# Chapter IV

# RATES, RATIOS AND OTHER INDICES

In migration analysis, there are special problems associated with the construction of rates and other indices. The principal reason for this is that each move involves two areas—origin and destination. The discussion below indicates some guidelines for coping with these problems and gives some illustrations of measures that are useful for the study of migration.

## Migration rates and ratios

In considering the various alternatives for rate or ratio computation, it should be understood that much depends upon the nature of the problem. Different problems may call for different kinds of rates. Available alternatives may result in trivial differences (for example, if the migration interval is short), in which case the population base that is the most convenient is the one to be preferred. But where differences are substantial, especially where the pattern of differences either over time or between population groups is affected, careful consideration should be given both to selecting the type of rate that is most suitable for the problem at hand, and to the interpretations that are made of the rates that are used.

In general, a rate expresses the number of events or the number of persons having a given characteristic as a proportion of the population exposed to risk during a specified time interval. A migration rate is, then, the number of migrants (or the number of migrations) related to the population that could have performed the migrations during the given migration interval. The equation is written in algebraic form, as follows:

$$m = \frac{M}{P} \cdot k \qquad (26)$$

where $m$ = the rate of migration for the specified migration interval

$M$ = the number of migrations or the number of persons migrating during the interval

$P$ = the population exposed to the likelihood of migration during the interval

$k$ = a constant, usually 100 or 1,000.

Both the selection of appropriate rate bases and the interpretation of the rates depend in part upon the nature of the available data (e.g., how a migrant is defined) and in part upon the object of the analysis. Thus, in the equation given above, if $M$ refers to migrations, the rate gives a measure of the incidence of moves in $P$ and it must be understood that, since one person may move more than once, some members of the population will appear more than once in the numerator. If, on the other hand, $M$ refers to migrants, $M/P$ becomes a probability rate and gives a measure of the proportion moving *at least once* during a given migration interval. If one had a distribution of migrants by number of migrations, he could calculate separate probabilities (*a*) of migrating once only, (*b*) of migrating more than once, (*c*) of migrating specific numbers of times and (*d*) for migrants, of migrating again after having migrated once. In each case:

$$M/P + N/P = 1$$

where $M$ refers to persons who performed the specified moves, $N$ to persons who did not, and $P$ to persons who could have done so ($M+N = P$). For (*d*) above, $P$ would represent all migrants, $M$ would represent migrants who moved more than once, and $N$ migrants who moved once only.

In most of the discussion to follow, it will be assumed that the data available are of the type obtained in population censuses and surveys; that is, refer to migrants as enumerated at the census or survey date and therefore exclude migrant deaths as well as allowing only one move per migrant.

### *Migration streams*

For migration streams, the population at risk is the population in the area of origin. The equation may be written as:

$$m_{ij} = \frac{M_{ij}}{p_i} \cdot k \qquad (27)$$

where the first subscript refers to the area of origin and the second to the area of destination. This procedure expresses all streams as probabilities of moving from a given origin to a given destination. Such rates may also be said to measure the attraction that the area of destination exerts upon the population at origin. They are useful in various types of analysis and for projections.

The particular form and time reference of $p_i$ that is appropriate depends upon the characteristics of the particular $M_{ij}$. Thus, for migration data based on residence at a fixed prior date, the exposed population is the population of $i$ at that date who survived to the census date and the rate expresses the probability that persons living in $i$ at the first date (time $t$) and surviving to the second date (time $t+n$) will be living in $j$ at the second date. The equation may be written as:

$$m_{ij} = \frac{M_{ij}}{p_{i,t+n} - M_{\cdot i} + M_{i\cdot}} \cdot k \qquad (28)$$

where $M_{\cdot i}$ refers to all in-migrants to $i$ ($\sum_j M_{ji}$) and $M_{i\cdot}$ refers to all out-migrants from $i$ ($\sum_j M_{ij}$). This procedure confines the measure to persons who were alive at the first date and survived to the second date. For some purposes, it may be considered desirable to base the rate upon the actual population at time $t$, including persons who died during the interval. In that case, the rate measures the probability that persons living in $i$ at time $t$ will *survive and be living in* $j$ at time $t+n$.

For lifetime migration streams, equation (28) is equally appropriate. Here, $M_{ij}$ refers to persons born in $i$ and living in $j$, $M_{\cdot i}$ to lifetime in-migrants to $i$, and $M_{i\cdot}$ to lifetime out-migrants from $i$. The rate measures the probability that persons born in $i$ and surviving will be living in $j$ at time $t+n$.

For data on duration of present residence cross-classified by place of last residence, rates for specific migration intervals ending at the census date (e.g., durations of five years or less, ten years or less etc.) can be calculated by the same formula. For these, $M_{ij}$ refers to persons whose last moves occurred within the specified interval, originating in $i$ and terminating in $j$. The base population for this rate does not refer to a specific point in time; it is composed of persons who resided in $i$ throughout the entire interval plus all persons whose last move (made after time $t$) originated in $i$, regardless of where they resided between time $t$ and time of last move. This is a genuine probability measure which relates last moves made between time $t$ and $t+n$ to the population at risk.

