# CHAPTER 12

# Correlation: Measuring Relations

***After reading this chapter you should be able to do the following:***

1. *Give examples from everyday experience that demonstrate correlation between variables.*
2. *Discuss the essential characteristics of the correlational relation: direction and degree of relation.*
3. *Make a scatter-plot diagram for some paired observations and explain how the chart shows direction and degree of a relation.*
4. *Explain the purpose and fundamental characteristics of correlation coefficients.*
5. *Compute and interpret the Pearson r for a group of paired observations.*
6. *Compute and interpret the Spearman rho.*
7. *Discuss some ways in which correlation coefficients are used to evaluate tests.*
8. *Compute and discuss some uses of the point biserial correlation coefficient.*

## Relations Among Variables

There are probably many good ways to describe the goals of science. One might be to say that science attempts to discover relations among natural phenomena, to describe them accurately, and then, ultimately, to explain why they exist. The correlational methods we cover in this chapter and the next address the first two of these goals, discovering and describing relationships.

If you reflect for a minute on any of the science courses you have taken, you should see easily that most of the basic "laws" that are taught as items of scientific knowledge are really statements about relationships among variables. In chemistry, the pressure of a gas is related to its volume and temperature; in economics, the market price of a commodity is related to its supply and the demand for it; in psychology, the amount of information remembered in a memory study is related to the amount of time spent studying it, and so on.

In the natural sciences, many of the lawful relationships among correlated variables are known with such precision that they can be stated quantitatively. If one variable has this value, then another variable must have that value. Einstein's famous mass-energy equation, for instance, shows the relationship between the amount of energy contained in a bit of matter (one variable) and its mass (essentially, its weight, the second variable):

$$E = mc^2$$

This relationship is stated quantitatively in that it specifies values for each variable involved: the energy (in ergs) contained in a body is equal to its mass (in grams) multiplied by the squared speed of light (in centimeters per second).

This particular law was stated theoretically many years before it was verified experimentally. Very often, however, science progresses in the opposite direction: relationships may be discovered through observation, and only after many repeated observations are the relationships known with precision. It is this latter kind of scientific progress, forming laws after collecting many, many observations, that calls for statistical analyses such as correlational methods. This is particularly true in the behavioral sciences.

Economists may suspect that the rate of inflation is related to the prime lending rate; psychologists may suspect that a person's reaction to a persuasive speech is related to the identity of the speaker. But these relations are much harder to discover than physical laws, much harder to summarize precisely, for several reasons. As discussed in the earlier chapters, measurement in the behavioral sciences is often prone to larger error than measurement in the physical sciences, and furthermore, we are often less sure that observable indicators (such as test scores) correspond closely to hypothesized constructs such as intelligence, anxiety, attitude strength, personality traits, and so on. These problems contribute to the difficulty of determining relations among variables; they account for the fact that the behavioral sciences cannot yet summarize many relations with neat mathematical laws.

This does not mean that we cannot hope to learn something about relations among behavioral phenomena, even though these relations are complex or obscure. **Correlational** methods are techniques designed specifically for identifying relations among variables; correlational methods give us a means of describing and measuring relations even in situations where the relations are difficult to see.

## Kinds of Relations

Here we will confine our discussion to relations between two variables. Two variables are related when changes in the value of one are systematically related to changes in the value of the other. There are two things we will want to know about relations: direction and degree.

When we speak of **direction** we mean that relations between variables are either positive or negative. A **positive relation** means that as values of one variable increase, values of the other tend also to increase. For human beings, the relation between height and weight is positive. People who register high values on the height variable tend also to register high values on the weight variable. There is probably also a positive relation between the number of shots a basketball player takes in practice and his shooting percentage during game play. Those who get lots of practice probably have higher shooting percentages than those who get little practice. There is also a positive relation between your scores on examinations in any class and the grade you ultimately receive; those who receive high exam scores tend to receive high grades.

A **negative relation**, as you might suspect, indicates the opposite relation between variables. High values on one variable tend to be accompanied by low values on the other. As the interest rate for mortgages goes up, the number of new home loans issued in a given time period tends to go down. There are volumes of statistics that suggest a negative relation between the number of cigarettes smoked over a lifetime and age at death.[1]

Relations between variables can also be characterized by **degree**. **Degree of relation** refers to the extent that observed values adhere to the designated relation. The highest degree of relation is a **perfect relation**, in which knowing the value of one variable determines exactly the value of the other. To give a trivial example, your age in years is perfectly correlated with your age in months. If you register 20.5 on the "years" variable, then you must register 246 on the "months" variable. Physical laws such as $F=MA$, $y=\frac{1}{2}gt^2$ and $E=mc^2$ are further examples of perfect relations. In a perfect relation there are no exceptions to the rule.

At the other extreme is the lowest degree of relation, the **zero relation**; this is just another way of saying that no relation exists between variables. You can probably think of any number of these, such as the relation between the number of coyotes living in Montana during any given year and the number of yellow taxicabs in New York City during the same year.

The vast majority of relations studied in the behavioral sciences range in degree between zero and perfect. The relation between intelligence scores of parents and those of their children is positive but not perfect. That is, parents who score high on intelligence tests tend to have children who score high, and vice versa, but there are many exceptions to this rule.[2]

Correlational methods offer a means of determining whether or not such relations exist, that is, whether variables are "co-related" or **correlated**. They can tell us the direction of the relation and its degree. Quite often we simply cannot ascertain this information by casual observation; quite often the analysis surprises us by showing little or no relation between variables. You might expect, for example, that people who get high grades in college would also tend to have more successful careers and more prestigious professions in later life, but available statistics suggest there is no relation between college grades and later success.[3] The only answer to questions about many such relations comes from empirical evidence; that is, collect the data and apply correlational analysis.

## Scatter-Plot Diagrams

When a single variable is under study, the frequency distribution graph (histogram or frequency polygon) provides a group picture of the data. From visual inspection of the frequency polygon, one gets a rough idea of central tendency and variability. Similarly, when one is inter ested in the relation between two variables, there is a graphical technique that can give a rough idea of the direction and degree of the relation. This is the **scatter-plot diagram**.

To make a scatter plot we need a number of **paired observations** on two variables. Usually, one variable is designated *X* and the other *Y*. Usually, the variable that is suspected to be the **independent** or **explanatory** variable is designated *X*, and the variable that is suspected to be the **dependent** or **response** variable is designated *Y*. We can make paired observations on these if in the group of interest there is one *Y* for every *X* and vice versa.

Do people who spend much time studying get better grades than those who spend little time studying? Suppose we collected some paired observations on the two variables mentioned here (grades and time spent studying) and present the results in a scatter plot.

Ten people volunteer to provide us with data, and for each person we record the number of hours per week spent studying (*X* or independent variable) and the current grade-point average (*Y* or dependent variable). These data are summarized in Table 12.1. Simply looking at Table 12.1 may or may not suggest an answer to the question of whether or not there is a relation between *X* and *Y*, but the scatter plot in Figure 12.1 clearly does. Notice the following characteristics of the scatter plot:

1. Each variable is represented by one of the axes; each axis is scaled to include only the range of values observed on one variable. Thus, the *X* axis covers only the range of hours spent studying, from 6 to 16; the *Y* axis covers the grade-point averages from 2.1 to 3.9.

**Table 12.1** *Paired Observations on Grade-point Averages and Hours per Week Spent Studying for 10 People.*

| | HOURS PER WEEK STUDYING (*X*) | GRADE-POINT AVERAGE (*Y*) |
|---|---|---|
| Bob | 12 | 3.00 |
| Carol | 9 | 2.30 |
| Ted | 6 | 2.10 |
| Alice | 12 | 3.70 |
| Cleo | 8 | 2.60 |
| Julius | 13 | 3.40 |
| Mark | 14 | 3.80 |
| Anthony | 7 | 2.40 |
| Mickey | 10 | 3.10 |
| Minnie | 16 | 3.90 |

2. Both axes are about the same length, making the diagram square in shape.
3. Each point represents one pair of observations. The point representing observations on Cleo, for instance, is labeled to indicate that for this pair of values, $X$=8 and $Y$=2.6.
4. The dots are not connected with a line (as they would be in a frequency polygon).

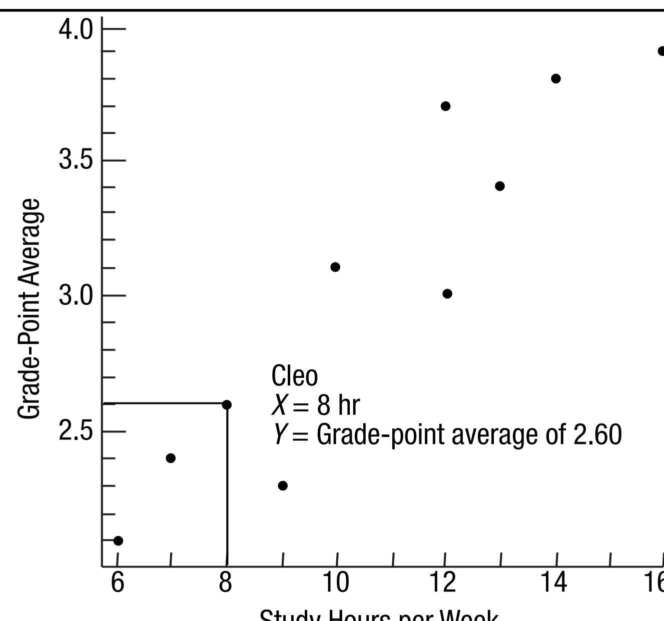
In order to get a better idea of the message conveyed by this graph, draw a light pencil line around all of the points (this is not part of the scatter plot, just an aid to interpreting it). Notice that the group of dots tends to rise toward the right; this indicates a *positive* relation—values of both variables tend to increase together. Scatter plots indicating positive relations are also shown in Figures 12.2(a), (b), and (c). (The $r$ values shown will be explained shortly.)

The body of dots slopes in the opposite direction when the two variables are related negatively. If the value of one tends to decrease as the value of the other increases, the group of points will form an oval sloping down toward the right, as in Figures 12.2(d), (e), and (f). Thus, the direction of the slope indicates the direction of the relation.

The scatter plot also permits rough estimation of *degree* of relation. If all the dots line up perfectly in a straight line, the relation between $X$ and $Y$ is a perfect relation. Figure 12.2(a) shows a perfect positive relation, and 12.2(d) shows a perfect negative relation. The more the dot pattern differs from a straight line, the lower the degree of relation.

