

Article from SIRS Discoverer Database; (ProQuest)

Lexile: 1210L

Encyclopedia Britannica
Copyright © 2006 Encyclopedia Britannica, Inc.

statistics

Anyone who listens to the radio, watches television, and reads books, newspapers, and magazines cannot help but be aware of statistics, which is the science of collecting, analyzing, presenting, and interpreting data. Statistics appear in the claims of advertisers, in predictions of election results and opinion polls, in cost-of-living indexes, and in reports of business trends and cycles. Every science depends to some extent upon the gathering of data and the interpreting of the data by statistical methods. On the basis of statistics, important decisions are made in the fields of government, industry, and education. Even the average person bases many personal decisions on information that has been supplied by statisticians.

The term statistics is used in two different ways. When used in the plural, it refers to numerical data. For example, one would say, 'Statistics show highway accidents to be caused by' When used in the singular, it means statistical methods, as in 'Statistics is the body of principles and methods that has been developed for collecting, analyzing, presenting, and interpreting large masses of numerical data.' Without statistical treatment of data, there would be no way to put the 'facts' together to see what they mean.

The results of statistical investigations may sometimes be stated in a single sentence, as in weather forecasting. Often, however, they are presented in the form of numerical tables, as in census reports, or are shown pictorially in the form of graphs or charts (see [graph and chart](#)).

How Statistical Data Are Collected

Statistical data are usually collected in one of the following ways: (1) by consulting existing source material, such as periodicals and newspapers, or reports from industries, government agencies, and research bureaus; (2) by setting up a survey and collecting data at first hand from individuals or organizations; and (3) by conducting scientific experiments and measuring or counting under controlled conditions.

Basic information must be collected in such a way that it is accurate, representative, and as comprehensive as possible. Statistical treatment cannot in any way improve the basic validity or accuracy of the raw data. Methods of collecting data are therefore basic to the whole field of statistics.

Populations and Samples

The term population ordinarily means the whole number of people in a city, county, state, or country. The statistician speaks of a population of automobiles, salaries, accidents, ballots, blood pressures, or any other characteristic. For the statistician a population-also called a universe-means the entire group of items in the class being considered.

Usually it is not possible to gather observations from all the possible cases in a population. Some populations are infinite. For example, if you should want to count how many times 'heads' will turn up when a penny is tossed, you would have to set a limit to the number of throws because even a billion would not exhaust the infinite universe of possible trials. (See also [probability](#).)

Other populations, though finite, are so large that it would take too long or cost too much to collect data on each unit in them. Every 10 years the United States government conducts a census of the entire population of the United States, but this is a gigantic and costly undertaking (see [census](#)). The statistician usually gets information from a relatively small number of cases, called a sample. From the measurements or observations of the individuals sampled, generalizations can be made about the population from which the sample was selected.

The individuals in a sample must be representative of the larger population; otherwise the conclusions drawn from the sample would not be valid for the larger population. For example, one cannot draw valid conclusions about the probable outcome of a national election from interviewing a sample of 10,000 registered voters unless they are representative of the people who will actually vote in the election.

The size of the sample is a factor also. Other things being equal, a larger sample is better than a smaller one. However, excellent results can be obtained with small samples that are properly set up.

Most public opinion surveys are conducted on samples that are made as representative as possible by means of stratified sampling techniques. For example, a national polling sample can be set up by first dividing the whole country into various geographic areas and then dividing each area into strata, or layers, according to the degree of urbanization. (See also [public opinion poll](#).)

The interviewers who go out to talk to people are assigned certain areas. Each interviewer is instructed to interview a specified number of people in certain categories, such as different socioeconomic levels and different age groups. This is quota sampling. Its purpose is to make the proportions in the sample of persons interviewed the same as they are in the general population. Quota sampling results in remarkably accurate forecasts even though the actual samples are small as compared with the total population.

Basic to all sampling techniques is the idea of randomness, which means that samples are drawn in such a way as to insure that any unit in the population is equally likely to be included. This is essentially what one does in shuffling a deck of cards before dealing hands in a card game. The dealer is simply insuring that any card or combination of cards is equally likely to end up in one hand as it is in another. Interestingly enough, much of the theory of sampling was developed from so-called games of chance. Random sampling is used a great deal in scientific experiments. Usually the cases to be included are selected by means of some device such as a table of random numbers or a computer.

