTICS	FOR	Costa	Rica,	BY	PROVINCES,	1950
----------	------------	-----------------	-----	-------	-------	----	------------	------

\* Including consensually married persons.

In summary, the small average size of household in the province of Limón seems consistent with a census count, as complete there as in the remainder of Costa Rica. Further checks can be made, of course, not only by re-enumerating the province or a sample of it, but also by investigating other aspects of its social, economic and ethnological structure, factors which could not be included in this brief example.

# COMPARISON OF RESULTS OF DIFFERENT TESTS

As the examples given above imply, the results of the types of tests described in the present section are often not sufficient in themselves to prove that the census count is either correct or in error, or to give an estimate of the amount of error. For the most part, they serve as indicators of possible errors that deserve investigation by means of more rigorous checks. It is often possible, however, to get more definite indications by comparing the results of different tests of the type described. For example, if a review of the census methods and evidence on the quality of performance gives rise to a suspicion that the results for a given administrative area may have been inaccurate, and if a comparison with non-census data or an analysis of the household figures suggests a possible deficiency in the count for the same area, the inference to be drawn from the two tests combined is far more positive than the result of either test considered alone.

# C. Consistency of totals at successive censuses

When the total population of a country is known for several successive censuses, better estimates of the probable accuracy of each census, including the most recent one, are possible than if data were available for only one census.

The guiding principle to be followed in comparing the results from two or more successive censuses, is that population changes normally proceed in an orderly manner. This is to say that in the absence of unusual events the rate of increase for the whole country and for each of its parts, can be expected to change only gradually in successive inter-censal periods, and to follow a fairly constant trend. When such an orderly pattern is not observed, the deviations should be explainable in terms of known events, such as the curtailment of immigration, the presence of famine, or any other event. Deviations from the pattern which cannot be so explained constitute a warning of possible errors; and the presumption of error is greatly strengthened if the results of other tests, however inconclusive in themselves, are found to point in the same direction.

#### Analysing the rate of change between two census dates

When the total population of a country is known for two successive census dates a test which can be applied consists of comparing the observed rate of change with the rates of change observed in other countries. For this purpose comparisons should be made with such other countries as seem to have approximately comparable demographic, social and economic conditions. Actually, of course, no two countries have identical conditions; nevertheless, countries which have fairly similar social and economic conditions should have fairly similar rates of population growth.

This procedure can be illustrated with data for Angola and neighbouring areas:

Area	Dates	Average annual rate of increase <sup>1</sup>
Angola	1940/1950	0.96
Mozambique South Africa, excl.	1940/1950	1.18
Europeans	1946/1951	2.03
Bechuanaland Swaziland	1936/1946 1936/1946	1.02 1.68

Although the average annual rate of population increase in Angola is lower than that in any of the other areas with which it is compared, its rate is not so extremely different as to lead to suspicion of the census counts. The result of the comparison is compatible with an accurate census both in Angola and in the other areas.

A second illustration is provided by data for Martinique and other island territories in the lesser Antilles. The population of Martinique, according to censuses, increased from 1910 to 1921 at an annual geometric rate of 2.66 per cent. Such a rate of population growth is high but is known to have occurred in some areas. Population changes during the same period in other islands of the region were as follows, according to censuses of 1911 and 1921:

	Per cent per annum
Barbados	0.93
Leeward Islands	0.90
Bahamas	0.53
Turks and Caicos Islands	0.17
Windward Islands	. +0.42
Guadeloupe	+0.79
Trinidad and Tobago	. +0.92

In some of the territories, population decreases were registered while in others the increases were only moderate. The decreases, or low rates of increase, may have been caused by the great influenza epidemic fol-lowing the First World War, which took a heavy toll of lives in this region. The epidemic may have affected some islands more than others. Differences in population growth may also have arisen through migration, both between these islands, and from these islands to other regions. However that may be, the average rate of population change did not exceed one per cent (plus or minus) per annum in any territory except Martinique, where an exceedingly high rate of growth was observed. How could this be? Did Martinique completely escape the influenza epidemic? Was there a large immigration into Martinique? If there is no such explanation, the conclusion must be either that the 1910 census of Martinique was incomplete or that the 1921 figure was inflated.

In some cases it may not be helpful to compare the rate of population growth in the country in question with that of other countries. It may be that conditions are so unusual as to preclude any useful comparison. Or it may be that no census figures are available for other countries with approximately similar social and economic conditions. Even so, it is possible to set probable upper and lower limits for the average annual rate of increase.

If the country's population changes only through natural increase, it is very unlikely to have an average

<sup>&</sup>lt;sup>7</sup> Geometric rate of increase, per annum.

annual rate of growth exceeding 3 per cent. This would be the result of a high birth rate (say 40 or more per 1,000) and a very low death rate (say 10 or less). On the other hand, it is only in unusual circumstances that the population would be likely to decline without heavy emigration. In fact, nearly all observed rates of natural increase in the various countries of the world in modern times have been in the range from zero to 3 per cent. If, in any given country, the rate of population change approaches or exceeds these limits without large-scale immigration or emigration, the question must be raised as to whether there is some explanation for such an unusual rate, or whether the census counts were in error.

