

12.5: Interpretation of r-Values

The symbol for the results of a correlation test is r . The result is known as an r -value, a *correlation coefficient*, or simply as the obtained value for the test. This is easy to remember because r is used to summarize the *relationship* between two variables. The names of variables, or placeholders for their names, are often shown as subscript. Thus, r_{XY} simply indicates that a correlation is being computed between one variable, referred to as X , and another, which is referred to as Y . We can see this reflected in the symbol-formatted hypotheses in the prior section. If the names of the variables are known, those can be used instead of these generic placeholders. For example, $r_{\text{sleep_stress}}$ could be used when a correlation is being computed between hours of sleep and level of stress. An underscore or hyphen can be used to distinguish between the two variable names in the subscript for ease of comprehension.

The r -value provides two important summaries about the relationship between two variables. It indicates:

1. The direction of the relationship and
2. The strength of the relationship.

Let's take a moment to review each of these and how they are interpreted from the r -value.

Direction of a Correlation

Correlations summarize linear relationships. When a relationship exists between two variables, that relationship can be either positive or negative.

Positive correlations refer to patterns where relatively high values of one variable tend to co-occur with relatively high values of the other variable and relatively low values of one variable tend to co-occur with relatively low values of the other variable. Another way to say this is that scores for two variables have a tendency to increase or decrease together. Positive correlations can also be called direct correlations.

Negative correlations refer to patterns where relatively high values of one variable tend to co-occur with relatively low values of the other variable and relatively low values of one variable tend to co-occur with relatively high values of the other variable. Another way to say this is that scores for one variables have a tendency to decrease when scores from the other variable increase. Negative correlations can also be called indirect correlations or inverse correlations.

Patterns Indicated by Correlation Coefficients

Positive Correlations	Negative Correlations
XY and XY	XY and XY

Patterns Indicated by Correlation Coefficients

↑↑ and ↓↓	↑↓ and ↓↑
-----------	-----------

Let's consider a few examples to put these two possible directions into context. Consider the possible correlation that could exist between sleep and quiz scores. The sample is comprised of college students and the two variables being measured are hours of sleep and quiz scores. If there is a positive correlation between these two variables, it would mean that as sleep increased across the sample, quiz scores also tended to increase; thus, when this is the case higher scores tend to co-occur with higher scores and lower scores tend to co-occur with lower scores. When the two variables are trending in the same way, their relationship is positive.

In contrast, consider a possible correlation that could exist between caffeine consumption and sleepiness among adults. The sample would be comprised of adults and the two variables being measured would be caffeine and sleepiness, each measured quantitatively. If there is a negative correlation between these two variables, it would mean that as caffeine intake increased across the sample, sleepiness tended to decrease. In this example, higher scores tend to co-occur with lower scores and lower scores tend to co-occur with higher scores. Because they are trending in opposition to each other (in opposite directions), their relationship would be negative.

Direction Using Graphs

Data for correlations can be graphed using a scatterplot (also known as a scattergram or scatter dot graph). Data from the first quantitative variable are plotted using the x-axis and data from the second quantitative variable are plotted using the y-axis. The x-

axis runs horizontally and the y-axis runs vertically. Thus, the two scores for each case are used together as a coordinate pair to place a dot on the two-dimensional graph. Coordinate pairs are written generally as (x, y) indicating that the first value in each pair specifies location on the x-axis and the second specifies location on the y-axis.

If there is a relationship between the variables, the data should approximate the shape of a line either angling up or down from left to right. The angle of a linear graph is known as its slope. In correlation, the slope is used to determine whether a relationship is positive or negative. If the graph angles up from left to right, the slope and the corresponding correlation will be positive. In contrast, if the graph angles down from left to right, the slope and corresponding correlation will be negative.

Let's take a look using an example. Suppose a researcher collected data from 10 college students about their hours of sleep and scores on a quiz to test the hypothesis that sleep would positively relate to quiz scores. The scores are shown in Data Set 12.1 below. The first column is not a test variable. Instead, it shows anonymous names (or identification numbers) for each case in the sample. These are sometimes used in research to help connect scores across variables for participants. Each row represents a different participant (i.e. case). The second column shows the data for sleep hours and the third column shows the data for quiz scores.