Many sampling procedures are loaded in some way that may influence the results. For example, a sample of persons who volunteer for a study may or may not be typical of the whole group. It is sometimes found that people who send back questionnaires immediately give different replies from those who send them back after several follow-up letters. Opinions expressed in letters to the editor on current issues would also be an example of nonrandom and nonrepresentative samples.

Two Kinds of Data

The types of data that can be handled statistically occur in one of two forms. One kind is obtained by counting. The other is obtained from measurements.

Discrete Data

Numbers obtained by counting a small group—such as the members of a family—are exact. If a family has three members and another child is born, the number leaps from three to four without passing through any intermediate stages, such as 3.12. A succession of such numbers is called a discrete, or discontinuous, series.

Suppose a market-research analyst wants to study consumer preferences for certain types of breakfast foods, popcorn, or soft drinks. The researcher might ask a number of people to state their brand preferences, or determine what brands of soda pop they drink or what kinds of breakfast foods or popcorn they have on their kitchen shelves. The categories will be expressed by a word or a phrase, and the data will be an exact enumeration of the number of cases in each category. When graphed, discrete data are usually represented by bars separated from one another to suggest the discontinuity between the categories (see [graph and chart](#)).

Continuous Data

The second type of data consists of measurements that fall along a continuous scale, such as distance in feet or meters, weight in pounds or kilograms, temperature in degrees, time in minutes, hours, or years, and grades made on examinations. The measurements are usually obtained by using some sort of measuring instrument, such as a ruler, a scale, a thermometer, or a school test. This type of data is called continuous, because all gradations are possible between the lowest and highest in the series. The classes are expressed numerically rather than with a word or phrase and are not definite and distinct from one another.

Some types of data can be expressed as either discrete or continuous. For example, measurements of intelligence are frequently expressed in terms of 'average,' 'above average,' and 'below average' instead of in numerical scores, which fall on a continuous scale. Psychological test scores are often expressed in terms of age norms or grade norms. A time series also may be expressed as a discrete series, as in the United States Census, which is taken at 10-year intervals. Actually, of course, the population changes are gradual.

The Reliability and Validity of Measurements

It is possible to be 100 percent accurate when counting a small group. Even in counting, however, errors occur if the numbers handled are large. Measurements, on the other hand, are never 100 percent accurate; there is always some residual error that cannot be eliminated. Generally, however, they are sufficiently accurate and dependable if the measuring instrument itself is accurate—that is, has been calibrated properly—and the person using the instrument is a skilled observer and a careful recorder.

The reliability of a measuring instrument or of a test refers to how consistently it measures similar kinds of things, such as events or sizes of groups of people. A reliable watch will show the same time from one day to the next day (that is, every 24 hours) even though it might always be a little fast or a little slow as compared with the correct time. An accurate watch, on the other hand, gives correct time (within certain limits) according to some standard, such as Coordinated Universal Time. Likewise, an accurate thermometer is one that shows the correct temperature (within very close limits) so that one does not have to add or subtract a certain number of degrees to arrive at a correct reading.

The validity of a measuring instrument refers to whether or not it measures what it is supposed to measure. A thermometer is a valid instrument for measuring temperature, but one does not use it to measure humidity, wind velocity, time, or distance.

For public opinion polling on political issues, the measuring device has to be built. The measuring device that is used most often is a questionnaire. Generally it consists of a series of printed questions with spaces for writing in the answers to the questions that the interviewer asks. Poorly written questions that are difficult to understand lead to worthless information. In order to guard against this, a great deal of time is spent in perfecting the questionnaire before the survey is begun. The interviewers are carefully selected and are trained to secure and record the information, because they too could be a source of error. Often the interviewer conducts a small pilot study to test the questionnaire so that sources of bias and ambiguity can be removed prior to the questionnaire's use in the much larger, final survey.

Frequency Distribution Tables

The statistician works with large masses of data. Before any conclusions can be drawn from such data, it must be condensed and arranged in usable form. Almost all tabulations that one sees are grouped in one way or another. One way to summarize and describe a mass of statistical data is to make a frequency distribution table.

Table I lists average grade-level reading scores made by a group of 88 sixth-grade students. The scores have not been arranged in any order. It is extremely difficult to draw any conclusions on the basis of these figures except to say that relatively few of the sixth graders are at a sixth-grade level (scores 6.0 to 6.9) in their reading ability. It would be impossible, using this haphazard arrangement, to answer readily any of the following questions:

What is the **range** in reading ability among these students, from highest to lowest? How well do they read as a group? What is the average grade-level score? Do the scores seem to cluster at one or two grade levels, or are they scattered widely? What proportion is below grade level? What proportion is above grade level? What **range** of scores includes the middle half? How would a pupil with a grade-level score of, say, 6.5 compare with the rest of the group?