For some purposes, a measure that expresses the migrant stream as a ratio to the population at destination may be indicated. The ratio may be written as:

$$m_{ij} = \frac{M_{ij}}{p_{j,t+n}} \cdot k \qquad (29)$$

It expresses migrants as a proportion of the population of $j$ at time $t+n$. It is not to be regarded as an "at-risk" rate unless we accept the notion that the population at destination is exposed to the risk of receiving in-migrants. This interpretation of the term diverges from its usual meaning, which implies that the exposed population must be capable of performing the acts or experiencing the events represented in the numerator of the rate, in short, is capable of being in the numerator. Non-migrant persons at destination are incapable of being in-migrants. We must therefore conclude that ratios based on the receiving population are not rates in the probability sense.[1] However they do give a measure of the *impact* of migration upon the receiving population and in this sense may be useful analytical tools. They also permit analysis of the composition of the population with respect to migration status. But in using them, it should be kept in mind that they constitute "a relative frequency statement, which must be handled with caution and whose range of permissible inferences is restricted".[2]

For net streams ($M_{ij} - M_{ji}$) and for gross interchange or turnover ($M_{ij} + M_{ji}$), a rate base that combines the populations of $i$ and $j$ is appropriate. This combination may be the sum or the average of the two populations, preferably the latter, since it will yield a rate level that is comparable with that of its stream components. The composition and time reference of the base population should be the same for both $p_i$ and $p_j$:

$$m_{(ij-ji)} = \frac{M_{ij} - M_{ji}}{.5(p_{i,t+n} - M_{\cdot i} + M_{i\cdot}) + .5(p_{j,t+n} - M_{\cdot j} + M_{j\cdot})} \cdot k \qquad (30)$$

This rate expresses the net stream or net shift as a proportion of the population within which the shift occurred, or as a proportion of the average of their populations.

### *In-migration, out-migration, net migration*

Principles analogous to those discussed above apply with respect to the calculation of rates of in-migration, out-migration and net migration for component areas. Here we consider each area in relation to all other areas combined rather than taking them in pairs. For area $i$, in-migration is the sum of all incoming streams ($M_{\cdot i} = \sum_j M_{ji}$); out-migration is the sum of all outgoing streams ($M_{i\cdot} = \sum_j M_{ij}$) and net migration is the difference between the two ($M_{\cdot i} - M_{i\cdot}$). A probability rate of out-migration would relate $M_{i\cdot}$ to the population of the area of origin ($p_i$). A probability rate of in-migration to $i$ would relate $M_{\cdot i}$ to the population of the remainder of the country ($P - p_i$). These procedures will yield rates of in-migration with a general level that is very much lower than that of rates of out-migration, since the relative difference between the two population bases will be much greater than the relative difference between the two numerators. The two sets of rates would therefore not be directly comparable as to levels, but one could study the two rate distributions and obtain insights into differences between areas as revealed by them.

For rates of net migration, the logically consistent base is the sum of the populations of the two areas concerned, $p_i + (P - p_i)$, i.e., the entire population of the country. For a given migration interval, such rates will bear the same relationship to one another as do the amounts of net migration. They therefore do not have any analytical value for a single migration interval. They are, however, potentially useful for studying time trends for individual areas.

In determining the form and time reference of the population base that is appropriate, the same principles

[1] For a more extended discussion, see C. Horace Hamilton, "Practical and mathematical considerations in the formulation and selection of migration rates", *Demography* (Chicago), vol. 2, 1965, pp. 429-443. See also Ralph Thomlinson, "The determination of a base population for computing migration rates", *Milbank Memorial Fund Quarterly* (New York), vol. XL, 1962, pp. 356-366.

[2] William Haenszel, "Concept, measurement and data in migration analysis", *Demography* (Chicago), vol. 4, No. 1, 1967, p. 255.

as those developed for stream rates are applicable. Thus, rates of out-migration from $i$ would have the same base as equation (28), that is, $(p_{i,t+n}-M_{.i}+M_{i.})$. For in-migration, the base would be the complement of the above: $P_{t+n}-(p_{i,t+n}-M_{.i}+M_{i.})$. The base for the rate of net migration would then be: $P_{t+n}$.

It should be noted again that it is also possible to take the view that deaths should not be excluded from the base even though they are excluded from the numerator. Thus, Shryock calculates "at-risk" rates of intra-area migration and out-migration based on $p_{i,t}$.[3] Such rates are measures of the probability of *migrating and surviving* to the end of the migration interval. It should also be noted that there may be problems connected with the estimation of $p_{i,t}$ if the migration data refer to a migration interval that does not coincide with the intercensal interval. This will always be the case with data of the duration-by-place-of-last-residence type, which, though susceptible of analysis in terms of migration intervals, do not have an exposed-to-risk population that can be referred to a definite initial date.

The above approach to rate computation for component areas is not the one most commonly used. Instead, it is customary to regard in-migration, out-migration and net migration as "attributes" of the given area and to base all three rates on some measure of that area's population. Such rates would take the following general forms:

$$\text{In-migration: } \frac{M_{.i}}{p_i}k \qquad (31)$$

$$\text{Out-migration: } \frac{M_{i.}}{p_i}k \qquad (32)$$

$$\text{Net migration: } \frac{M_{.i}-M_{i.}}{p_i}k \qquad (33)$$

The particular measure of $p_i$ that should be used is a matter of some disagreement. According to one approach, the first of these rates should be based on the population of $i$ at time $t+n$ (that is, should include in-migrants and exclude out-migrants), that the second should be based on the survivors at time $t+n$ of the population residing in $i$ at time $t$ (that is, should include out-migrants and exclude in-migrants), and that the third should be based on an average of these two quantities.[4] These may be written as follows:

$$m_{.i} = \frac{M_{.i}}{p_{i,t+n}}k \qquad (34)$$

$$m_{i.} = \frac{M_{i.}}{p_{i,t+n}-M_{.i}+M_{i.}}k \qquad (35)$$

$$m_{(.i-i.)} = \frac{M_{.i}-M_{i.}}{p_{i,t+n}-.5(M_{.i}-M_{i.})}k \qquad (36)$$

These procedures give the rate of in-migration the same base as in equation (29), out-migration, the same base as in equation (28), and net migration an average of the two.

Another approach emphasizes the convenience of using the same base for all three rates. In this approach, an average base seems desirable, inasmuch as it contains one half of net migration. This base is identical with that in equation (36) above. Such rates may be regarded as measures of the effects, or of the relative importance of migration with regard to the population of $i$. They can thus be manipulated in the same way as amounts. Still another approach bases the rates on the non-migrant population: $p_{i,t+n}-M_{.i}$. The resulting ratios give a measure of the impact of migration upon the non-migrant segment of the population.[5]

### *Indirect measures of net migration*

The preceding discussion has assumed the availability of statistics on gross migration. But since most of the migration data currently available are estimates of net migration derived by indirect methods, it is important to examine the special problems that arise in the calculation of rates for data of those types. In this presentation, it is assumed that area rates of migration would not be based on the total population, except for special analysis of historical series, and that what is wanted is a consistent and logically defensible base for computing area-specific rates of net migration, some of which will be positive and some of which will be negative.