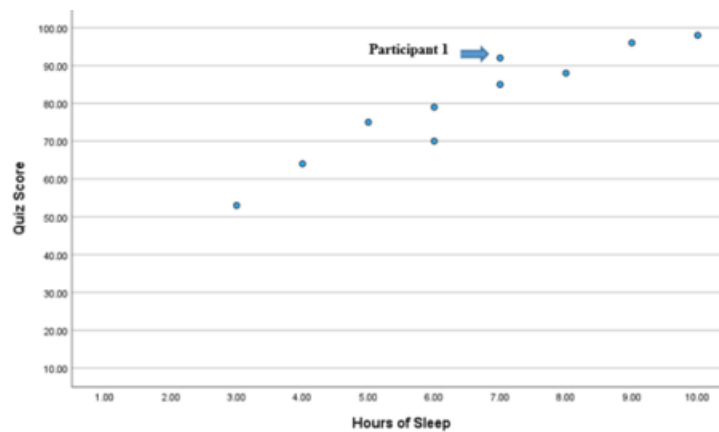
Data Set 12.1. Hours of Sleep and Quiz Scores ($n = 10$)

Participant Number	Sleep Hours	Quiz Score
1	7	92
2	8	88
3	9	96
4	6	70
5	6	79
6	4	64
7	5	75
8	10	98
9	3	53
10	7	85

Before we graph the data, we can use a simple method to get a sense of the possible direction of the correlation. To do so, we identify the higher and lower scores in each variable to see what the pattern is between the variables, if any. The five highest hours of sleep are 10, 9, 8, 7 and 7 and the lowest are 3, 4, 5, 6 and 6. The highest five quiz scores are 98, 96, 92, 88 and 85 and the lowest are 53, 64, 70, 75, and 79. In Data Set 12.1, the five highest scores for the X variable occurred in the same cases as the five highest scores for the Y -variable. Consistent with this, the lowest five scores for the X -variable occurred in the same cases as the five lowest scores for the Y -variable. Thus, a positive pattern is apparent between the variables so we should expect to see a positive slope when graphed.

We can also see this pattern by graphing. Below is a scatterplot for Data Set 12.1. Each dot represents a case or participant. For example Participant 1 had an x -value of 7 and a y -value of 92 giving them a coordinate pair of $(7, 92)$. You can see their location marked on the graph as an example. The location is over 7 along the x-axis and up 92 along the y-axis. The same process can be used to identify each participant on the graph.

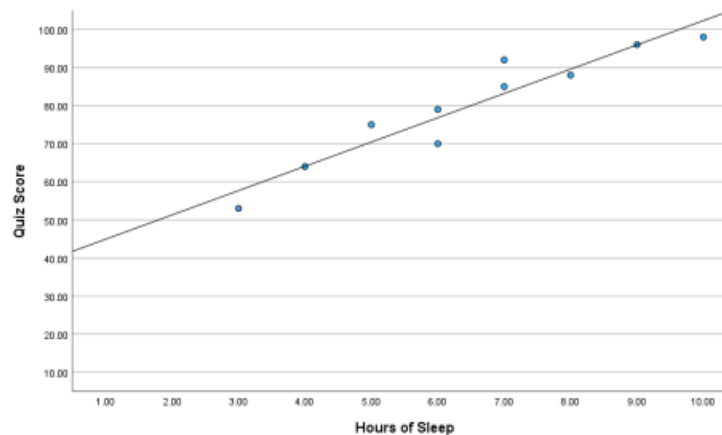
If two variables are correlated, we should see the lines approximating a straight line either sloping up or down from left to right. When we look at the pattern of the dots (which represent the data) in Graph 12.1, we can see that they are approximately following the path of a straight line angling up.



