What is Demography?

Demography is the science of population. Like most other sciences, demography may be defined narrowly or broadly. The narrowest sense is that of "formal demography." Formal demography is concerned with the size, distribution, structure, and change of populations. **Size** is simply the number of units (persons) in the population. **Distribution** refers to the arrangement of the population in space at a given time, that is, geographically or among various types of residential areas. **Structure**, in its narrowest sense, is the distribution of the population among its sex and age groupings. **Change** is the growth or decline of the total population or of one of its structural units. The components of change in total population are births, deaths, and migrations. In analyzing change in structure, however, we have to include the transition from one group to another. In the case of age, this is expressed as a simple function of time.

A broader sense includes additional characteristics of the units. These include ethnic characteristics, social characteristics, and economic characteristics. Ethnic characteristics like race, legal nationality, and mother tongue shade into social characteristics. Other examples of social characteristics are marital and family status, place of birth, literacy, and educational attainment. Economic characteristics include economic activity, employment status, occupation, industry, and income, among others. Other characteristics that might be encompassed are genetic inheritance, intelligence, and health; but the usual sources of demographic data, such as censuses, seldom deal with these directly. Furthermore, demography may look beyond the basic personal units to such customary social groupings as families and married couples.

The widest sense of demography extends to applications of its data and findings in a number of fields including the study of problems that are related to demographic processes. These include the pressure of populations upon resources, depopulation, family limitation, eugenics, the assimilation of immigrants, urban problems, legislative apportionment, manpower, and the mal-distribution of income.

Hauser and Duncan regard the field of demography as consisting of a narrow scope – demographic analysis – and a wider scope – population studies. "Demographic analysis is confined to the study of components of population variation and change. Population studies are concerned not only with population variables but also with relationships between population changes and other variables – social, economic, political, biological, genetic, geographical, and the like. The field of population studies is at least as broad as interest in the 'determinants and consequences of population trends.'"

Demographic Data and their uses

Counts of persons are obtained from censuses and sample surveys and from the files of continuous population registers. Counts of events are obtained from registered vital events (births, deaths, marriages, divorces, etc.), and also from continuous population registers. Sometimes censuses and surveys inquire about events, e.g., the number of children women have borne in the preceding 12 months. Any of these counts may be shown in the form of multiple classifications, e.g., population by sex and age for urban areas or deaths by age and cause. Various demographic measures, such as percentages, ratios, and averages may be derived from them. Furthermore, numbers of registered vital events can be related to the corresponding population to produce vital rates, for example, the number of deaths per 1,000 of the population or the number of births to married women 20 to 24 years old. Such vital rates can also be derived from continuous population registers without the use of other sources of demographic data.

The resulting demographic statistics can then be used to describe the distribution of the population in space, its density and degree of concentration, the fluctuations in its rate of growth, its movements from one area to another, and the force of natality, nuptiality, and mortality within it. These demographic statistics have many and increasingly varied applications. The fields of application include public health; local planning for land use, school and hospital construction, public utilities, etc.; marketing; manpower analysis; family planning programs; land settlement; immigration and emigration policy; and many others. An analysis of current demographic levels and past trends is the necessary first step in the construction of population forecasts that in turn form the underpinning of national plans for economic development and other programs, including explicit population policies in some cases.

Sample Surveys, Registration Systems, and Other Sources

Current sample surveys like censuses represent periodic stock-takings, whereas registration systems (universal population registers, vital records, etc.) and immigration and emigration control systems involve the continuous recording of data even though the published statistics are extracted only at periodic intervals. Thus, some of the desiderata in census data are not applicable to registration data.

Sample Surveys:- Unlike censuses, many surveys are taken under private auspices. Although a periodic sample survey constitutes a continuing "micro-census" of a country or other area, a one-time survey can provide useful data for an ad hoc purpose. Surveys taken as census pretests or as experiments in survey methodology need not necessarily result in published tables based on the substantive demographic data. In other respects (individual enumeration, universality, and simultaneity), the scope of the survey and the methods of recording the data should be subject to the same principles as are censuses.

Registration System:- In the interests of national uniformity, it is best for a register to be operated by an agency of the national government. In some countries with a federal form of government or at least with relatively autonomous state governments, however, good vital statistics have been produced under an arrangement whereby the states handle the registration process and the national government exercises a coordinating function, setting standards and publishing national reports, for example.

The equivalent of the desideratum "individual enumeration" in the census or survey is the individual vital record. As Hauser and Duncan put it, "To provide adequate data for demography, a vital registration system must include procedures to assure the filing of a uniform record for every vital event – for example, live birth, death, stillbirth, and marriage: to provide for complete and usable answers to the inquiries on the record form; and to enable the information in the record form to be processed for purposes – that is, edited, coded, tabulated, and presented, preferably through some central office which provides vital statistics for the nation and its subdivisions on a comparable basis. Principles and procedures for achieving these objectives have been evolved over the years."