#### *VS estimates*

As indicated earlier, the Vital Statistics method yields estimates of the net balance of migration (which is the same, whether migrants or migrations are considered) and includes the net balance resulting from the migration of persons who died. An appropriate base for this rate is the average population, usually estimated as $.5(p_{i,t}+p_{i,t+n})$, an approximation to the population at the midpoint of the migration interval. This base contains half of migrants and their deaths, plus non-migrants and half of their deaths.

#### *CSR estimates*

Forward census survival ratios yield estimates of the net balance of surviving migrants and are directly comparable with the net balances obtained from census data on residence at a fixed prior date. An appropriate base for them is therefore the same as that given in equation (36), namely: $p_{i,t+n}-.5(M_{.i}-M_{i.})$.

It has been shown that when the relative error in net migration is the same as the relative error in $p_{i,t+n}$ that error will vanish if the rate is based on $p_{i,t+n}$. This holds true also for a population base in the above form, since the relative error in $p_{i,t+n}$ will be the same as that in $p_{i,t+n}-.5(M_{.i}-M_{i.})$.

Estimates of net migration derived by reverse survival ratios contain intercensal deaths to both the components,

[3] Henry S. Shryock, Jr., *Population Mobility Within the United States* (Community and Family Study Center, University of Chicago, 1964), chap. 6.

[4] See C. Horace Hamilton, op. cit., pp. 429-443.

[5] Ratios of this type have been used by Ann R. Miller in "The migration of employed persons to and from metropolitan areas of the United States", *Journal of the American Statistical Association* (Washington, D.C.), vol. 62, December 1967, pp. 1418-1432.

non-migrants and net migration. Net migration thus may be expressed in terms of forward estimates as $(M_{\cdot i}-M_{i\cdot})\,1/S$ where $S$ is the survival ratio. The indicated base is then: $[p_{i,t+n}-.5(M_{\cdot i}-M_{i\cdot})]\,1/S$. It is at once apparent that:

$$\frac{M_{\cdot i}-M_{i\cdot}}{p_{i,t+n}-.5(M_{\cdot i}-M_{i\cdot})}=\frac{(M_{\cdot i}-M_{i\cdot})\,\frac{1}{S}}{[p_{i,t+n}-.5(M_{\cdot i}-M_{i\cdot})]\,\frac{1}{S}} \quad (37)$$

and that identical relative errors will still be cancelled.

Estimates of net migration derived by the average method may be expressed in terms of "forward" estimates, as follows:

$$(M_{\cdot i}-M_{i\cdot})\,\frac{1+S}{2S}$$

and the corresponding estimates, derived by the application of the forward and reverse methods is given by:

$$\frac{(M_{\cdot i}-M_{i\cdot})\,\frac{1+S}{2S}}{[p_{i,t+n}-.5(M_{\cdot i}-M_{i\cdot})]\,\frac{1+S}{2S}} \quad (38)$$

and again, identical relative errors will cancel.

Before concluding this chapter, it should be emphasized that although other forms of rate bases may be considered acceptable for rates of net migration from various points of view, the fundamental problem of finding a base is usable for computing both rates of net in-migration and rates of net out-migration seems to be most nearly solved by using a base that is not "weighted" in favour of either in-migration or out-migration, but that contains one half of each. The solution of basing rates on the non-migrant population, a solution that is perhaps the least biased in this respect, is not feasible with estimates derived by indirect methods because the magnitude of this segment of the population cannot be determined.

### *Specific rates*

Because the propensity to migrate varies sharply with age and is likely to differ considerably by sex, it is desirable to calculate rates that are specific for these characteristics and, indeed, for other characteristics if the needed classifications are available. The principles and procedures to be followed in selecting suitable population bases are the same as those given above in general terms. The chief concern here is to maintain cohort identity between numerator and denominator. Thus, equation (28) may be rewritten to indicate age-specificity, as follows:

$$m_{ij}(x)=\frac{M_{ij}(x)}{p_{i,x+n,t+n}-[M_{\cdot i}(x)-M_{i\cdot}(x)]}\cdot k \quad (39)$$

where $M(x)$ refers to migrants who were $x$ years of age at time $t$ (aged $x+n$ at time $t+n$). The other symbols are as previously defined.

### *Problems of annualizing period rates*

It is a common practice to express amounts or rates of change as annual averages when the period to which the data refer is more than one year. In general, this practice renders data for time periods of differing lengths reasonably comparable. It is not, however, an appropriate procedure for most types of migration data. Unless the migration measure is a count, or estimate, of all moves of the given type made during the given interval, an average obtained by dividing the amount or the rate for an interval longer than a year by the number of years in the interval will understate the actual annual amounts to a degree that tends to increase as the length of interval increases. The census approach to the measurement of migration identifies migrants on the basis of one past residence only, and allows only one move per migrant. As a result, the count of migrants for a long interval (for example, ten years) will be less than the sum of the numbers that would be obtained if the count were made at the end of each one-year interval that forms part of the longer period. In this respect, migration data differ from statistics of births and deaths in which events are additive and the sums of the numbers for individual years are equal to the numbers for the period as a whole. (Averaging migration data for a series of one-year migration intervals is, of course, not subject to the objections indicated.)

These observations apply to statistics of gross migration (migration streams, in-migrants, out-migrants, total migrants within an area or a country etc.). Measures of net migration may be averaged without danger of understatement provided the balance of migrant deaths is included in the estimate. This is true because, as indicated earlier, the balance of migrations equals the balance of migrants for any given migration interval.