Table I.

Table I.

Table II.

Table II.

In order to make such data usable, the statistician ordinarily groups them into classes. This has been done in Table II. All the possible scores, from the highest to the lowest, are written in the stub—the vertical column at the left (Grade Score).

In the next column are tabulated the number of times each score occurs. Technically, this number is called the frequency. (In statistical work, the letter *f* means frequency.) The tabulating was done by taking each score shown in

and placing a tally mark (/) opposite that score value in

Table I.

Table I.

. The tally marks were then changed to numbers. Notice that some scores did not occur at all and others occurred more than once.

Grade Score	f	Grade Score	f	Grade Score	f
11.2	1	8.8	1	6.4	1
11.1	1	8.6	1	6.2	1
11.0	1	8.4	1	6.0	1
10.9	1	8.2	1	5.8	1
10.8	1	8.0	1	5.6	1
10.7	1	7.8	1	5.4	1
10.6	1	7.6	1	5.2	1
10.5	1	7.4	1	5.0	1
10.4	1	7.2	1	4.8	1
10.3	1	7.0	1	4.6	1
10.2	1	6.8	1	4.4	1
10.1	1	6.6	1	4.2	1
10.0	1	6.4	1	4.0	1
9.9	1	6.2	1	3.8	1
9.8	1	6.0	1	3.6	1
9.7	1	5.8	1		
9.6	1	5.6	1		
9.5	1	5.4	1		
9.4	1	5.2	1		
9.3	1	5.0	1		
9.2	1	4.8	1		
9.1	1	4.6	1		
9.0	1	4.4	1		
8.9	1	4.2	1		
8.8	1	4.0	1		
8.7	1	3.8	1		
8.6	1	3.6	1		
8.5	1				
8.4	1				
8.3	1				
8.2	1				
8.1	1				

Table II.

shows that the reading ability of these sixth-grade pupils varies from 3.6 (third-grade level) to 11.2. The **range** between 3.6 and 11.2 is wide-7.6 grade levels.

Grade Score	f	Grade Score	f	Grade Score	f
11.2	1	8.8	1	6.4	1
11.1	1	8.6	1	6.2	1
11.0	1	8.4	1	6.0	1
10.9	1	8.2	1	5.8	1
10.8	1	8.0	1	5.6	1
10.7	1	7.8	1	5.4	1
10.6	1	7.6	1	5.2	1
10.5	1	7.4	1	5.0	1
10.4	1	7.2	1	4.8	1
10.3	1	7.0	1	4.6	1
10.2	1	6.8	1	4.4	1
10.1	1	6.6	1	4.2	1
10.0	1	6.4	1	4.0	1
9.9	1	6.2	1	3.8	1
9.8	1	6.0	1	3.6	1
9.7	1	5.8	1		
9.6	1	5.6	1		
9.5	1	5.4	1		
9.4	1	5.2	1		
9.3	1	5.0	1		
9.2	1	4.8	1		
9.1	1	4.6	1		
9.0	1	4.4	1		
8.9	1	4.2	1		
8.8	1	4.0	1		
8.7	1	3.8	1		
8.6	1	3.6	1		
8.5	1				
8.4	1				
8.3	1				
8.2	1				
8.1	1				

Table II.

Table II.