Graph 12.1. Hours of Sleep and Quiz Scores ($n = 10$)

To know whether the correlation is positive or negative we can visually assess the slope. We can see that the line is angling up from left to right, which means that the slope is positive. This is because slope refers to how much the graph rises along the y-axis relative to how much it increases as it runs (i.e. moves) across the x-axis. This is often summarized by saying “slope equals rise over run.” Because we read the x-axis from left to right, the run is increasing (meaning it has a positive value). When we read the graph from left to right, the slope is the change in the line relative to the y-axis. If the dots are going up from left to right, the relationship is positive. If the dots were going down from left to right, the relationship would be negative.

To make the slope clearer, a fit line is often added to a graph. A fit line can also be called a line of best fit or slope line in correlation (in regression this same line is called a regression line). A fit line is balanced to approximate the location of the dots with as little error as possible. Below is the scatterplot for Data Set 12.1 with a fit line added. Notice that we can clearly see that the line is angling up (or rising) from left to right and that the correlation is, therefore, positive.



Graph 12.2. Hours of Sleep and Quiz Scores with Fit Line

Magnitude of a Correlation

Correlations have a magnitude, also known as a strength. Thus, magnitude indicates how strongly the two variables are related to each other. The clearer and more consistent the pattern between the variables, the stronger their correlation. The strength is estimated using r values and can be understood visually using scatterplots.

Magnitude Using r

The strength of a correlation is summarized using the number of an r -value without consideration for the sign of the value (because the sign is the direction and the value is the strength). All correlation strengths are between .00 and 1.00. They cannot exceed these boundaries, in absolute value. When there is no correlation between two variables, the correlation strength is .00. When there is a moderate correlation between two variables, the correlation strength will be at or around .50. When there is a perfect correlation

between two variables, which is the strongest a correlation can be, the correlation strength will be 1.00. Thus, the closer r is to .00, the weaker it is and the closer r is to 1.00, the stronger it is.

It can be useful to interpret and describe the strength of a correlation. Though there are no definite rules on how the strength of a specific r -value must be described, there are general guidelines that can be used. These provide approximate ranges and how their strengths can be described, but we must keep in mind that these are guidance and that the most appropriate wording can depend upon the implications of the correlation or other considerations in research and theory relative to the variables. Thus, the cutoffs and terms used here should be considered useful but not obligatory or absolute in their meaning.

Table 12.1. Guidelines for Interpreting the Strength of an r -Value

r -value	Description of Strength
1.00	Perfect
~ .80 to .99	Very Strong
~ .60 to .79	Strong
~ .40 to .59	Moderate
~ .20 to .39	Weak
~ .00 to .19	No Correlation to Very Weak

Remember, strength is determined by the magnitude of the result, regardless of its direction. Here are some examples of r -values, their magnitudes, and the description of their strengths following the guidelines:

r -value	Magnitude	Description of Strength
.93	.93	Very Strong
-.27	-.27	Weak
-1.00	1.00	Perfect
.46	.46	Moderate
.11	.11	Very Weak
-.11	.11	Very Weak
.68	.68	Strong
.00	.00	No Correlation

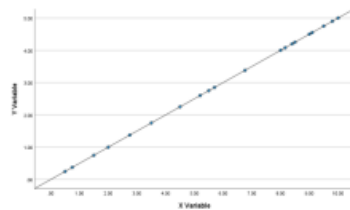
Magnitude Using Graphs

The magnitude can also be roughly estimated by looking at the nature of the graph, though this does take some practice. However, reviewing graphs can help us get a sense of the pattern between the variables and a better understanding of what a correlation is computing. The values used for the x-axis, the y-axis, and the sample size can all impact the way a scatterplot looks, making it hard to compare graphs which vary on any or all of these three things. Using different directions of the slope (positive compared to negative) can also make comparisons a bit more challenging. Therefore, we will review several graphs with different magnitudes while keeping the axes, direction (positive), and sample sizes consistent to allow us to more easily focus on the magnitudes.