An illustration can be found in data for Honduras. According to the census counts of 1926 and 1930, the population increased 5.8 per cent per year. Such a rate could not have been attained by the excess of births over deaths. Only an extraordinary immigration into the country could have caused it. If there was no such immigration, it is clear that the census figures are erratic. Possibly the earlier census was incomplete and the later one complete, or the first was correct and the second unduly inflated, or more likely, both were in error. The inconsistencies in census data for Honduras are treated in more detail in appendix C.

With some information, however approximate, regarding the conditions of mortality and fertility in the country, the limits of the rate of growth may be defined more closely. For example, if there is reason to believe, even without reliable vital statistics, that the birth rate is normally at least 30 per 1,000 (corresponding to moderately high fertility) and that the death rate does not normally exceed 20 per 1,000 (corresponding to fairly good control of major infectious diseases), then, in the absence of large-scale emigration, any average rate of growth below one per cent per annum shown by the censuses gives a strong presumption of error.

# COMPARING RATES OF CHANGE DURING INTERVALS BETWEEN THREE OR MORE CENSUSES

If population counts are available for three or more successive censuses it becomes possible to make a more accurate evaluation by comparing the successive rates of growth. Again, the same principle is followed, namely, that the pattern of population growth should be regular except in so far as it can be shown that changes in the circumstances may have led to departures from the pattern.

An example of this test is afforded by Guatemalan census figures. The average annual rates of increase were as follows:

1880 to	1893	0.84
1893 to	1921	1.36
1921 to	1940	2.67
1940 to	1950	-1.63

The rate of increase in the period 1921 to 1940 was improbably high in comparison with the earlier period; furthermore, it approaches the upper limit of about 3 per cent per year, which was previously noted. Finally, the census counts show a decrease in population in the next decade, 1940 to 1950. The latter also seems improbable, since, as was mentioned previously, declines are unusual. These observations strongly suggest that the census count of 1940 was exaggerated; or, which is less probable, that both the 1921 and 1950 counts, and perhaps also the 1893 enumeration, were grossly deficient. No evidence is available to show that there were such changes in social and other conditions, or such large migrations, as could have led to these diverse rates of change.

As a further check, the average annual rate of change for the period 1921 to 1950 can be calculated. This rate is 1.16 per cent. The fact that it is below the rate for 1893 to 1921 is difficult to explain except on the hypothesis that the census figures for 1893, 1921, or 1950, as well as 1940, are in error (some exaggeration in 1921, or some deficiency in 1893, 1950, or both).

Another example of this test is afforded by data for the Indian peninsula (the combined areas of India and Pakistan). The average annual rates of increase for the inter-censal periods were as follows:

	Fer cen	
1901 to	1911	5
1911 to	1921	9
1921 to	1931	2
1931 to	1941	1

Inspection of these rates suggests, at first glance, a serious under-count at the 1921 census. It is known, however, that the Indian peninsula suffered an enormous number of deaths as a result of the influenza epidemic following the First World War, and this is a sufficient explanation for the abnormally low rate of population increase during the decade 1911 to 1921.

#### POPULATION TOTALS FOR SUBDIVISIONS OF A COUNTRY

The type of tests previously described, for analysing the changes in a nation's total population, can be applied in principle to the analysis of changes in the various subdivisions. Here too, we expect to find "orderly" patterns of population change, both within the same subdivision in successive inter-censal periods, and among different subdivisions in any period. Any dissimilarities should be explainable in terms of known conditions. As a practical matter it is well known that there is considerable diversity in the rates of population change among the various parts of any nation. Accordingly, the problem becomes one of trying to distinguish between changes which are explainable in terms other than errors in the statistics and those which are not. It should be noted that although these procedures may reveal the presence of errors and in some cases indicate their order of magnitude, they provide no basis for exact estimates of the size of the errors. Definite estimates must be made by other methods, some of which are described in chapters II and III.

#### COMPARING SUCCESSIVE CENSUSES

With this method alone, final evaluation of one census count on the basis of preceding counts cannot ordinarily be made until one or more subsequent enumerations are available for purposes of comparative analysis. Preliminary analysis can be made, however, without waiting for subsequent counts. Actually, it is a part of normal census procedure, as soon as an area has been enumerated, to calculate the percentage change in its population since the previous census date and compare it with the changes in surrounding areas and during previous inter-censal periods. Any unusual rate of change should be questioned immediately, and if necessary, the areas should be re-enumerated, either on a complete or sample basis.

