DOCUMENT RESUME

ED 279 729 TM 870 178

| | |
|---|---|
| AUTHOR | Johnston, Denis F. |
| TITLE | Guide to the Interpretation of Indicator Data. Draft. Technical Report No. 21. |
| INSTITUTION | American Institutes for Research in the Behavioral Sciences. Palo Alto, CA. Statistical Analysis Group in Education. |
| SPONS AGENCY | National Center for Education Statistics (ED), Washington, DC. |
| REPORT NO | AIR-87500-9-81-TR |
| PUB DATE | Sep 81 |
| CONTRACT | 300-80-0823 |
| NOTE | 39p. |
| PUB TYPE | Guides - Non-Classroom Use (055) |
| EDRS PRICE | MF01/PC02 Plus Postage. |
| DESCRIPTORS | Charts; *Data Interpretation; Educational Research; *Error of Measurement; Higher Education; *Measurement Techniques; Secondary Education; *Statistical Analysis; *Statistical Data; Tables (Data); *Visual Aids |
| IDENTIFIERS | *Educational Indicators |

ABSTRACT

This guide is designed to assist those readers of "The Condition of Education" and similar reports who may lack experience or confidence in reading and understanding statistical information. It serves four purposes (1) identifies and describes the principal features of statistical tables and charts; (2) presents a few illustrations of how indicators and other statistical measures are constructed; (3) alerts the reader to both the strengths and weaknesses of statistical data by describing and illustrating the nature and effects of sampling and non-sampling variability; and (4) offers a few precepts and suggestions with respect to the interpretation of statistical data. Numerous examples and illustrations are given in each section to clarify what is said and familiarize the reader with the techniques to use in extracting information from both tabular and graphic data presentations. (JAZ)

ED279729

Technical Report No. 21

# Guide to the Interpretation of Indicator Data

(DRAFT)

Denis F. Johnston



Prepared by

SAGE
STATISTICAL ANALYSIS GROUP IN EDUCATION

For the

National Center for Education Statistics

American Institutes for Research

Box 1113, Palo Alto, California 94302

Technical Report No. 21

# Guide to the Interpretation of Indicator Data

(DRAFT)

Denis F. Johnston

The Statistical Analysis Group in Education
Area I: Indicator Development
American Institutes for Research
P.O. Box 1113
Palo Alto, California 94302

This work was done under Contract No. 300-80-0823 with the National Center for Education Statistics, Department of Education. The content does not necessarily reflect the position or policy of either agency, however, and no official endorsement should be inferred.

September 1981

Summary

SAGE Technical Report No. 17, Workplans for Developing Educational Indicators, called for the preparation of a brief guide to interpreting indicator data that would be written for nonstatisticians and designed for inclusion in The Condition of Education. This guide has been prepared and is presented in the following report. Four major sections are included in the guide: (1) How to Read Statistical Tables and Charts, (2) Statistical Indicators and Other Measurements, (3) Sources of Error in Statistical Data, and (4) Some Notes on Data Interpretation. Numerous examples and illustrations are given in each section to clarify what is said and familiarize the reader with the techniques to use in extracting information from both tabular and graphic data presentations.

TABLE OF CONTENTS

## Guide to the Interpretation of Indicator Data

This guide is designed to assist those readers of The Condition of Education and similar reports who may lack experience or confidence in reading and understanding statistical information. The guide has four main objectives. First, it identifies and describes the principal features of statistical tables and charts so that the reader may develop greater facility in their use. Second, it presents a few illustrations of how indicators and other statistical measures are constructed. Third, it alerts the reader to both the strengths and weaknesses of statistical data by describing and illustrating the nature and effects of sampling and non-sampling variability. Finally, it offers a few precepts and suggestions with respect to the interpretation of statistical data.

### How to Read Statistical Tables and Charts

#### Statistical Tables

All statistical tables include five parts: a title; a boxhead; a stub; a source note; and of course, a set of data in the body of the table. In addition, most tables include a sixth feature: one or more headnotes or footnotes. In general, there are two kinds of statistical tables: reference tables, designed as convenient repositories of detailed statistical information, and summary or text tables, designed to complement the narrative treatment of a given subject by providing pertinent data that can be seen at a glance. Table 1, adapted from the 1981 edition of The Condition of Education (p. 64), is used in the following paragraphs to illustrate these features.

The title. A good title tells the reader what is presented in the table, how that subject-matter is classified, where the data were collected (i.e., what universe they represent), and when the data were obtained (i.e., to what time period they relate). The subject-matter of

Table 1

Public and Private Elementary/Secondary School Enrollment by Metropolitan Status and Racial/Ethnic Group: October 1979

| letropolitan status and racial/ethnic group | Total enrolled (1,000s) | Enrolled in public schools | Enrolled in private schools: Total | Religiously affiliated | Un-affiliated | Affiliation not reported |
|---|---|---|---|---|---|---|
| | | (Percent distribution) | | | | |
| l, metropolitan | 28,435 | 87.7 | 12.3 | 10.4 | 1.6 | 0.3 |
| hite | 22,730 | 86.3 | 13.7 | 11.5 | 1.8 | .3 |
| lack | 5,027 | 93.8 | 6.2 | 5.3 | .7 | .2 |
| lispanic[1] | 2,457 | 90.3 | 9.7 | 8.6 | .8 | .3 |
| otal, central city | 11,106 | 84.0 | 16.0 | 13.5 | 2.1 | .4 |
| White | 7,154 | 79.6 | 20.4 | 17.2 | 2.7 | .5 |
| Black | 3,608 | 92.7 | 7.3 | 6.1 | .9 | .3 |
| Hispanic[1] | 1,507 | 89.1 | 10.9 | 10.3 | .3 | .3 |
| otal, outside central city | 17,329 | 90.0 | 10.0 | 8.4 | 1.3 | .2 |
| White | 15,575 | 89.4 | 10.6 | 8.9 | 1.4 | .3 |
| Black | 1,419 | 96.5 | 3,5 | 3,2 | .3 | .0 |
| Hispanic[1] | 950 | 92.1 | 7.9 | 6.0 | 1.7 | .2 |

panics may be of any racial group.

: Details may not add to totals because of rounding.

CE: U.S. Department of Commerce, Bureau of the Census, Current Population Survey, unpublished tabulations.

the table should be mentioned first in the title; it is often the "dependent" variable. (Subject-matter components or categories will usually appear in the boxhead of the table.) The remaining parts of the title should identify the categories of the universe or population for which the subject-matter is shown: these are usually the "independent" variables whose components appear in the stub of the table. Finally, the title should specify the geographic and temporal loci of the data.

In Table 1, the subject-matter presented is persons enrolled in public and private elementary or secondary schools. This information is presented according to the metropolitan residence and racial or ethnic classification of the enrollees. By implication the data are for the entire United States, and they were collected with reference to October 1979. (Note that the data might have been collected later in 1979 or even in 1980, but the data collection procedure asked about enrollment in October 1979.)

The boxhead. The boxhead of a statistical table contains the column headings. Normally, these headings will delineate the main components or categories of the subject-matter presented. In Table 1, reading from left to right, the boxhead describes the classifications or breakdowns of the population--that is, the groups whose school enrollment is described in the table. It then identifies the column of data providing figures on total enrollment, followed by a percentage distribution according to two dimensions: whether the schools are public or private, and, if private, whether they are religiously affiliated or not. The last column presents the percentage of persons enrolled in private schools whose religious affiliation status was not reported.

The stub. The stub identifies the classifications or categories of the universe or population to which the given subject-matter relates. In this example, the population of school enrollees in metropolitan areas of the United States is classified according to two dimensions: by residence in central cities or outside of central cities and by racial or ethnic classification (white, black, or Hispanic).