In the grade-point/study hours example [Figure 12.1], the correlation between variables is relatively high; this is reflected in a thin, compact oval. Scatter plots showing

**Figure 12.1**




*Scatter-plot diagram for the paired observations presented in Table 12.1. The body of dots suggest that as hours per week of studying increase, grade-point averages increase.*

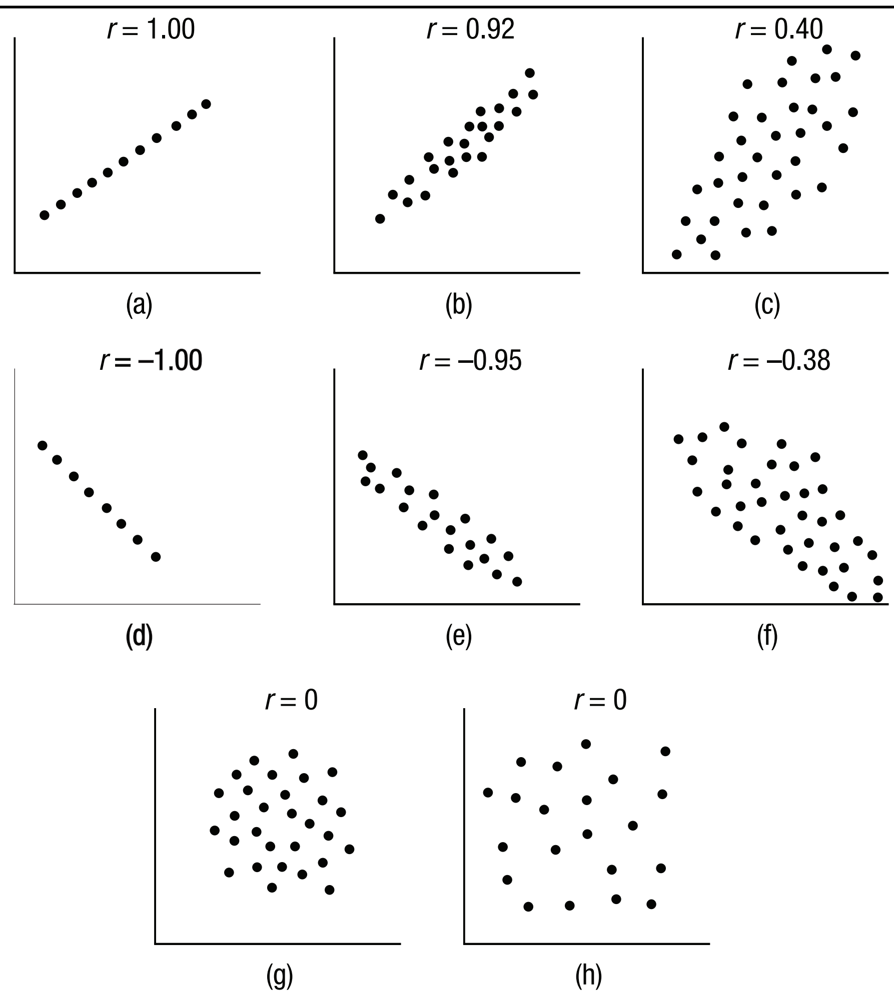
other high correlations are pictured in Figures 12.2(b) and (e). The scatter plot indicates a low-degree relation with a relatively fat oval. Scatter plots represent low-degree relations in 12.2(c) and (f).

Finally, when no relation exists between the variables, the scatter plot will take the general shape of a circle [Figure 12.2(g)] or some other irregular form [for example, Figure 12.2(h)], in which there is no discernible upward or downward slope.

## Correlation Coefficients

The scatter plot is a quick and ready method for picturing correlational relations, but it is not the endpoint of correlational analysis. Statistics designed to specify precisely the direction and degree of relations are known as **correlation coefficients**.

Figure 12.2



There are many available correlation coefficients, and their respective computational formulas are quite different; each is designed for particular kinds of data and particular kinds of variables. However, all of them share some important characteristics.

Most correlation coefficients range in value from –1.00 to + 1.00. The computational methods that produce them have been designed to produce coefficients with negative values when a negative relation exists and positive values when a positive relation exists. So all you have to do to determine the direction of a relation is to look at the *sign* of the correlation coefficient describing the relation. Coefficients such as –0.90, –0.37, and –0.66 indicate negative relations, and coefficients like +0.72, +0.62, and +0.21 indicate positive relations.

The various correlation coefficients have also been designed so that the *degree* of the relation is specified by the size of the coefficient, irrespective of its sign. Perfect relations produce correlation coefficients of +1.00 (perfect positive relation) or –1.00 (perfect negative relation); the larger the size, the higher the degree of correlation. Values of +0.97, –0.92, +0.83, and +0.92 appear when the degree of correlation is high; values such as +0.09, –0.23, +0.31, and –0.14 turn up when computed for variables correlated to a low degree. Correlation coefficients for the relations graphed in Figure 12.2 are shown on the respective scatter plots as $r$ values.

We will begin by considering the most useful correlation coefficient: the **Pearson product-moment correlation coefficient**.

## Pearson Product-Moment Correlation Coefficient

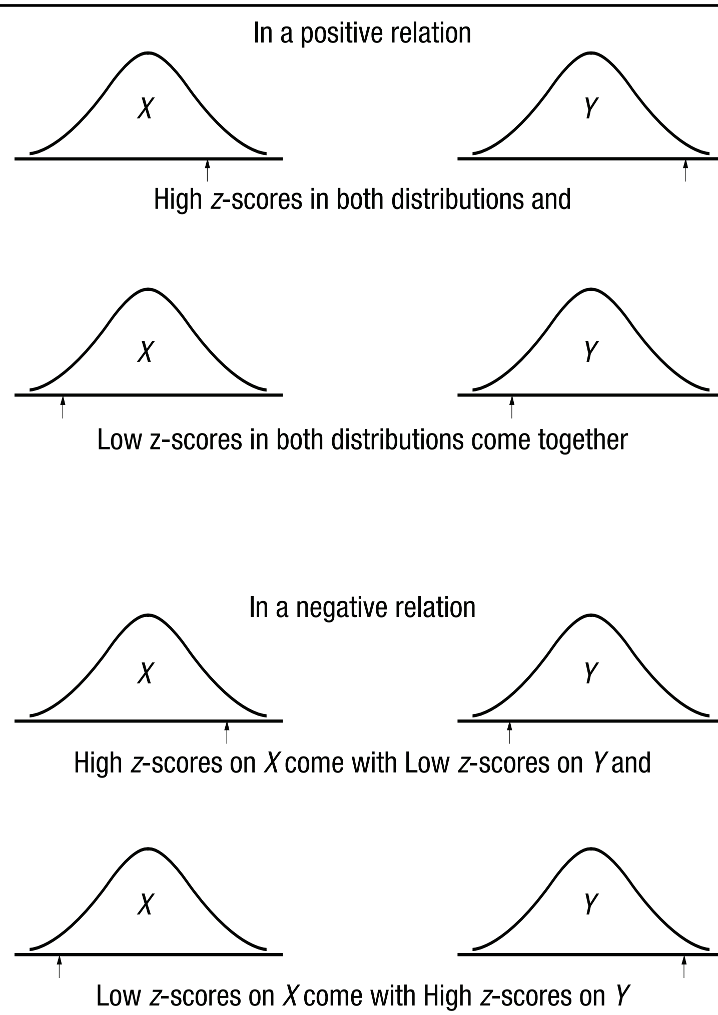
The Pearson product-moment correlation coefficient has such a long name that we will call it either the Pearson $r$, or simply $r$. In descriptive statistics $r$ is the symbol for this statistic.

In order to appreciate the message conveyed by the Pearson $r$ and its derivation, consider again the meaning of relation between two variables. In the grade-point/study hours example, observations on one variable range between 2.1 and 3.9, while on the other variable the numbers range from 6 to 16. A measure of correlation must be scale-independent in that the size of the numbers registered on the scale should not be reflected in the correlation statistic. We might even be searching for relations between variables in which observations on one variable are on the order of 0.00003 and observations on the other variable are numbers like 6 million, and so on. When two variables are associated in a positive relation, we would like to find a statistic that says, "Those who score relatively high on $X$ tend to score relatively high on $Y$ *and* those who score relatively low on $X$ tend to score relatively low on $Y$;" the important concern is the *relative* positions within the two groups. This should suggest to you that $z$-scores might be useful here.

Figure 12.3 shows how the preceding statement can be rephrased in terms of $z$-scores. The distributions suggest that we are searching for a statistic that says, "In a positive relation those with high *z*-scores in the X distribution receive high *z*-scores in the *Y*, and those with low *z* values in the *X* distribution obtain low *z* values in the *Y*." Conversely, the statistic should also be designed to say that in negative relations high *z*-scores in the *X* distribution are associated with low *z*-scores in *Y*, and low *z*-scores in *X* are associated with high *z*-scores in *Y*. The degree to which individual paired observations conform to these rules—the degree of the relation—should also be reflected in the statistic.

The Pearson *r* is designed to do all of this. However, before it is applied to a set of paired observations, several conditions must be met. These are sometimes referred to as the assumptions underlying the Pearson *r*. This means that, when you summarize your data with a Pearson *r*, and report that statistic, it is assumed that the following conditions have been met (otherwise the statistic can present a misleading picture of the relationship between variables):

**Figure 12.3**



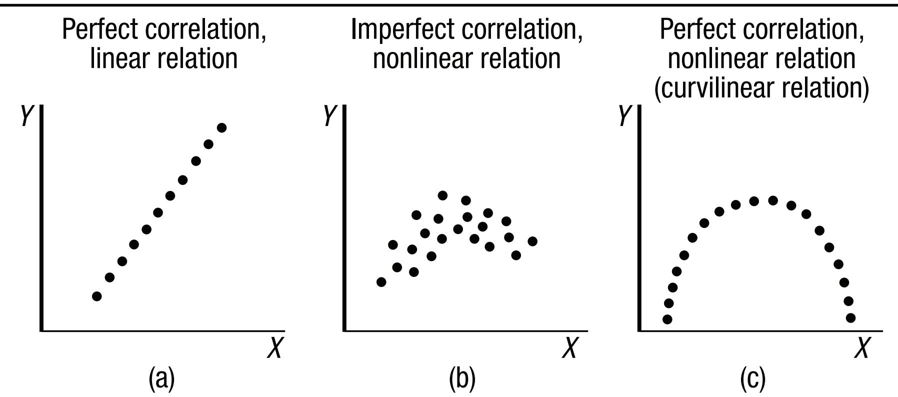
1. We must have at least interval-scale data on both variables. *z*-scores and statistics derived from them make use of distances on the measurement scale. These distances have meaning only for interval-, ratio-, or absolute-scale measurements.
2. We assume that the relation between *X* and *Y* is approximately linear. One meaning of linear relation is that the graph of the relation between *X* and *Y* is a *straight line*.[4] Figure 12.4 shows several *X*–*Y* relations, but only those in 12.4(a) are linear, since only these are represented by straight lines. Notice that the relation in 12.4(c) is a perfect relation—specifying a value of *X* determines exactly what the value of *Y* will be. Graph 12.4(c) is not, however, a linear relation. Although the correlation is perfect, the relation will not be accurately measured or detected by the Pearson *r*. Figure 12.4(c) is a *curvilinear* relation; the Pearson r is a measure of *linearity* in a relation.