Let us turn now to an example of how the rates of change can be analysed for several successive censuses. In the United States there was serious under-counting in the Southern States at the time of the 1870 census. This deficiency became evident when the percentage increase in population for the decade 1860 to 1870 was compared with the changes during the preceding and following decades. The percentage increases in population for the South as compared with the remainder of the country were as follows:

Period	South	Remainder of United States	Ünited States
1850 to 1860	23.2	43.0	35.4
1860 to 1870	10.8	29.6	23.0
1870 to 1880	34.1	28.0	30.0
1880 to 1890	21.2	27.3	25.3

The rate of increase in the South in the decade 1860 to 1870, which encompassed the Civil War and the ensuing reconstruction period, seems much too low in comparison with the preceding decade. The rate of increase for the period 1870 to 1880 seems abnormally high in comparison with the preceding and following decades. The population changes in the remainder of the United States seem much smoother, in contrast. The differences are of such magnitude that, even with no other information, it would appear highly probable that the enumeration of 1870 in the South was seriously incomplete. Taken together with the knowledge that conditions in the South in 1870, during the aftermath of the Civil War, were not favourable to accurate census work, the figures give convincing evidence of a major error. This under-counting was not recognized as an established fact, however, until after the 1890 census had been taken.<sup>8</sup>

# COMPARING POPULATION CHANGES WITH SOCIAL AND ECONOMIC CONDITIONS

Procedures for evaluating diverse rates of change in terms of differences in the social and economic conditions in a country can be illustrated with selected census data for Puerto Rico. No attempt is being made here to evaluate the censuses of 1940 and 1950 for the whole island; rather examples of a few selected municipalities have been chosen to show how such comparisons can be made and conclusions drawn regarding the probable completeness of the census counts.

The San Juan metropolitan area (see map on page viii) increased in population by about 62 per cent in the decade 1940 to 1950, as compared with a rate of growth of 18.3 per cent for the entire island. Such a large increase in this area is consistent with the economic and other developments which occurred there since 1940. These included: (a) about a doubling in the volume of manufacturing activity (as measured by the increase in the number of production workers); (b) substantial increases in commercial and transportation activities, both directly and indirectly related to the growth of manufacturing; (c) substantial increases in governmental activities and numbers of employees; this increase was in part related to the needs of the

<sup>8</sup>1950 U.S. Census of Population, U.S. Summary, Bulletin P-C 1, p. xxviii, Washington, 1953. Second World War and in part came about when Puerto Rico achieved self-government and the types and amounts of government services were greatly expanded.

Another factor which was probably important was the building of the new highway along the northern coast during this decade (see map). As a result the San Juan area became a focal point for a much larger volume of traffic. Furthermore increased employment opportunities became available in connexion with the road-in its construction and maintenance and in the operation of roadside garages and refreshment stands et cetera. Even more important as a source of employment and population support was the fact that this new road made it possible for commercial establishments and factories to be located near the San Juan port, but on the outskirts of the metropolitan area. In the absence of improved transportation some of these establishments probably would never have been opened. Other new roads were put in south of the city of San Juan-into Rio Piedras, Guaynabo and Bayamon, and eastward into the northern part of Carolina. At present the entire metropolitan area is served by excellent transportation facilities, much superior to those available in 1940. Catano, across the harbour from San Juan, has excellent ferry service, and now contains large residential areas.

Proceeding westward from metropolitan San Juan, the census figures show that Toa Baja and Dorado both grew substantially, by 38 and 24 per cent, respectively. Both of these municipalities probably benefited from the developments in the San Juan area. Distances are short enough and transportation facilities good enough so that a person may live near the main highway in either of these municipalities and have his employment in metropolitan San Juan; from Bayamon to Dorado, for example is hardly more than fifteen minutes by private or public automotive transportation, the total distance being hardly ten miles. That the highway is an important factor influencing population growth in these two municipalities is suggested by the fact that those barrios through which the main highway passes, grew the most rapidly, as follows:

	Per cent inco population, 19	rease in 940-1950
Toa Baja: Total	38.1	ļ
Remainder		5
Total	23.9	)
Remainder	34.4	5

Continuing westward along the northern coast the next three municipalities and their rates of population growth in the decade 1940 to 1950 are:

Per cent
15 3
10.0
25.2
43.4
27
J./

In all three of these municipalities those parts through which the main highway passes grew more rapidly than the other parts. All three municipalities are in part within the zone of influence of the San Juan metropolitan area; numbers of people travel to work regularly between Manati and the metropolitan area. It can then be asked why were the differences in population growth so large, and is there a possibility of an undercount in the 1950 census in Vega Alta, and more particularly in Manati? Both areas produce sugar and pineapples, and neither has much manufacturing. In both areas there were very significant increases in the production of both these crops during the decade. Presumably, then, both areas experienced increases in employment opportunities and should have grown at about the same rate. Why they grew at such different rates, the present analysis does not reveal. Further investigation, including consideration of earlier census data, the returns for smaller subdivisions of these municipalities, and any available information regarding the performance of enumerators in this area in both 1940 and 1950 is required.

Another municipality which had a relatively high rate of population growth during the decade, was Ponce on the southern coast. This grew by 20.6 per cent. This growth seems possible since there was considerable growth in manufacturing, and large increases in sugar and coffee production. Furthermore, Ponce is advantageously situated with respect to transportation, having both main highways and a seaport.

Guanica, to the west of Ponce, increased in population by 23.2 per cent during this decade. No reason for such an increase is readily apparent. This municipality is largely dependent on raising and processing sugar cane; there was but a moderate increase in the volume of such activity, however, and not enough to account for the growth in population. Further analysis of the possibility that either the 1940 or 1950 census counts may have been in error, is indicated.