But the calculation of annual averages in the attempt to render comparable the measures for intervals of differing lengths must be approached with caution, no matter what the nature of the migration data. Precautions are especially necessary in the comparison of age-specific rates of migration. Because the propensity to migrate differs strongly with age, the rate obtained for any age groups is closely linked to the exposure interval. If, for example, one wishes to compare rates of the age group 20-24 years (age at end of interval) for two migration intervals, one ten years in length and the other five years in length, he will not achieve comparability by dividing the first rate by 10 and the second rate by 5 (or by using some more elaborate technique) to arrive at an annual average. The reasons are that, in the first case, he is averaging the experience of a cohort that was moving from ages 10-14 to ages 20-24, while, in the second, he is averaging the experience of a cohort moving from 15-19 to 20-24 years. Rates between ages 10-14 and 15-19 are likely to be much lower than rates between 15-19 and 20-24. Before comparisons are undertaken, the ages should be adjusted to reflect the average age during the interval and perhaps some interpolative procedures will be called for in order to approximate identical age groups. With these manipulations, it should not be lost to sight that possibly doubtful assumptions are being made about the regularity of change in migration behaviour over the interval and that fine comparisons are probably not justifiable.

For migration intervals significantly longer than a decade, it is doubtful that "annualization" of migration data should be attempted at all.

## INDEX OF REDISTRIBUTION

Net migration and natural increase or natural change are the components of population growth and redistribution. Their effects are not always synchronous nor are they mutually exclusive. The contribution of natural change to area growth is usually positive. The contribution of migration may be either positive or negative. Furthermore, natural change affects the contribution that migration makes to population change; migration, in turn, has an effect upon the contribution of natural change. In short, there is interaction between migration and natural change. To develop the complexities of this interaction by factoring population change would, however, take us beyond the scope of this Manual.[6] The present chapter is, therefore, limited to the contribution of internal migration to population redistribution.

Inasmuch as the algebraic sum of areal gains and losses through internal migration is zero, measures of redistribution due to migration ($R_M$) are obtained by summing net changes of like sign, which is the same as taking one half the sum of all changes without regard to sign. Thus:

$$R_M = \sum_{+}(M_{.i}-M_{i.}) = \frac{\sum_{i}|M_{.i}-M_{i.}|}{2} \quad (40)$$

where $(M_{.i}-M_{i.})$ refers to the measure of net change due to migration and the symbol $\sum_{+}$ indicates the summation of those net changes having a positive sign. This number can be expressed as a rate of redistribution, or a rate of displacement due to migration, by relating it to the total population within which the displacement occurred. An appropriate base is the average population. The rate ($r_M$) may be written as:

$$r_M = \frac{R_M}{.5(P_t+P_{t+n})} \cdot k \quad (41)$$

An illustration of the application of equations (40) and (41) is given in table 38. Appropriate modification of the above formulae will yield measures of redistribution due to natural increase and also total net redistribution.

The measure of redistribution is specific for the class of area to which the data apply, as interprovincial, interstate, intercounty etc. Different systems of areas yield different amounts and different rates. Smaller areas will yield indices at least as high as do larger areas, but generally higher. This characteristic of the measure means that international comparisons are hazardous.

Bachi has demonstrated that if all mobility is taken into account—movement within component areas as well as between them—the resultant measures of redistribution will be unaffected, or almost unaffected, by the class of geographical unit used. His techniques represent a centrographic approach to the measurement of redistribution. He makes use of measures of central tendency (the "mean centre" or "centre of gravity" of the population) and dispersion (the "standard distance" of the population from its centre). Briefly, the basic procedures are as follows:

(*a*) The location of the centres of the smallest areal units for which data are available is expressed in degree of latitude on the horizontal co-ordinate ($X_i$) and, in degrees of longitude, on a vertical co-ordinate ($Y_i$);

[6] See K. C. Zachariah, op. cit., pp. 191-196. Also, K. E. Vaidyanathan, op. cit., pp. 113-125.

TABLE 38. ILLUSTRATION OF COMPUTATION OF INDEX OF REDISTRIBUTION: NATIVES, UNITED STATES OF AMERICA, 1940-1950

| *Subregions* | *Intercensal net migration (thousands)* | |
|---|---|---|
| *North East* | | |
| N-1 | −91 | |
| N-2 | +7 | |
| N-3 | −310 | |
| N-4 | +334 | $R_M$ = 3,948,00 |
| *South* | | |
| S-1 | −322 | Average population = 0.5 ($P_{1940}+P_{1950}$) |
| S-2 | −1,249 | |
| S-3 | −877 | = 129,535,000 |
| *North Central* | | |
| C-1 | +373 | |
| C-2 | +590 | $r_M = \frac{3,948,000}{129,535,000}$ = 3.1 per cent |
| C-3 | −453 | |
| *West* | | |
| W-1 | −56 | |
| W-2 | +160 | |
| W-3 | +3,074 | |
| Sum of net gains (or losses) | ±3,948 | |

SOURCE: H. T. Eldridge and D. S. Thomas, *Population Redistribution and Economic Growth, United States, 1870-1950*, vol. III, *Demographic Analysis and Interrelations* (Philadelphia, American Philosophical Society, 1964), tables 1.33, p. 111, A1.12, p. 252.

(*b*) The centre of population ($\bar{X}$, $\bar{Y}$) is determined by the means of $X$ and $Y$ weighted by the population ($p_i$) of the areal units. Thus:

$$\bar{X} = \frac{\sum_i p_i X_i}{\sum_i p_i} \quad (42)$$

$$\bar{Y} = \frac{\sum_i p_i Y_i}{\sum_i p_i} \quad (43)$$

(*c*) The standard distance ($d$) is obtained as follows:

$$d = \sqrt{\frac{\sum_i p_i(X_i-\bar{X})^2}{\sum_i p_i} + \frac{\sum_i p_i(Y_i-\bar{Y})^2}{\sum_i p_i}} \quad (44)$$

These measures can be calculated for the migrant population before and after migration, or for migrants and non-migrants (or the general population) at various points in time. Comparison of the results yields information on the prevailing directions of migration and its effect upon population spread.[7]

### Indices of migration differentials and selectivity

One of the advantages of census data on migration is that all characteristics required in the census for the general population of the country are available for migrant segments. It is, therefore, potentially possible to analyse such characteristics of the migrants as sex, age, marital status, educational attainment, occupation, industry, and in fact all personal and household characteristics that were covered in the census. These data open up the broad field of analysis of migration selectivity and differentials.