Incidentally, there are some instances in the behavioral sciences where *curvilinear* relations appear. Figure 12.4(c) might represent, for instance, the relation between people's hand strength and chronological age. If the vertical axis represents strength and the horizontal axis represents age, Figure 12.4(c) would suggest that hand strength increases up to a certain age and then decreases.[5]

Most scatter plots, as you know, do not appear as straight lines; but the Pearson *r* is appropriate if it is assumed that the underlying relation between *X* and *Y* is essentially linear. The relation is probably linear if the scatter plot resembles those in Figures 12.5(a), (b) or (c). The bends in the ovals of scatter plots 12.5(d), (e) and (f) suggest nonlinear relations.

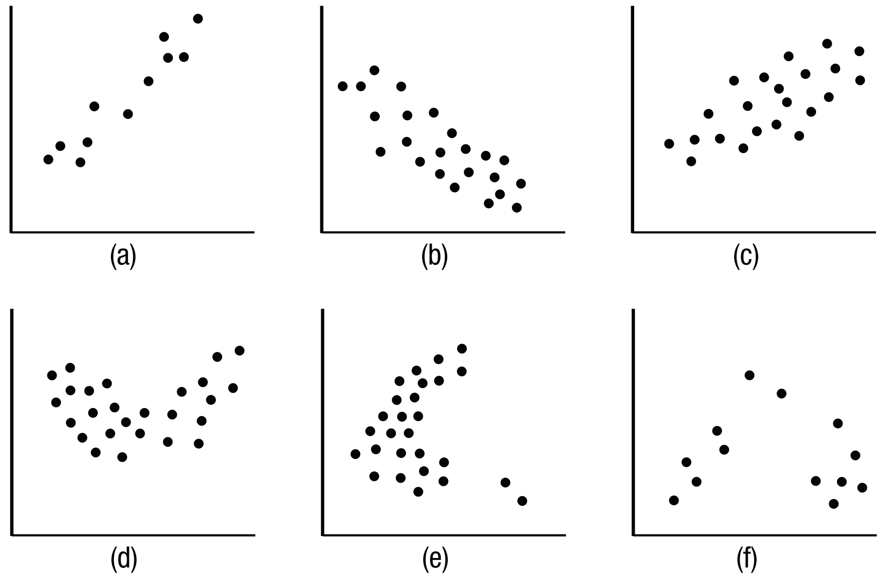
In most situations there is probably little harm in computing a Pearson r for data that depart somewhat from the linear relation; the consequences of doing so are a cor-

**Figure 12.4**




*Three kinds of X-Y relations: Plot (a) appears to show a perfect linear relation, since all the points fall on a straight line. Plot (b) shows an imperfect correlation, since specifying a value of X does not determine exactly the value of Y. Plot (c) appears to represent a perfect correlation, since each value of X is associated with one value of Y. Plot (c) does not show a linear relation, however, since the scatter plot points do not fall on a straight line; it represents a typical curvilnear relation.*

**Figure 12.5**




*Linear and nonlinear X-Y relations. Although none of the scatter plots show a perfect linear relation, it is probably safe to assume that plots (a), (b), and (c) represent essentially linear relations. The shapes of scatter plots (d), (e), and (f) suggest that the underlying relations are nonlinear.*

relation coefficient with a spuriously low size. The Pearson $r$ measures only linear relations; relations that exist but that are nonlinear do not show up in the value of $r$.

3. When computing the Pearson $r$, we assume **homoscedasticity** of variances. This means essentially that the width of the scatter-plot oval is relatively uniform throughout its length. Figure 12.6 shows one scatter plot with homoscedasticity and two without it. In a homoscedastic relation the variance of $Y$ values above one interval on the $X$ axis is about the same as the variance of $Y$ values above any other interval on the $X$ axis, such as in Figure 12.6(a). Naturally, no real scatter plot

**Figure 12.6**

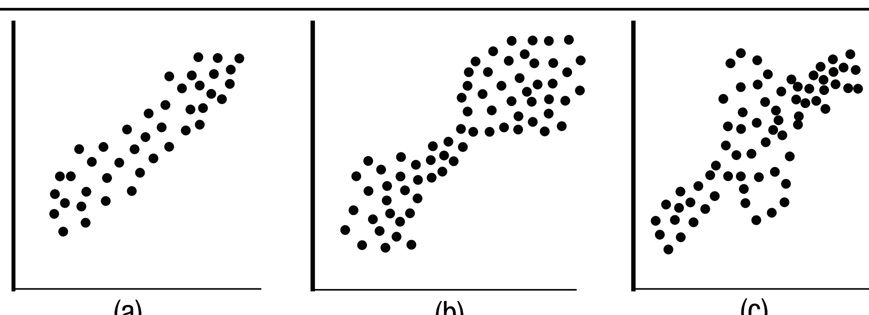



*Illustration of homoscedasticity and nonhomoscedasticity of variances. Scatter plot (a) has about the same width all along its length. Scatter plots (b) and (c) are much wider in some places than others. Thus, (a) exhibits homoscedascity of variances while (b) and (c) do not.*

meets this condition exactly, but scatter plots such as 12.6(b) and 12.6(c) depart sufficiently from homoscedasticity to make the use of the Pearson *r* inappropriate. If one has at least interval-scale data on both variables, and if inspection of the scatter plot suggests that conditions of linearity and homoscedasticity are met, the Pearson *r* will yield useful information about the relationship between the variables.

Just as we did with the standard deviation, we will first compute the Pearson *r* by means of its definition formula. The definition formula is more cumbersome to work with than the (optional) computational formula, but it shows clearly how *z*-scores are related to the value of *r*. Defined verbally, the Pearson *r* is the mean of the "cross-products" of the *z*-scores. We take each pair of observations, multiply the *z*-score in the *X* distribution by the *z*-score in the *Y* distribution, add these cross -products, and divide by the number of pairs of observations. The statistic obtained has all the characteristics of the correlation coefficient described earlier. In symbols:

$$r_{XY} = \frac{\sum_{i=1}^{n} z_{X_i} z_{Y_i}}{n}$$

Do not let all of the subscripts in the above expression confuse you—they are there to indicate what each symbol represents. It may be useful for some people to take this formula part by part:

$r_{XY}$ is the correlation coefficient *r* between two variables.

$z_{X_i}$ represents the *z*-score associated with $i^{th}$ individual score on the *X* variable.

$z_{Y_i}$ represents the *z*-score associated with $i^{th}$ individual score on the *Y* variable.

*n* represents the number of paired observations.

Obtaining a value of *r* does not require any computational feats beyond those already covered in Chapter 4; it is an exercise in careful bookkeeping, however, since we will have to obtain many *z*-scores and keep track of several other statistics as well.

The data from Table 12.1 have been reproduced in Table 12.2, together with some other values needed to find the value of r. U sing the preceding formula, compute r in steps as follows:

1. First find the mean and standard deviations of the *X* and *Y* distributions (mean $X = \overline{X}$, mean $Y = \overline{Y}$; standard deviation for $X = s_X$ and for $Y = s_Y$). These are computed exactly as described in Chapters 3 and 4. Of course, *Y* values do not enter into computation of the *X* statistics, and vice versa. The computed values for these statistics are shown at the bottom of Table 12.2.
2. Use the means and standard deviations to find a *z*-score for each *X* and for each *Y* score within their respective distributions. The *z*-score for *X* = 12, for instance, is:

$$z_{X_i} = \frac{X_i - \overline{X}}{s_X} = \frac{12 - 10.70}{3.07} = \frac{1.30}{3.07} = +\ .423$$

3. For each pair of observations multiply the $z$-score for the $X$ value ($z_X$) by the $z$-score for the $Y$ value ($z_Y$) to obtain the cross-product; these are listed in the last column of Table 12.2. For example, the first person listed registered a $z$-score of +0.423 on the $X$ variable and a $z$-score of –0.048 on the $Y$. The product of these two values is –0.0203.

4. Find the mean cross-product. This is done by adding the cross-products and dividing by the number of paired observations. The result is the Pearson $r$, here equal to +0.93 ($r$ is usually specified to two decimal places).

For the data collected, the correlation between hours per week studying and grade point averages turned out to be +0.93. But what does this mean? Now that we have the correlation coefficient, what do we do with it?

**Table 12.2** *Computation of Pearson r*

| $i$ | $X_i$ | $X_i - \overline{X}$ | $z_{Xi}$ | $Y_i$ | $X_i - \overline{X}$ | $z_{Yi}$ | $z_{Xi}\, z_{Yi}$ |
|---|---|---|---|---|---|---|---|
| 1 | 12 | +1.3 | +0.423 | 3.00 | –0.03 | –0.048 | –0.0203 |
| 2 | 9 | –1.7 | –0.554 | 2.30 | –0.73 | –1.166 | +0.6460 |
| 3 | 6 | –4.7 | –1.531 | 2.10 | –0.93 | –1.485 | +2.2735 |
| 4 | 12 | +1.3 | +0.423 | 3.70 | +0.67 | +1.070 | +0.4526 |
| 5 | 8 | –2.7 | –0.879 | 2.60 | –0.43 | –0.687 | +0.6039 |
| 6 | 13 | +2.3 | +0.749 | 3.40 | +0.37 | +0.591 | +0.4426 |
| 7 | 14 | +3.3 | +1.075 | 3.80 | +0.77 | +1.230 | +1.3222 |
| 8 | 7 | –3.7 | –1.205 | 2.40 | –0.63 | –1.006 | +1.2122 |
| 9 | 10 | –0.7 | –0.228 | 3.10 | +0.07 | +0.112 | –0.0255 |
| 10 | 16 | +5.3 | +1.726 | 3.9 | +0.87 | +1.390 | +2.3991 |

$\sum X_i = 107$ $\quad \sum z_{X_i} = 0.0$ $\quad \sum Y_i = 30.30$ $\quad \sum z_{Y_i} = 0.0$ $\quad \sum z_{X_i} z_{Y_i} = 9.3063$

$\overline{Y} = 3.03$

$\overline{X} = 10.70$

$$r_{XY} = \frac{\sum z_{X_i} z_{Y_i}}{n} = \frac{9.3063}{10} = +\ 0.93$$

$s_X = 3.07$ $\quad s_Y = 0.626$

$n$ = Number of paired observations = 10

## Interpreting the Pearson $r$

Interpreting the sign of the correlation coefficient is easy. The plus sign indicates a positive relation—those who study more tend to have higher grade-point averages.