Turning now to municipalities which lost population, or gained but very little, it is found that the majority are in the interior mountainous areas where the main economic activities consist of raising coffee and/or tobacco. This group includes: Maricao, Lares, Adjuntas, Jayuya, Ciales, Morovis, Barranguitas, Comerio and Cidra. The development of these activities has not been sufficient to support large population increases; by 1940 or earlier these municipalities already had enough people to supply all the labour needed for these two crops. Furthermore, these municipalities are all poorly supplied with roads. Of this group, Adjuntas is the only one through which a main highway passes; however, since the area is mountainous the highway is a difficult one to travel. This lack of transportation, in turn, is probably one of the factors contributing to the lack of population growth and, hence, to the lack of population growth.

There is no apparent explanation for the low rate of population growth in Yabucoa, in the southeast corner (see map). This municipality's economy is based largely on sugar and tobacco; the production of both these crops increased considerably during the decade. Yabucoa is also advantageously situated with respect to highway transportation.

This procedure of examining each municipality in considerable detail and relating the changes in population size as shown by the census enumerations to social and economic changes can be continued indefinitely. Data on all the various types of agricultural production, on the opening of new factories, on the building of new roads and power stations and dams et cetera, together with vital statistics, the tabulations of the census data on population characteristics, the evidence of the census schedules, and the results of earlier censuses, can all be introduced into the final evaluation. It is apparent that this procedure calls for detailed knowledge of the country and its parts as well as of the census materials and cognate statistics, if a proper evaluation of the adequacy of census returns is to be made.

It should be understood in making such comparisons that internal migration plays an important part in bringing about the differences in rates of population growth in the subdivisions of an area such as Puerto Rico. The expansion of economic opportunity in the form of manufacturing, cane-sugar production, coffee raising, dairy farming, trade et cetera, was uneven in different parts of the island. As job openings appeared in a specific area, workers and their families migrated into the area, and the inter-censal rate of population growth increased. Differences in natural increase, the excess of births over deaths, may also be an important factor, in some circumstances, producing regional variations in the rates of growth within a country, but natural increase is unlikely to respond so greatly and so quickly to changes in economic opportunity as to produce such large differences in inter-censal rates of change as are observed in Puerto Rico.

# D. Consistency of census totals with vital statistics and migration statistics

# THE BALANCING EQUATION

The *balancing equation*, which is to be considered in this section, is a procedure which, if properly used, permits a more precise quantitative evaluation of the accuracy of a census count than can usually be made by such methods as described in the preceding sections. This procedure, employing vital and migration statistics, brings much more evidence to bear and provides a much more definitive and quantitative answer on just how accurate and complete a given census enumeration probably is. In this section only the completeness of the head count is considered; in subsequent chapters the accuracy of data on age and sex composition and of the vital and migration statistics will also be included. However, in practice a final determination of the completeness and accuracy of a census count, using the balancing equation, cannot be made until all of the components-population by age and sex, births, deaths and migration-have been tested.

#### DEFINITION AND LIMITATIONS

The principle of the balancing equation is essentially very simple. In any interval of time, such as between two censuses, the population of a country can increase or decrease only as a result of births and deaths and movements across the country's boundaries. Births and immigration add to the population, and deaths and emigration subtract from it. Accordingly, if data are available from two censuses, and the numbers of births, deaths and in- and out-migrants are known, then the equation must balance exactly, *if* all the data are perfectly accurate. The equation is very simple:

the population at the second census  $(P_1)$  equals:

- the population at the first census  $(P_0)$ 
  - plus the number of births during the inter-censal period (B)

plus the number of immigrants in the inter-censal period (I)

minus the number of deaths during the inter-censal period (D)

minus the number of emigrants during the intercensal period (E)

$$P_1 = P_0 + B + I - D - E$$

Completely accurate census, vital and migration statistics have never been attained in any country, and never will be. Accordingly, this equation never balances out exactly; some unaccounted-for residual always remains. If the residual is large, at least one of the components contains a sizeable error, and further investigation has to be conducted in an effort to determine where the error lies. Sometimes the major part of the error may be found in one set of data; generally it is found that there are errors of varying size in all of the components of the equation. By testing each of the components separately by means of the balancing equation, the approximate size of the error in each of the components can be determined.

Even if the residual is small, however, there is no definite proof, until all of the components have been separately evaluated, that the censuses, vital and migration statistics are correct. This is because of *compensating errors*. For example, under-registration of births may be compensated for by under-registration of deaths; incompleteness of enumeration at one census may be compensated for by incompleteness at the other; incompleteness of census enumeration may be compensated for by under-registration of births or deaths, or by incomplete recording of migrants.

One further caution is required. Compensating errors may sometimes be of such a nature that, even after studying the various components as thoroughly as possible by means of the balancing equation and finding that the data are all in good agreement, it may still not be justified to conclude that they are all reasonably correct in absolute terms. For example, the application of various tests may make it appear that the birth statistics are consistent with the census enumeration of young children and therefore probably correct, that the migration statistics tally with the census returns on the foreign-born population and thus seem to be correct, and that the entire equation balances with but a very small residual, implying that the death statistics and the census enumerations were also correct. Yet, the censuses, vital statistics and migration data may all fail to cover some part of the population-for example, persons with no settled abode, or primitive tribes, possibly not under effective administration, or the population living in a frontier region which is difficult of access, or an outcast or criminal segment which defies census enumeration and reports none of its activities to the authorities. Generally, however, such groups constitute but a small proportion of any country's population.