Migrants tend to be different from the parent population in a number of characteristics; that is, they are not a random sample. For example, there may be an unduly large proportion of young adults among migrants. Such differences between characteristics of migrants (at the time of out-migration) and of the population from which they originate are called migration selectivity or *origin differentials*. They arise from the fact that the rate of out-migration is not the same in all the population subgroups.

Even if migrants were not different from the parent population, they might still be different from the population which they enter. The differences between the characteristics of migrants and non-migrants at the destination are called *destination differentials*. They arise because of the fact that the rate of in-migration at the place of destination is not the same for all population subgroups.

Following are some procedures that are developed for the measurement of migration differentials.

[7] See Roberto Bachi, "Standard distance measures and related methods for spatial analysis", Regional Science Association, *Papers*, vol. X, *Zurich Congress*, 1962, pp. 83-132; also "Statistical analysis of geographical series", *Bulletin de l'Institut international de la statistique* (The Hague), vol. 36, No. 2, 1958, pp. 229-240 (reprinted in Brain J. L. Berry and Duane F. Marble, *Spatial Analysis* (Englewood Cliffs, N.J., Prentice-Hall, Inc., 1968), pp. 101-109).

Migration differentials may be measured in a number of ways, but all the methods are based on the frequency distributions of migrants and non-migrants at the place of destination with respect to the particular characteristic under investigation. The differences in the patterns of the two distributions measure the magnitude of the differentials. Two common procedures are: (1) in terms of differential proportions and (2) in terms of differential ratios (or rates).

The two methods cited yield different measures of differentials but identical indices.

Let:

$M_1, M_2, \ldots M_n$ represent the distribution of migrants at the place of destination with respect to some characteristic, and

$N_1, N_2, \ldots N_n$ represent the distribution of non-migrants in the same area with respect to the same characteristic.

A measure of migration differentials by the differential proportions method is given by:

$$\left(\frac{M_i}{M} - \frac{N_i}{N}\right)$$

where $M = \sum_i M_i$ and $N = \sum_i N_i$ — $i(= 1, 2 \ldots n)$ denotes the category under investigation

An index of migration differentials by this method is obtained by dividing the differences in the proportions between migrants and non-migrants by the proportion for the non-migrants. Thus, we have:

$$IMD_i \text{ (procedure 1)} = \left[\left(\frac{M_i}{M} - \frac{N_i}{N}\right) \Big/ \frac{N_i}{N}\right] \cdot k \quad (45)$$

A measure of the differential by the ratio method is given by:

$$\left(\frac{M_i}{N_i} - \frac{M}{N}\right)$$

where $M$, $N$, $M_i$, and $N_i$ have the same meaning as above.

An index of migration differentials by this method is given by:

$$IMD_i \text{ (procedure 2)} = \left[\left(\frac{M_i}{N_i} - \frac{M}{N}\right) \Big/ \frac{M}{N}\right] \cdot k \quad (46)$$

It can easily be shown that equations (45) and (46) are identical.

$$\left[\left(\frac{M_i}{M} - \frac{N_i}{N}\right) \Big/ \frac{N_i}{N}\right] \cdot k = [(M_iN - N_iM)/MN_i] \cdot k \quad (45)$$

$$\left[\left(\frac{M_i}{N_i} - \frac{M}{N}\right) \Big/ \frac{M}{N}\right] \cdot k = [(M_iN - N_iM)/MN_i] \cdot k \quad (46)$$

However, if we express the difference between the proportions of migrants and non-migrants as a ratio to the proportion for the total population (i.e., $P_i/P$)

instead of to the proportion for the non-migrants (as in equation (45)), the resulting index is given by:

$$IMD_i \text{ (procedure 3)} = \left[\left(\frac{M_i}{M} - \frac{N_i}{N}\right)\Big/\frac{P_i}{P}\right] \cdot k \qquad (47)$$

where $P_i$ is the total population in category $i$ and $P = \sum_i P_i$.

An example illustrating the computational procedures for the measurement of destination differentials with respect to industrial affiliation of male migrant workers in Greater Bombay is given in table 39. The proportions of migrant male workers in each industry category in column (5) are compared with those of non-migrants in column (6) to derive measure of migration differentials in column (8). In calculating the indices of migration differentials, the denominators may be taken as proportions of non-migrant male workers in each industry category (procedure 1), which yield indices given in column (9). On the other hand, we may take as denominators the proportions of total male workers in Bombay in each industry category (procedure 3), which yield the indices shown in column (12). The latter procedure is preferable when the non-migrants form a small proportion of the total population, as they do for Greater Bombay.

The indices by the first two methods for each industry category, shown in columns (9) and (11), are identical.

TABLE 39. PROCEDURE FOR MEASURING DESTINATION DIFFERENTIALS: AN EXAMPLE WITH RESPECT TO INDUSTRY GROUPS, MALE WORKERS IN GREATER BOMBAY, 1961