Interpreting the size of the coefficient is a more subtle matter. Because correlation coefficients look like proportions, many people respond to them as if they were proportions, thinking that a correlation of 0.70 is twice as high as a correlation of 0.35. This is unfortunate, because $r$ itself is not a proportion. Interpretation of the magnitude of r depends somewhat on the number of observations,[6] but more important for our immediate purposes, on the value of $r^2$. $r^2$ is a proportion, specifically the proportion of variation in the scatter plot accounted for by a linear equation, but we will have to wait until the next chapter to appreciate that definition completely. For the moment, we will do best simply to characterize some sizes (absolute values) of $r$ as being high and others as being low. These labels are presented with values of $r^2$ in Table 12.3.

The most common misuse of correlation coefficients occurs when people assume that a high correlation between variables proves that a cause-and-effect relationship also exists between these variables. For instance, there have been several recent medical studies showing a positive correlation between coffee consumption and heart attacks. But this evidence, by itself, does not prove that coffee drinking *causes* heart problems any more than it proves that heart problems cause coffee drinking. It is entire-

**Table 12.3** *A Rough Guide to Degree of Correlation Indicated by Different Sizes of $r$'s Absolute Value and $r^2$*

| $\|r\|$ | $r^2$ | DEGREE OF CORRELATION (ROUGH GUIDE) |
|---|---|---|
| 1.00 | 1.00 | Perfect |
| 0.98 | 0.96 | High |
| 0.95 | 0.90 | |
| 0.90 | 0.81 | |
| 0.85 | 0.72 | |
| 0.80 | 0.64 | |
| 0.70 | 0.49 | Moderate |
| 0.60 | 0.36 | |
| 0.50 | 0.25 | |
| 0.40 | 0.16 | Low |
| 0.30 | 0.09 | |
| 0.20 | 0.04 | Negligible |
| 0.10 | 0.01 | |

ly possible that a third factor causes changes in both variables. Perhaps people who drink excessive amounts of coffee also tend to be people with sedentary jobs, which would mean that many of them do not get enough proper exercise. At any rate, the discovery of a high correlation coefficient in no way establishes that changes in one variable cause changes in another variable. However, correlational methods are very useful for *predicting* the value of one variable if the value of another correlated variable is known. In fact, all of Chapter 7 will deal with the subject of making predictions from correlational information.

## Obtaining *r* Through Covariation

Computing *r* with the definition formula and methods shown above is a rather laborious process because one must compute means, standard deviations, and then *z*-scores for each *X* and *Y* score. The many computational steps in that definition formula require an excessive amount of time for both human and electronic computers to execute and they increase the likelihood of rounding error and other mistakes. So, now that you know how *z*-scores are involved in the value of *r*, we will cover a simpler definition formula. Of course, many people will consider this section optional.

This alternative approach makes use of a statistic called **covariation**. The covariation for a set of *X* and *Y* scores is obtained by taking the *X* deviation score for each pair, $(X_i - \overline{X})$, multiplying it by the *Y* deviation score for that pair, $(Y_i - \overline{Y})$, and then adding up all such cross-products for the entire group. In symbols:

$$XY \text{ covariation} = \sum_{i=1}^{N} (X_i - \overline{X})(Y_i - \overline{Y})$$

The Pearson *r* is then obtained as follows:

$$r_{XY} = \frac{XY \text{ covariation}}{\sqrt{X \text{ variation}}\,\sqrt{Y \text{ variation}}}$$

$$= \frac{\sum(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum(X_i - \overline{X})^2}\,\sqrt{\sum(Y_i - \overline{Y})^2}}$$

To show how this approach works, Table 12.4 extracts some of the information from Table 12.2 and shows the necessary computational steps. Notice that here, unlike the example in Table 12.2, the cross-products in the last column are cross-products of deviation scores rather than cross-products of *z*-scores.

The steps in computing *r* with the covariation approach may be summarized:

1. Obtain the values of $\overline{X}$ and $\overline{Y}$.

$$\overline{X} = \frac{\sum X_i}{n} = \frac{107}{10} = 10.7 \text{ and } \overline{Y} = \frac{\sum Y_i}{n} = \frac{30.30}{10} = 3.03$$

**Table 12.4** *Steps in Computing r with the Covariation Approach*

| $i$ | $X_i$ | $X_i - \overline{X}$ | $(X_i - \overline{X})^2$ | $Y_i$ | $Y_i - \overline{Y}$ | $(Y_i - \overline{Y})^2$ | $(X_i - \overline{X})(Y_i - \overline{Y})$ |
|---|---|---|---|---|---|---|---|
| 1 | 12 | +1.3 | 1.69 | 3.00 | –0.03 | 0.0009 | –0.0390 |
| 2 | 9 | –1.7 | 2.89 | 2.30 | –0.73 | 0.5329 | +1.2410 |
| 3 | 6 | –4.7 | 22.09 | 2.10 | –0.93 | 0.8649 | +4.3710 |
| 4 | 12 | +1.3 | 1.69 | 3.70 | +0.67 | 0.4489 | +0.8710 |
| 5 | 8 | –2.7 | 7.29 | 2.60 | –0.43 | 0.1849 | +1.1610 |
| 6 | 13 | +2.3 | 5.29 | 3.40 | +0.37 | 0.1369 | +0.8510 |
| 7 | 14 | +3.3 | 10.89 | 3.80 | +0.77 | 0.5929 | +2.5410 |
| 8 | 7 | –3.7 | 13.89 | 2.40 | –0.63 | 0.3969 | +2.3310 |
| 9 | 10 | –0.7 | 0.49 | 3.10 | +0.07 | 0.0049 | –0.0490 |
| 10 | 16 | +5.3 | 28.09 | 3.90 | +0.87 | 0.7569 | +4.6110 |

$\sum X_i = 107$ $\quad \sum(X_i - \overline{X})^2 = 94.10$ $\quad \sum Y_i = 30.30$ $\quad \sum(Y_i - \overline{Y})^2 = 3.9210$

$\sum(X_i - \overline{X})(Y_i - \overline{Y}) = 17.8900$

$$\overline{X} = \frac{\sum X_i}{n} = \frac{107}{10} = 10.70 \qquad \overline{Y} = \frac{\sum Y_i}{n} = \frac{30.30}{10} = 3.03$$

$$r = \frac{\sum(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum(X_i - \overline{X})^2}\sqrt{\sum(Y_i - \overline{Y})^2}} = \frac{17.8900}{\sqrt{94.10}\sqrt{3.9210}} = +0.93$$

2. Obtain a deviation score for each $X_i$ value within the $X$ distribution and each $Y_i$ within the $Y$ distribution. For the first $X$ and $Y$ pair:

$$X_1 - \overline{X} = 12.0 - 10.7 = 1.3 \text{ and } Y_1 - \overline{Y} = 3.00 - 3.03 = -0.03$$

3. Square the deviation scores:

For $i$ =1, $(X_1 - \overline{X})^2 = (1.3)^2 = 1.69$ and $(Y_1 - \overline{Y})^2 = (-0.03)^2 = 0.0009$

4. Obtain the separate $X$ variation and $Y$ variation as the sums of the squared deviation scores:

$$SS_x = \sum (X_1 - \overline{X})^2 = 94.10 \text{ and } SS_y = \sum (Y_1 - \overline{Y})^2 = 3.9210$$

5. Obtain the cross-product of $(X_i - \overline{X})$ and $(Y_i - \overline{Y})$ deviation scores for each pair of deviation scores. For $i$ = 1,

$$(X_i - \overline{X})(Y_i - \overline{Y}) = (1.3)(-0.03) = -0.0390$$

6. The sum of these cross products is the *covariation*.

$$\sum(X_i - \overline{X})(Y_i - \overline{Y}) = 17.8900$$

7. The Pearson $r$ is then obtained as

$$r_{XY} = \frac{XY\text{ covariation}}{\sqrt{X\text{ variation}}\sqrt{Y\text{ variation}}} = \frac{\sum(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum(X_i - \overline{X})^2}\sqrt{\sum(Y_i - \overline{Y})^2}}$$

$$= \frac{17{,}8900}{\sqrt{94.10}\sqrt{3.9210}} = +0.93$$

The value of $r$ obtained in this way must equal the $r$ obtained from the $z$-score formula in the preceding section.

## The Computational Formula for $r$

The simplest and most direct method of computing $r$ from raw scores has the most complex formula (the $i$ subscripts are omitted because it is clear they are implied):

$$r = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{n}}{\sqrt{\sum X^2 - \frac{(\sum X)^2}{n}}\sqrt{\sum Y^2 - \frac{(\sum Y)^2}{n}}}$$

This is the formula commonly used when computers are programmed to compute correlations because it is faster than the previous methods shown. You will also find it fastest and easiest in hand or small calculator computations because only raw scores are used in the process and it actually has the fewest number of steps involved.

Again, a table will be helpful in showing how to use the formula. We will use the $X_i$ and $Y_i$ values from the preceding examples, their squared values, and the cross products of the raw scores. These are shown in Table 12.5.