#### TREATMENT OF MIGRATION

The element of migration in the balancing equation often presents a difficulty because many countries do not have comprehensive and reliable statistics on the flow of migrants across their borders, and migratory movements within the country are even less likely to be

adequately recorded. However, even though such statistics are lacking or seriously defective, the available sources of information will be sufficient as a rule to show whether or not migration has taken place on a large enough scale to have a significant effect. If not, the balancing equation may usefully be applied without regard to migration.

It is primarily in the case of statistics for subdivisions of countries that the lack of adequate migration statistics interferes with the use of balancing equations. In most countries, the volume of internal movements is so large, at least for the majority of their subdivisions, that in the absence of a proper measure of it the balancing equation is inapplicable.

# Examples of the use of the balancing equation

The procedures and examples used in this chapter are dictated both by the purpose and by the nature of the available data. The purpose is to evaluate the total count of two successive censuses by means of the balancing equation; evaluation of the vital and migration data and evaluation by age and sex are reserved for subsequent chapters. Because the results of the balancing equation in the form used here cannot be interpreted until the remaining statistics are also evaluated separately, no definite conclusions can be drawn in these examples regarding the accuracy of any particular components. Some tentative inferences can be made, however, in certain cases.

Since migration statistics for all countries are not available, examples of the use of the balancing equation are presented for some countries having such data, and for some not having them. For all of the examples shown, the equations were computed without migration statistics; and for those countries having such data, additional computations were made which included the migration balance.

The specific procedures employed are as follows:

(1) By subtraction of successive census totals, the apparent inter-censal increase is determined.

(2) The apparent natural increase—or excess of births over deaths—for the inter-censal period is then determined on the basis of the reported birth and death statistics. (Where necessary, adjustments for a fraction of a year are made.)

(3) Subtraction of the apparent natural increase from the apparent inter-censal increase provides a measure of the *discrepancy*.

For countries with migration statistics, the following additional steps are involved:

(4) The migratory balance is computed for the intercensal period by subtracting recorded departures from recorded arrivals.

(5) This migratory balance is added to the apparent natural increase to determine that population increase which can be accounted for by means of available vital and migration statistics.

(6) The residual is obtained by subtracting the figure obtained in step 5, from the apparent inter-censal increase obtained in step 1.

The above steps are illustrated in detail for the following areas:

# 1. Puerto Rico

Sten	(1)	Population according to census of 1 April.	
Dtop	(-)	1950	2.211.000
		Population according to census of 1 April.	,,-
		1940	1.869.000
		Apparent inter-censal increase	342,000
S4	(2)	Number of births reported in inter-censel	•
Step	(2)	nominal	830,000
		Number of deaths reported in inter-censal	000,000
		Number of deaths reported in inter-censar	289.000
		Apparent natural increase	541,000
		Apparent natural increase	341,000
Step	(3)	Apparent inter-censal increase	342,000
-		Apparent natural increase	541,000
		Discrepancy	- 199,000
Sten	(4)	Number of recorded arrivals in inter-	
Dtop	( )	censal period	627,000
		Number of recorded departures in inter-	•
		censal period	808.000
		Migratory balance	- 181.000
-		Migratory Datance	541.000
Step	(5)	Apparent natural increase	101,000
		Migratory balance	181,000
		Population increase accounted for	360,000
Sten	(6)	Apparent inter-censal increase	342,000
- top	(-)	Population increase accounted for	360,000
		Residual	- 18,000

The residual of 18,000, or the amount by which the equation fails to balance, is relatively small in this case, and amounts to only 0.8 per cent of the 1950 population as recorded in the census enumeration. It should be noted that some of this residual is explained by the fact that deaths to members of the armed forces are excluded from the reported death statistics. Hence, the amount of natural increase as recorded in step 2 should be somewhat smaller than the figure shown. This, in turn, would decrease the size of the "Population increase accounted for" in step 5, and lead to a smaller residual in step 6.

The residual, expressed as a percentage of the recorded inter-censal increase, amounts to 5.3 per cent. As will be seen from the other examples to be presented, the equation for Puerto Rico balances out fairly well better than in some countries but not as well as in some others. Without further analysis of the components of the balancing equation, and without applying the balancing equation to specific age-sex cohorts, it can be concluded that the balancing equation gives no evidence of major errors.

# 2. Thailand

Censuses were conducted in 1937 and 1947. Vital statistics are available for these years, together with an estimate of the net immigration: The balancing equation for the period 1937 to 1947 then is:

Consus count 1047	17,443,000
Census count, 1937	14,464,000
Apparent inter-censal increase	2,979,000
Recorded natural increase in inter-censal period	2,569,000
Migratory halance	100,000
Population increase accounted for	2,669,000
Residual (excess of census increase): Number Per cent of inter-censal increase Per cent of 1947 population	310,000 10.4 1.8

The relatively small residual suggests that the components of the equation are not subject to very large errors; in anticipation of material to be presented in the chapter on vital statistics, however, it should be pointed out that there were actually rather large compensating errors in the reports of births and deaths. Furthermore, the data on migratory balance are meagre and unsatisfying; the figure of 100,000 presented above for net in-movement is an estimate pieced together from incomplete records.

