Chapter 7

Association Between Two Continuous Variables: The Pearson Correlation

Do you believe in UFOs? Maybe you have seen a UFO or perhaps know someone who has been abducted by aliens? Do you think there is proof that aliens have visited Earth? Rock-and-roller Sammy Hagar (from *Van Halen*, *Montrose*, and his solo work) reported recently that he was visited by aliens when he was in his late teens! Individuals who feel strongly that aliens have visited Earth also tend to read books about UFOs and aliens. And, according to research by Patry and Pelletier (2001), there is an association – a correlation – between the amount one reads about UFOs and aliens, and having strong beliefs that aliens have visited Earth.

Notice we are not suggesting that reading about UFOs and aliens causes one to believe aliens roamed Earth. Instead, we are merely pointing out that as one reads about these topics, there also appears to be a strong belief that aliens have visited us in the past. And, if individuals do not think it is likely aliens traveled here, their reading list probably does not include UFO and alien literature. There is an *association* between these two phenomena.

The focus of this chapter is on measuring the association between two variables using the Pearson Product Moment Correlation Coefficient. For example, is relationship security associated with jealousy? Is the amount of time college students spend text messaging their friends associated with scores on a measure of extroversion? Is church attendance related to ratings of happiness? Is reading about UFOs associated with the belief that aliens visited Earth?

The research situation is one where we are interested in discovering if two variables are related *and* the two variables are continuous or discrete. The Pearson correlation coefficient is a statistical tool for examining such relationships. It also tells us the strength of the relationship between the two variables. Sometimes two variables are very strongly related and sometimes

they are less strongly related. For example, family satisfaction is strongly related to general happiness; work satisfaction is also related to happiness, but the relationship is less strong.

In this chapter, we provide three examples. *Case Study 7.1* is similar to our example in the opening paragraph, but assesses the association between UFO belief and trust in authority; do individuals who have a strong belief in UFOs also tend to distrust authority? *Case Study 7.2* concerns baseball, and evaluates the run production by a team during a Little League season: Is there an association between the decline in runs scored by the "Cubs" and season progression? The third example, *Case Study 7.3*, focuses on text messaging and extroversion (which was first introduced in Chapter 4). Are higher levels of text messaging associated with higher levels of extroversion? Also in the chapter, we cover assumptions of the Pearson correlation, including normality and linearity, and integrate testing these assumptions in *Case Study 7.3*. The end of the chapter provides correlation examples using both IBM SPSS and SAS.

Before we move forward, spend a few moments with Exploration 7.1. Sometimes issues addressing association are right under our noses, but we never notice.

Exploration Task 7.1: Exploring Trends Over Time

Correlation is about movement – as phenomena increase or decrease, other phenomena increase or decrease. A fun way to illustrate this is by using Google Trends http://www.google.com/trends to evaluate patterns of movement over time. Google keeps track of everything searched on their search engine by date, and you can enter search terms to get a sense of popularity over time. For example, call up Google Trends on your computer or smartphone and type in the search word "mittens". What do you see? You should note increases for the search term "mittens" in the winter months, and decreases in summer months. Type in the baseball team "Kansas City Royals" and you will see increases during baseball season, and declines in the off season. Search for the term "Google" and you will note an almost straight-line increase in the occurrence of that search term over time. For the term "reality

television" a gradual decline is noted since 2004, while for "Kim Kardashian" a gradual increase is noted since 2006. Give these a try – or try your own search terms!

Case Study 7.1: Association Between Personal Belief in UFOs and Trust in Authority

Unidentified flying objects (UFOs) are fascinating! You may or may not believe in UFOs, but regardless they are an interesting topic for discussion. Let's say a friend of yours drops by and casually mentions he saw an odd object in the night sky. Blinking lights, hovering in the distance, then suddenly moving erratically at incredible speeds, then whoosh it was gone. He says it must have been a UFO, no doubt flown by aliens to investigate our planet. You counter that maybe it was a reflection, or an illusion, or perhaps an aircraft. The next day an article in the newspaper notes that although others saw something that night, a Government spokesperson denies anything was in the sky. You mention this to your fiend, who states, "Well, what do you expect – those folks in charge deny everything!"

