The calculation of the index of dissimilarity on a computer terminal

JERRY W. WICKS

Department of Sociology, Bowling Green State University
Bowling Green, Ohio 43403

Description. The index of dissimilarity measures the difference between two relative percentage distributions over a particular group of categories by first summing the differences between the relative frequencies in each category, and second, by dividing the total sum in half in order that the final index value will fall somewhere between 0 and 100.

The formula for the index is:

$$ID = \tfrac{1}{2}\Sigma / \, r_{2a} - r_{1a} /.$$

The index is a measure with numerous applications (See Bogue, 1969, p. 117; and Shyrock, Siegel, et al., 1973, p. 232-233) and is extensively used in situations in which a considerable number of distributions must be compared with one another or with some "standard" distribution. One of the more interesting utilizations of the index has been in determining the degree of racial and ethnic segregation and assimilation in metropolitan areas of the United States (Duncan & Duncan, 1955). When used for this purpose, the index may be interpreted as determining the extent to which a group in residential areas is separated from the rest of the population. The score ranges from "0" (perfect integration) to "100" (complete segregation) and indicates the percent of either population which would have to be redistributed to achieve complete integration.

Input. The investigator must specify the number of units (e.g., census tracts) in the study, enter the total population of each unit, and then enter the population in each of the units for the group being compared.

Output. The program provides the option of either printing total counts on data entered and the index score; or a complete listing of all data entered, relative percentage distributions of both groups, and their differences, and the index score.

Computer and language. The program (Dissim) is written in Call-os Fortran IV and is now in use on an IBM 360/75 with teletypewriter remote facilities.

Limitations. There are two basic limitations. As now written, the user is limited to 500 categories. This is easily altered by making appropriate changes in the DIMENSION statement. Percentages may not be entered since only "raw" totals are acceptable as data. All subtraction and calculation of percentages is completed by the program.

Availability. A listing of the program may be obtained without charge by contacting the author at: Department of Sociology, Bowling Green State University, Bowling Green, Ohio 43403.

REFERENCES

Bogue, D. J. *Principles of demography.* New York: Wiley, 1969.
Duncan, O. D., & Duncan, B. A methodological analysis of segregation indexes. **American Sociological Review, 20,** 210-217, 1955.
Shryock, H. S., Siegel, J. S., et al. *The methods and materials of demography.* U. S. Bureau of the Census. Washington, D. C: U. S. Government Printing Office, 1973.

AGEBLEND: A Fortran IV program to detect digit preference and avoidance in census age data

JERRY W. WICKS

Department of Sociology, Bowling Green State University
Bowling Green, Ohio 43403

Description. There are many deficiencies in census data. While defects in data vary in severity and type from population to population, one deficiency that can be particularly serious and is fairly common (Stockwell & Wicks, 1974) is the tendency for people, when reporting their ages or dates of birth, to select certain preferred digits. This selection process results in the "heaping" of persons throughout an age distribution.

Detecting this deficiency in census data is essential for the demographer. One method available for uncovering digit preference and avoidance is Myers' (1940) index of preference. Basically, Myers' technique involves deriving a weighted sum of the number of persons reporting ages at each of 10 digits $(0,1,2, \ldots, 9)$. If the total of any digit surpasses 10% of the total "blended" population, overselection is indicated. Conversely, if the total of any digit is less than 10% of the total "blended" population, underselection is indicated. The index of preference is obtained by taking one-half of the absolute sum of the differences from 10% of the total blended population for each of the 10 terminal digits. The range of the index is 0 (no age heaping) to 90 (all ages reported at a single digit, such as 5) (Shryock & Siegel, 1973; 204-209).

Input. The user must specify the upper limit of his age data (i.e., 59, 69, 79, etc.), and then, beginning with age 10, enter the population enumerated at each age.

Output. The user may select one of two tables for output. The complete table includes all age data entered, the blended population sum for each digit, the deviation of percent from 10.00, and the index of preference. The short table presents the blended population sum for each digit, the deviation of percent from 10.00, and the index score.

Computer and language. The program is written in Call-os Fortran IV and is currently in use on an IBM 360/75 with teletypewriter remote facilities.

Limitations. If the complete table is requested for output, the terminal must have a carriage width of 130 characters.

Availability. A listing of the program may be obtained without charge by contacting the author at: Department of Sociology, Bowling Green State University, Bowling Green, Ohio 43403.

REFERENCES

Myers, R. J. Errors and bias in the reporting of ages in census data. Transactions of the Actuarial Society of America, 41, 304-315, 1940.
Shryock, H. S., Siegel, J. S., et al. *The methods and materials of demography* (Vol. 1). U. S. Bureau of the Census. Washington, D. C: U. S. Government Printing Office, 1973.
Stockwell, E. G., & Wicks, J. W. Age heaping in recent national censuses. Social Biology, 21, Summer 1974, pp. 163-167.