#### 3. Dominican Republic

Censuses were conducted in 1920, 1935 and 1950. A complete series of vital statistics covering these 30 years was not found; birth records were missing for the years 1920 to 1924, 1929, and 1931 to 1934, while death records were missing for the years 1920 to 1923 and 1929 to 1934. Data for these missing years were estimated. No statistics on migration were found. The balancing equation for the period 1920 to 1935 is:

Census count. 1935	1,479,417
Census count, 1920	894,665
Apparent inter-censal increase	584,752
Recorded natural increase in inter-censal period	426,384
Discrebancy:	
Number	158,368
Per cent of inter-censal increase	27.1
Per cent of 1935 census count	10.7
The balancing equation for the period 1935 to	1950 is:
Census count 1950	2.135.872
Census count 1935	1,479,417
Apparent inter-censal increase	656,455
Recorded natural increase in inter-censal period	706,442
Discrepancy:	
Number	- 49,987
Per cent of inter-censal increase	- 7.6
Per cent of 1950 census count	- 2.3

Any migratory balance during either of these two periods was probably small in comparison with the discrepancies. In order to determine whether the discrepancies are due chiefly to errors in the census enumerations or the vital statistics, it is necessary to test the accuracy of the latter.

#### 4. Fiji Islands

Censuses were taken in 1936 and 1946, and vital statistics were reported for the entire decade. Data on migration, however, were found only for the years 1936 to 1938 and 1945. The balancing equation for this decade is:

Census count 1946	259,638
Census count, 1936	198, <b>379</b>
Apparent inter-censal increase	61,259
Recorded natural increase in inter-censal period	58,578
Migratory balance	2,007
Population increase accounted for	60,585
Residual (excess of census increase) Number	674 1 2
Per cent of 1946 census count	0.3

Since complete data on the migratory balance were not found, it is difficult to determine the probable degree of accuracy of the statistics. The residual is so small, however, that it does not seem likely that the censuses contained any very large errors. Even with an error of several thousand in the migratory balance, the equation would still balance out fairly well. To be sure, compensating errors may be present.

#### 5. Mexico

Although a number of censuses have been	taken in
Mexiico, the equation will be applied only to t	he period
1940 to 1950. The result is as follows:	
Census count, 1950 (preliminary figure)	25,781,000
Census count, 1940,	19,654,000
Apparent inter-censal increase	6,127,000
Recorded natural increase in inter-censal period	5,770,323

Discrepancy:	457 000
Number	357,000
Per cent of inter-censal increase	5.8
Per cent of 1950 population	1.4
Recorded migratory balance	- 122,000
Population increase accounted for	5,648,000
Residual:	479.000
Number	78
Fer cent of inter-censal increase	1.0
Per cent of 1950 population	1.9

The discrepancy (if migration statistics are not taken into account) is not very large in proportion to the population. If the recorded migratory balance is included in the computations, a somewhat larger residual is obtained. Whether or not the migration statistics are accurate, it is very probable that during this period more persons left the country than entered it. The analysis shown here does not yield any clues as to whether the residual should be mainly attributed to defects in the vital statistics or to differences in the completeness of the two census enumerations. However, further tests of the Mexican statistics (to be presented in a later chapter) suggest that the statistics of births and deaths are fairly accurate, so that probably no very large residual would have resulted from inaccuracies in the vital statistics.

#### 6. Egypt

Although several censuses have been taken in this country, the balancing equation will be applied only to the period 1937 to 1947. For this period vital statistics are available, but no migration data were found. The balancing equation is:

Census count, 1947	18,967,000
Census count, 1937	15,921,000
Apparent inter-censal increase	3,046,000
Recorded natural increase in inter-censal period	2,543,000
Discrepancy:	
Number	503,000
Per cent of inter-censal increase	16.5
Per cent of 1947 population	2.7

In the absence of migration data, the balancing equation is incomplete. However, unless there were a net in-migration of some 500,000 persons, the equation would not balance. For fuller investigation of the data for this country, it would be necessary to obtain some estimates of the probable volume and direction of migration, and to analyse the vital statistics component of the balancing equation separately. Furthermore, the balancing equation should be applied separately to the various age-sex groups as explained in chapter III. It would then be possible to reach a conclusion as to probable accuracy of the census counts.

# E. Direct checks on the accuracy of census totals

Direct checks consist primarily in re-enumeration of at least a part of the population, in such a way as to provide a measure of the numbers of persons who were omitted from or erroneously included in the census count. The list of persons recorded in the re-enumeration is compared with the record of the census, and any discrepancies are investigated in order to determine whether or not the persons concerned should have been included in the census. Some of the ways in which such a check can be made are described below. The examples presented here refer to entire countries, but the same procedures are evidently applicable to particular subdivisions of a country, and it may be efficient, in some instances, to limit their use to those subdivisions in which errors are suspected on the basis of other tests.

