TOPIC 49 THE PEARSON CORRELATION COEFFICIENT

When we want to examine the relationship between two quantitative sets of scores (at the interval or ratio levels; see Topic 41), we compute a correlation coefficient. The most widely used coefficient is the **Pearson product-moment correlation coefficient**, whose symbol is r (usually called the **Pearson *r***).

Consider the scores in Table 1. As you can see, the employment test scores put participants in *roughly* the same order as the ratings by supervisors. In other words, those who have high employment test scores (e.g., Joe and Jane) have high supervisors' ratings, *and* those who have low test scores (e.g., John and Jake) have low supervisors' ratings. This illustrates what we mean by a **direct relationship** (also called a **positive relationship**).

Table 1
Direct Relationship; r = .89

Employee	Employment Test Scores	Supervisors' Ratings
Joe	35	9
Jane	32	10
Bob	29	8
June	27	8
Leslie	25	7
Homer	22	8
Milly	21	6
Jake	18	4
John	15	5

Notice that the relationship in Table 1 is not perfect. For example, although Joe has a higher employment test score than Jane, Jane has a higher supervisors' rating than Joe. If the relationship were perfect, the value of the Pearson r would be 1.00. Being less than perfect, its actual value is .89. As you can see in Figure 1, this value indicates a strong, direct relationship.

In an **inverse relationship** (also called a **negative relationship**), those who are high on one variable are low on the other. Such a relationship exists between the two sets of scores in Table 2. Individuals who are high on self-concept (such as Joe and Jane) are low on depression while those who are low on self-concept (such as Jake and John) are high on depression. However, the relationship is not perfect. The value of the Pearson r for the relationship in Table 2 is –.86.

Table 2
Inverse Relationship; r = –.86

Employee	Self-Concept Scores	Depression Scores
Joe	10	2
Jane	8	1
Bob	9	0
June	7	5
Leslie	7	6
Homer	6	8
Milly	4	8
Jake	1	9
John	0	9

The relationships in Tables 1 and 2 are strong because they are near 1.00 and –1.00, but in each case, there are exceptions, which make the Pearson *r*s less than 1.00 and –.100. As the number and size of the exceptions increase, the values of the Pearson r become closer to 0.00. A value of 0.00 indicates the complete absence of a relationship. (See Figure 1 below.)

It is important to note that a Pearson r is *not* a proportion and *cannot* be multiplied by 100 to get a percentage. For example, a Pearson r of .50 does not correspond to 50% of any characteristic of the data. To think about correlation in terms of percentages, we must convert Pearson *r*s to another statistic called the **coefficient of determination**, whose symbol is r^2, which indicates how to compute it: simply square r. Thus, for an r of .50, r^2 equals .25. If we multiply .25 by 100, we get 25%. What does this mean? Simply this: A Pearson r of .50 is 25% better than a Pearson r of 0.00. Table 3 on the next page shows selected values of r, r^2, and the percentages you should think about when interpreting a value of r.[1]

–1.00	INVERSE RELATIONSHIP			0.00	DIRECT RELATIONSHIP			1.00
⇧	⇧	⇧	⇧	⇧	⇧	⇧	⇧	⇧
perfect	strong	moderate	weak	none	weak	moderate	strong	perfect

Figure 1. Values of the Pearson *r*.

[1] Note that the procedure for computing a Pearson r is beyond the scope of this book.

Table 3
Selected Values of r and r^2

r	r^2	Percentage better than zero[1]
.90	.81	81%
.50	.25	25%
.25	.06	6%
–.25	.06	6%
–.50	.25	25%
–.90	.81	81%

[1]Also called *percentage of variance accounted for* or *percentage of explained variance.*

EXERCISE ON TOPIC 49

1. "Pearson *r*" stands for what words?

2. When the relationship between two variables is perfect and inverse, what is the value of *r*?

3. Is it possible for a negative relationship to be strong?

4. Is an *r* of –.90 stronger than an *r* of .50?

5. Is a relationship "direct" *or* "inverse" when those with high scores on one variable have high scores on the other *and* those with low scores on one variable have low scores on the other?

6. What does an *r* of 1.00 indicate?

7. For a Pearson *r* of .60, what is the value of the coefficient of determination?

8. What do we do to a coefficient of determination to get a percentage?

9. A Pearson *r* of .70 is what percentage better than a Pearson *r* of 0.00?

Question for Discussion

10. Name two variables between which you would expect to get a strong, positive value of *r*.

For Students Who Are Planning Research

11. Will you be reporting Pearson *r*s? If so, name the two variables that will be correlated for each value of *r*.