The source note. This note informs the reader of the source or sources of the data presented. If proper procedure is followed, the source that is reported will be the original source rather than the place where the information was actually found. Even if the information is actually obtained from an almanac or statistical compendium, the source from which the almanac obtained the data should be given, not merely the name of the almanac or other secondary data source. In Table 1, the reader is informed that the data are from the Bureau of the Census of the U.S. Department of Commerce, that they were collected as a part of the Bureau's Current Population Survey, and that they are unpublished.

The data. Each number in the body of a statistical table is positioned at the intersection of two identifiers: the heading of the column in which it appears and the label to the left of the row in which it is placed. In Table 1, two kinds of numerical data are shown: absolute numbers and percentages. The cells of a table may of course contain other kinds of information as well, such as letter codes, brief labels, or even descriptive sentences or phrases. To return to our example, we learn (from Table 1, row 8, columns 1 and 3) that 10.9 percent of the 1,507,000 Hispanics enrolled in public or private elementary or secondary schools in the Nation's central cities were enrolled in private schools.

Additional notes. These notes may appear at the head of the table (headnotes), just below the title, or at the foot of the table (footnotes), just above the source note. In general, these notes provide warnings, qualifications, and reminders that assist the reader in understanding and interpreting the data presented. When these qualifiers can be described briefly, they may be included in the column headings or stub labels to which they pertain. In Table 1, the heading of column 1 informs the reader that the enrollment figures shown are rounded off to the nearest thousand persons. In addition, the spanner just below the column headings ("Percentage Distribution") indicates that the remaining data in the table are in the form of percentages. Percentage distributions may be arrayed vertically (in a column) or horizontally (along a row). The position of the total number on which the distribution is based (as in Table 1) or the total of 100 percent (not included in Table 1) informs the reader of its direction; if the total appears at the head of a column, or 100 percent appears at the foot of a column, the percentage distribution runs vertically; if the total appears on the left of a row or 100 percent appears at the left or right of a row, the distribution runs horizontally.

Our table contains two additional notes, each of which provides information needed to avoid confusion. First, the fact that the ethnic category "Hispanic" may include persons of any race means that the three racial/ethnic categories shown are not mutually exclusive and therefore are not additive. For example, the enrollment of a white Hispanic will be included in the table twice, once under whites and once under Hispanics. The numbers presented in this table contain a further source of possible confusion, since the totals include persons of other races not shown separately (e.g., Asians, American Indians). Thus, the totals are smaller than the sum of the three groups (because Hispanics are counted twice), but larger than the sum of whites and blacks alone.

The second note in Table 1 contains the familiar reminder that when totals and component figures are rounded independently, the sum of the rounded components may not be exactly equal to the rounded totals. For the same reason, percentages rounded to the nearest tenth of a percent do not always add exactly to 100. Furthermore, percentages and other derived statistics are sometimes calculated from the original, unrounded figures; these may therefore differ from the values that one would obtain from the rounded figures shown in the table.

One additional feature of our example is often a source of misunderstanding: subtotals may be included in a column or row of figures, together with their component elements. Adding the elements in the column or row without regard to the subtotals (i.e., treating subtotals as component elements) results in a total that is too high, often a multiple of the total shown. The clue to the presence of subtotals is (or should be) their indentation in the stub.

The most important function of statistical tables is to provide an array of information in as compact a form as possible. Because so much information is contained in a small space, it is essential to read a table in its entirety before trying to make sense of the details it contains.

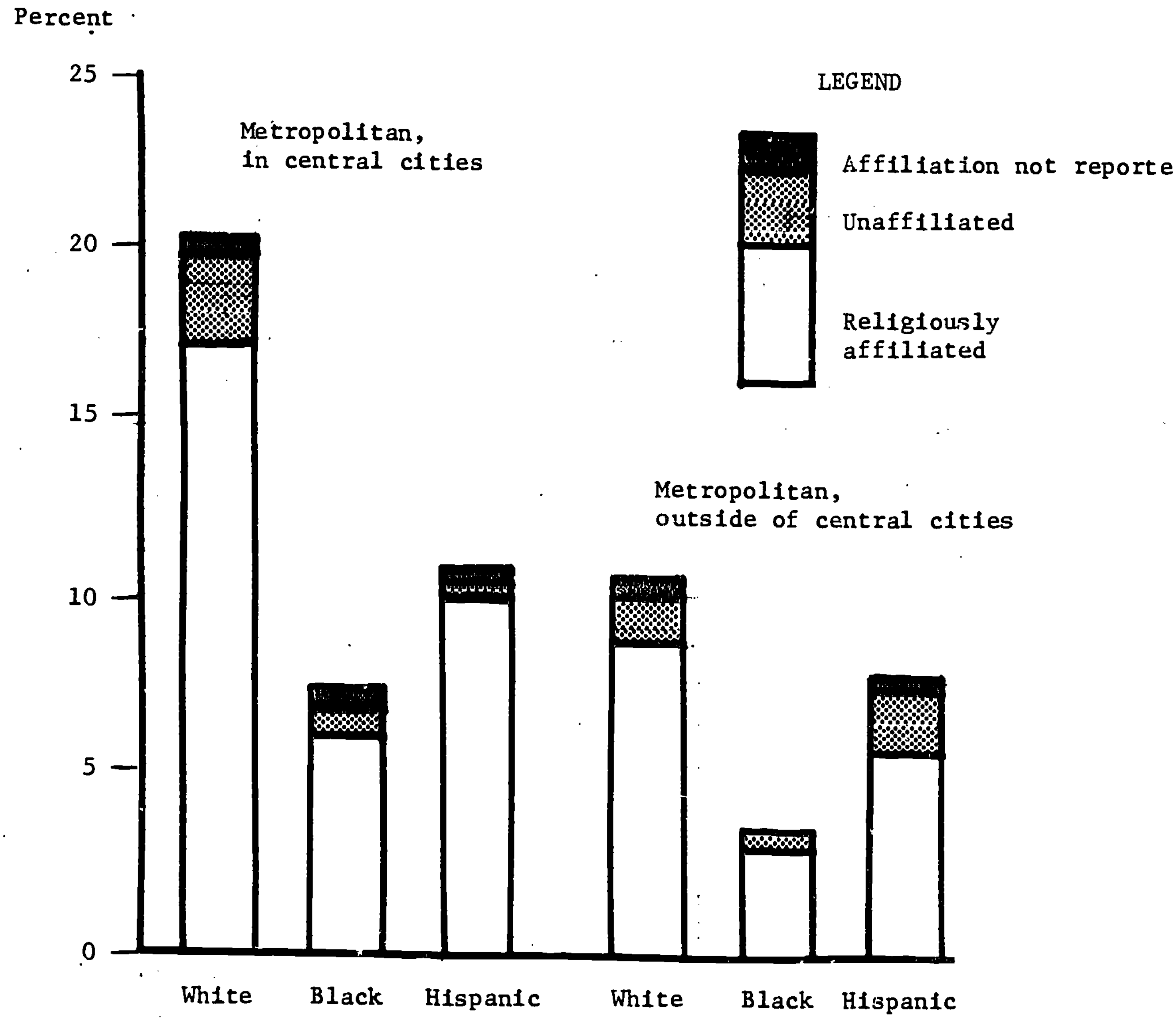
## Charts and Other Graphic Displays

In many cases, graphic displays of statistical data serve to complement data presented in tabular form. They call attention to the salient features of the data and thus facilitate comprehension of their overall significance. Graphic displays can also be designed to portray magnitudes and relationships that might otherwise be submerged among the detailed statistics in data tables. The most serious limitation of charts or graphs, as compared with statistical tables, is their inability to include more than a summary or a small subset of the data contained in a typical statistical table of the same size. But that limitation is also a source of their principal advantage over tables: with a well-designed chart, the reader can see the forest as a whole, rather than all the individual trees, or can compare a few individual trees while ignoring the rest. In the following example (Chart 1), drawn from data in the preceding table, the focus of attention is on the differences among the three racial/ethnic groups with respect to the percentages of students enrolled in private schools.