A registration system should likewise have "universality within a defined territory." The principle of "simultaneity" does not apply, of course; but it suggests another criterion that is pertinent to registration, namely, a specified maximum interval of time between the occurrence of the event and its recording. For most demographic purposes, moreover, the data should be tabulated as of the date of occurrence, not the date of recording. On the other hand, tabulation by the place of occurrence is less useful than tabulation by the place of residence of the person concerned.

Again, "defined periodicity" would not apply to the continuous recording of data, but it does suggest to us that compilations of the records for statistical purposes should be made periodically. (The date – day, month, year – of both occurrence and registration should be on the record form.) The usual interval for publication purposes is the year; but some series should be published quarterly, or even monthly. Finally, the compilers of statistics from registers, like the agencies that collect and tabulate census data, are obligated to publish, evaluate, and, to some extent, analyze them.

The Balancing Equation

The most basic method of demography is the decomposition of population change into its components, or, conversely, the synthesis of the components to estimate the total population change. Schematically, we may express this process in terms of the fundamental equation

$$P_t - P_0 = B - D + I - O, \ldots\ldots\ldots\ldots..(1)$$

where P_t is the population at the end of the period, P_0 that at the beginning of the period, B is births, D is deaths, I is in-migration, and O is out-migration.

This simple equation, which is called the "balancing equation" (or the "inflow-outflow relationship" or the "component equation") has many forms and many uses. To be exactly true (i.e., represent a necessary relationship), it must apply to a fixed territory and there must be no measurement errors. In fact, the equation may be used to estimate the net error in this system of demographic statistics. If we find that the right-hand side differs from the left-hand side by an amount e, then we can write,

$$P_t - P_0 = B - D + I - O + e, \ldots\ldots\ldots\ldots(2)$$

Here e can be called the "residual error" or the "error of closure." On the basis of additional knowledge about the accuracy of the various terms, one may be able to decide whether e can be attributed as a measurement error almost wholly to a particular term in the equation. For example, if there is evidence that the right-hand terms are all based on very accurate registration data and the population figures come from successive censuses, then e would represent the relative accuracy of coverage of the two censuses. If e is positive, P_t is more nearly complete than P_0; if e is negative, then the reverse would be true.

Let us consider some other illustrations of the uses of (1) or its variations. Suppose that a country has adequate vital statistics and statistics on immigration and emigration. Then t years after the last census, it is desired to make a postcensal of the current national population. We have

$$P_t = P_0 + B - D + I - O, \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3)$$

In this form, the equation may be thought of as the “basic estimating equation,” which uses a straightforward bookkeeping procedure.

If we are interested in projecting the population to a future date, we can use equation (3) in principle by making assumptions about the future births, deaths, and migration. Especially in the case of births and deaths, however, these assumptions are ordinarily made in the form of fertility and mortality rates, not in the form of the absolute numbers of births and deaths.

For another application, suppose that we have two successive population counts for a subnational area (province, county, commune, etc.). We also have vital statistics on births and deaths but no statistics on internal migration (or on the extent to which external migration affects the individual subnational areas). Then we may write

$$M = I - O = (P_t - P_0) - (B - D) \ldots\ldots\ldots\ldots\ldots\ldots\ldots(4)$$

where M is the net migration to or from the area. In other words, to estimate the intercensal net migration for the subnational area, we subtract the natural increase, B-D, from the total population change, $P_t - P_0$.

However, we should mention another kind of elaboration of the balancing equation, namely, its use for a population subgroup, such as the male population, the female population of childbearing age, the native population, or university graduates. For some subgroups, the males in the native population, for example, we simply have to obtain the corresponding components, i.e., statistics on births, deaths, and migration for that subgroup. This restriction may be expressed by using the superscript i, to denote the subgroup, thus:

$$P_t^i - P_0^i = B^i - D^i + I^i - O^i \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(5)$$

Methods of Estimation from Sample Registration Data

Chadrasekharan-Deming’s Method:- The logic of considerations suggests that, other things being equal, the best results will be obtained if comparisons are performed at the level of the smallest unit, i.e., for individuals or for individual events. When the results of two systems – such as a sample survey and sample vital registration – are matched on such a level, it is possible to obtain a numerical estimate of the degree of completeness of both systems and hence to estimate the true total number of events, on the basis of assumptions described below. Case-by-case matching of data from a registration system and a survey was employed in connection with the 1940 and 1950 Censuses of the United States and the Current Population Survey in 1969-70 (to measure completeness of birth registration, or of both infant underenumeration and birth underregistration), but the technique of estimating the total number of events was refined and tested by Chandrasekharan and Deming.