Notice first that each radical sign ( $\sqrt{\ }$ ) in the denominator contains a variation ($SS$) for one of the variables. That is:

$$\sum (X_i - \overline{X})^2 = \sum X^2 - \frac{\left(\sum X\right)^2}{n} \text{ and } \sum (Y_{i+} - \overline{Y})^2 = \sum Y^2 - \frac{\left(\sum Y\right)^2}{n}$$

We have already met these parts of the computational formula in Chapter 4, as part of the computational formula for $s$. From Table 12.5, these variations are easily found:

$$\sum(X_i - \overline{X})^2 = 1239 - \frac{(107)^2}{10} = 94.10$$

**Table 12.5** *Values Needed for the Computational Formula Method of Finding r*

| FROM TABLE 12.2 | | | | |
|---|---|---|---|---|
| $X_i$ | $Y_i$ | $X_i^2$ | $Y_i^2$ | $X_iY_i$ |
| 12 | 3.00 | (12)2 = 144 | (3.00)2 = 9.00 | (12)(3.00) = 36.00 |
| 9 | 2.30 | 81 | 5.29 | 20.70 |
| 6 | 2.10 | 36 | 4.41 | 12.60 |
| 12 | 3.70 | 144 | 13.69 | 44.40 |
| 8 | 2.60 | 64 | 6.76 | 20.80 |
| 13 | 3.40 | 169 | 11.56 | 44.20 |
| 14 | 3.80 | 196 | 14.44 | 53.20 |
| 7 | 2.40 | 49 | 5.76 | 16.80 |
| 10 | 3.10 | 100 | 9.61 | 31.00 |
| 16 | 3.90 | 256 | 15.21 | 62.40 |
| $\sum X_i = 107$ | $\sum Y_i = 30.30$ | $\sum X_i^2 = 1239$ | $\sum Y_i^2 = 95.73$ | $\sum X_iY_i = 342.10$ |

and

$$\sum (Y_i - \overline{Y})^2 = 95.73 - \frac{(30.30)^2}{10} = 3.921$$

$$\sum (\mathrm{Y}_i - \overline{Y})^2 = 95.73 - \frac{(30.30)^2}{10} = 3.921$$

The numerator of the computational formula for $r$ is simply the computational formula for the covariation statistic introduced in the preceding section:

$$\sum (X_i - \overline{X})(Y_i - \overline{Y}) = \sum XY - \frac{(\sum X)(\sum Y)}{10} = 342.10 - \frac{(107)(30.30)}{10}$$

When these elements, $X$-variation, $Y$-variation, and $XY$ covariation, are arranged as shown, we obtain the same Pearson $r$:

$$\frac{\sum XY - \frac{(\sum X)(\sum Y)}{n}}{\sqrt{\sum X^2 - \frac{(\sum X)^2}{n}}\sqrt{\sum Y^2 - \frac{(\sum Y)^2}{n}}} = \frac{17/89}{\sqrt{94.10}\sqrt{3.92}} = 0.93$$

## Test of Significance for the Population Correlation Coefficient: ρ

We are now ready to perform a test of the significance of the linear relationship between the variables *X* and *Y*. The null hypothesis states that there is no linear relationship between the two variables, while the alternative hypothesis claims that there is a linear relationship between the two variables. The population parameter of interest is ρ, the population correlation coefficient between the two variables. In terms of this parameter, the null hypothesis claims that $\rho = 0$, while the alternative hypothesis claims that $\rho \neq 0$. Symbolically, this is stated as:

$H_0$: $\rho = 0$

$H_1$: $\rho \neq 0$

We will assume that the following assumptions have been satisfied: (1) that the sample data of the ordered pairs (*X*,*Y*) are from a simple random sample, and (2) both *X* and *Y* values are sampled from populations that have normal distributions. The test statistic assuming that $H_0$ is true, is given by:

$$t = \frac{r}{\sqrt{\frac{1 - r^2}{n - 2}}}$$

has a *t*-distribution with degrees of freedom = $n - 2$. So, given a level of significance equal to α, the decision rule is:

Reject $H_0$ if | observed *t* | > critical *t*-value. For level of significance $\alpha = 0.05$, the critical *t*-value for a two tailed test with $df = 8$, is 2.306 (Table B, Appendix). In our example on the relationship between hours of study and grade point average, we found the correlation coefficient $r = 0.93$. The observed test statistic is given by:

$$t = \frac{r}{\sqrt{\frac{1 - r^2}{n - 2}}} = \frac{0.93}{\sqrt{\frac{1 - 0.93^2}{8}}} = \frac{0.93}{0.13} = 7.156$$

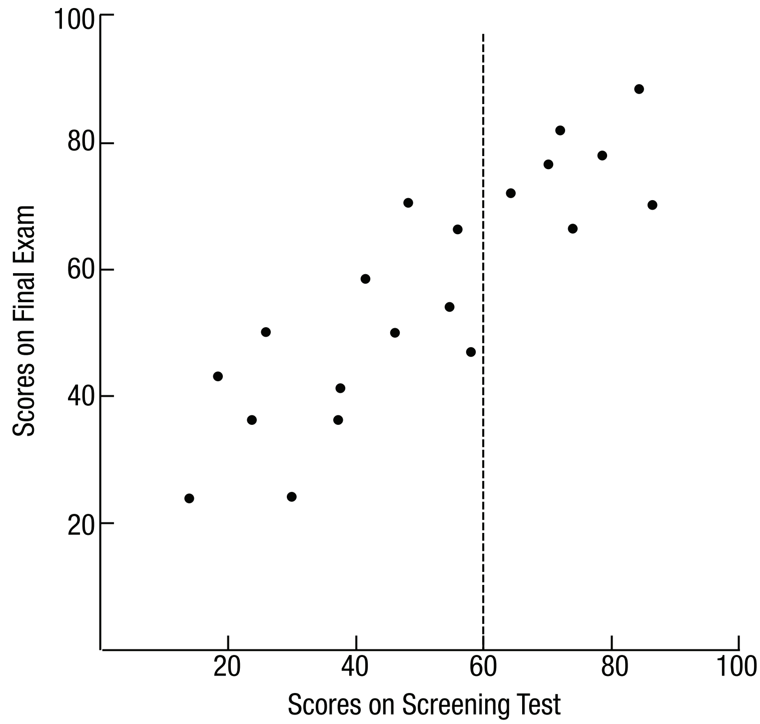
Since the | observed *t* | is greater that the critical *t* value, we reject the null hypothesis. Note that the null hypothesis is rejected at level of significance of $\alpha = 0.001$(critical *t*-value = 5.041 for $\alpha = 0.001$, $df = 8$, two-tailed test). In fact, the *p*-value calculated using Minitab is 0.0000893. The conclusion is that the correlation coeffient between the two variables is significant.

## Pitfalls in the Use of *r*

A recurring theme in this book is that statistics can be dangerous if used blindly. In order to use and interpret statistics correctly, you must be able to do more than apply formulas to data. This is as true with *r* as it is with any other statistic.

Consider two kinds of situations where *r* can mislead you about the nature of a relationship. The first of these derives from the scatter plot shown in Figure 12.7. Here, an instructor gave a "screening test" to all 20 members of his statistics class at the very first

**Figure 12.7**

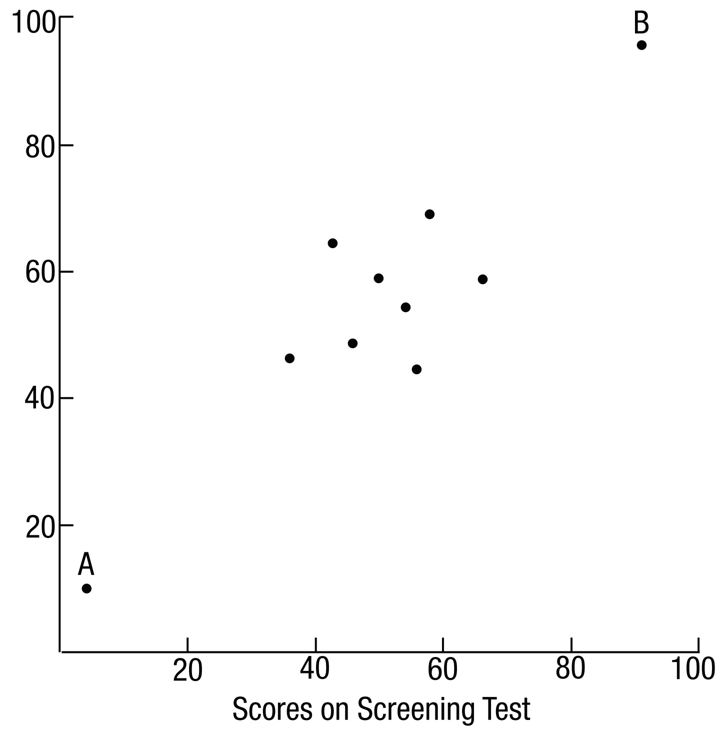



*Scatter plot showing the effects of restricted range on the value of r. The Pearson r between variables is +0.90 for the whole scatter plot, but only +0.64 for the values to the right of the dotted line.*

session. The final exam scores for these same people, taken at the end of the term, were then paired with the screening test scores, with the resulting scatter plot shown in Figure 12.7. The Pearson *r* between screening test scores and final exam scores is +0.90, and one could therefore conclude that there is a high positive relationship between these variables. However, suppose now that he had admitted only the top seven students to class—the people with screening test scores to the right of the dotted line in Figure 12.7. This process *restricts the range* of the first variable (screening test scores), and we now have a much smaller scatter plot made up of only 7 points.[7] It may be instructive for you to draw a pencil line around the complete scatter plot, and then the smaller scatter plot; the smaller scatter plot's line is less elongated. Not surprisingly, the *r* for these seven people is lower than the overall *r*: +0.64. This shows how arbitrarily restricting the range of one of the variables in a correlational relationship can produce a spuriously low *r*.

There are also times when including all of the data produces a spuriously high *r*. One of these is shown in Figure 12.8. This, again, represents a screening test/final exam situation like the one just described. The *r* for these data is +0.90, and from this statistic one could again conclude that there is a strong positive relationship between the two variables. Notice here, though, that two people, marked *A* and *B*, by themselves account for the slope and apparent narrowness of the scatter plot. Draw a pencil line around the

**Figure 12.8**




*Scatter plot illustrating how a spuriously high value of r can appear when only a few points account for most of the linearity in a relationship. With all individuals, the Pearson r between variables +0.90; without A and B, the value of r is only +0.28.*

scatter plot with all points included. You may recognize this situation as a violation of the homoscedasticity assumption. Now draw another line excluding *A* and *B*. Without *A* and *B* in the picture, the value of *r* between variables for the remaining eight people is only +0.28.

What do you report if your data appear in a form such as that shown in Figure 12.8? One approach would be to compute *r* without *A* and *B*, and then be sure to mention in your report that you are excluding two sets of paired observations because they are anomalous. In any case, the decision to include or exclude those people is based on information other than that obtained from the scatter plot; it is simply a situation calling for judgment on the part of the statistician.