If a chart is presented by itself, that is, unaccompanied by a supporting statistical table, it should include the same components as a table (i.e., a complete title and source note, labels equivalent in function to the boxhead and stub, and additional notes when needed). The preferred practice, as done in The Condition of Education, is to accompany charts with the statistical tables from which the charts are derived. When the tables and charts are linked, many of the details provided in the former need not be duplicated in the latter. This educational indicators report has also adopted another convention that facilitates interpretation of the information presented in graphic displays: the data that are selected from the tables for graphic representation are printed in boldface type.

Chart 1

Private Elementary/Secondary School Enrollment
as a Percentage of Total Enrollment, by Metropolitan Status
and Racial/Ethnic Group: October 1979




NOTE: Details may not add to totals because of rounding.

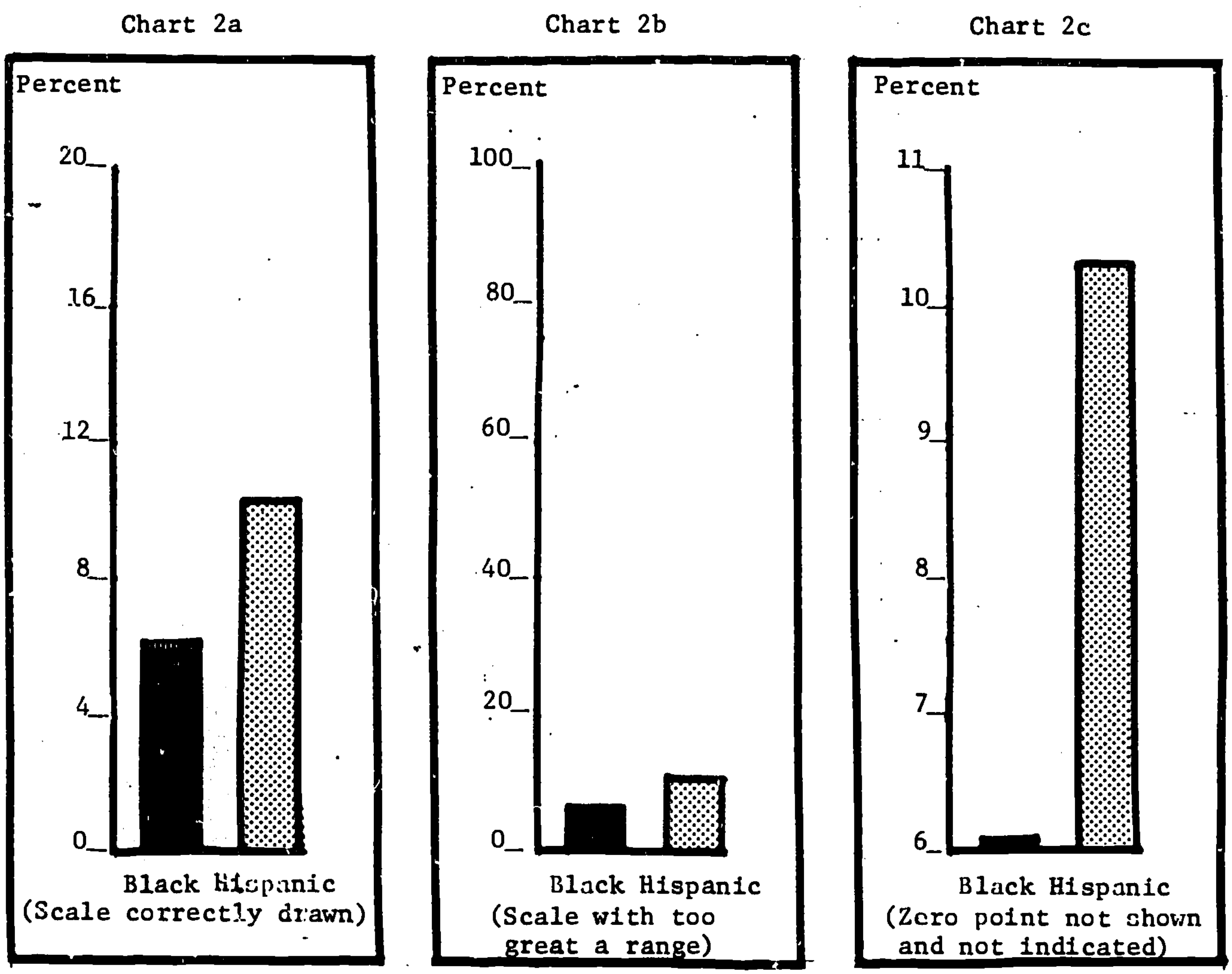
SOURCE: U.S. Department of Commerce, Bureau of the Census, Current Population Survey, unpublished tabulations.

Despite their structural similarities, graphs differ from tables in several ways. First, there are several types of charts and an almost unlimited variety of charting techniques and styles. The most common types are bar charts, circle or pie charts, and line charts. Bar charts (as shown in Chart 1) present the magnitude of some variable in terms of the height of vertical bars or the length of horizontal bars. Circle or pie charts indicate the proportions of some total that fall into various categories by the size of the "piece of pie" assigned to each category. Line charts display one or more series of values as connected points, usually denoting successive observations of some variable made at intervals over time. It is conventional to plot such line charts with time represented by regular intervals along the horizontal, or X-axis, and the values of the measured variable plotted along the vertical, or Y-axis. It is also conventional to connect the points representing individual observations with lines (hence the term "line charts"), despite the fact that nothing is known about the actual values during periods between observations. This practice helps the reader to visualize general trends, and that is the main purpose of graphic presentations.

Every mode of graphic presentation entails some risk of misrepresentation or misunderstanding. With bar charts, the most common problem results from the selection of different measurement scales to present the same data. Examinations of Table 1 and Chart 1 reveal a substantial difference between blacks and Hispanics in central cities with respect to the percent enrolled in religiously affiliated private schools--6.1 and 10.3 percent of total enrollment, respectively. Chart 2 illustrates how differently these findings can be made to appear by changing the measurement scale. Chart 2a is drawn using an appropriate scale of measurement, thus facilitating both the understanding of the enrollment status of each group and their comparison. Chart 2b, in contrast, is poorly designed, since the scale is too small (i.e., covers too wide a range of values) for the range of data presented; most of the space available for the presentation is wasted. Chart 2c is also improperly designed because the label along the Y-axis fails to remind the reader that the zero point is not shown. Scales that do not include the zero point should be avoided if possible. However, when all the values of the given variable fall within a narrow

Chart 2

Private, Religiously Affiliated Elementary/Secondary School Enrollment of Blacks and Hispanics in Central Cities, as a Percent of Total Enrollment: October 1979




NOTE: Details may not add to totals because of rounding.

SOURCE: U.S. Department of Commerce, Bureau of the Census, Current Population Survey, unpublished tabulations.

range, it becomes necessary to use a very large scale (one that covers a narrow range of values) to differentiate among small differences. In such cases, it may be impossible to show the true zero point on the scale. Instead, a break in the axis should be shown to alert the reader. It is also sometimes necessary to include bars on the same chart for comparative purposes, despite the fact that one or more of these bars have values that exceed the range of the scale. In such cases, a break should be shown on the bar in question to indicate that its height should not be compared visually to the heights of the other bars. In addition, the actual numerical value for the broken bar should also be shown.