Your friend has always been a big believer in UFOs. Heck, he reads a lot of books about UFOs and aliens, and you already know that such a reading list is associated with UFO beliefs! But his statement makes you further wonder about other factors that might be associated with UFO beliefs. You do a quick search of the research literature, and find that UFO beliefs correlate with issues such as fantasy proneness, rural living, and lack of religious beliefs. But your friend's statement suggests something else might be associated with his beliefs. Indeed, possibly your friend is distrustful of authority, as evidenced by his reaction to the Government spokesperson's comment.

To investigate whether UFO belief is associated with trust in authority, you decide to ask a sample of 20 college students a few simple questions about these topics. Your research hypothesis reads: H_1: *There will be a negative association between personal belief in UFO encounters and trust in authority, with those reporting stronger belief in UFOs also reporting*

lower levels of trust in authority. The null hypothesis is H_0: *Rho* = 0. We use *Rho* to write the formal null hypothesis for correlation. The null hypothesis states that the resulting correlation between UFO belief (X) and trust in authority (Y) will be zero, indicating no association.

Since your study addresses such broad topics as UFOs and trust in authority, you can select students from an introductory psychology course as study participants. Once recruited for the study, students are provided a single item measuring UFO belief, rated on an 7-point scale. They are asked how much they 1 (disagree) to 7 (agree) with the statement, "I believe UFOs visit our planet on a regular basis." Higher scores indicate a greater belief in UFOs. They are also given a single item to measure trust in authority. They are asked how much they 1 (disagree) to 7 (agee) with the statement, "Authority figures (such as a Government officials) are completely truthful when asked about events of public concern." Higher scores indicate greater trust in authority. Table 7.1 displays data from the 20 individuals for both questions.

Table 7.1: UFO Belief (X) and Trust in Authority (Y) data for 20 students

Participant #	UFO Belief (X)	Trust in Authority (Y)
1	5	2
2	5	3
3	5	4
4	4	4
5	4	4
6	3	3
7	3	4
8	5	2
9	4	5
10	2	4
11	4	2
12	4	4
13	7	2
14	3	5
15	2	4
16	4	3
17	1	5
18	5	4
19	6	3
20	6	3

Summary Statistics: Means, Variances, and Standard Deviations

A good place to start when conducting research is to evaluate summary statistics for the variables. The summary statistics of interest will be the mean, variance, and standard deviation. These summary statistics are listed in Table 7.2.

Mean. The mean value (*M*) is a measure of central tendency and introduced in Chapter 2. It represents the average score across all the cases. The average score for UFO Belief is 4.10, indicating our sample on average is moderately believes in UFOs. The average score on Trust in Authority is 3.50, indicating on average a moderate level of trust.

Variance. The sample variance (s^2) is a measure of the degree of dispersion with a variable, and was also covered in Chapter 2. Using data from Table 7.2, the sample variance for both variables is easily calculated. As seen in the calculations, the variances are rather small.

Variance for UFO Belief: $$s^2_X = \frac{\sum(X_i - M_X)^2}{n-1} = \frac{41.8}{19} = 2.20$$

Variance for Trust in Authority: $$s^2_Y = \frac{\sum(X_i - M_Y)^2}{n-1} = \frac{19}{19} = 1.00$$

Standard deviation. The sample standard deviation (*s*) is the square-root of the variance, and is a measure of the average spread of scores. The average spread of scores is close to 1.0 for the UFO Belief, and exactly one for Trust in Authority.

Standard Deviation for UFO Belief: $$s_X = \sqrt{s^2_X} = \sqrt{2.20} = 1.48$$

Standard Deviation for Trust in Authority: $$s_Y = \sqrt{s^2_Y} = \sqrt{1.00} = 1.00$$

Table 7.2: UFO Belief (X) and Trust in Authority (Y) sample data: Summary statistic calculations