The essential features of the Chandrasekharan-Deming procedure may be summarized as follows (using as an example statistics of births). Suppose that births are recorded for a given year in a sample vital registration system and in a corresponding sample survey (conducted at the end of the year) in which a question on births during the 12-month period preceding the survey is asked. Suppose, furthermore, that the two sets of birth records so obtained are matched event by event. From the matching procedure, the following classification of these events may be obtained:

C= number of events recorded in both registration and survey

N_1 =events recorded only by registration

N_2 =events recorded only by the survey

X= events missed by both systems

The classification may be represented in the following schematic table:

	In registration system	Not in registration system	Total
In sample survey	C	N_2	S
Not in sample survey	N_1	X^*	
Total	R		N^*

*Estimate.

An estimate of the total number of events N may be then obtained as

$$N=C+N_1+N_2+X$$

In turn, X is estimated as N_1N_2/C, thus giving the expression for the estimated total number of events as

$$N=C+N_1+N_2+ N_1N_2/C$$

which may also be written as

$$N=\frac{(C+N_1)(C+N_2)}{C}=\frac{RS}{C}$$

where R denotes the number of events recorded by registration and S denotes the number of events recorded by the survey.

From rearranging the last formula, it can be seen that the Chandrasekharan-Deming formula estimates the completeness of the coverage of the registration system as the match rate of the survey, and estimates the completeness of the coverage of the survey as the match rate of the registration:

$$\frac{R}{N}=\frac{C}{S} \text{ and } \frac{S}{N}=\frac{C}{R}$$

In reviewing estimates based on a single system (survey), it has been emphasized that comparisons of estimates based on alternative estimating procedures constitute essential checks on the quality of the results obtained. It is the great merit of the dual systems of estimating vital events that such checking is a built-in feature of the estimating procedure. Unfortunately, the simplicity of the estimating formulas conceals a number of difficulties in the practical application of the method. The nature of the major problems will be discussed below briefly.

Suppose that in a dual system the number of births in a given year and in a given geographic area is found to be 1200 when recorded by birth registration, whereas the number registered in a retrospective survey is 1300. Suppose also that subsequent individual matching of the births recorded in the two systems is successful in 900 instances. Using the notation given above, we have:

$$R=1200,\ S=1300,\ C=900,\ N_1=300,\ N_2=400$$

Applying the Chandrasekharan-Deming technique, we estimate the number of births missed by both registration and survey as

$$X=\frac{300\times 400}{900}=133$$

Hence the estimated total number of births is obtained as

N=900+300+400+133=1733

Inserting these figures in our schematic table, we have:

	In registration system	Not in registration system	Total
In sample survey	900	400	1300
Not in sample survey	300	133*	433*
Total	1200	533*	1733*

*Estimate.

In other words, the completeness of the registration of births is estimated as 69.2 percent $\left(\frac{1200}{1733}\ or\ \frac{900}{1300}\right)$ and the completeness of the listing of births in the survey is estimated as 75.0 percent $\left(\frac{1300}{1733}\ or\ \frac{900}{1200}\right)$.

The validity of the above estimates will depend, however, on the fulfillment of the following main conditions:

(1) the matching procedure successfully identifies all true matches and, conversely, only true matches are identified as matches.

(2) All events identified in either of the two systems are true events, i.e., occurred in the population under investigation and in the appropriate time period.

(3) The two systems are independent, i.e., the probability of an event being omitted from one system is not related to the chance of the event being omitted from the other system.

Evaluation and Adjustment of Age Data

Concepts and Types of Age Errors:

The errors in the reporting of age have probably been examined more intensively than the reporting errors for any other question in the census. Three factors may account for this intensive study: Many of these errors are readily apparent, measurement techniques can be more easily developed for age data, and the actuaries have had a special practical need to identify errors and to refine the reported data for the construction of life tables. Errors in the tabulated data on age may arise from the following types of errors of enumeration: Coverage errors, failure to record age, and misreporting of age. There is some tendency for the types of errors in age data to offset one another; the extent to which this occurs depends not only on the nature and magnitude of the errors but also on the grouping of the data.

The defects in census figures for a given age or age group due to coverage errors and misreporting of age may each be considered further in terms of the component errors. Coverage errors are of two types. Individuals of a given age may have been missed by the census or erroneously included in it (e.g., counted twice). The first type of coverage error represents gross **underenumeration** at this age and the balance of the two types of coverage errors represents **net underenumeration** at this age.