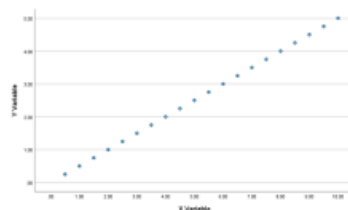
We will look at graphs starting from a perfect correlation and gradually move to no correlation so you can observe how magnitude is reflected visually. Each dot represents a case (or participant). The more closely the dots approximate a straight line, the stronger the relationship is. This means that when a fit line is used, dots are closer to the line when the correlation is stronger. When the dots are more spread out such that they do not tend toward forming a line, the correlation is weaker. This means that when a fit line is used, dots are further from the line when the correlation is weaker.

Thus, strength can be assessed visually by looking at how close the dots are to forming a perfectly straight line that angles up or down from left to right. Take a look at Graphs 12.3 and 12.4 below. Both of these have a perfect correlation. In Graph 12.3, we see a perfectly uniform line angling up from left to right. Even without a fit line added, the straight line is clear. A fit line has been

added to Graph 12.4 to illustrate that it is also perfectly straight, despite the data points being unevenly spaced along the line. Both of these are perfect, positive correlations.



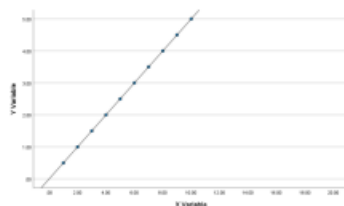
Graph 12.3. Perfect, Positive Correlation
($r = 1.00$)



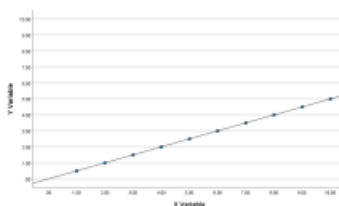
Graph 12.4. Perfect, Positive Correlation
($r = 1.00$)

It is important to note that a common mistake people make is assuming the steepness of the line represents the magnitude; however, steepness can be a bit misleading. This is because the steepness of the line will look different dependent upon the anchors used to label the x-axis and y-axis. Thus, it is possible to make the same data look steeper or flatter by changing the height of the y-axis or the width of the x-axis to compress or stretch the look of the graph. Therefore, we will focus on how close the dots are to the line rather than how steep or flat the line looks when visually assessing magnitude.

Let's take a moment to look at how the steepness can be misleading. The two examples below show how Graph 12.3 would look if the x-axis anchors were changed and if the y-axis anchors were changed. Notice that the line seems steeper in one and less steep in the other but that the locations represented by each dot and how straight and perfect the lines are have not changed. All three versions of Graphs 12.3 were made using the same data set and all have the same correlation coefficient of $r = 1.00$. Thus, the steepness of the slope alone is not a great indicator of strength. Instead, strength is assessed visually by looking at how closely the dots approximate a line and checking that the line is not completely horizontal. As long as a line is not completely flat (meaning as long as it is not completely parallel to the x-axis), it is possible for the two variables to be related.



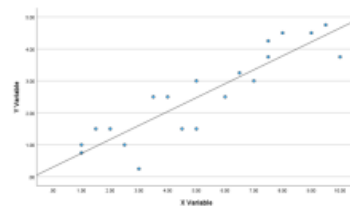
Graph 12.3 with Altered X-Axis
($r = 1.00$)



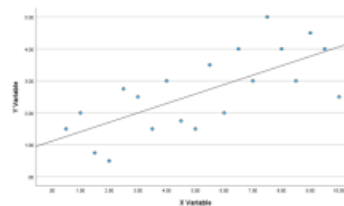
Graph 12.3 with Altered Y-Axis
($r = 1.00$)

To get a sense of how to estimate the strength of a correlation based on a graph, we will look at several scatterplots with the same sample size but with varying magnitudes. Because the visual steepness of the line can be impacted by changes to the anchors on the axes (as we just saw), the same values are used for both the x- and y-axes for each comparison graph.