Occasionally, it is useful to display rates of change in some variable in such a manner that a constant rate of change (e.g., 50 percent increase per year) appears as a straight line. This can be done by employing a logarithmic scale on the Y-axis. (A graph with a logarithmic scale for one of its two dimensions is called semilogarithmic.) With such a scale, constant rates of change are represented as constant slopes of the line on the graph (i.e., constant increases in height on the Y-axis per unit of time), despite the fact that the amounts of change vary with the change in the base (e.g., 50 percent increase from an initial value of 100 is much more than a 50 percent increase from 10). Scales of this type also permit data series with widely different magnitudes to be included on the same chart, thus facilitating their comparison.

The reader may encounter a large variety of more complex charts, such as three-dimensional representations and maps that show the geographic distribution of two variables simultaneously. Attempts are also made to dress up the charts with attractive symbols or other decorative embellishments. Whatever their esthetic qualities, such practices may require some extra effort on the part of the reader to properly interpret the findings. Although in graphs as in verbal communications the ratio of "noise" to "signal" should be kept to a minimum, it is up to the reader, finally, to find the message in any form of communication. In The Condition of Education, the reader's search for the message may begin with the charts; next, the statistical table on which a given chart is based should be consulted; third, the accompanying text may provide further insight; finally, the reader can consult the sources and additional references cited in the tables or listed in the appendix.

## Statistical Indicators and Other Measurements

All statistical data consist either of basic counts or derived measures. Basic counts are enumerations of any discrete entities--objects, events, characteristics, and so on--that can be observed and differentiated and are deemed worthy of concern. Derived measures are statistics obtained by performing some calculation or operation upon the basic counts. The diversity of derived measures is very large, ranging from some simple statistics, such as measures of central tendency (e.g., arithmetic means, medians), ratios, proportions, rates, and percentage distributions, to more complex calculations, such as measures of dispersion (e.g., the standard deviation and coefficient of variation), index numbers, standardization procedures, tests of significance, estimates of sampling variability, and the like.

In reading statistics of any kind, one is required to grasp two areas of meaning and their interrelation: the concept being measured and the reported magnitude or observed value. The concept refers to the phenomenon of interest and the operational procedure through which it is observed and measured. The magnitude refers to the value of a particular statistic and its significance in relation to comparable values. Note that neither the concept nor the magnitude can be understood in a statistical sense except in reference to the real world they purport to reflect. For example, it does not suffice to have a reasonable sense of how the "years of school completed" by members of a particular population group comes to be known and recorded. Equally important is the understanding that this statistic is being used as an indicator of educational attainment--a general concept that is both more interesting and much harder to gauge. Likewise, the fact that 52 percent of white males 25 to 29 years old in March 1979 had completed one or more years of college has little significance as an isolated datum. However, it acquires far more meaning when it can be related to corresponding data for other population groups--for example the 30, 32, and 44 college completion percentages, respectively, for the corresponding age groups of black males, black females, and white females. One reason for the common failure to understand the significance of both concepts and measurements is that their definitions are usually elliptical and thus do not distinguish clearly between the object of our interest (e.g., educational attainment) and the object that is actually counted or measured (e.g., years of school completed).

The reader of statistical data is likely to encounter a wide variety of statistical measures; new measures and measurement techniques are being developed continuously. It is therefore useful to recognize at least the basic types of measurement, their principal uses, and their limitations and possible misuses.

## Descriptive Comparisons

These are everyday statistics designed to weave a meaningful picture of conditions in a particular area of interest. Most common among these are ratios, proportions, percentages, and rates. A ratio expresses the relative magnitude of two quantities. In a basket containing 6 apples and 3 oranges, the ratio of apples to oranges is 6 to 3, or 2 to 1, or just 2.0. A proportion expresses the share of a particular item in relation to the total of which it is one component. In the above example, the proportion of fruit in the basket that are apples is 6 out of 9, or 2 out of 3, or .67. Percentages are a convenient means of expressing the distribution of elements or components in two or more populations so that their relative frequencies can be compared easily.

As demonstrated by the following data, if one wishes to compare the proportion of boys in two different schools, their absolute numbers must be expressed in percentage terms: the school with the larger absolute number of males (i.e., the larger school) actually has the lower proportion of males.

| | SCHOOL A | | | SCHOOL B | | |
|---|---|---|---|---|---|---|
| | Total | Male | Female | Total | Male | Female |
| Number . . . . . . . | 264 | 136 | 128 | 583 | 229 | 354 |
| Percent . . . . . . | 100.0 | 51.5 | 48.5 | 100.0 | 39.3 | 60.7 |

From the same data, one might describe the ratio of boys to girls in each school in terms of the conventional sex ratio: the number of males per 100 females: (136/128) x 100 = 106 in School A and (229/354) x 100 = 65 in School B.

Rates express the frequency of occurrence of some event or phenomenon during a specified time period. For example, the "dropout" rate in a particular class during a specified school year would be the number of dropouts during the year per 100 or per 1,000 of "the population at risk"--that is, persons enrolled in that class at the beginning of the year. Rates also measure the change occurring in some phenomenon over a specified time interval: if the enrollment in School B increased from 456 a year ago to 583 at present, the rate of increase is (583/456) x 100 - 100 = 27.8 percent per year.

The potential misuses or misinterpretations of these simple comparative measures can be suggested by citing the most common source of error: changing the base of the comparison makes the comparison invalid. For example, the percentage of college students receiving federal assistance can be computed using the total number of students enrolled in two- and four-year universities and colleges as a base. If this same base is used for successive years, comparisons may be made of the annual rates of change in the indicator. In contrast, if a different base was used in one of those years (e.g., the total number of male students in two- and four-year institutions), the value for that year could not be used in calculating annual rates of change. Because they are derived using different base values, the two indicators describe different percentages and, therefore, are not comparable. Whenever percentage distributions are shown, the base on which they are calculated should also be given, so that the absolute numbers represented by particular percentages can be estimated.

## Measures of Central Tendency

The nonstatistical reader should be familiar with at least three such measures: the mode, the mean, and the median. (The "average" value can mean any one of these three measures.) Consider the frequency distribution in Table 2. The mode is the simplest of these measures; it is defined as the interval containing the largest frequency of cases. In Table 2, there is no modal frequency, since each state is in a separate category. But if the number of students in each state was the basis for grouping the states, a modal frequency could readily be determined by inspection. For example, if the 50 states (plus the District of Columbia) were classified according to the following intervals (200,000 or more; 100,000 to 199,999; 50,000 to 99,999; 20,000 to 49,999; and under 20,000), the number of states in each category would be 3, 9, 12, 11, and 16, respectively. In this case, the modal frequency is 16 and the corresponding modal interval is "under 20,000."

The median can also be determined by inspection in Table 2, since the states are ranked in descending order according to the number of students enrolled in special programs. The median value is the one that divides a ranked distribution into two equal parts, such that half the cases fall above that value and half below it. Given 51 cases (50 states plus the District of Columbia), our median value corresponds to the 26th state (Colorado) and is therefore 44,274.

The mean (arithmetic mean) is readily calculated from the same table, since it is simply the sum of all the values, divided by the number of cases: 3,409,672 divided by 51 equals 66,856 as the mean value.