In addition, the ages of some individuals included in the census may not have been reported, or may have been erroneously reported by the respondent, erroneously estimated by the enumerator, or erroneously allocated by the census office. A complete array of census reports of age in comparison with the true ages of the persons enumerated would show the number of persons at each age for whom age was correctly reported in the census, the number of persons incorrectly reporting “into” each age from lower or higher ages, and the number of persons incorrectly reporting “out of” each age into higher or lower ages. Such tabulations permit calculation of measures of **gross misreporting** of age, referred to, in general, as **response variability** of age. If, however, we disregard the identity of individuals and allow for the offsetting effect of reporting “into” and reporting “out of “ given ages, much smaller errors are found than are shown by the gross errors based on comparison of reports for individuals. Such **net misreporting** of a characteristic is, in general, referred to as **response bias**. The combination of net underenumeration and net misreporting for a given age is termed **net census undercount** (net census overcount, if the number in the age is overstated) or **net census error**.

For example, the group of persons reporting age 42 in the census consists of (1) persons whose correct age is 42 and (2) those whose correct age is over or under 42 but who erroneously report age 42. The latter group is offset partly or wholly by (3) the number erroneously reporting “out of” age 42 into older or younger ages. The difference between groups (2) and (3) represents the net misreporting error for age 42. In addition, the census count at age 42 is affected by net underenumeration at this age, i.e., by the balance of the number of persons aged 42 completely omitted from the census and the number of persons aged 42 who are erroneously included in the census.

We shall consider the types of deficiencies in census tabulations of age under four general headings: (1) Errors in single years of age, (2) errors in grouped data, (3) reporting of extreme old age, and (4) failure to report age.

Single Years of Age

Measurement of Age and Digit Preference:- The tendency of enumerators or respondents to report certain ages at the expense of others is called **age heaping**, age preference, or **digit preference**. The latter term refers to preference for the various ages having the same terminal digit. Age heaping is most pronounced among populations or population subgroups having a low educational status. The causes and patterns of age or digit preference vary from one culture to another, but preference for ages ending in "0" and "5" is quite widespread. In some cultures certain numbers are specially avoided, e.g., 13 in the West and 4 in the Orient. Heaping is the principal type of error in single-year-of –age data, although single ages are also affected by other types of age misreporting, net underenumeration, and nonreporting or misassignment of age. Age 0 is underreported often, for example, because "0" is not regarded as a number by many people and because parents may tend not to think of newborn infants as regular members of the household. We shall confine ourselves to the topic of age heaping, that is, age preference or digit preference.

Population of the Philippines, by Single Years of Age: 1960
(Source: United Nations, *Demographic Year Book*, 1962, table 6)

Age	Number	Age	Number
Total	27,087,685	26 years	358,549
Under 1 year	786,464	27 years	376,221
1 year	888,180	28 years	395,766
2 years	963,230	29 years	300,610
3 years	969,309	30 years	535,924
4 years	965,232	31 years	222,086
5 years	957,698	32 years	318,481
6 years	928,673	33 years	246,260
7 years	938,899	34 years	233,700
8 years	841,636	35 years	401,936
9 years	702,492	36 years	242,659
10 years	841,356	37 years	242,462
11 years	581,400	38 years	316,210
12 years	796,786	39 years	225,207
13 years	619,293	40 years	434,156
14 years	596,592	41 years	126,632
15 years	565,714	42 years	217,881
16 years	566,942	43 years	169,167
17 years	538,891	44 years	151,142
18 years	651,318	45 years	319,118
19 years	491,441	46 years	160,329
20 years	565,801	47 years	160,855
21 years	494,895	48 years	237,287
22 years	515,823	49 years	155,094
23 years	456,892	50 years	313,636
24 years	425,212	51 years	78,534
25 years	522,203	52 years	128,935

Age	Number
53 years	93,279
54 years	95,715
55 years	163,093
56 years	87,754
57 years	71,828
58 years	93,049
59 years	72,206
60 years	275,436
61 years	31,299
62 years	49,634
63 years	40,154
64 years	34,381
65 years	102,440
66 years	26,445
67 years	35,311
68 years	40,711
69 years	20,921
70 years	136,771
71 years	13,000
72 years	28,017
73 years	16,662
74 years	14,490
75 years	50,558
76 years	15,010
77 years	11,878
78 years	23,353
79 years	9,212
80 years	73,741
81 years	5,532
82 years	9,331
83 years	5,653
84 years	5,089
85 years	18,604
86 years	4,803
87 years	5,617
88 years	4,388
89 years	4,000
90 years	57,111

In principle, a post-enumeration survey or a sample reinterview study should provide considerable information on the nature and causes of errors of reporting in single ages. A tabulation of the results of the check re-enumeration by single years of age, cross-classified by the original census returns for single years of age, could not only provide an indication of the net errors in reporting both of specific terminal digits and of individual ages but could also provide the basis for an analysis of the errors in terms of the component directional biases characteristic of reporting at specific terminal digits and ages. In practice, however, the size of sample of the reinterview survey ordinarily precludes any evaluation in terms of single ages.