## Spearman Rank-Difference Correlation Coefficient

There are many times, of course, when you may want to compute correlation statistics but do not have data that meet the assumptions necessary for the Pearson *r*—particularly with respect to the assumption about interval-level data or better. Scores on essay exams, statements of attitude strength, and confidence ratings, for instance, are often more appropriately treated as ordinal-level data rather than as interval- or ratio-level data. In this situation, you may wish to compute the Spearman rho rather than the Pearson *r*.

The **Spearman rho** (or more simply, rho) is also applicable when there are very many pairs of observations and you must compute a coefficient without the help of an electronic computer; rho is faster and easier to compute than *r* for most data. And rho is often the correlation coefficient of choice when there are one or two anomalous observations, perhaps in a situation such as that illustrated in Figure 12.8.

Rho does everything a correlation coefficient should do, and it can be interpreted almost in the same way that *r* is. The difference, however, is that rho makes use of ranks rather than actual values of measurements.

Return to the question of whether the variables "hours per week spent studying" and "grade-point averages" are related. We might want to compute a Spearman rho instead of a Pearson *r* for the data in Table 12.1 (1) because we are in a hurry or because an electronic calculator isn't available; (2) we do not have at least interval-level observations for all the people involved. Ted, for instance, might not want to tell us what his grade-point average is, except to say that it is below everyone else's. In this case, we can still rank-order all observations on the *Y* variable from high to low, even though we do not know Ted's exact *Y* value.

The first step in computing the Spearman rho is to rank-order the subjects on both variables. They may be ranked from high to low or from low to high, but both variables must be ranked in the same direction. For the people listed in Table 12.1, the rank orders on each variable, together with the *X* and *Y* values, are listed in Table 12.6. There are several points to consider in this listing:

1. People are ranked separately on *X* and *Y* variables.
2. The rank orders on *X* differ slightly from the rank orders on *Y*.
3. Two people are tied for the fourth and fifth ranks on the X variable. When ties occur, both people receive the *mean* of the two tied ranks. Here, Bob and Alice are tied for fourth and fifth ranks, and the mean rank is (4+5)/2=4.5. If more than two people were tied for a given rank, the mean of all tied ranks would be used for all of them. Notice that the rank after Alice is 6.

Spearman rho specifies the correlation between ranks on variable *X* and variable *Y* for the people observed. In order to compute the statistic we will have to determine the *difference* between ranks (*D*) for each person. The correlation coefficient itself is computed as follows: rho is equal to 1, minus six times the sum of the squared differences between ranks divided by $n(n^2–1)$, where $n$ is the number of paired observations. In symbols:

$$\text{rho} = 1 - \frac{6\sum D^2}{n(n^2 - 1)}$$

**Table 12.6** *Data Prepared for Computation of the Spearman rho*

| RANK ON *X* | PERSON | *X* VALUE | RANK ON *Y* | PERSON | *Y* VALUE |
|---|---|---|---|---|---|
| 1 | Minnie | 16 | 1 | Minnie | 3.90 |
| 2 | Mark | 14 | 2 | Mark | 3.80 |
| 3 | Julius | 13 | 3 | Alice | 3.70 |
| 4.5 | Bob | 12 | 4 | Julius | 3.40 |
| 4.5 | Alice | 12 | 5 | Mickey | 3.10 |
| 6 | Mickey | 10 | 6 | Bob | 3.00 |
| 7 | Carol | 9 | 7 | Cleo | 2.60 |
| 8 | Cleo | 8 | 8 | Anthony | 2.40 |
| 9 | Anthony | 7 | 9 | Carol | 2.30 |
| 10 | Ted | 6 | 10 | Ted | 2.10 |

This formula looks a little strange perhaps, but it has been designed to produce a statistic that obeys all the requirements for a correlation coefficient. The number 6 is a constant, always present regardless of the situation in which rho is applied.

Table 12.7 summarizes the steps in computing rho for these data.

1. List the paired observations. Here, each pair of observations represents one person, so the people are listed in the first column.
2. List the ranks each person obtained on variable $X$ and variable $Y$. *Do not list the $X$ values or the $Y$ values.* (Using the $X$ and $Y$ values at this point is the most frequent mistake students make in computing the Spearman rho. This statistic is concerned only with *ranks*.)
3. Obtain the difference between ranks, $D$, for each person. Subtract the rank on $Y$ from the rank on $X$ to obtain each person's $D$. The $D$ values are presented in the fourth column. (As a check on your computations, the sum of the $D$ values should be equal to zero.)
4. Square the $D$ values. The fifth column of Table 12.7 lists $D^2$ for each person.
5. Add the $D^2$ values and insert them into the formula.

The obtained value for the Spearman rho is +0.92, very close to the obtained value of +0.93 for the Pearson $r$ for the same data (except when there are many tied ranks). The difference between rho and $r$ is due to the fact that rho ignores distances on the measurement scales and deals only with the rank orders of observations; thus it utilizes less information in the data than does the Pearson $r$. However, the difference between rho and $r$ is usually small, and rho offers a quick means of computing correlation that is very close to the value of the product-moment correlation coefficient. Of course, if the

**Table 12.7** *Computation of the Spearman rho*

| | RANK ON *X* | RANK ON *Y* | DIFFERENCE BETWEEN RANKS (*D*) | $D^2$ |
|---|---|---|---|---|
| Minnie | 1 | 1 | 0.0 | 0.00 |
| Mark | 2 | 2 | 0.0 | 0.00 |
| Julius | 3 | 4 | –1.0 | 1.00 |
| Bob | 4.5 | 6 | –1.5 | 2.25 |
| Alice | 4.5 | 5 | +1.5 | 2.25 |
| Mickey | 6 | 5 | +1.0 | 1.00 |
| Carol | 7 | 9 | –2.0 | 4.00 |
| Cleo | 8 | 7 | +1.0 | 1.00 |
| Anthony | 9 | 8 | +1.0 | 1.00 |
| Ted | 10 | 10 | 0.0 | 0.00 |
| | | | $\sum D = 0$ | $\sum D^2 = 12.50$ |

$$\text{rho} = 1 - \frac{6\sum D^2}{n(n^2 - 1)} \qquad (n\text{=Number of paired observations})$$

$$= 1 - \frac{6(12.50)}{10(100 - 1)} = 1 - \frac{75}{990} = 1 - 0.076 = 0.92$$

data are not interval level or above, rho is definitely better than,. because statements about r and underlying variables are not correct (see Chapter 1).

## The Point Biserial Correlation Coefficient

One point mentioned in Chapter 1 was that very often in the behavioral sciences, the variables we are interested in can be observed, at best, at the nominal or ordinal levels. The **point biserial correlation coefficient** (or $r_{pb}$) is one statistic that can be used when data on one variable are nominal or ordinal—*if that variable is truly dichotomous*. That is, the nominal or ordinal variable must be able to take only two values. We will use $r_{pb}$for finding the correlation between a continuous variable and one that is really dichotomous.

There are many ready examples of dichotomous variables: sex (male/females); performance on a single test item (correct/incorrect); ownership of an automobile (yes/no). There are times when we will want to know if the status of the dichotomous variable is correlated with the values of a continuous variable. We might, for instance, want to know if car ownership is correlated (positively or negatively) with grades that college students receive. To show how $r_{\text{pb}}$ can be used, we will resurrect our ten students again for an illustration of test-item analysis.

Consider the problems facing a teacher in making up a multiple-choice test. The test should consist of good items, but what can be used as criteria for "goodness?" One characteristic of a good test item is that it discriminates among students who know the material and those who do not. The teacher would hope that those who do well on the overall test would tend to get the item right, and those who do poorly should tend to get that specific item wrong. Our correlation coefficient gives us a measure of how well an item does this.

Table 12.8 presents our ten students again, their scores on a final examination in English, and indicates whether or not they got multiple-choice item #13 correct. These were last year's scores. The teacher wants to determine if the same item should be used again on this year's test. A quick look at the table suggests that some of the students who scored highest overall missed the item. We can see if there really is a negative correlation between overall scores and performance on #13 by computing $r_{\text{pb}}$:

$$r_{\text{pb}} = \frac{\overline{X}_p - \overline{X}_q}{s_X}\sqrt{pq}$$

where $\overline{X}_p$ = the mean exam score for those passing the item

$$= \frac{63 + 76 + 69 + 70}{4} = 69.50$$

$\overline{X}_q$ = mean exam score for those failing the item

**Table 12.8** *Data for Computation of the Point Biserial r Coefficient*

| STUDENT | SCORE ON ENGLISH FINAL EXAM | PERFORMANCE ON TEST ITEM # 13 |
|---|---|---|
| Bob | 86 | Wrong |
| Carol | 71 | Wrong |
| Ted | 63 | Correct |
| Alice | 79 | Wrong |
| Cleo | 76 | Correct |
| Julius | 82 | Wrong |
| Mark | 69 | Correct |
| Anthony | 62 | Wrong |
| Mickey | 70 | Correct |
| Minnie | 97 | Wrong |

*Some statisticians argue that another correlation coefficient, the biserial r, is more appropriate than the point biserial r for evaluating test items this way. Their view is that the dichotomous correct/incorrect variable really represents a continuous underlying variable. (See, for instance, Q. McNemar, *Psychological Statistics*, 4th ed. [New York: John Wiley & Sons, 1969].)

$$= \frac{86 + 71 + 79 + 82 + 62 + 97}{6} = 79.50$$

$s_X$ = the standard deviation of the exam scores
$= 10.29$

$p$ = the proportion of people in the group passing the item

$$= \frac{4}{10} = 0.40$$

$q$ = the proportion of people in the group failing the item

$$= \frac{6}{10} = 0.60$$

From these data, $r_{pb}$ is easily computed as:

$$r_{pb} = \frac{69.50 - 79.50}{10.29}\sqrt{(0.60)(0.40)} = (-0.97)(0.4899) = -0.48$$

Thus, we have a moderate negative correlation between overall performance (the continuous variable) and performance on item 13.* This indicates that the better students tended to miss the item, and the poorer students tended to get it right. Such evidence would be grounds for not using the item again, or else modifying it to be a better discriminator.

Incidentally, $r_{pb}$ does not behave exactly like $r$ and rho, in that it is not always possible to get values of +1.0 or –1.0, even when, say, a test item is a perfect discriminator. The values +1.0 and –1.0 can appear only when $p$ and $q$ each equal 0.50.