The properties of these measures indicate some of their limitations. Neither the mode nor the median can be estimated precisely if the interval in which they fall is broad. In addition, if different intervals have the

Table 2

Public Elementary/Secondary Students Participating in Special Education Programs, by State: Fall 1978

(Ranked by number of students participating)

| State | Number | State | Number |
|---|---|---|---|
| TOTAL | 3,409,672 | Colorado | 44,274 |
| California | 305,883 | Arizona | 39,091 |
| Texas | 262,214 | Washington | 37,845 |
| Illinois | 203,512 | Mississippi | 34,151 |
| Ohio | 170,888 | Arkansas | 34,064 |
| New York | 153,682 | Utah | 32,533 |
| Michigan | 147,901 | Iowa | 31,281 |
| Pennsylvania | 143,775 | Kansas | 31,226 |
| Florida | 127,121 | Oregon | 25,791 |
| Massachusetts | 118,851 | Nebraska | 21,440 |
| North Carolina | 103,332 | West Virginia | 19,707 |
| New Jersey | 102,761 | New Mexico | 19,380 |
| Tennessee | 102,182 | Maine | 17,885 |
| Missouri | 99,860 | Delaware | 13,990 |
| Georgia | 84,643 | Rhode Island | 13,682 |
| Maryland | 84,435 | Idaho | 13,520 |
| Indiana | 83,083 | Hawaii | 9,886 |
| Louisiana | 80,845 | Nevada | 9,836 |
| Wisconsin | 78,158 | Alaska | 8,884 |
| Virginia | 72,374 | Vermont | 8,276 |
| Minnesota | 71,488 | South Dakota | 7,869 |
| South Carolina | 68,218 | New Hampshire | 7,567 |
| Kentucky | 64,448 | Montana | 7,537 |
| Connecticut | 62,777 | Wyoming | 7,371 |
| Alabama | 62,226 | North Dakota | 6,592 |
| Oklahoma | 44,796 | District of Columbia | 6,541 |

SOURCE: U.S. Department of Education, Office for Civil Rights, State, Regional, and National Summaries of Data from the Fall 1978 Civil Rights Survey of Elementary and Secondary Schools, 1980. As presented in The Condition of Education, 1981 Edition, Table 6.1.

same frequency, the distribution does not contain any single mode. The median is also less capable than the mean of being manipulated algebraically. But the critical limitation these measures of central tendency share is that they all represent some distribution of values by means of a single value that may be more-or-less representative. Unwarranted inferences drawn from hasty examination of averages are legend. People can drown in lakes whose average depth is 18 inches, and families can suffer extreme deprivation in communities whose average family income is over $25,000 per year. The point is that one should always look at the range of values (i.e., their distribution) and not just at their central tendency.

## Measures of Dispersion

As noted above, measures of dispersion are important supplements to measures of central tendency, since they express the degree to which the different values of a distribution are widely dispersed or closely clustered about the average for that distribution. Certain measures of dispersion normally go with certain measures of central tendency because of the ways in which they are computed: the interquartile range corresponds to the median, and the standard deviation corresponds to the (arithmetic) mean.

The first and third quartiles of a distribution are calculated in the same manner as the median (which is the second quartile, or midpoint). The first quartile is the value of the case that divides a distribution into two unequal parts--three-fourths falling above it and one-fourth below it. The third quartile is the value of the case such that one-fourth of the distribution falls above it, and three-fourths below. The value of the third quartile minus that of the first quartile is the interquartile range.

To compute the standard deviation of a distribution, one must first compute the mean value for the distribution. Then the deviations of each of the values in the distribution from that mean value are calculated and squared. The sum of these squared deviations from the mean is then divided by the number of elements in the distribution. The square root of the resultant quotient is the standard deviation. The wider the dispersion of values around their mean, the greater the value of the interquartile range and standard deviation. As the mean is more sensitive to extreme or out-of-range values than the median, so the standard deviation is more sensitive to extreme values than is the interquartile range. In comparing sets of data, a measure of relative dispersion that is independent of the data measurement units is valuable. Just as percentage distributions allow comparisons of relative frequencies for different sets of data, the coefficient of variation (CV) allows comparisons of dispersion among different sets of data. The CV of a distribution is its standard deviation divided by its mean times 100 (to express the CV as a percentage). A CV of 10 percent means that the standard deviation is one-tenth the size of the mean.

A basic reason for being concerned with the distribution of data values is to determine how much confidence can be placed in a single average value as being representative of the distribution. If a class of students scores 78 on the average on an examination (i.e., the mean score is 78), the representativeness of the score 78 for that class is greater if the corresponding standard deviation is 2 than if it is 10. Assuming a reasonably normal, or symmetrical, distribution about the mean value of 78, a standard deviation of 2 implies that about two-thirds (68 percent) of the students in the class scored within 2 points of 78 (i.e., between 76 and 80). Where the standard deviation is 10, in contrast, about two-thirds of the class scores would be scattered between 68 and 88.

## Index Numbers

An index number is a computational device for measuring differences in the magnitude of a group of related variables. Percentage distributions are a common form of indexing, inasmuch as they express proportions of different total amounts as proportions of a common total of 100. The Consumer Price Index (CPI) is probably the best known index number. Problems of its technical composition aside, the CPI can be used to transform consumer income figures for different years from the current dollars in terms of which they are originally reported to constant dollars, thereby removing the effects of price changes or inflation from the figures compared. Tables or charts reporting dollar values for an indicator over time that have not been corrected for inflation (i.e., expressed in constant rather than current dollar values) will be misleading, as the worth of the dollar has decreased steadily over time.

Another form of indexing is the transformation of score values by means of a standardization procedure that expresses them in terms of a common group mean and standard deviation. An example in the 1981 edition of The Condition of Education (Table 2.27) is the expression of the average performance scores of a number of racial/ethnic groups of high school sophomores and seniors in terms of deviations from the average score and standard deviation of the total population tested.

The main purpose of all the adjustments and manipulations called for by indexing and standardization procedures is to permit more precise and legitimate comparisons among sets of data by restricting such comparisons to a single variable while controlling for the effects of other variables. For example, if we wish to compare the educational attainment of two communities, we might do so by calculating the average (median or mean) years of school completed by their adult populations. But if we know or suspect that younger adults are likely to have more formal education than older people, we might want to make our comparison more precise by standardizing both communities for possible differences in the age distribution of their adult populations. Without such standardization, neither the similarities nor the differences observed among different sets of data can be interpreted with confidence. For example, the proportion of elementary or secondary school enrollees who were enrolled in special programs for gifted and talented children in the fall of 1978 varied from a low of 0.4 percent (in New Hampshire, South Dakota, and the District of Columbia) to a high of 4.4 percent in Nebraska (as reported in The Condition of Education, 1981 edition, Table 2.12). It is difficult to interpret such findings without standardizing for the possible effects of such relevant characteristics as the resources devoted to gifted and talented programs in the different areas, the criteria employed in determining eligibility to participate, and so on.

## Indicators

In general, indicators are statistical measures that are recognized as reflecting significant aspects of some phenomenon of interest. Given the common need to ascertain the direction of development or change in such fundamental phenomena as enrollment rates, achievement scores, and the like, many indicators are in the form of time series of observations. In addition, indicators should be capable of aggregation and disaggregation along pertinent dimensions for analytic purposes. For example, school district data are often more useful for policy decisions if they are aggregated, or combined, and reported at the state level. Likewise, national estimates often better serve decisionmaking if they are disaggregated, or divided, and reported for smaller geographic units (e.g., regions, states) or population subgroups (e.g., whites, blacks, Hispanics). Statistical data pertaining to education, particularly the data included in The Condition of Education, provide many examples of the wide variety of estimates and measures that may serve as indicators of aspects of education.