Indexes of age preference:- In place of sample reinterview studies, various arithmetic devices have been developed for measuring heaping on individual ages or terminal digits. These devices depend on some assumption regarding the form of the true distribution of population by age over a part or all of the age range. On this basis an estimate of the true number or numbers is developed and compared with the reported number or numbers.

The simplest devices assume, in effect, that the true figures are **rectangularly** distributed (i.e., that there are equal numbers in each age) over some age range (such as a 3-year, 5-year or 11-year age range) which includes and, preferably, is centered on the age being examined. For example, an index of heaping on age 30 in the 1960 census of the Philippines may be calculated as the ratio of the enumerated population aged 30 to one-third of the population aged 29, 30, and 31 (per 100):

$$\frac{P_{30}}{\frac{1}{3}(P_{29}+P_{30}+P_{31})}\times 100$$

=

$$\frac{535{,}924}{1/3(300{,}610+535{,}924+222{,}086)}\times 100$$

=151.9 …………. (1)

or, alternatively, as the ratio of the enumerated population aged 30 to one-fifth of the population aged 28, 29, 30, 31, and 32 (per 100):

$$\frac{P_{30}}{\frac{1}{5}(P_{28}+P_{29}+P_{30}+P_{31}+P_{32})}\times 100$$

$$=\frac{535{,}924}{1/5(395{,}766+300{,}610+535{,}924+222{,}086+318{,}481)}\times 100$$

$$=151.1 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(2)$$

In this case the two indexes are approximately the same whether a 3-year group or a 5-year group is used; both indicate considerable heaping on age 30. The higher the index the greater the concentration on the age examined; an index of 100 indicates no concentration on this age. Particularly where the age range is large (e.g., 11 years), the assumption regarding the true form of the distribution may alternatively be regarded as an assumption of linearity (that is, that the true figures form an arithmetic progression, or that they decrease by equal amounts from age to age over the range) if the age under consideration is centered in the age range selected.

Whipple's index:- Indexes have developed to reflect preference for or avoidance of a particular terminal digit or of each terminal digit. For example, employing again the assumption of rectangularity in a 10-year range, heaping on terminal digit "0" in the range 23 to 62 may be measured by comparing the sum of the populations at the ages ending in "0" in this range with one-tenth of the total population in the range:

$$\frac{\sum(P_{30}+P_{40}+P_{50}+P_{60})}{\frac{1}{10}\sum(P_{23}+P_{24}+P_{25}+\ldots+P_{60}+P_{61}+P_{62})}\times 100 \ldots\ldots\ldots\ldots\ldots\ldots(3)$$

Similarly, employing either the assumption of rectangularity or of linearity in a 5-year range, heaping on multiples of five (terminal digits "0" and "5" combined) in the range 33 to 62 may be measured by comparing the sum of the populations at the ages in this range ending in "0" or "5" and one-fifth of the total population in the range:

$$\frac{\sum(P_{25}+P_{30}+\ldots+P_{55}+P_{60})}{\frac{1}{5}\sum(P_{23}+P_{24}+P_{25}+\ldots+P_{60}+P_{61}+P_{62})}\times 100$$

$$=\frac{\sum_{23}^{62} P_{a\,ending\,in\,0\,or\,5}}{1/5\sum_{23}^{62} P_a}\times 100 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(4)$$

For the Philippines in 1960, we have,

$$\frac{2{,}965{,}502}{1/5(9{,}506{,}437)}\times 100=156.0$$

The corresponding figure for the United States in 1960 is 100.9. This measure is known as Whipple's index. It varies between 100, representing no preference for "0" or "5" were reported. Accordingly, the Philippine figures show the U. S. figure. In fact, the population tabulated at these ages for the Philippines may be said to overstate the corresponding unbiased population by more than 50 percent.

The choice of the range 23 to 62 is largely arbitrary. In computing indexes of heaping, the ages of early childhood and extreme old age are often excluded since they are more strongly affected by other types of errors of reporting than by preference for specific terminal digits and the assumption of equal decrements from age to age is less applicable.

The procedure described can be extended to provide an index for each terminal digit (0, 1, 2, etc.). The population ending in each digit over a given range, say 23 to 82, or 10 to 89, may be compared with one-tenth of the total population in the range, as was done for digit "0" above, or it may be expressed as a percentage of the total population in the range. In the latter case, an index of 10 percent is supposed to indicate an unbiased distribution of terminal digits, and hence, presumably, accurate reporting of age. Indexes in excess of 10 percent indicate a tendency toward preference for a particular digit, and indexes below 10 percent indicate a tendency toward avoidance of a particular digit.