## Evaluating Tests with Correlation Coefficients

Anyone who has received a low grade in school, failed a driver's license exam, or not obtained a desired job may have spent a few moments reflecting on the quality of the tests or examinations that led to those consequences. How do we know that tests measure what they are supposed to measure; that the test scores do not reflect errors in measurement?

These questions should interest everyone who lives in a society as test-dependent as ours. Tests determine who gets into college, into graduate school, into the army; who receives professional licenses and certifications; who gets certain jobs; and so on. It seems that for every important event in life there is a test lurking nearby.

Test construction is a very subtle and by no means certain business. In the course of an hour, a teacher cannot ask the student everything he or she is supposed to have learned during the year and so must select only a few questions from a very long list of potential questions. Similarly, an employer may want to know if a job applicant will perform successfully on the job, but the only way to determine this with certainty is to hire the applicant for a period of time. Instead, the employer administers a test that requires the applicant to demonstrate only a few skills. In either case, simply making up a list of questions does not ensure that the test will accomplish its purpose—measuring academic achievement or predicting job performance. Tests must be evaluated before they can be trusted, and correlational methods are especially useful for doing this.

There are many aspects of a test that should be evaluated if it is to be used in important situations or given to many people. Two of these are the test's **reliability** and its **validity**.

A test is reliable if it is *consistent*. Although we take many tests only once, we want a test that would produce the same score every time if we took it more than once. We do not want a test score that is influenced by the weather, by what the person taking the test had for breakfast, or by the fact that the test taker is not feeling well at the time he takes the test. In short, we do not want a test score to reflect fluctuating, extraneous variables.

The most direct approach to assessing the reliability of a test is to administer the test to a group of people, wait a period of time, and then readminister the same test to the same people. The correlation coefficient between scores from the first and second test sessions is known as the **reliability coefficient**, and the method described is known as the **test-retest method** of measuring reliability. If the test is highly reliable, that is, if scores do not reflect chance factors to a very high degree, then the obtained $r$ will be high.

Obviously, the test-retest method of measuring reliability will not be suitable for some forms of tests, such as, for example, intelligence tests. IQ is known to change with age, so if the waiting period between the first test and the retest is very long, it is likely that the characteristic being measured will change and that the obtained reliability of the test will be spuriously low. On the other hand, if the waiting period between test sessions is short, and if exactly the same questions are asked each time, then it is possible that answers in the retest will be contaminated by the practice received on the same questions a short time ago. Thus, the test-retest method of measuring reliability can be used only where it is certain that both test sessions produce unbiased measures of the tested characteristic.

Two commonly used methods of getting around the problems inherent in the test-retest method are the **split-half method** and the **comparable forms method**. These techniques also use correlation coefficients as coefficients of reliability.

In the split-half method the test is administered to a group of people only once. Unbeknownst to the people taking the test, the questions are divided into two groups (say, one group consists of even-numbered questions and the other group consists of odd-numbered questions). The correlation obtained between the two groups of test items is taken as a measure of the test's reliability.[8]

When the test has been administered to many people, and when the test constructor has good statistics on people's responses to many different questions, it may be possible to construct two comparable forms of the same test. The two forms may contain different questions, but they must be constructed so that any person will achieve the same score using either form. This requires large amounts of statistical information on the questions used and the performances of different kinds of people, and thus comparable forms are more likely to be issued by large commercial testing organizations than by an individual classroom teacher. Nonetheless, when comparable forms are available, a measure of reliability can be obtained by administering *both* forms to the same group of individuals. Each person will thus have a score on both forms, and the between obtained scores on both halves will constitute a measure of reliability.

Many commercially made tests are furnished with an instruction book that contains reliability information; usually this includes a description of the method used to obtain reliability coefficients and of individuals involved as subjects. In any case, the coefficients provide a precise answer to the questions, "How precise is this test" and "How large an error is typical for this test?"

An issue completely unrelated to the reliability of the test is expressed by the question, "Does the test measure what it is supposed to measure?" A test can be perfectly reliable—that is, a person could receive the same score every time he takes it—but still be invalid. It might not measure what it was intended to measure at all. This is known as the problem of construct validity or, more simply, **validity**. A test is valid if test scores reflect the underlying characteristic that they are supposed to. If the test is supposed to predict performance on a specific job, then it is valid only to the extent that it actually predicts that performance.

Correlation can also be used to measure **validity** if some independent criterion or measure is also available. Suppose, for example, that an aptitude test is supposed to measure ability of people to perform on a factory job. The extent to which the test is valid, that is, the extent to which it really measures that ability, can be assessed by correlating aptitude-test scores with some independent criterion of job performance. Similarly, college entrance exams are supposed to measure a student's aptitude for college-level work. But do they? A way to test the validity of entrance exams would be to compute the correlation coefficient between entrance-exam scores and grade-point averages obtained 4 years later. Correlation coefficients used in this way are called **validity coefficients**. Correlation coefficients are thus highly useful in evaluating tests.[9] Look for them on booklets that come with the tests or in the articles written about various tests.

## SUMMARY

Every person or object can be measured on an almost unlimited number of variables. Some variables seem to be related so that knowing an individual's status on one allows us to predict his status on another. This relation is called correlation, and the methods discussed in this chapter are designed to detect correlation and measure it.

Correlation may have two directions: positive or negative. When two variables are positively correlated, an increase in the value of one tends to be accompanied by an increase in the value of the other, and vice versa. A negative relation exists when an increase in the value of one tends to be accompanied by a decrease in the value of the other, and vice versa. Thus, people's heights and weights are positively correlated, since taller people tend to have higher weights than shorter people. A negative correlation might be represented by the relation between prevailing mortgage interest rates and the number of new homes built in a given year; when the value of one variable is high, the other tends to be low.

Correlation may also be characterized by degree. This refers to the extent to which observed values adhere to the designated relation. The highest degree of relation is a reflect relation, in which the value of one variable determines exactly the value of the other. At the other extreme is the lowest degree of correlation, a zero relation, in which values on one variable are completely unrelated to values on the other. Most relations studied in the behavioral sciences fall between perfect and zero in degree.

Some feeling for the direction and degree of a relation may be obtained by visually inspecting the scatter-plot diagram. To make a scatter plot, paired observations on two variables are graphed. Usually one variable is arbitrarily designated *X* and the other *Y*. Such paired observations can be represented as points in a two-dimensional graph if there is a *Y* value corresponding to every *X* value. The direction of the slope of scatter-plot points indicates the direction of the relation; the extent to which scatter-plot points fall into a narrow line of deviate from such a line indicates the degree of the relation.

Correlation coefficients are statistics designed to specify quantitatively the direction and degree of a relation. Many of these have been developed, but all share some characteristics. The value of most correlation coefficients will be between −1.00 and +1.00. A positive value indicates a positive relation, a negative value indicates a negative relation. The size or absolute value of the coefficients describes the degree of relation.

The Pearson product-moment correlation coefficient, or the Pearson *r*, is one of the most frequently used correlation coefficients. It is defined as the mean cross-product of the *z*-scores for pairs of observations. One should not compute the Pearson *r*, however, unless conditions of homoscedasticity of variances and a linear relation between variables are met. Also, the Pearson *r* is appropriate only for interval-level data or higher.

When observations on one or both variables are at the ordinal level, the Spearman rank-difference correlation coefficient, or the Spearman rho, may be used. It is computed from the differences in each individual's ranks on the variables and is easier to calculate than the Pearson *r*.

Precise evaluation of the obtained correlation coefficients requires more sophisticated procedures than can be covered here. However, we can roughly categorize different values as "high," "moderate," "low," or "negligible." The key to rough evaluation is to square the coefficient. The squared *r* indicates the proportion of one variable's variation accounted for by the other variable's variation.

The final correlation coefficient introduced in this chapter is the point biserial correlation coefficient, or $r_{pb}$.This statistic gives you an index of the relationship between a continuous variable and a dichotomous variable; for instance, between the overall scores people obtain on an examination and their performance (right or wrong) on a single test item.

A high correlation between two variables in no way proves that a cause-and-effect relation exists between them. It is entirely possible that a third factor causes both variables to vary together.

One of the most useful applications of correlation coefficients is the area of test evaluation. Correlation measures can be used to indicate whether a test is consistent, that is, reliable, and whether it really measures what it is supposed to measure, that is, whether or not it is valid.

## KEY CONCEPTS

- correlated variables
- correlational analysis
- direction of relation—positive or negative
- degree of correlation
- perfect relation
- scatter-plot diagram
- paired observations
- correlation coefficients
- Pearson product-moment correlation coefficient (Pearson $r$)
- linear relation
- homoscedasticity
- covariation
- Spearman rank-difference correlation coefficient (Spearman rho)
- split-half method
- comparable forms method
- point biserial correlation coefficient
- correlation versus cause and effect
- reliability
- validity
- test-retest method

## PROBLEMS

1. Indicate whether the following pairs of variables would most likely exhibit positive correlations or negative correlations:

   a. Outdoor temperature and the number of people at the beach.
   b. Body weights of adult men and the amount of time it takes each man to run 1,500 meters.
   c. IQ scores of high-school students and their grade-point averages.
   d. Room temperature and the amount of time it takes an ice cube to melt.
   e. Home prices and the annual incomes of the people who buy them.

2. Workers in a factory are rated annually by their supervisors in two categories. Quality of workmanship is rated on a 5-point scale, with a score of 5 being the highest possible, and a score of 0 being the lowest possible. Productivity is rated on a point scale with possible scores ranging from 0 (no productivity) to 1,000 (highest possible productivity). Scores for eight workers are given below:

| WORKER | QUALITY RATING | PRODUCTIVITY RATING |
|---|---|---|
| 1 | 2.1 | 411 |

| | | |
|---|---|---|
| 2 | 3.7 | 510 |
| 3 | 3.6 | 430 |
| 4 | 1.0 | 250 |
| 5 | 1.9 | 240 |
| 6 | 3.7 | 637 |
| 7 | 3.9 | 637 |
| 8 | 3.9 | 712 |

a. Prepare and interpret a scatter plot diagram for these data.
b. Compute and interpret the Pearson r between scores on the two rating scales.