Grade Score	f	Grade Score	f	Grade Score	f
11.2-11.0	3	8.8-8.6	2	6.4-6.2	2
11.0-10.8	1	8.6-8.4	1	6.2-6.0	1
10.8-10.6	1	8.4-8.2	1	6.0-5.8	1
10.6-10.4	1	8.2-8.0	1	5.8-5.6	1
10.4-10.2	1	8.0-7.8	1	5.6-5.4	1
10.2-10.0	1	7.8-7.6	1	5.4-5.2	1
10.0-9.8	1	7.6-7.4	1	5.2-5.0	1
9.8-9.6	1	7.4-7.2	1	5.0-4.8	1
9.6-9.4	1	7.2-7.0	1	4.8-4.6	1
9.4-9.2	1	7.0-6.8	1	4.6-4.4	1
9.2-9.0	1	6.8-6.6	1	4.4-4.2	1
9.0-8.8	1	6.6-6.4	1	4.2-4.0	1
8.8-8.6	1	6.4-6.2	1	4.0-3.8	1
8.6-8.4	1	6.2-6.0	1	3.8-3.6	1
8.4-8.2	1	6.0-5.8	1		
8.2-8.0	1	5.8-5.6	1		
8.0-7.8	1	5.6-5.4	1		
7.8-7.6	1	5.4-5.2	1		
7.6-7.4	1	5.2-5.0	1		
7.4-7.2	1	5.0-4.8	1		
7.2-7.0	1	4.8-4.6	1		
7.0-6.8	1	4.6-4.4	1		
6.8-6.6	1	4.4-4.2	1		
6.6-6.4	1	4.2-4.0	1		
6.4-6.2	1	4.0-3.8	1		
6.2-6.0	1	3.8-3.6	1		
6.0-5.8	1	3.6-3.4	1		
5.8-5.6	1	3.4-3.2	1		
5.6-5.4	1	3.2-3.0	1		
5.4-5.2	1	3.0-2.8	1		
5.2-5.0	1	2.8-2.6	1		
5.0-4.8	1	2.6-2.4	1		
4.8-4.6	1	2.4-2.2	1		
4.6-4.4	1	2.2-2.0	1		
4.4-4.2	1	2.0-1.8	1		
4.2-4.0	1	1.8-1.6	1		
4.0-3.8	1	1.6-1.4	1		
3.8-3.6	1	1.4-1.2	1		
3.6-3.4	1	1.2-1.0	1		
3.4-3.2	1	1.0-0.8	1		
3.2-3.0	1	0.8-0.6	1		
3.0-2.8	1	0.6-0.4	1		
2.8-2.6	1	0.4-0.2	1		
2.6-2.4	1	0.2-0.0	1		
2.4-2.2	1	0.0-0.0	1		
2.2-2.0	1	0.0-0.0	1		
2.0-1.8	1	0.0-0.0	1		
1.8-1.6	1	0.0-0.0	1		
1.6-1.4	1	0.0-0.0	1		
1.4-1.2	1	0.0-0.0	1		
1.2-1.0	1	0.0-0.0	1		
1.0-0.8	1	0.0-0.0	1		
0.8-0.6	1	0.0-0.0	1		
0.6-0.4	1	0.0-0.0	1		
0.4-0.2	1	0.0-0.0	1		
0.2-0.0	1	0.0-0.0	1		
0.0-0.0	1	0.0-0.0	1		

Table III.

Table III.

Table III is further condensed by grouping the classes to the nearest grade level. Thus all scores representing sixth-grade reading ability-from 5.5 through 6.4-are put together. These two figures are called the class limits, and the distance between them is called the class interval. Notice that all the class intervals are the same size.

One should remember that when a grouped frequency distribution is used, all information about specific individuals is lost. The unit classification in

is more precise, but the class interval is usually preferred because it shows more clearly the overall pattern of the group.

Reading Grade Level	f	Grade Score	f
10.0-11.2	3	8.8-8.6	2
8.8-10.0	1	8.6-8.4	1
7.6-8.8	1	8.4-8.2	1
6.4-7.6	1	8.2-8.0	1
5.2-6.4	1	8.0-7.8	1
4.0-5.2	1	7.8-7.6	1
2.8-4.0	1	7.6-7.4	1
1.6-2.8	1	7.4-7.2	1
0.4-1.6	1	7.2-7.0	1
0.0-0.4	1	7.0-6.8	1
	1	6.8-6.6	1
	1	6.6-6.4	1
	1	6.4-6.2	1
	1	6.2-6.0	1
	1	6.0-5.8	1
	1	5.8-5.6	1
	1	5.6-5.4	1
	1	5.4-5.2	1
	1	5.2-5.0	1
	1	5.0-4.8	1
	1	4.8-4.6	1
	1	4.6-4.4	1
	1	4.4-4.2	1
	1	4.2-4.0	1
	1	4.0-3.8	1
	1	3.8-3.6	1
	1	3.6-3.4	1
	1	3.4-3.2	1
	1	3.2-3.0	1
	1	3.0-2.8	1
	1	2.8-2.6	1
	1	2.6-2.4	1
	1	2.4-2.2	1
	1	2.2-2.0	1
	1	2.0-1.8	1
	1	1.8-1.6	1
	1	1.6-1.4	1
	1	1.4-1.2	1
	1	1.2-1.0	1
	1	1.0-0.8	1
	1	0.8-0.6	1
	1	0.6-0.4	1
	1	0.4-0.2	1
	1	0.2-0.0	1
	1	0.0-0.0	1

Table IV.