Case #	**UFO Belief (X)**	$X - M_X$	$(X - M_X)^2$	**Trust in Authority (Y)**	$Y - M_Y$	$(Y - M_Y)^2$
1	**5**	0.90	0.81	**2**	-1.50	2.25
2	**5**	0.90	0.81	**3**	-0.50	0.25
3	**5**	0.90	0.81	**4**	0.50	0.25
4	**4**	-0.10	0.01	**4**	0.50	0.25
5	**4**	-0.10	0.01	**4**	0.50	0.25
6	**3**	-1.10	1.21	**3**	-0.50	0.25
7	**3**	-1.10	1.21	**4**	0.50	0.25
8	**5**	0.90	0.81	**2**	-1.50	2.25
9	**4**	-0.10	0.01	**5**	1.50	2.25
10	**2**	-2.10	4.41	**4**	0.50	0.25
11	**4**	-0.10	0.01	**2**	-1.50	2.25
12	**4**	-0.10	0.01	**4**	0.50	0.25
13	**7**	2.90	8.41	**2**	-1.50	2.25
14	**3**	-1.10	1.21	**5**	1.50	2.25
15	**2**	-2.10	4.41	**4**	0.50	0.25
16	**4**	-0.10	0.01	**3**	-0.50	0.25
17	**1**	-3.10	9.61	**5**	1.50	2.25
18	**5**	0.90	0.81	**4**	0.50	0.25
19	**6**	1.90	3.61	**3**	-0.50	0.25
20	**6**	1.90	3.61	**3**	-0.50	0.25
Sum	82.00		41.80	70		19
Mean (M)	4.10			3.50		
Sample Variance	2.20			1.00		
Sample Standard Deviation	1.48			1.00		

Descriptive Data Plots

Histograms. Individual plots can be made for both UFO Belief and Trust in Authority. Figures 7.1 and 7.2 illustrate these variables using histograms to get a sense of their distributions. Getting to know the shape of the variable distributions is an important part of the research process because the graphs tell us a great deal about our data.

Figure 7.1: Histogram of UFO Belief (Draft Graph)

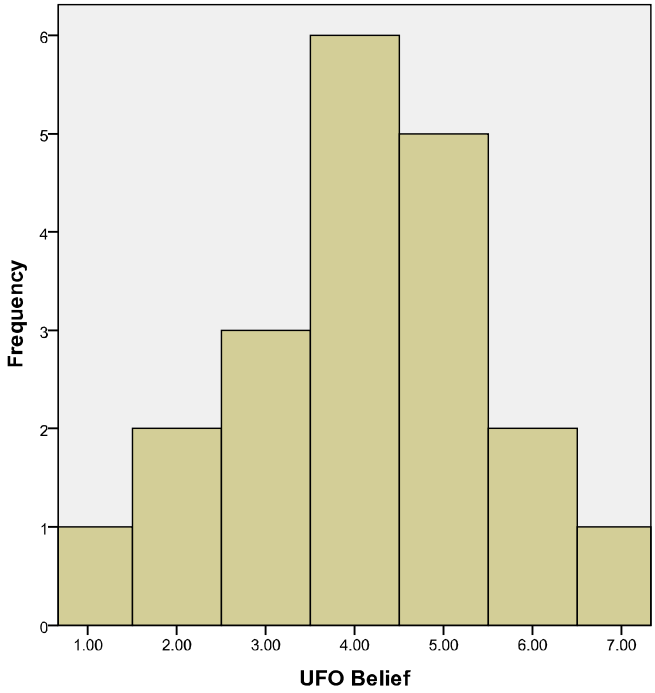
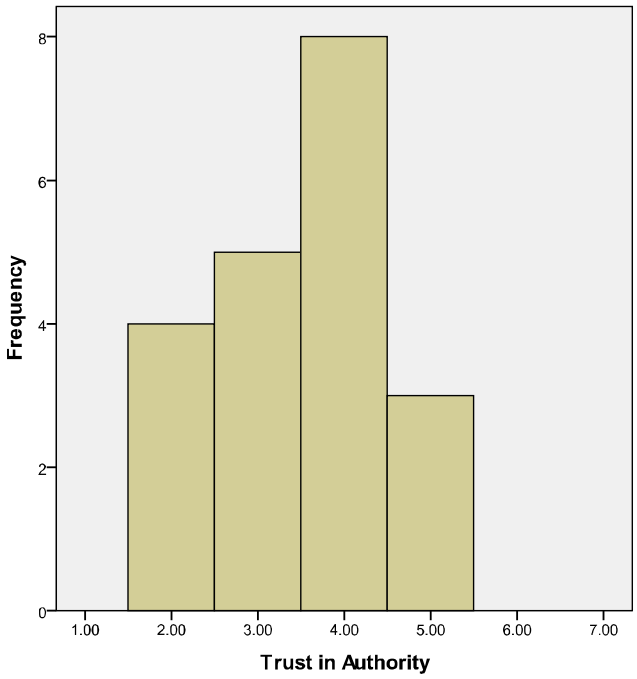