Myers' blended method:- Myers has developed a "blended" method to avoid the bias in indexes computed in the way just described that is due to the fact that numbers ending in "0" would normally be larger than the following numbers ending in "1" to "9" because of the effect of mortality. The principle employed is to begin the count at each of the 10 digits in turn and then to average the results. Specifically, the method involves determining the proportion which the population which the population ending in a given digit is of the total population ending in a given digit is of total population 10 times, by varying the particular starting age for any 10-year age group. The following Table shows

Calculation of Preference Indexes for Terminal Digits by Myers' Blended Method, for the Philippines: 1960

Terminal digit, a	Population with terminal digit a		Weights for ..		Blended population		Deviation of percent from 10.00=(6)-10.00
	Starting at age 10+a (1)	Starting at age 20+a (2)	Column 1 (3)	Column 2 (4)	Number = (1)x(3)+(2)x(4) (5)	Percent Distribution (6)	(7)
0	3,176,821	2,335,465	1	9	24,196,006	16.06	6.06
1	1,553,378	971,978	2	8	10,882,580	7.22	2.78
2	2,064,888	1,268,102	3	7	15,071,378	10.00	-----
3	1,647,360	1,028,067	4	6	12,757,842	8.47	1.53
4	1,556,321	959,729	5	5	12,580,250	8.35	1.65
5	2,143,666	1,577,952	6	4	19,173,804	12.72	2.72
6	1,462,491	895,549	7	3	12,924,084	8.58	1.42
7	1,443,063	904,172	8	2	13,352,848	8.86	1.14
8	1,762,082	1,110,764	9	1	16,969,502	11.26	1.26
9	1,278,691	787,250	10	0	12,786,910	8.49	1.51
Total	(X)	(X)	(X)	(X)	150,695,204	100.00	20.07
Summary index of age preference= Total/2	(X)	(X)	(X)	(X)	(X)	(X)	10.04

the calculation of the indexes of preference for terminal digits in the age range 10 to 89 for the Philippine population in 1960 based on Myers' blended method. In this particular case, the first starting age was 10, then 11, and so on, to 19. The abbreviated procedure of calculation calls for the following steps:

Step (1) Sum the populations ending in each digit over the whole range, starting with the lower limit of the range (e.g., 10, 20, 30, …., 80; 11, 21, 31, …., 81).

Step (2) Ascertain the sum excluding the first population combined in step (1) (e.g., 20, 30, 40, …., 80; 21, 31, 41, …, 81).

Step (3) Weight the sums in steps (1) and (2) and add the results to obtain a blended population (e.g., weights 1 and 9 for the 0 digit; weights 2 and 8 for the 1 digit).

Step (4) Convert the distribution in step (3) into percents.

Step (5) Take the deviation of each percent in step (4) from 10.0, the expected value for each percent.

The results in step (5) indicate the extent of concentration on or avoidance of a particular digit. The weights in step (3) represent the number of times the combination of ages in step (1) or (2) is included when the starting age is varied from 10 to 19. Note that the weights for each terminal digit would differ if the lower limit of the age range covered were different. For example, if the lower limit of the age range covered were 23, the weights for terminal digit 3 would be 1 (col. 1) and 9 (col. 2) and for terminal digit 0 would be 8 (col. 1) and 2 (col. 2).

The method thus yields an index of preference for each terminal digit, representing the deviation, from 10.0 percent, of the proportion of the total population reporting on the given digit. A summary index of preference for all terminal digits is derived as one-half the sum of the deviations from 10.0 percent, each taken without regard to sign. If age heaping is nonexistent, the index would approximate zero. This index is an estimate of the minimum proportion of persons in the population for whom an age with an incorrect final digit is reported. The theoretical range of Myers' index is 0, representing no heaping, and 90 which would result if all ages were reported at a single digit, say zero. A summary preference index of 10.0 for the Philippines in 1960 is obtained.

Grouped Data

Types of Error and Methods of Measurement:- As indicated earlier, several important types of errors remain in age data even when the data are grouped. In addition to some residual error due to digit preference, 5-year or 10-year data are affected by other types of age misreporting and by net underenumeration. Absolute net underenumeration would tend to cumulate as the age band widens. On the other hand, the percent of net underenumeration would be expected to vary fairly regularly over the age distribution, fluctuating only moderately up and down. Absolute net age misreporting error and the percent of net age misreporting error should tend to take on positive and negative values alternatively over the age scale, dropping to zero for the total population of all ages combined. For the total population, therefore, net census error and net underenumeration are identical. In general, as the age band widens, net age misreporting tends to become less important and net underenumeration tends to dominate as the type of error in age data.