Table IV.

Table IV is a summary made from

. It shows the data arranged in three groups according to categories that describe reading ability.

Frequency Distribution Graphs



Figure A.

Figure A.

Figure A shows

graphed in two ways. At the left is the frequency scale. Above each class interval a line is drawn on the horizontal scale at a level corresponding to the frequency of that interval. The resulting stair-step pattern is called a histogram. Connecting the centers, or midpoints, of the class intervals by straight lines produces a frequency polygon. Notice that the frequency polygon gives the impression that the class intervals are continuous. Even a casual examination of either of these graphs, or curves, gives some idea of the general characteristics of the distribution.

In statistics considerable attention is paid to the form of such curves. The distribution is said to be bilaterally symmetrical if it can be folded vertically so that the two halves of the curve are essentially the same. If the curve is lacking in symmetry, the distribution is said to be skewed. The so-called normal curve has a bell shape and is perfectly symmetrical.

Measures of Average, or Central Tendency

The statistician uses frequency tables to carry on further computations. Usually the statistician seeks to find some one number that will represent all the data in some definite way. One method of summarizing data is to calculate the average of the group. The statistician uses three kinds of averages. Each kind represents the group in a different way.

Arithmetic Mean

Suppose you are interested in selecting better students for a technical school with the aim of decreasing the proportion of students who fail or drop out before they finish the course. It is decided to give all entering students a mechanical aptitude test and then follow up later to see whether the test actually predicts anything about success in the school.

Table V shows the results that might have been obtained in such a study. The criterion of success is simply graduation. Before deciding to use the aptitude test for selection, however, the averages and variabilities of the two groups must be studied.

Aptitude Score	Frequency		
	Graduates	Non-graduates	%
60-69	2	1	4.2
50-59	3	1	5.8
40-49	22	4	27.7
30-39	14	7	21.2
20-29	1	1	3.4
10-19	1	1	3.4
0-9	1	1	3.4
Total	45	16	100.0

Table V.

Table VI shows very clearly that the students who graduated had a higher average score than those who did not. This is true whether one compares the modes, the medians, or the means. Note that 90 percent of the graduates exceeded the mean for the total group, while only 15 percent of the nongraduates exceeded it. In addition, while there is considerable variation in each group, there is greater variability among the nongraduates than among the graduates. There is even greater variation in the combined group.

Statistical Results	Graduates	Non-graduates	Combined Group
N	45	16	61
Measure of Central Tendency			
Mode	60	60	60
Median	62.5	40	57.5
Mean	62.2	40.7	57.7
Measure of Variability	(45-60)	(16-40)	(61-60)
Interquartile Range	55	40	55
Standard Deviation	11.8	14.7	14.2
Standard Error	1.9	3.6	2.3
Standard Deviation	7.8	8.9	10.2
Standard Error	1.3	2.2	1.7
Standard Deviation	1.8	3.4	2.3
Standard Error	0.3	0.6	0.4

Table VI.

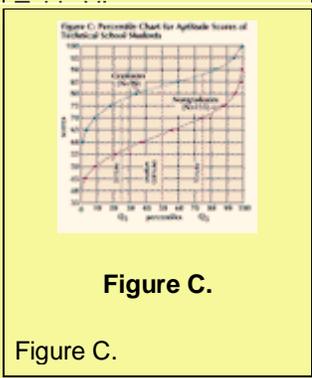
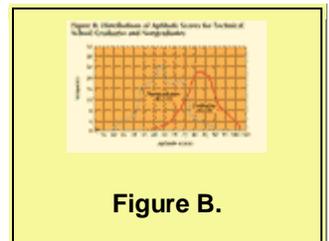
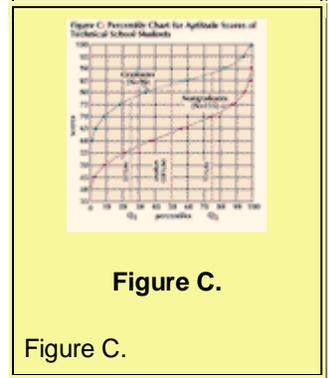


Figure B shows two simple frequency polygons on the same chart. Figure C shows the two distributions in terms of cumulated percentage frequencies. The vertical distance between the two curves shows that the graduates scored distinctly higher than did the nongraduates all along the line. Any score equivalent (such as the median score, or 50th percentile) can be obtained by running up from the percentile scale to the curve and across to the score scale.



actually constitutes a set of norms for this test, because any applicant's score can be evaluated in terms of how the applicant compares with either group.