Many indicators are conceptually simple and straightforward, such as trends in the percent of persons of a certain age group who are enrolled in school or trends in the years of school completed (educational attainment) of the adult population. Such data, collected over successive years and disaggregated by sex, race, ethnicity, place of residence, or other pertinent background characteristics, yield important information on differential educational opportunities in our society. Other indicators may involve more elaborate data manipulation. For example, the "retention" rate involves calculating the percentage of students enrolled in a specified grade in one year who are enrolled in the next higher grade in the following year. If one wishes to carry out this estimate without distortion due to persons who remain in school but failed to be promoted or who took up enrollment in another school, it is necessary to trace the same individuals from year to year. In general, measures of educational inputs (e.g., information on facilities, resources, personnel, and budgeting) are more susceptible to straightforward estimation than are indicators of the educational process (what goes on in the classroom) or of educational outcomes (e.g., what was learned, or how students later perform in a variety of situations). Of course, some "input" measures, such as those reflecting motivation of students, their health status, and other background characteristics, may also be difficult to measure.

More complex indicators of education include attempts to measure educational achievement by means of scores obtained on a variety of standardized tests. If a composite measure of educational achievement is desired, it is necessary to combine the scores on a number of separate tests in some manner. For example, standardized scores of white and Asian-American secondary school seniors, measured in the spring of 1980 and shown for eight subject-areas, make up the following data:

| Subject-area | Whites | Asian Americans |
|---|---|---|
| 1. Vocabulary, Part 1 | 51.4 | 50.2 |
| 2. Vocabulary, Part 2 | 51.3 | 50.5 |
| 3. Reading | 51.5 | 50.3 |
| 4. Math, Part 1 | 51.5 | 54.2 |
| 5. Math, Part 2 | 50.9 | 55.4 |
| 6. Mosaic Comparison, Part 1 | 50.9 | 52.4 |
| 7. Mosaic Comparison, Part 2 | 51.0 | 54.6 |
| 8. Three Dimensional Visualization | 51.0 | 55.2 |
| Average (with equal weights) | 51.2 | 52.8 |
| First weighted average[1] | 51.2 | 52.2 |
| Second weighted average[2] | 51.2 | 53.4 |

[1]Assigning double weight to items 1 to 3.
[2]Assigning double weight to items 4, 5, and 8.

Source: The Condition of Education, 1981 edition, Table 2.27

As these data show, alternative methods of combining the scores on the eight tests yield different results, at least for the Asian-American group. Its mean score of 52.8 is 1.6 points higher than that of the white group. But if double weight is given to the tests of reading and vocabulary, that advantage is reduced to 1.0 points. If, alternatively, double weight is given to mathematics and three-dimensional visualization, that advantage is increased to 2.2 points.

This simple illustration is designed to demonstrate both the procedures employed in constructing composite indicators and the judgments entailed along the way. Since education, like beauty and other virtues, is a many-splendored thing, it is susceptible to measurement in many ways. The careful reader should not allow the ingenuity or complexity of index construction to obscure the fact that no measure, however subtle and refined, can wholly reflect such complex phenomena as are observed in the educational system. In examining any indicator, however complex its construction, the reader should seek to understand what was actually measured, how that measure was made and combined or otherwise transformed, and what is the correspondence between the actual measure or indicator and the phenomenon it purports to reflect. Only then can the significance of the findings be ascertained.

## Sources of Error in Statistical Data

All statistics, however collected, are subject to some risk of inaccuracy or error. Some kinds of information are inherently difficult to obtain because of their complexity or because they relate to sensitive topics or issues. Other kinds of information are subject to certain biases or distortions because of selective perceptions, faulty memories, and so on. Errors in recording or coding information can occur at any time. Even those events whose reporting is mandatory or administratively routine, such as vital statistics or school enrollment figures, may not be fully or accurately recorded or reported. In addition, many long-standing data series, such as the administrative records for particular school districts, may be subject to serious distortions because of poorly documented changes over time in concepts, definitions, classifications, or data collection and recording procedures.

Statistical errors are of two kinds: sampling error (or sampling variability) and nonsampling error. Whenever a population is observed by selecting a sample of its component elements for analysis, the findings of that analysis are subject to sampling variability. Such variability can only be avoided by observing or measuring every element in the given population--as when a complete count or census is taken. However, all statistical data, whether obtained by sampling or by complete enumeration, are subject to some risk of nonsampling variability. Such variability stems from all the things that can go wrong in any data collection operation, excepting sampling variability. Some of these nonsampling errors are random in occurrence, inasmuch as they result from such factors as varying interpretation of questions by respondents, interviewers, coders, or other data processors. The effect of such errors is to reduce the precision of the survey estimates. Some nonsampling errors are also nonrandom in occurrence. Examples include nonresponses (because nonrespondents may tend to be different from respondents), incorrect responses, or undercoverage of certain segments of the population under study.

## Nonsampling Errors

As is shown in the following listing, important sources of nonsampling variability may be found at each stage of the data collection and analysis sequence.

Faulty sample design or execution. Failure to include certain segments of the population of interest in the sample design, inability to locate some of the respondents designated for inclusion in the sample, or the refusal of some respondents to answer the questions may all seriously bias the results obtained, since the nonrespondents may differ from the respondents in ways that significantly affect the findings.

Faulty questionnaire design. Poor question wording or improper question sequences may also bias the results by failing to elicit correct answers or by failing to allow for the full range of possible responses.

Poor communication. Respondents may misinterpret questions or related instructions because of educational handicaps or cultural or linguistic barriers to effective communication.

Deliberate falsification. Respondents may falsify their responses, particularly in sensitive areas where the questions asked may arouse suspicion, fear, or hostility or where they wish to conceal illegal or otherwise questionable activities or situations.

Faulty recollection. Respondents may not recall events or details, or they may experience distorted or biased recollection. Such distortions may sometimes be introduced by the respondent's subconscious desire to conform to the interviewer's apparent expectations.

Recording errors. Interviewers, coders, and other data processors may record the information they receive incorrectly.

Editing and imputation errors. Procedures designed to adjust the data for apparent inconsistencies, implausible entries, or to fill missing entries by some imputation procedure (which estimates what the response would have been from other characteristics of the nonrespondent) may themselves introduce additional bias or imprecision. Although the intention of these procedures is to reduce bias and increase precision, the procedures may be performed incorrectly (e.g., insufficient or inadequate data may be used as a basis upon which to impute missing values).

Analytic or interpretive errors. The analysis and interpretation of statistical data are highly complex operations that contain many risks of bias or error, ranging from simple errors in transcription or interpretation to subtle distortions in judgment introduced by the analyst's personal values or prejudices.

## Sampling Errors

In general, sampling variability can be estimated with far greater reliability and precision than can nonsampling error. Sampling variability is the discrepancy that can occur between the estimate obtained from a sample of the population under study and the "true" value that would have been obtained from a complete and error-free coverage of that population. Such variability is not due to faulty sample design or faulty survey methods. Any sample, however flawlessly selected, yields values that are likely to differ somewhat from those of a complete count. However, it is possible to estimate the likelihood that the given sample result falls within a specified range of the "true" population value, provided the sample in question is a random (or probability) sample. That condition is satisfied only if the sample was selected by means of a chance device (such as a table of random numbers) whereby each and every element in the population has a known chance of being included in the sample. The most common measure of sampling error is the standard error of the estimate (SE). The chance or probability that a given sample estimate falls within one SE of the corresponding "true" population value is about two out of three (68 percent). The chance that such a sample value falls within two SE's of the "true" value is about 19 out of 20, or 95 percent.

An alternative measure of sampling variability is the relative standard error, or coefficient of variation (CV), which was introduced earlier. The CV in this case is the standard error of a sample estimate expressed as a percent of the estimate itself. The CV can be used in the same manner as the SE to estimate the likelihood that the sample value falls within a specified range of the "true" population value. For example, if the CV is estimated as 2.4 percent, the chances are about two out of three that the sample estimate falls within 2.4 percent of the "true" value and about 19 out of 20 that it falls within 4.8 percent.