Age ratio analysis:- The quality of the census returns by age groups may also be evaluated by comparing so-called **age ratios**, calculated from the census data, with expected or standard values. An age ratio may be defined as the ratio of the population in the given age group to one-third of the sum of the populations in the age group itself and the preceding and following groups, times 100. The age ratio for a 5-year age group, ${}_5P_a$, is defined then as follows:

$$\frac{{}_5P_a}{1/3({}_5P_{a-5} + {}_5P_a + {}_5P_{a+5})} \times 100$$

Barring extreme fluctuations in past births, deaths, or migration, the three age groups should form a nearly linear series. Age ratios should then approximate 100.0, even though actual historical variations in these factors would produce deviations from 100.0 in the age ratio for most ages. In as much as, over a period of nearly a century, most countries have experienced not only minor fluctuations in populations in population changes but also major upheavals, age ratios for some ages may deviate substantially from 100.0 even where reporting of age is good. The assumption of an expected value of 100.0 also implies that coverage errors are about the same from age group to age group and that age reporting errors for a particular group are offset by complementary errors in adjacent age groups. In sum, age ratios serve primarily as measures of net misreporting, not net census error, and they are not to be taken as valid indicators of error for particular age groups.

U.N. Index:- The regularity of the change in the expected sex ratio from age to age which we have just seen provides a basis for elaborating the age-accuracy index based solely on age ratios described earlier to incorporate some measure of the accuracy of sex ratios. The United Nations has proposed such an **age-sex accuracy index**. In this index the mean of the differences from age to age in reported sex ratios, without regard to sign, is taken as a measure of the accuracy of the observed sex ratios, on the assumption that these age-to-age changes should approximate zero. The U. N. age-sex accuracy index combines the sum of (a) the mean deviation of the age ratios for males from 100.0, (b) the mean deviation of the age ratios for females from 100.0 and (c) three times the mean of the age-to-age differences in reported sex ratios. In the U. N. procedure, an age ratio is defined as the ratio of the population in a given age group to one-half the sum of the populations in the preceding and following groups. The calculation of the U. N. age-sex accuracy index is illustrated in the table for Greece in 1961. Applying the U. N. formula, we have: 10.4+9.1+3(5.4)=35.5. Comparable indexes for a few other countries are:

Country (census year)	U. N. age-sex accuracy index
United States (1960)	12.2
Sweden (1963)	15.1
Philippines (1960)	32.8
Greece (1961)	35.5
Taiwan (1964)	49.3
Turkey (1960)	70.6

Census age-sex data were described by the United Nations as “accurate,” “inaccurate,” or “highly inaccurate” depending on whether the U. N. index was under 20, 20 to 40, or over 40.

Calculation of the United Nations Age-Sex Accuracy Index, for Greece: 1961

Age	Population		Analysis of sex ratios		Analysis of age ratios			
	Male	Female	Ratio	Successive differences	Male		Female	
			$[(1)\div(2)]\times 100=$	$\Delta(3)=$	Ratio	Deviations from 100 (5)-100=	Ratio	Deviations from 100 (7)-100=
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Under 5 years	493,600	385,100	128.17	(X)	(X)	(X)	(X)	(X)
5 to 9 years	364,900	346,900	105.19	+22.98	83.69	-16.31	93.43	-6.57
10 to 14 years	378,400	357,500	105.85	-0.66	111.94	+11.94	109.65	+9.65
15 to 19 years	311,200	305,200	101.97	+3.88	84.06	-15.94	84.72	-15.28
20 to 24 years	362,000	363,200	99.67	+2.30	110.27	+10.27	107.01	+7.01
25 to 29 years	345,400	373,600	92.45	+7.22	97.57	-2.43	101.32	+1.32
30 to 34 years	346,000	374,300	92.44	+0.01	117.17	+17.17	116.80	+16.80
35 to 39 years	245,200	267,300	91.73	+0.71	87.89	-12.11	87.01	-12.99
40 to 44 years	212,000	240,100	88.30	+3.43	85.90	-14.10	91.15	-8.85
45 to 49 years	248,400	259,500	95.72	-7.42	111.99	+11.99	109.01	+9.01
50 to 54 years	231,600	236,000	98.14	-2.42	102.82	+2.82	101.55	+1.55
55 to 59 years	202,100	205,300	98.44	-0.30	105.73	+5.73	97.79	-2.21
60 to 64 years	150,700	183,900	81.95	+16.49	99.11	-0.89	111.32	+11.32
65 to 69 years	102,000	125,100	81.53	+0.42	87.14	-12.86	84.76	+15.24
70 to 74 years	83,400	111,300	74.93	+6.60	(X)	(X)	(X)	(X)
Total (irrespective of sign)	(X)	(X)	(X)	74.84	(X)	134.56	(X)	117.80
Mean	(X)	(X)	(X)	5.35	(X)	10.35	(X)	9.06

Index=3 times mean difference in sex ratios plus mean deviations of male and female age ratios = 3X5.35+10.5+9.06=35.46

$$\text{Ratio}=\frac{{}_5P_a}{1/2({}_5P_{a-5}+{}_5P_{a+5})}\times 100 .$$

Source: United Nations, *Demographic Yearbook*, 1962, table 5.