Figure 7.2: Histogram of Trust in Authority (Draft Graph)

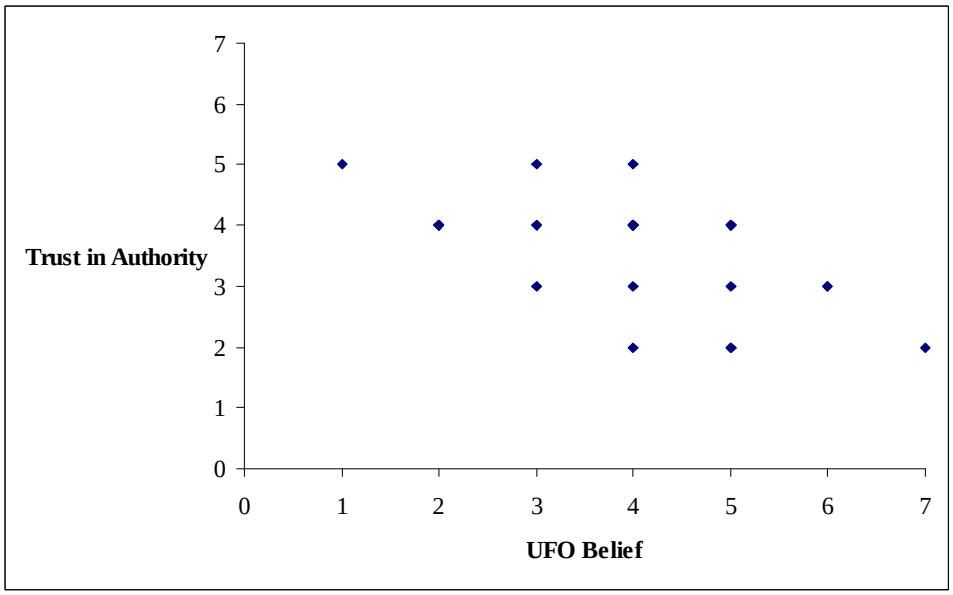


Both graphs appear to be normally distributed, which is an assumption of the data that should be met prior to calculating a Pearson correlation (we cover later in this chapter the

assumptions of the Pearson correlation). In addition, we can assess in these graphs the general spread of scores.

Bivariate scatterplots. The bivariate scatterplot is a great way to illustrate two variables simultaneously in one graph (the term "bivariate" means "two variables" – i.e., bi–variate), and was first covered in Chapter 3. Bivariate scatterplots are formed by plotting values for each variable – one along the bottom axis (known as the X axis) of the graph, and one along the left axis (known as the Y axis). Figure 7.3 is the bivariate scatterplot of the UFO Belief and Trust in Authority scores. Notice that as UFO Belief increases, Trust in Authority decreases. This indicates an association between the two variables. Although the association has not been formally tested with a test statistic, the bivariate scattterplot can be used to make a general determination of the association between the two variables.

Figure 7.3: Bivariate scatterplot of UFO Belief and Trust in Authority (Draft Graph)



Sometimes with bivariate scatterplots, it is helpful to place a straight line through the plotted values. By doing this, one can easily discern the direction of association between the two variables. Associations between two variables can be positive, negative, or no association. If the

line is in an upward direction  it suggests a **positive association** (as one variable increases, so does the other). A downward direction  suggests a **negative association** (as one variable increases, the other decreases). A line that is flat  suggests **no association** (a variable moving up or down has little or no influence on the other variable).

In looking at Figure 7.3, take a moment and draw a line through the UFO Belief and Trust in Authority data points. The line is downward  , suggesting a negative association. Although we cannot say whether this association is beyond what is expected by chance occurrence, it does suggest that as UFO Belief increases, Trust in Authority decreases.