3. Twenty-five students from a small high school's graduating class take a college entrance exam, and all are admitted to the state university. After two years, the university registrar prepares a list of their current grade-point averages (GPA) and their entrance exam scores:

| STUDENT | EXAM SCORE | GPA | STUDENT | EXAM SCORE | GPA | STUDENT | EXAM SCORE | GPA |
|---|---|---|---|---|---|---|---|---|
| 1 | 515 | 2.49 | 9 | 760 | 3.13 | 18 | 740 | 3.39 |
| 2 | 640 | 3.45 | 10 | 680 | 2.90 | 19 | 620 | 2.62 |
| 3 | 315 | 2.95 | 11 | 300 | 2.02 | 20 | 157 | 1.78 |
| 4 | 300 | 2.41 | 12 | 580 | 2.85 | 21 | 463 | 2.94 |
| 5 | 812 | 3.42 | 13 | 395 | 2.50 | 22 | 189 | 2.26 |
| 6 | 241 | 2.73 | 14 | 200 | 2.51 | 23 | 550 | 3.08 |
| 7 | 665 | 3.18 | 15 | 760 | 3.68 | 24 | 405 | 2.76 |
| 8 | 158 | 2.05 | 16 | 480 | 3.22 | 25 | 420 | 2.48 |
| | | | 17 | 521 | 2.50 | | | |

Prepare a scatter plot for these data and, from it, indicate whether or not you think entrance exam scores and college GPAs are correlated for these students. Indicate the direction of the relationship and characterize the degree as either zero, slight, moderate, or extremely high.

4. What is the value of the Pearson *r* if
   a. *ZXZY* = –230 and *N* = 1,000
   b. *ZXZY* = –92 and *N* = 110
   c. The covariation is 421, the *X*-variation is 371, and the *Y*-variation is 483.
   d. The co variation is –7.31, the *X*-variation is 9.20 and the *Y*-variation is 8.37.

5. Compute the Pearson *r* for the data in problem 3 above using each of the three methods presented in this chapter. (It might be interesting to record the amount of time required for each of the three calculations; the computational formula method should be faster than the other methods.)

6. Compute and interpret the Spearman rho for the data given in problem 2.

7. Demonstrate how restricting the range can change the value of $r$. Do this by computing $r$ (use anyone computational procedure you prefer) only for the people who scored between 400 and 600 on the exam data given in problem 3 above. If one is concerned only with this middle group, do the exam scores tell very much about what a person's GPA is likely to be?

8. Prepare a scatter plot for each of the following sets of paired observations:

a.

| $x$ | $y$ |
|---|---|
| 107 | 3.94 |
| 109 | 5.68 |
| 120 | 9.75 |
| 87 | 6.12 |
| 116 | 7.40 |
| 75 | 3.71 |

b.

| $x$ | $y$ |
|---|---|
| 12 | 650 |
| 7 | 460 |
| 9 | 640 |
| 3 | 350 |
| 1 | 300 |
| 4 | 520 |

c.

| $x$ | $y$ |
|---|---|
| 120 | 93 |
| 156 | 75 |
| 102 | 122 |
| 119 | 76 |
| 137 | 138 |
| 104 | 104 |

d. From your scatter plots, which of these shows the lowest degree of correlation?

9. Compute the Pearson $r$ for each set of paired observations in problem 8 above.

10. Listed below are three items of financial information about ten people: their annual incomes, the value of the automobiles they own, and the amount of money each spends on entertainment each week:

| INDIVIDUAL | ANNUAL INCOME | WEEKLY ENTERTAINMENT EXPENSES | AUTOMOBILE VALUE |
|---|---|---|---|
| 1 | $22,000 | $59 | $8,000 |
| 2 | 14,000 | 30 | 5,500 |
| 3 | 21,000 | 50 | 7,000 |
| 4 | 26,500 | 68 | 10,000 |
| 5 | 16,000 | 36 | 6,800 |
| 6 | 22,500 | 56 | 9,800 |
| 7 | 19,000 | 42 | 5,800 |
| 8 | 25,700 | 70 | 8,000 |
| 9 | 22,000 | 53 | 5,600 |
| 10 | 19,200 | 45 | 8,700 |

a. Prepare two scatter-plot diagrams, one to show the relationship between annual incomes and weekly entertainment expenses, and the other to show the relationship between annual incomes and automobile values. From your scatter plots, make a verbal comparison of the two scatter plots.

b. Compute the Pearson *r* associated with each of the scatter plots you prepared for part a. Interpret the two obtained *r*s in your own words.

11. Are students' performances in algebra related to their overall grade-point averages? Are students' grade-point averages related to their art grades? Six students happened to be taking both algebra and art one semester. Their averages and grades in these courses are presented in the table. Compute the Spearman rho to examine the relation between grade-point averages and algebra grades; compute also the rho between grade-point averages and art grades. Use the letter grades to rank-order each student's performance in the two courses. Interpret each correlation coefficient in your own words.

| | OVERALL GRADE-POINT AVERAGE | GRADE IN ALGEBRA | GRADE IN ART |
|---|---|---|---|
| Richard | 3.65 | B+ | A |
| Elizabeth | 3.70 | C– | A |
| Natalie | 2.75 | A | C |
| Robert | 2.50 | C | C+ |
| Jennifer | 3.00 | B | B |
| Elliot | 2.10 | A+ | D |

12. Some competitive athletic events, such as diving and gymnastics, are scored by judges who arbitrarily assign numerical ratings to individual performances. Listed in the table are ratings assigned to seven divers by three judges. Consider these ratings to be ordinal-level measurement and compute the Spearman rho between each pair of judges (that is, 1 and 2, 2 and 3, 1 and 3). Which pair of judges are closest in agreement? Which pair is least in agreement?

| | | RATINGS | |
|---|---|---|---|
| DIVER | JUDGE 1 | JUDGE 2 | JUDGE 3 |
| Dan | 7.9 | 8.2 | 9.8 |
| Steve | 8.9 | 9.8 | 7.8 |
| Gordon | 9.8 | 8.7 | 8.1 |
| Marty | 6.1 | 7.7 | 7.7 |
| Terry | 7.5 | 8.1 | 7.9 |
| Jim | 7.4 | 7.9 | 8.0 |
| Randy | 7.5 | 7.8 | 9.1 |

13. Do college students who own cars get different grades from students who don't? Answer this question after computing the point biserial correlation coefficient for the following data:

| STUDENT | GPA | OWN CAR? | STUDENT | GPA | OWN CAR? |
|---|---|---|---|---|---|
| 1 | 3.50 | Yes | 6 | 2.10 | Yes |
| 2 | 3.40 | Yes | 7 | 2.60 | Yes |
| 3 | 3.80 | No | 8 | 3.00 | No |
| 4 | 2.70 | Yes | 9 | 3.10 | No |
| 5 | 1.90 | No | 10 | 2.50 | Yes |

14. The final exam in a German class included a multiple-choice section. The table below shows the overall exam scores for ten students and their individual pelformances on three of the multiple-choice items. In order to evaluate the degree to which each of these items discriminated between high scorers and low scorers, compute $r_{pb}$ between overall scores and each item. Which item discriminates best? Which item has a negative $r_{pb}$ and should probably not be used the next time the test is given?

| | | PERFORMANCE | | |
|---|---|---|---|---|
| STUDENT | OVERALL SCORE | ITEM 7 | ITEM 9 | ITEM 14 |
| Bruce | 84 | Right | Right | Wrong |
| Gail | 81 | Right | Wrong | Wrong |
| Sandy | 73 | Wrong | Wrong | Right |
| Ken | 58 | Wrong | Wrong | Right |
| Eva | 68 | Right | Right | Wrong |
| Jean | 98 | Right | Right | Right |
| Rody | 93 | Right | Wrong | Wrong |
| David | 75 | Wrong | Right | Wrong |
| Cynthia | 62 | Wrong | Wrong | Right |
| Mary Ann | 47 | Wrong | Right | Right |

15. Explain in your own words why a high degree of correlation between two variables cannot, by itself, be taken as evidence that changes in one variable cause (or produce) changes in the other.

## NOTES

1. See, for example, R. Pearl, "Tobacco smoking and longevity," *Science* 87 (1938): 216-217
2. Many studies designed to examine this question have been reviewed by L. Erlenmeyer-Kimling and L. F. Jarvik, "Genetics and Intelligence: A Review," *Science* 142 (1963): 1477-1479.
3. This information will be comforting to some and disturbing to others. You may wish to read more on the subject in D. McClelland, "Testing for Competence Rather than for Intelligence," *American Psychologist* 28 (1973): 1-14.
4. More formally, the relation between $X$ and $Y$ is linear if it is of the form $Y = a + bX$, where $a$ and $b$ are constants.

5. For an introduction to correlation with nonlinear relations, see J. P. Guilford and B. Fruchter. *Fundamental Statistics in Psychology and Education*, 5th ed. (New York: McGraw-Hill 1973), pp. 285-293.
6. The size of $N$ and the magnitude of $r$ allow us to make a decision about the significance of $r$. Significance is a technical term used in inferential statistics (Chapters 8-15), and it need not concern us as long as we restrict ourselves to descriptive statistics.
7. It should be noted that a different result would probably have appeared if the instructor randomly selected seven students (rather than picking the seven highest). Random selection is covered in Chapter 9, but we can say here that it means selecting seven people in such a way as to ensure that each of the original twenty has an equal chance of being selected. With random selection, one is not arbitrarily restricting one variable's range.

8 If one simply computes the value of $r$ between scores obtained on two halves of the test, $r$ will be spuriously low. (Its value is influenced by the number of observations. The number of observations in each group of items is only half that of the full test.) Using the split-half technique, the Spearman-Brown prophecy formula is applied:

$$\text{Reliability coefficient} = \frac{2r}{1 + r}$$

where $r$ = Pearson $r$ between scores on two halves of test.

9. Reliability, validity, and other statistical concepts in testing are covered in A. Anastasi, *Psychological Testing*, 4th ed. (New York: Macmillan, 1976).
10. Darrell Huff, *How to Lie with Statistics* (New York: W. W. Norton, 1954), Chapter 8.
11. Huff suggests that older women were raised during a time when toeing out was encouraged; young women today are encouraged to walk with a different posture.
12. W. A. Wallis and H. V. Roberts, *The Nature of Statistics* (New York: Free Press, 1962), p. 108.
13. Ibid., p. 108.
14. Huff, *How to Lie With Statistics*, p. 90.
15. From P. Milvey, "Getting to the Heart," *Runner's World*, 12 (April 1977): 27-31. Re~ printed with permission from *Runner's World*, Mountain View, California.
16. M. Rodin and B. Rodin, "Student Evaluation of Teachers," *Science* 177 (1972): 11641166.
17. Ibid., p. 1166.
18. L. Giambra. "Mathematical Background and Grade-Point Average as Predictors of Course Grade in an Undergraduate Behavioral Statistics Course," *American Psychologist* 25 (1970): 366-367.