An important consideration, and one reason why such checks have not often been made, is the question of cost. The cost of a properly planned check, however, can be very modest in comparison with the cost of the census itself, and it can greatly enhance the value of the census results, either by confirming their accuracy or by providing the basis for a correction.

# TIMING OF A DIRECT CHECK

Because of the necessity of name-by-name checking, direct checks are carried out most efficiently directly after a census enumeration. The mechanics of checking the names obtained by the re-questioning with the names on the census schedule become quite complicated and laborious, particularly if people migrate. Ideally, immediately after a small area has been enumeratedas for example one block in a city, or one small villagethe procedures for a direct check should be inaugurated in that area; thus the problem of searching the census records for the persons found in the check enumeration is greatly facilitated. The longer the time allowed to lapse before the check is instituted, the more the problems of identifying persons and deciding who should have been counted is complicated by births and deaths, by changes in address, by changes of name on the part of women marrying et cetera.

#### Selection of areas to be re-surveyed

There is no need to re-enumerate the entire population of the country in order to check the census count. Depending on the precise objective of the check, the re-enumeration can be carried out for a selected sample of the entire country, for one or several special areas of the country, or for samples of such special areas. Any combination of these procedures can also be made.

Re-enumeration of a sample of the entire country is desirable if the chief objective is to establish some measure of the over-all accuracy of the census count. Scientific sampling techniques make it possible to select a relatively small sample in such a manner that the most representative results are obtained at smallest cost. A sample of the entire country is also sufficient if there is no special reason to suspect that the census enumeration may have been particularly inaccurate in some special part of the country, or among some special segment of its population.

The various tests described in sections A and B of this chapter are relevant to determining whether or not the census results for particular areas should be questioned. By these or other tests, areas may be selected for re-enumeration. Because of the desirability of conducting checks as soon after the census as possible, it is important that census results should be analysed from this point of view within the shortest possible time.

It should be noted that direct checks conducted in selected areas only, whether on a sample basis or by complete re-enumeration, do not provide an accurate measure of the accuracy of the census for the entire country. In the "suspected" areas, the relative magnitude of errors in the census is likely to be greater than in areas not so suspected. A combination of intensive checks in suspected areas with checks of small samples for the remainder of the country is perhaps the most effective method of checking over-all accuracy.

# COMPARING CENSUS RECORDS WITH LISTS OF NAMES

A method of direct checking which does not require re-enumeration and which can be undertaken even after a considerable lapse of time is the matching of census records with other lists of persons who were living in the area at the time of the census, or with samples from such lists. The kinds of lists that may be used for this purpose include the names of pupils enrolled in school, lists of taxpayers, registers of voters et cetera. The comparison may be made at any time after the census, even if it is already too late to proceed to a sample re-enumeration. It is important, however, to make every effort to ascertain that the persons on such lists actually lived in the specified area at the time of the census.

This method of checking, though it may be less expensive, is normally less satisfactory than reenumeration, for a variety of reasons. It is not always possible to be certain that the persons named on the lists were actually living in the specified area at the time. Such lists usually refer only to portions of the population living in the area; for example, persons who were not on the taxpayers' list are not represented. The extent of coverage of the list may not always be known; for example, the list of schoolchildren may include only a part of all the children enrolled in school. A list of names generally does not provide any information about the families of these individuals, nor very much information about their personal characteristics. A re-enumeration, on the other hand, attempts to locate all persons (or a correct sample of all persons) who were in the area at the time of the census enumeration; in addition to verification of the head count, it also provides an opportunity to determine whether their personal characteristics (age, occupation et cetera) have been correctly reported.

For these reasons, it is generally preferable to check the accuracy of a census by sample re-enumeration rather than by reference to existing lists of persons, unless too much time has elapsed to make such reenumeration still useful for the purpose of checking.

# EXAMPLE OF PARTIAL RE-ENUMERATION IN COSTA RICA

At the time of the 1950 Costa Rican census, efforts were made to resurvey those parts of the country where the enumeration was deemed suspect, for one reason or another. Included in this group were those areas which lacked good maps, and those for which there were any doubts as to the quality of the returns. For census purposes Costa Rica was divided into 316 districts; a post-enumeration check was made in 135 of these. This check was not carried out uniformly throughout the country; the Province of Guanacaste was resurveyed in its entirety whereas in the Province of Cartago only 9 of the 39 districts were revisited, and in Limón, only 2 of the 10 districts.