Table 3 and Chart 3 are designed to illustrate the effects of sampling variability on actual data obtained from the Current Population Survey conducted monthly by the Bureau of the Census. The data shown in the illustration were collected in March 1979. The basic counts shown in

Table 3

Persons 25 to 64 Years Old with Four Years or More of College Education and Employed in Professional, Technical or Kindred Occupations, by Sex, Race, and Hispanic Origin: March 1979

(Numbers in thousands)

| Sex, race and Hispanic origin | (1) Total employed in PT&K | (2) College graduates employed in PT&K | (3) (2) as a percent of (1)[2] | (4) Approximate Standard Error | (5) |
|---|---|---|---|---|---|
| | | | | For number of college graduates in PT&K | For percent of employed who are college graduates |
| MALE | | | | | |
| White | 7,001 | 5,015 | 71.6 | 52 | 0.8 |
| Black | 338 | 180 | 53.2 | 13 | 4.8 |
| Hispanic origin[1] | 169 | 103 | 60.9 | 10 | 6.4 |
| FEMALE | | | | | |
| White | 4,830 | 3,068 | 63.5 | 48 | 1.0 |
| Black | 483 | 286 | 59.2 | 17 | 3.5 |
| Hispanic origin[1] | 102 | 51 | 50.0 | 8 | 8.4 |

[1]May be of any race.

[2]Calculated from rounded numbers shown in table.

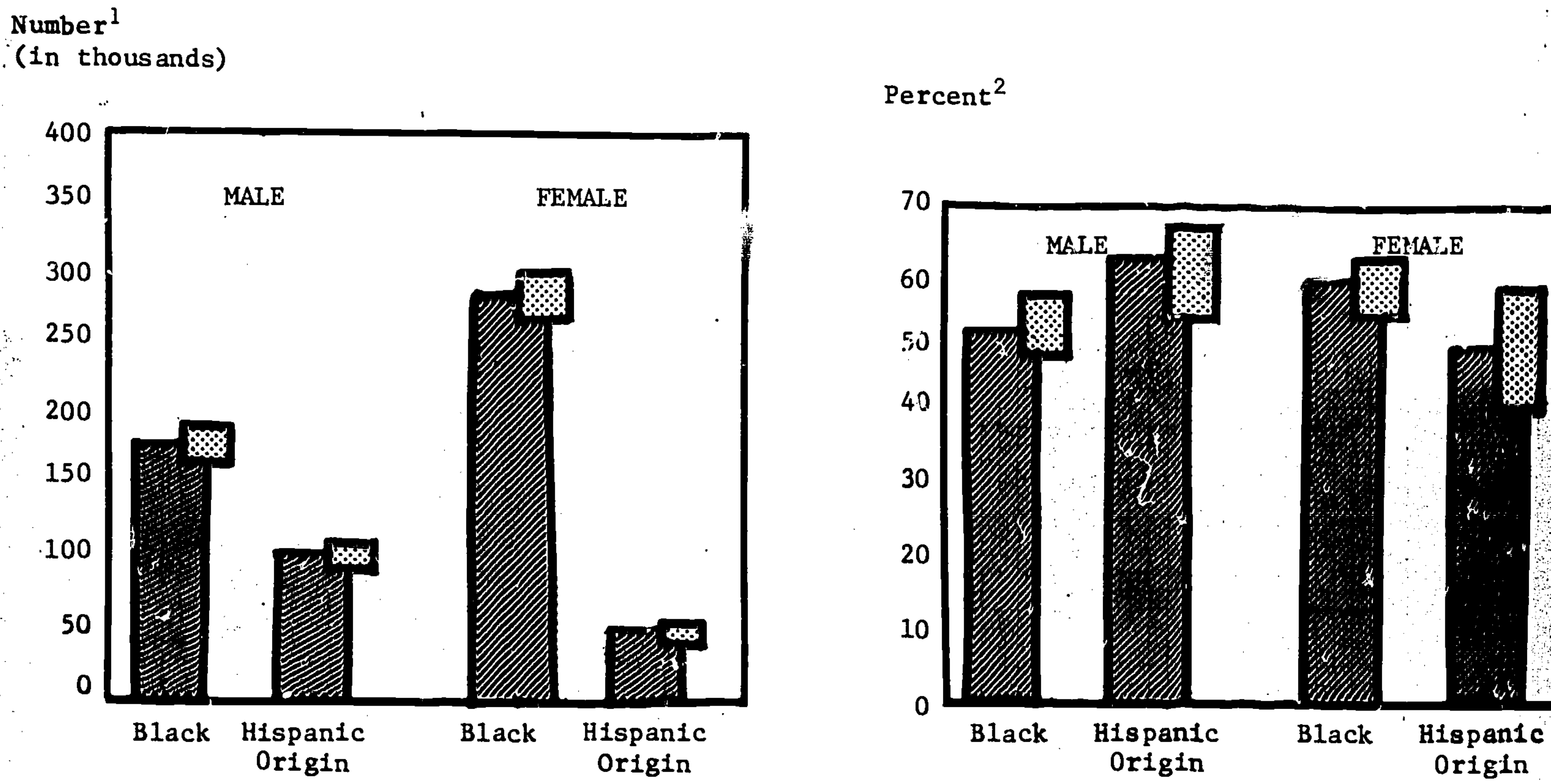
NOTE: Underscored values are shown on the accompanying charts.

SOURCE: U.S. Department of Commerce, Bureau of the Census, Current Population Reports, Series P-20, No. 356, "Educational Attainment in the United States: March 1979 and 1978," Tables 5 and B-1 to B-4.

31

Chart 3

Blacks and Persons of Hispanic Origin, 25 to 64 Years Old, with Four Years or More of College, Employed in Professional, Technical, or Kindred Occupations, by Sex: March 1979




[1]Number of persons in the specified category who were employed in the specified occupation group in March 1979.

[2]Percent of college graduates among persons in the specified group who were employed in the specified occupation group in March 1979.

NOTE: Vertical bars denote reported values; shaded rectangles denote the range of values obtained by adding and subtracting one standard error.

the table are the number of whites, blacks, and persons of Hispanic origin (by sex) who were employed in professional, technical, or kindred occupations in March 1979 (see column 1). In the next column is shown the number of persons so employed who had completed four or more years of college. These numbers are also expressed as a percentage of the former (column 3). The last two columns of the table show the approximate standard errors for both the number and the percentage of college graduates in that occupation group. These SE's were estimated from appendix tables provided in the source report identified in the table.

The corresponding chart presents both the number and the percent of college graduates in this occupation group among blacks and persons of Hispanic origin, by sex. These sample estimates are represented by vertical bars. The shaded rectangles partially superimposed on the bars represent the range of values obtained by adding plus or minus one standard error to the sample estimates. The chances are about two out of three that the "true" population value (i.e., the value that would be obtained by averaging the results of all possible samples from the given population) falls within this range. For example, in March 1979, 180,000 black male college graduates were reported as employed in this occupation group. That figure is our sample estimate. Given a standard error of 13,000, we have two chances out of three (a 68 percent probability) that the actual number so employed was between 167,000 and 193,000, or 180,000 plus or minus 13,000. By the same token, there is a 95 percent probability that the "true" number falls between 154,000 and 206,000, or 180,000 plus or minus 26,000.