Measures of Relationship

When data are obtained for two or more traits on the same sample, it may be important to discover whether there is a relationship between the measures. For example, statisticians may try to answer questions such as the following: Is there a relationship between a person's height and weight? Can one judge a person's intelligence from any physical characteristic? Is personality related to job success? How consistent are repeated measures of achievement in school? Is income related to how far a person went in school? Can one predict a person's reading comprehension from reading speed or word comprehension?

These questions are examples of correlation, or relationship, problems. In every case there has to be a pair of measurements for each person in the group before one can measure the correlation. For example, to determine the correlation between height and weight for high-school students, each student's height and weight must be known. By tabulating each pair of measurements on a scattergram, or scatter diagram, a visual idea of the correlation is possible.

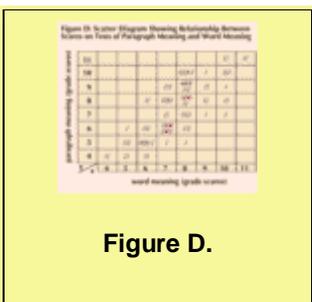


Figure D.

Figure D, a scatter diagram, shows the paired grade-level scores on a test of paragraph meaning and a test of word meaning for a group of sixth-grade pupils. The vertical axis (y) is laid off in terms of grade level for the paragraph-meaning test scores. The horizontal axis (x) is laid off in terms of grade level for the word-meaning test scores. Each tally mark represents both scores for one pupil. For example, one pupil scored 8 on word meaning and 5 on paragraph meaning. The two scores are represented by a single tally mark placed in the square that is directly above the 8 on the horizontal scale and across from the 5 on the vertical scale.

Table VII is a contingency table that shows the scores grouped by class intervals, with numerals in place of the tally marks. In both the scatter diagram and the contingency table, the scores tend to fall into a straight band that rises from left to right. It is evident that there is a decided trend toward higher scores on paragraph meaning to go with higher scores on word meaning. This is called positive correlation. Note, however, that the correlation is not perfect. For example, 10 pupils who scored at the sixth-grade level for paragraph meaning scored at the seventh-grade level for word meaning, as indicated in

Paragraph Meaning	Word Meaning				Total
	6-6	6-7	6-8	6-9	
10-11	1	1	1	1	4
9-10	1	1	1	1	4
8-9	1	1	1	1	4
7-8	1	1	1	1	4
6-7	1	1	1	1	4
Total	5	5	5	5	20

Table VII.

Table VII.

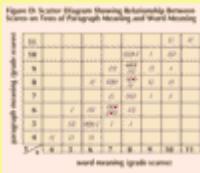


Figure D.

Figure D.

Sometimes there are negative correlations. This means that higher scores for one variable tend to be associated with lower scores for the other variable. Zero correlation indicates that there is no relationship between the two; knowing a person's score or rank on one variable would not enable the prediction of the person's score on the other variable.

The statistician is not satisfied to indicate correlation in a general way but wants a single number that will show the amount of relationship. This precise measure involves the calculation of the correlation coefficient (r), which expresses numerically the actual degree or intensity of relationship. The correlation coefficient runs from -1.00 (perfect negative correlation) through 0 (zero correlation) to $+1.00$ (perfect positive correlation). The method of obtaining the correlation coefficient is beyond

the scope of this article.

High correlations-whether positive or negative-are extremely useful because they enable statisticians to make accurate predictions. Zero correlations-which will not predict anything-are also useful. They may show, for example, that one cannot judge a person's intelligence from head size. In this case the correlation between head size and intelligence is close to zero. However, a high correlation between two traits does not necessarily mean that one trait caused the other trait.

The amount of evidence required to prove a cause-effect relationship between two traits is much greater than that needed to simply show a correlation. The size of the correlation coefficient as computed for the data shown in

is 0.76 . Since this correlation is not extremely high, a good statistician would bear this in mind and proceed with caution in predicting one variable from the other.

The Manipulation of Statistics



Figure E.

Figure E.

One of the chief problems with statistics is the ability to make them say what one desires through the manipulation of numbers or graphics. Graphs are frequently used in newspapers and in the business world to create a quick and dramatic impression. Sometimes the graphs used are misleading. It is up to the reader to be alert for those graphs that are designed to create a false impression. A false impression can be presented, as in Figure E for example, by changing the scales in a graph or chart or by omitting a portion of the items in the sample.



Figure D.

Figure D.