There were 6,878 persons located in the postenumeration check who had not been recorded at the census. These comprised 0.86 per cent of the total population of the country. It was calculated that if the entire country had been resurveyed the number of omissions which would have been uncovered would have amounted to but 2.01 per cent of the total population. This number was so small that no efforts were made to revise the census count to include an estimate of omitted persons in the country as a whole. Instead the final census count includes only the 793,997 persons originally enumerated plus the 6,878 omissions actually discovered in the post-enumeration check, for a total of 800,875. For all of these people the required data on personal and family characteristics were available so that they could be included in the detailed tabulations; it was not necessary to adjust the tabulations so as to include estimates of other persons presumably omitted, classified by their characteristics.<sup>9</sup>

#### EXAMPLE OF THE POST-ENUMERATION SURVEY IN THE UNITED STATES

Following the 1950 census, the United States Census Bureau undertook a re-enumeration on a sample basis. This sample was designed to provide an estimate of the number of households completely missed, and the number of individuals within enumerated households who had been missed or enumerated erroneously. A sample of 3,500 small areas representing the entire United States was carefully chosen, and re-enumerated meticulously by interviewers who were especially selected and given more intensive training than had been received by the average census enumerator. In these areas the enumerators obtained the addresses of all households that could be located and the names and characteristics of their members. These data were checked against the original census schedules for those 3,500 areas. The post-enumeration check covered a total of 22,000 households.

The check indicated a net under-enumeration of about 1.4 per cent in the original 1950 census count. This was the net result of erroneous omissions and erroneous inclusions, as follows:

- erroneously omitted, 3,400,000 persons, or 2.3 per cent of the census count
- erroneously included, 1,309,000 persons, or 0.9 per cent of the census count
- net error..... 2,091,000 persons, or 1.4 per cent of the census count

The great majority of those missed were the members of households that were entirely omitted:

erroneously	omitted	3,400,000 persons
in missed	households	2,416,000 persons
in enumer	rated households	984,000 persons

Of the persons erroneously included, the great majority should have been enumerated in another enumeration district:

erroneously included	1,309,000 persons
anywhere	198,000 persons
enumeration district	1,111,000 persons

Even this elaborate post-enumeration check, however, probably failed to obtain a complete estimate of the number and proportion of persons missed, since the

<sup>9</sup> Costa Rica, Dirección General de Estadística y Censos, Censo de Población de Costa Rica, San José, 1953, pp. 2, 3. check itself contains some errors. A small-scale field check was carried out following the post-enumeration check, and indicated that the original post-enumeration check contained errors which were, in general, in the direction of underestimating the number of erroneously omitted persons. Some of the errors found in the original post-enumerating check were of the type commonly found in any census enumeration. For example, in some cases, it was not possible to locate a person thought to be living in a given household; in other cases, although the person was located, no interview could be obtained. Infants, who are always under-enumerated in the census, were also under-enumerated in the original postenumeration check, as other data indicated. Although the magnitude of the errors in the check is largely unknown, those errors which could be identified were all in the direction of under-estimating the number of persons erroneously omitted. Altogether, these various checks indicate that the estimated net deficiency of 1.4 per cent in the original census enumeration is the minimum.10

In short, the United States experience indicates that, as repeated checks are employed, the size of the errors in the enumeration is progressively reduced, but that perfection is never attained. Some people are always missed no matter how carefully a census enumeration is conducted. Since any post-enumeration check survey is somewhat expensive, it is evident that a point is eventually reached at which additional cost of an additional improvement in accuracy exceeds the value of the improvement.

# SAMPLE VERIFICATION OF THE CENSUS COUNT IN INDIA

After the Indian census of 1 March 1951, the Government of India "requested all State Governments to carry out a sample verification of the 1951 Census Count according to a scheme framed by the Registrar General."<sup>11</sup> The re-enumeration of samples of the population was carried out by most Indian States, in the majority of cases during the latter part of the year 1951. In general, the scheme provided for re-enumeration of one-tenth of one per cent of occupied houses in every census tract, by verification officers specially appointed from among qualified civil officials. The purpose of the operation was to discover, by visits to carefully selected households, any cases of "clear omissions", "fictitious entry", or "erroneous count of visitors and absentees", as well as any evidence of omission of entire households in the vicinity of households selected. The selection of households was to be made from census lists and had to conform rigidly to specified procedures.

The details of re-enumeration procedures were not uniform in every State. Enumerators were given special training and their attention was directed particularly to the kinds of census errors which they might be expected to encounter. In at least one State (Mysore), the verification officers were also informed that the lists given to them, which had been copied from census returns, contained some "ghost" entries, that is to say, entries of fictitious persons purposely added to the list and made to look entirely plausible. These "ghosts" were added to the lists in order to increase the vigilance of the verification officers, and most "ghosts" were indeed recognized as such when the field visits were completed.

The area in which the verification was carried out contained 81 per cent of India's population. Within this area, clear evidence was found of omissions amounting to 1.3 per cent of the population, and of faulty inclusion of persons amounting to 0.4 per cent of the population, in addition to various minor errors consisting of omission or double counting of temporary visitors and absentees. As a net result of all these errors, the census count appeared to have fallen short of the truth by 1.1 per cent for all of India (1.0 per cent in rural, and 1.4 per cent in urban areas). The percentage varied among the several regions.

<sup>&</sup>lt;sup>10</sup> United States, Bureau of the Census, 1950 Census of Population, United States Summary, Bulletin P-C 1, pp. xxix, ff., Washington, 1953.

<sup>&</sup>lt;sup>11</sup> India, Registrar General, Sample Verification of the 1951 Census Count (Census of India, Paper No. 1, 1953), New Delhi, 1953, p. 3.