<u>Exploration Task 7.2: Making a Scatterplot</u>

Making a scatterplot is a fun and easy way to evaluate visually how two variables move together. Ask a few of your fellow classmates these questions, then plot their responses to create a scatterplot.

Directions: Use the following scale to indicate how much you like or dislike a particular music artist:

1 2 3 4 5

Dislike Like

#1: How much do you like the music of Lady Gaga? _____ (enter a number from the scale)

#2: How much do you like the music of Britney Spears? _____ (enter a number from the scale)

With these two items, make a scatterplot to see if there is a visual relationship between liking the music of Lady Gaga and Britney Spears. Whether your small sample of classmates loves Gaga, or hates Gaga, you should start seeing a pattern as you plot the values. Plot the responses, then draw a straight line that best "fits" the plotted responses. Does there appear to be a positive  or negative  *association between the responses based on the scatterplot? What is the visual association between liking the music of Lady Gaga and Britney Spears?*

The Pearson Correlation Coefficient

We just completed examining our two variables of interest – UFO Belief and Trust in Authority -- through descriptive statistics and visually through histograms and bivariate scatterplots. We can now formally assess the null hypothesis regarding the association between these two variables. To make such an assessment, a Pearson correlation coefficient will be used.

The formal name for the Pearson Correlation is the *Pearson Product Moment Correlation Coefficient*, and was named after Karl Pearson who developed the measure. The Pearson Correlation evaluates how variables *move* together, and assesses whether this joint movement exceeds what is expected by chance occurrence. The formula for the Pearson Correlation is noted below in Formula 7.1.

$$r = \frac{\sum (X_i - M_X)(Y_i - M_Y)}{\sqrt{\sum (X_i - M_X)^2 \sum (Y_i - M_Y)^2}}$$ *Formula 7.1: Pearson Correlation Coefficient*

The *r* is the symbol for the correlation value. The Σ sign stands for summation of terms. *X* will be a score on one of the variables. *Y* will be a score on the other variable. The subscript "*i*" refers to row numbers or participant numbers (e.g., participant 1, participant 2, etc.). M_X and M_Y are the mean values for the *X* and *Y* variables.

An alternative formula using population standard deviations is noted in Formula 7.2.

$$r = \frac{\frac{\sum XY}{N} - (M_X)(M_Y)}{(\sigma_X)(\sigma_Y)}$$ *Formula 7.2: Correlation formula using means and population standard deviations*

Here, X and Y represent the scores on variables, and M_X and M_Y are the mean values. In addition, the population standard deviations are used in the denominator. These were introduced

in Chapter 2, and use N instead of N-1 in the variance formula, then converted to the population standard deviation via square root. *XY* indicates that scores on X and Y are multiplied together.

Formulas 7.1 and 7.2 provide exactly the same measure of association. **We think Formula 7.1 better illustrates how the correlation coefficient works, and it is this formula we primarily use in the chapter.** However, we do illustrate Formula 7.2 for Case Study 7.3 at the end of the chapter.

Correlation coefficients can range in value from -1.0 to +1.0. Values cannot exceed -1 or 1, and if they do, then a calculation mistake has occurred. For example, a calculated *r* value of 1.72 indicates a calculation error has occurred. Calculated *r* values close to zero indicate little or no association between two variables, and *r* values further away from zero in either direction indicate a stronger association. For example, an *r* value of -0.50 is interpreted as a stronger association than an *r* value of 0.33 (-0.50 is further away from zero than 0.33). A positive correlation value indicates that movement across both variables is positive; as X increases, Y increases  (or as X decreases, Y also decreases). However, if the sign is negative for a Pearson correlation, that indicates the variables are moving in opposite directions; as X increases, Y decreases  (or as X decreases, Y increases).

Exploration Task 7.3: Exploring Scatterplot Shapes Based on Correlation

What might a scatterplot look like for a correlation of -0.30? How about a scatterplot for a correlation of 0.60? The Consortium for the Advancement of Undergraduate Statistics Education provides a program to explore what these might look like. On your computer or smartphone, enter http://tinyurl.com/43oa99w and you will be provided with a scatterplot screen. Hit "play movie" and watch scatterplots form based on correlations starting from -1.0 to +1.0. Or, enter your own correlation value and see a scatterplot based on hypothetical data.