Very strong correlations have dots that are relatively close to the line, but without all dots falling in a perfectly straight line. Strong correlations have dots that still clearly trend in a line but with noticeably more distance between the dots and the line, on average. Note that it is the average distance between the dots and the line that matters. In keeping, notice that, overall, the dots are closer to the line and that the correlation coefficient (r -value) is stronger (greater) in Graph 12.5 than in Graph 12.6.

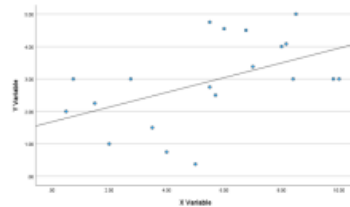


Graph 12.6. Very Strong, Positive Correlation
($r = .90$)

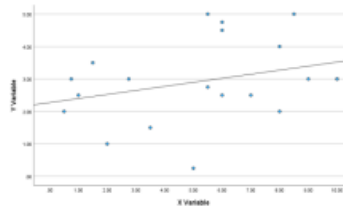


Graph 12.6. Strong, Positive Correlation
($r = .70$)

As the data, on average, are farther from approximating a straight line, the correlation is weaker. Notice that as we move in order from Graphs 12.5 to 12.8, the data are spread farther and farther from the fit line. The less the data approximate a straight line angling either consistently up or consistently down, the weaker the correlation is.

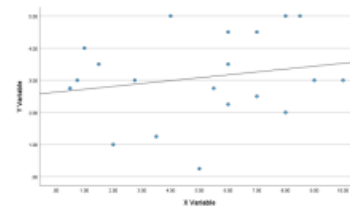


Graph 12.7. Moderate, Positive Correlation
($r = .50$)

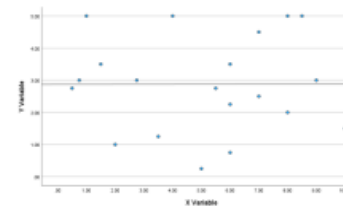


Graph 12.8. Weak, Positive Correlation
($r = .30$)

By the time we get to Graph 12.9, it would be very hard to even see whether the data are generally following a line without the fit line there to guide the eye, because the data are only very loosely following a positive sloping pattern. Graph 12.9 has a very weak correlation. In Graph 12.10 there is no relationship present. The data appear scattered about with no clear slope which they tended to follow. Thus, when there is no relationship between the variables, the fit line lies perfectly horizontal, parallel with the x-axis.



Graph 12.9. Very Weak, Positive Correlation
($r = .10$)



Graph 12.10. No Correlation
($r = .00$)

Interpreting r

Graphs provide us the ability to see the correlation visually whereas correlation coefficients (r values) summarize both the direction and strength of those correlations. These are two ways to represent the relationships between two quantitative variables. However, the brilliant construction of the correlation coefficient formula makes interpretations easy and allows for the easy comparison of different correlations. Thus, correlations are usually interpreted simply by looking at their r -values.

Let's take a look at some examples. Correlations of .75 and -.75 have the same magnitude but different directions. When $r = .75$, the correlation is strong and positive. When $r = -.75$, the correlation is strong and negative. A correlation of $r = -.93$ is very strong and negative. A correlation of $r = .28$ is weak and positive. A correlation of $r = .06$ is very weak and positive. We can imagine the approximate look of the graph, including whether it slopes up or down and how close the dots are to forming a line, simply by knowing the r -value.

Limitations

A major limitation of the correlation is that it cannot be used to determine cause-effect relationships. Just because two things are mathematically related, does not mean that either is the cause of the other. Therefore, though tempting, it is usually inappropriate to use causal language when interpreting the results of correlation (see Chapter 8 for a review of causal language).

Reading Review 12.1

1. What assumptions must be met before using a Pearson's Product Moment Correlation (PPMC)?
2. What is a non-directional hypothesis that could be tested using a Pearson's Product Moment Correlation (PPMC)?
3. What is the symbol used for correlation coefficients?
4. How would the strength and direction of each of the following correlation coefficients be described, using the general rules of thumb for strengths: -1.00, -.44, .59, .12, .86, -.70, .70

This page titled [12.5: Interpretation of \$r\$ -Values](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by .