Percent Changes by Age: - An important phase of the analysis of age data relates to the measurement of changes over time. Most of the methods of description and analysis of age data to be considered below are applicable not only to the comparison of different populations but also to the comparison of the same population at different dates.

Use of Indexes: - Comparison between two percentage age distributions is facilitated by calculating indexes for each age group or overall indexes for the distributions. Age distributions for different areas, for population subgroups in a single area, and for the same at different dates may be compared in this way.

Index of relative difference:- The magnitude of the differences between any two age distributions, whether for different areas, dates, or population subgroups, may be summarized in single indexes from the individual age-specific indexes as defined above, i.e., the index of relative difference and the index of dissimilarity. In the former procedure, (1) the deviations of the age-specific indexes from 100 are summed without regard to sign, (2) one-nth (n representing the number of age groups) of the sum is taken to derive the mean of the percent differences at each age, and (3) the result in (2) is divided by 2 to obtain the index of relative difference. The formula is:

$$\text{IRD}=\frac{1}{2}\times\frac{\sum\left|\left(\frac{r_{2a}}{r_{1a}}\times 100\right)-100\right|}{n}$$

To reduce the likelihood of very large percent differences at the oldest ages, which are given equal weight in the average, a broad terminal age should be used.

Index of dissimilarity: - Another summary measure of the difference between two age distributions – the index of dissimilarity – is based on the absolute differences between the percents at each age. In this procedure, the differences between the percents for corresponding age groups are determined, they are assumed without regard to sign, and one-half of the sum is taken. (Taking one-half the sum of the absolute differences is equivalent to taking the sum of the positive differences or the sum of the negative differences.) The general formula is then:

$$\text{ID}=\frac{1}{2}\sum|r_{2a}-r_{1a}|$$

It should be apparent that the magnitude of these indexes is affected by the number of age in the distribution as well as by the size of the differences and, hence, that the results are of greatest value in comparison with similarly computed indexes for other populations.

Age Dependency Ratios: - The variations in the proportions of children, aged persons, and persons of "working age" are taken account of jointly in the so-called age dependency ratio. The age dependency ratio represents the ratio of the combined child population and aged population to the population of intermediate age. One formula for the age dependency ratio useful for international comparisons relates the number of persons under 15 and 65 and over to the number 15 to 64:

$$\frac{P_{0-14}+P_{65+}}{P_{15-64}}\times 100$$

Summary Measures of Age Composition, for Various Countries: Around 1960.

Country and year	Median age	Percent of total population		Ratio of aged persons to children (per 100)
		Under 15 years	65 years and over	
	(1)	(2)	(3)	(4)
Chile (1960)	23.3	39.6	4.3	10.8
France (1961)	33.3	25.4	12.1	47.8
Ghana (1960)	18.4	44.6	3.2	7.1
Honduras (1961)	16.1	47.8	2.4	5.1
India (1961)	20.5	41.0	3.1	7.5
Iran (1966)	17.3	46.1	3.9	8.4
Italy (1961)	31.6	24.6	9.5	38.8
Japan (1960)	25.6	30.0	5.7	19.1
Sweden (1960)	36.2	22.0	12.0	54.4
Syria (1960)	17.2	46.3	4.7	10.1
Taiwan (1956)	17.9	44.2	2.5	5.6
USSR (1959)	26.6	30.4	6.2	20.5
United Arab Republic (1960)	19.4	42.7	3.5	8.1
United States (1960)	29.5	31.1	9.2	29.7
Venezuela (1961)	17.8	44.8	2.8	6.2
Yugoslavia (1961)	26.4	31.5	6.1	19.3

Source: Basic data from United Nations, *Demographic Year book*, 1962, table5; and 1964, table5.

Applying the formula to the data for India in 1961, we have:

$$\frac{180{,}019{,}000 + 13{,}468{,}000}{245{,}112{,}000} \times 100 = 78.9$$

Separate calculation of the child-dependency ratio, or the component of the age dependency ratio representing children under 15 (i.e., the ratio of children under 15 to persons 15 to 64), and the old-age dependency ratio, or the component representing persons 65 and over (i.e., the ratio of persons 65 and over to persons 15 to 64), gives values of 73.4 and 5.5.