| Industry group (1) | *Number (thousands)* Migrants (2) | Non-migrants (3) | Total (4) | *Percentage distribution* Migrants (5) | Non-migrants (6) | Total (7) | *Differential* (5)-(6) (8) | *Index (procedure 1)* (8)/(6) (9) | *Index Ratio* (2)/(3) (10) | *Index (procedure 2)* (11) | *Index (procedure 3)* (8)/(7) (12) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Agriculture and mining | 19.2 | 7.3 | 26.5 | 1.46 | 3.24 | 1.72 | −1.78 | −55 | 2.62 | −55 | −103 |
| Manufacture of textiles | 303.0 | 41.7 | 344.8 | 23.05 | 18.45 | 22.37 | 4.60 | 25 | 7.26 | 25 | 21 |
| Manufacture of metals and chemicals | 249.5 | 54.2 | 303.7 | 18.98 | 23.95 | 19.71 | −4.97 | −21 | 4.61 | −21 | −25 |
| Construction | 36.0 | 4.2 | 40.2 | 2.74 | 1.87 | 2.61 | .87 | 46 | 8.51 | 46 | 33 |
| Utilities | 19.2 | 5.2 | 24.4 | 1.46 | 2.31 | 1.59 | −.85 | −37 | 3.67 | −37 | −53 |
| Commerce | 240.9 | 45.8 | 286.7 | 18.32 | 20.23 | 18.60 | −1.91 | −9 | 5.26 | −9 | −10 |
| Transport | 152.2 | 28.2 | 180.3 | 11.57 | 12.45 | 11.70 | −.88 | −7 | 5.40 | −7 | −8 |
| Services | 292.1 | 39.0 | 331.1 | 22.21 | 17.25 | 21.48 | 4.96 | 29 | 7.49 | 29 | 23 |
| Activities not adequately described | 2.6 | 0.6 | 3.2 | .20 | .26 | .21 | − .06 | . | . | . | . |
| TOTAL | 1,314.7 | 226.2 | 1,540.9 | 100 | 100 | 100 | | | 5.81 | | |

SOURCE: Computed from data in K. C. Zachariah, op. cit., table 12.1, p. 241.

*Note:* Index (procedure 2) in column (11) is obtained as follows:

$$\left(\frac{\text{Specific industry ratio}}{\text{All industry ratio}} - 1\right) \cdot 100$$

To give a numerical example: the index of migration differentials for the service industry by the differential proportions method, shown in column (9), is:

$$\left[\left(\frac{M_i}{M} - \frac{N_i}{N}\right)\Big/\frac{N_i}{N}\right] \cdot k$$

$$= [(22.21 - 17.25)/17.25] \cdot 100 = 29$$

and the index of migration differentials for the same industry group by the ratio (or rate) method shown in column (11) is:

$$\left[\left(\frac{M_i}{N_i} - \frac{M}{N}\right)\Big/\frac{M}{N}\right] \cdot k$$

$$= [(7.48 - 5.81)/5.81] \cdot 100 = 29$$

If, however, the proportion of total male workers in each industry is used as the denominator (as in procedure 3) the resulting index for the service industry will be:

$$\left[\left(\frac{M_i}{M} - \frac{N_i}{N}\right)\Big/\frac{P_i}{P}\right] \cdot k$$

$$= [(22.21 - 17.25)/21.48] \cdot 100 = 23$$

as shown in column (12).

In accordance with our discussion earlier in this chapter an example is in table 40 illustrating the measurement of age selectivity of migrants in Japan (defined as those who, a year ago, resided in a place different from the place of enumeration in the 1960 census), by comparing the age composition of migrants with that of the total population of the country in 1960.

The formulae for the indices of migration selectivity or origin differentials by the differential proportions method and differential ratio method are the same as those given in equation (45) and (46). However, in this context:

$M_1, M_2, \ldots M_n$ represent the distribution of total migrants in the country with respect to some characteristic.

$M = \sum_i M_i$

$N_1, N_2, \ldots N_n$ are replaced by $P_1, P_2, \ldots P_n$ to represent the distribution of total population of the country with respect to the same characteristic, and

$P = \sum_i P_i$

The computational procedure for deriving indices of

TABLE 40. PROCEDURE FOR MEASURING MIGRATION SELECTIVITY WITH RESPECT TO AGE, JAPAN, 1959-1960

| Age in 1960 (1) | Interprefectural migrants, 1959-60: Number (2) | Interprefectural migrants, 1959-60: Percentage (3) | Total population, Japan, 1960: Number (4) | Total population, Japan, 1960: Percentage (5) | Procedure 1: Difference (3)-(5) (6) | Procedure 1: Index of selectivity (6)/(5).100 (7) | Procedure 2: Ratio (2)/(4).100 (8) | Procedure 2: Index[a] of selectivity (9) |
|---|---|---|---|---|---|---|---|---|
| 1-14 | 316,900 | 12.3 | 26,434,600 | 28.8 | −16.5 | −57.3 | 1.1988 | −57.3 |
| 15-19 | 684,900 | 26.5 | 9,257,500 | 10.1 | +16.4 | +162.4 | 7.3983 | +162.4 |
| 20-24 | 588,400 | 22.8 | 8,286,400 | 9.0 | +13.8 | +153.3 | 7.1007 | +153.3 |
| 25-29 | 394,800 | 15.3 | 8,220,700 | 9.0 | +6.3 | +70.0 | 4.8025 | +70.0 |
| 30-39 | 315,600 | 12.2 | 13,529,800 | 14.7 | −2.5 | −17.0 | 2.3326 | −17.0 |
| 40-49 | 137,700 | 5.3 | 9,839,100 | 10.7 | −5.4 | −50.5 | 1.3995 | −50.5 |
| 50-59 | 78,900 | 3.1 | 7,861,600 | 8.6 | −5.5 | −64.0 | 1.0036 | −64.0 |
| 60-69 | 41,400 | 1.6 | 5,105,600 | 5.6 | −4.0 | −71.4 | 0.8108 | −71.4 |
| 70-79 | 19,200 | 0.7 | 2,545,600 | 2.8 | −2.1 | −75.0 | 0.7542 | −75.0 |
| 80+ | 4,100 | 0.2 | 677,800 | 0.7 | −0.5 | −71.4 | 0.6048 | −71.4 |
| TOTAL 1+ | 2,581,900 | 100.0 | 91,758,700 | 100.0 | . | . | 2.8137 | . |

SOURCE: Japan, Bureau of Statistics, Office of the Prime Minister, *Population of Japan, 1960, Summary of results* (Tokyo, 1963), table 60, p. 542.