An important use of the standard error (SE) is to help us decide whether or not two sample values are "really" different from one another--that is, whether or not the difference between them is statistically significant. For example, 60.9 percent of male Hispanics and only 53.2 percent of male blacks employed in professional, technical, and kindred occupations are college graduates (as reported in Table 3). However, the SE's for those values indicate a range that overlaps. Hence, there is some chance that the "true" percentage among black males may be as high as that of Hispanic males. The prudent reader should, therefore, interpret these differences with extreme caution.

It is impracticable to estimate and display sampling variability for the diverse sets of data included in The Condition of Education, since the factors that determine this variability are unique to each statistic. Furthermore, as noted, even when estimates of sampling variability are available, the confounding effects of nonsampling variability are seldom known. In this situation, the reader should bear in mind that our weakest and least reliable data sets are often those whose error structures (both sampling and nonsampling) are virtually unknown and inestimable. This sometimes produces the paradoxical result that the best available data come to be viewed with the greatest suspicion precisely because so much is known and stated about their associated errors. In short, the reader must view the preceding discussion as merely illustrative of the band of uncertainty that surrounds all statistical data, particularly data whose associated uncertainty cannot be estimated. In other words, all statistical data must be viewed with caution; the precision of reported quantities may be more apparent than real; and even the best of statistics can only approximate the reality they seek to reflect.

Given all these disturbing caveats, why bother with statistics at all? Two reasons may be offered. First, in a world of uncertainty, properly understood statistics convey a fuller sense of reality precisely because their approximate nature is clearly expressed. Second, in a world of complexity, our statistics, for all their limitations, provide far more reliable guides to effective decisions than the available alternatives--at least in the long run.

## Some Notes on Data Interpretation

It is arguable that statistics, particularly the common descriptive variety, are more valuable for the questions they raise than for the questions they answer. The facts they represent are of course often useful in their own right; it may be important to know that school retention rates have been rising during the past four or five decades, for example. But practical concern for such trends can seldom rest with their mere description. The inevitable questions that are prompted by such findings are questions stemming from the practical concerns of policymakers and their advisors: What accounts for the observed phenomenon? Is it desirable or otherwise? In either case, what can be done to enhance or inhibit the given trend or development? If few of these questions can be dealt with solely on the basis of available statistics, none can be dealt with without taking statistical realities into account. Hence, the following observations are intended to help the reader to avoid some of the more common pitfalls in statistical inference.

## Causal Inferences

Descriptive statistics, particularly when they reveal significant or surprising conditions, changes, or differences between population groups or other entities, invite the reader to formulate explanations of their underlying causes. All such inferences, however plausible, must be regarded as purely speculative unless supported by additional information. The inference of causation where only statistical associations have been observed is a classical error in trying to interpret statistical findings.

## Aggregation Fallacy

The aggregation fallacy results from the tendency to attribute properties or characteristics observed among large areas or population groups to their individual components. For example, in March 1979, 67.7 percent of the adult population of the United States aged 25 years and over was reported to have completed four years of high school or more. The aggregation fallacy would prompt the false conclusion that 67.7 percent of the corresponding population in particular states or localities must also be high school graduates. In fact, aggregate statistics are a form of average; one cannot assume that every value of a distribution is equal to its average value.

## Significance

The problem of statistical significance is concerned with the question of deciding whether a given observation is sufficiently different from the value that might have occurred by chance that it reflects a real change or difference. This question arises in many forms, and the corresponding range of techniques in the field of statistical inference is quite broad. But the reader's concern with significance normally extends beyond significance in a purely statistical sense. It encompasses the everyday notion of significance as denoting importance. The question of importance, in turn, involves further questions of purpose and values--important for what or to whom. Tests of statistical significance cannot tell us whether given findings are important in the latter sense. Significance is also sometimes confused with saliency. However, the fact that a particular finding is outstanding or unexpected does not necessarily imply that it represents a real departure from normal levels or trends in the phenomenon under study. Aberrant values, or outliers, may reflect nothing more than random variations.

## Time Series

Our most reliable source of information on emerging trends that may presage future developments consists of series of observations or measurements taken at intervals through time. The analysis of time series has also given rise to a vast body of technical literature aimed at improving our ability to anticipate future trends. But the success of even the most sophisticated techniques in predicting the future depends heavily upon the stability of the trends on which the predictions are based. In most areas, changes in policy, unanticipated events, or underlying changes in the structure of relationships among critical variables can result in drastic changes in the direction of time series. In principle, many of these "disturbances" might have been anticipated if their underlying dynamics had also been subject to analysis in terms of appropriate time series. In the real world, however, the supply of pertinent information is limited.

In examining any time series of data, it is important to consider what period of time is relevant to the problem under consideration. If one's concern is with current trends in school enrollment, it is seldom necessary to examine trends since 1900 or even 1950. Because changes in annual enrollments occur gradually over time, it suffices for the study of current trends to look at data for only the past five or, at most, ten years. Extending this series back for more than ten years from the present will add little information that is pertinent to current trends and is likely to obscure our appraisal of current conditions. A more subtle question in this regard is that of data obsolescence. Even if a long time series has managed to preserve the same definitions of concepts and classifications and the same data collection procedures, the signficance (in terms of meaning or importance) of the phenomenon under observation may have undergone profound changes. This has been the case with both the high school drop-out rate and the proportion of high school graduates going on to college. In the past, dropping out of high school was the norm, while today, high school dropouts and those graduates who choose not to enroll in college are at a severe disadvantage in finding employment. A further problem that frequently confronts the time-series analyst is the treatment of aberrant values, or outliers. Similar to the cases of the mean and the standard deviation, many of the curve-fitting techniques used to extrapolate time series are strongly influenced by a few extreme values. One alternative is to "correct for" or omit outliers. But if these extreme values are adjusted or omitted, the resultant trend line may fail to reflect significant variation. Finally, it must be reiterated that most lengthy time series of data are plagued by poorly documented changes in definitions, classifications, coverage, and the like. Anyone who has attempted to reconstruct an historical series by working backward from the most recent set of information has soon discovered how quickly the wealth of information presently available diminishes to a slender thread. In short, most attempts to reconstruct the past, like those that seek to predict the future, entail heroic assumptions.

## Subjective Phenomena

A small but important portion of the information presented in The Condition of Education relates to public reactions to, feelings about, or attitudes toward certain features of our educational system. Such information, however obtained, is termed subjective because the views expressed denote inner, personal states or conditions that cannot be directly confirmed by independent observation. In contrast, objective information is susceptible, at least in principle, to independent verification by direct observation, examination of records, or other means. Hence, the interpretation of subjective data poses special difficulties. Further problems arise because of the general wording of many public opinion questions. For example, the proportion of parents with children in public schools who rated the public schools in their community with a grade of "C" or lower rose from 31 to 47 percent between 1974 and 1980 (The Condition of Education, 1981 edition, Table 1.15). Accepting these findings at face value, one can certainly conclude that general parental satisfaction with public schools declined markedly during this period. But the interpretation of these findings immediately gives rise to such obvious questions as whether the schools have in fact gotten worse or whether instead parental standards and expectations have risen; whether new demands have forced the schools to assume functions for which they were not designed; whether the decline in satisfaction extends to other public services in the community; and so on.

Despite these difficulties of interpretation, subjective conditions are vitally important in many areas of public policy inasmuch as the outcome of such policies is often affected by the way both the policies and the conditions they address are perceived. But if prudence dictates a continuing concern with subjective reactions and attempts to gauge them, prudence also suggests caution in the interpretation of opinion data.

As a final word, it must be emphasized that statistics, whether viewed as techniques of data analysis or as data sets, constitute tools that can greatly enhance our powers of observation and understanding, if properly employed. But as tools, they cannot replace judgment and they cannot dictate human decisions. Above all, their true significance can only be determined in terms of our values and aspirations.