[a] Derived as follows: (computed before rounding)

$$\left(\frac{\text{age-specific ratio}}{\text{all-ages ratio}} - 1\right)100 \qquad \text{Example: } \left(\frac{1.1988}{2.8137} - 1\right)100 = -57.3$$

selectivity is shown in table 40. In columns (3) and (5) are given the percentage distributions of migrants and of the total population in Japan as of 1960, by age group. In columns (7) and (9), the indices of selectivity are shown by the two methods described above. These indices reflect the highly selective nature of migration in the prime age groups.

An important point, which is sometime overlooked in the analysis of migration differentials and selectivity, is that these phenomena vary in magnitude as well as in direction in population subgroups, and consequently the over-all measures of differentials (that is, for the population as a whole) is as much a function of differentials within population subgroups as of the distribution of the total population among subgroups. Following the practice in other branches of demography, we may call the over-all measure *the crude index of migration differentials* and that for population subgroups, *the specific indices of migration differentials.* The fact that the specific indices can be quite different from the crude index is evident from the data given in table 41 where measures of migration differentials with respect to occupational groups in Greater Bombay are given for the male workers as a whole (column 2) and for subgroups by age and

TABLE 41. DESTINATION DIFFERENTIALS, BY OCCUPATIONAL GROUPS, MALE WORKERS, BY EDUCATIONAL CATEGORIES AND AGE, GREATER BOMBAY, 1961

| | Educational category and age group | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Total | | Illiterate | Literate without educational level | | | Primary or junior basic | | | Matriculation higher secondary | | | Degree or diploma | | |
| Occupational group | Standardized (1) | Crude (2) | all ages (3) | 15-34 (4) | 35-59 (5) | 60+ (6) | 15-34 (7) | 35-59 (8) | 60+ (9) | 15-34 (10) | 35-59 (11) | 60+ (12) | 15-34 (13) | 35-59 (14) | 60+ (15) |
| 1. Professional | 0.0 | −1.9 | 0.4 | −0.2 | −0.2 | 0.6 | −1.1 | 0.4 | 2.4 | 0.3 | 0.2 | −0.4 | −0.1 | −2.0 | −1.9 |
| 2. Administrative | 0.5 | −1.1 | 0.5 | −0.5 | 0.3 | −1.1 | 0.0 | 0.4 | 2.0 | 3.0 | −1.5 | −0.5 | 2.6 | 3.1 | 1.1 |
| 3. Clerical | −19.8 | −15.5 | −55.7 | −1.5 | −2.6 | −2.7 | −2.0 | −8.3 | −7.8 | −4.1 | −2.0 | −3.3 | 0.2 | 1.8 | −1.3 |
| 4. Sales | 3.1 | 1.6 | 7.8 | −1.1 | 1.8 | 5.3 | 1.2 | 6.8 | 5.1 | −1.3 | −1.0 | 3.7 | −3.0 | −1.8 | 1.9 |
| 5. Farmers | −1.2 | −1.1 | 0.8 | −4.8 | −3.1 | −3.5 | −1.1 | −1.5 | −1.5 | — | — | — | — | — | — |
| 6. Transport | −1.7 | −1.1 | −6.3 | 0.2 | 0.1 | 1.8 | 0.2 | 0.6 | 1.3 | 2.0 | 0.6 | 0.0 | 0.5 | −0.6 | 0.7 |
| 7. Craftsmen and labourers | 9.4 | 9.8 | 32.8 | 0.1 | 0.0 | −6.8 | −4.6 | −2.2 | −4.1 | −1.1 | 2.6 | 0.1 | −0.3 | −0.3 | −0.2 |
| 8. Service | 9.7 | 9.3 | 19.7 | 7.8 | 3.7 | 6.4 | 7.4 | 3.8 | 2.6 | 1.2 | 1.1 | 0.4 | 0.1 | −0.2 | −0.3 |
| 9. Coefficient of dissimilarity | 22.7 | 20.7 | 62.0 | 8.1 | 5.9 | 14.1 | 8.8 | 12.0 | 13.4 | 6.5 | 4.5 | 4.2 | 3.4 | 4.9 | 3.7 |
| 10. Percentage distribution of migrant workers | . | 100 | 32.0 | 16.7 | 8.9 | 1.4 | 16.6 | 7.7 | 1.3 | 7.8 | 2.8 | 0.6 | 3.3 | 0.9 | 0.2 |

SOURCE: Computed from *Census of India, 1961, Greater Bombay*, tables D-IV and B-VI.

*Note:* The figures in the table show the difference between the percentage of workers in an occupation among migrants, and that among non-migrants. Positive values indicate that there were relatively more migrants in a specified occupational group. Negative figures indicate that there were relatively fewer migrants.

education (columns 3 to 15). These differentials were derived from a series of percentage distributions by occupational group for male migrant workers and for male non-migrant workers in each of the categories shown in the headings of the table. It should be noted that in the census tables illiterate workers are not cross-classified by age.

A summary measure of migration differentials may be obtained by the coefficient of dissimilarity (comparable to the index of redistribution discussed in the preceding section). It is obtained by computing the differences between the percentage distribution of migrants and non-migrants and summing those of like signs.

The coefficients of dissimilarity given in row 9 show that the magnitude of differentials varies considerably from one subgroup to another. In general, the differentials in occupational composition decrease as the level of education increases. Maximum differentials are observed among illiterates and minimum among highly educated. The direction of differentials is not the same in all population subgroups. In row 5, for example, the crude differential of −1.1 indicates that there are relatively fewer farmers among migrants than among non-migrants in Greater Bombay. But for illiterate workers, the reverse relation holds. In this group, there are relatively more farmers among migrants than among non-migrants. Similar tendencies of conflicting differentials in population subgroups may be observed for other occupational divisions. The most striking example of variation in the magnitude of the differentials is given by the clerical workers among whom the crude differential is −15.5 percentage points, but the specific occupational differential ranges from +1.8 to −55.7.

Following the practice in other branches of demography, we may calculate from the specific differentials, standardized indices. These indices for each occupational division, computed on the basis of the percentage distribution of male migrant workers (shown in row 10) are given in column 1. Thus, for transport workers the standardized differential (−1.7) is obtained by the algebraic summation of the product of row 6 (columns 3-15) and row 10, divided by 100. The coefficient of dissimilarity for the standardized differentials is 22.7, which may be taken as a more refined measure of over-all migration differentials.

Let us now examine how far census data are useful in analysing migration selectivity and differentials. Broadly speaking, all types of measure (birth-place, place of last residence, duration of residence and place of residence $x$ years ago) provide materials for such study, but those which separate fixed-term migrants from lifetime or all-time migrants are the most satisfactory. Many of the characteristics of migrants tend to vary rather significantly by length of residence at destination. The sex or the race of a migrant does not change, of course, but his age increases by the same amount as the length of his residence. These are "fixed" characteristics which either do not change or, if they do, the amount and direction of change can be exactly calculated when the length of residence is known. This is not the case with social and economic characteristics such as marital status, education, economic status, industry, occupation etc. For these characteristics, census data are not satisfactory for the analysis of migration selectivity or origin differentials. It is the characteristics of migrants, as of the census date, that are recorded in the census. These characteristics will often be different from those existing at the time of out-migration from communities of origin. Information on characteristics before migration are a pressing need for refined analysis of the selectivity and concomitants of migration. In annex II, an example is given of how, under favourable survey conditions, selectivity can be measured for streams of migrants.

### Some other indices

In migration analysis, as in other fields of demographic research, it is often desirable to calculate relative rates or relative indices of various kinds which reflect variations in the intensity of migration while holding constant certain disturbing factors or certain characteristics of the populations involved. In general, these procedures yield expected frequencies or rates, which represent the stated assumptions and which may be compared with observed frequencies or rates, or from which may be calculated standardized measures that permit comparisons between population groups without the interference of the disturbing factors. An obvious example is standardization for age, a procedure which permits comparisons of all-ages rates of migration while holding constant the contribution that variations in age structure make to variations in levels of mobility.

There are various possible bases of varying degrees of specificity and complexity that may be used for the derivation of expected numbers. A few of the more common procedures are presented here by way of illustration.

#### *Index of preference*

If migration propensities were uniform, the number of out-migrants from $i$ would be $M(p_i/P)$. Similarly, the number of in-migrants to $j$ would be $M(p_j/P)$, where $M$ represents total migrants. The expected number of migrants from $i$ to $j$ will be $M \cdot (p_i/P \cdot p_j/P)$ and an index of preference or relative intensity (IPR) is:[8]

$$IPR = \frac{M_{ij}}{M\left(\frac{p_i}{P} \cdot \frac{p_j}{P}\right)} \cdot k \qquad (48)$$

This procedure takes $M$ as given even though it is known that the magnitude of $M$ is determined by varying propensities as observed in the population.

#### *Index of velocity*

Bogue, Shryock and Hoermann proposed a similar

[8] Several indices of "preference" are discussed in Henry S. Shryock, Jr., *Population Mobility Within the United States* (Community and Family Study Center, University of Chicago, 1964), pp. 267-269.

measure, which they called the "velocity" of migration streams, and which may be written as:[9]

$$IGV = \left(\frac{M_{ij}}{p_i \cdot p_j} \cdot P\right) \cdot k \qquad (49)$$

This gives the total rate of out-migration that $i$ will have if:

$$\frac{M_{i.}}{M_{ij}} = \frac{P}{p_j}$$

When this rate is expressed as a ratio to the general rate $M/P$, the resulting index is identical with (48) above and gives a measure of the relative intensity of $M_{ij}$, or the relation of $M_{ij}$ to the number that would be expected if migration were determined by population size at origin and destination.

*Index of net velocity*

Kono and Shio[10] have used a variant of the index of velocity, which is based on net streams between areal units. This is called the index of net migration velocity.

$$INV = (M_{ji} - M_{ij}) \frac{P}{p_i p_j} \qquad (50)$$

They have also used another measure taking distance into account:

$$\Delta_{ij} \cdot INV = (M_{ji} - M_{ij}) \frac{P}{p_i p_j} \Delta_{ij} \qquad (51)$$

Where $\Delta_{ij}$ stands for a measure of distance between areal units.

*Index of effectiveness*

An index used by both Thomas and Shryock,[11] which relates net migration to turnover, and is called by the latter an "effectiveness index", can be written as follows:

$$IE = \frac{|M_{.i} - M_{i.}|}{M_{.i} + M_{i.}} \cdot k \qquad (52)$$

Similar measures have been calculated for pairs of streams:

$$Ie = \frac{|M_{ij} - M_{ji}|}{M_{ij} + M_{ji}} \cdot k \qquad (53)$$

The "expected" value here is unity or $k$, if migration is completely effective; that is, if migration is all in one direction.

In using measures such as those described above, thought should be given to the degree to which they may defeat the purpose of understanding the causes and concomitants of migration by building into the measure assumptions that may themselves need testing, or that may obscure other relationships that need to be examined.

[9] See Donald J. Bogue, Henry S. Shryock, Jr. and Siegfried A. Hoermann, *Subregional Migration in the United States, 1935-1940*, vol. I, *Streams of Migration Between Subregions*, Scripps Foundation Studies in Population Distribution, No. 5, 1957, pp. 48-49. Also, see Donald J. Bogue, "Internal migration", in Philip M. Hauser and Otis Dudley Duncan (eds.), *The Study of Population* (Chicago, University of Chicago Press, 1959), pp. 503-504.

[10] Shigemi Kono and Mitsuru Shio, *Inter-Prefectural Migration in Japan, 1956 and 1961: Migration Stream Analysis* (New York, Asia Publishing House, 1965), p. 9.

[11] Dorothy S. Thomas, *Social and Economic Aspects of Swedish Population Movements, 1750-1933*, (New York, Macmillan Co., 1941), chap. 7; Henry S. Shryock, Jr., op. cit., chap. 9.