Numerator

The numerator of the Pearson correlation in Formula 7.1 is a measure of covariance or covariation between two variables. The covariance indicates how the variables covary, move, or change together. In other words, the numerator measures the amount of variation X and Y have in common. Covariance is an unbounded measure of association, while the Pearson correlation is a bounded measure of association ranging from -1.0 to 1.0.

For each variable, mean deviations for X and Y are calculated and multiplied together. The covariance will be larger if these deviation products are large and smaller if deviation products are small. Also, whether the resulting covariance (and eventual Pearson correlation) is positive or negative is driven by these deviation products. If the deviation products are mostly negative, then a negative covariance is derived. However, if the deviation products are mostly positive, then a positive covariance is produced.

The numerator for Formula 7.1 is derived by the following steps:

Steps in calculating the Numerator (i.e., covariance) of the Pearson Correlation

1) Take the difference of each case from its mean value on X and Y separately, creating a mean deviation
2) Multiply these deviations, creating a cross-product
3) Sum the cross-products – it is the sum of the cross-products that form the covariance

Denominator

The denominator of Formula 7.1 for the Pearson correlation is used to standardize the covariance. Covariances are dependent on the scale or value range of the variables, and are an unbounded measure of association. This makes the covariance hard to interpret. By standardizing the covariance, the resulting statistic – the Pearson correlation – is more interpretable and is given a bounded range of -1.0 to +1.0.

The denominator contains the sum of the squared deviations for scores on X and Y. These squared deviations are then multiplied and the square root is taken. Why is the square root taken? Taking the square root returns the squared deviations to their original scale of measurement. We square the deviations to remove the negative values, and then must reduce the squared deviations back to their original scale.

The steps for calculating the denominator for Formula 7.1 are as follows:

Steps in calculating the Denominator of the Pearson Correlation

1) Calculate mean deviations of each case for the variables
2) Square the mean deviations for each variable to remove the negative sign
3) Sum the squared deviations for each variable
4) Multiply the sum of the squared deviations for each variable, then take the square root of the product

The Pearson correlation is now calculated dividing the numerator by the denominator.

Assessing the Statistical Significance of the Pearson Correlation

Once a Pearson correlation value has been calculated, we assess whether the value exceeds what is expected by chance occurrence. This is done by finding a critical value at which we would accept the null hypothesis given a specific probability level. If the correlation value exceeds the critical value, then it's a significant finding and thus exceeds what is expected by chance occurrence. Note that as with any significance test, the critical value depends on sample size, alpha level, and whether it is a one- or two-tailed test.

To find a correlation critical value, we need to do three things. First, we need to adopt an alpha level – usually a .05 significance level will suffice, although more stringent alpha levels such as .01 are sometimes used. Second, we need to adopt a one- or two-tail test of significance based on how the null hypothesis is stated – usually a two-tail test is adopted given how the null hypothesis is stated. Third, we need to calculate *degrees of freedom*. Formula 7.3 is the formula

for the Pearson correlation degrees of freedom:

N – 2 = Degrees of Freedom *Formula 7.3: Degrees of Freedom*

Once these three steps have been completed, *Appendix R* is used to find the *r* critical value. In *Appendix R*, find the appropriate row using the degrees of freedom, then find the appropriate column based on the selected probability of .05 or .01, and a two-tail test. Once the column is located, the critical *r* value is found. If the Pearson correlation value exceeds the *r* critical value taken from *Appendix R*, we reject the null hypothesis and conclude that the correlation was not due to chance occurrence.

6 Steps for Determining the Significance of the Pearson Correlation

1) **Adopt a statistical significance level** (usually .05)
2) **Choose a one- or two-tailed test based on the null hypothesis**
3) **Calculate the degrees of freedom using *N* - 2**
4) **Use *Appendix R* to find the *r* critical value**. Note that Appendix R lists the critical *r* values as positive, but these values would also apply to negative associations (for example, a critical *r* value in Appendix R of 0.576 also reflects the value of -0.576).
5) **Ask whether the Pearson correlation value calculated for the study exceeds the critical *r* value**. If the study *r* value exceeds the critical *r* value – be it either negative or positive – then the finding is statistically significant, which allows for rejection of the null hypothesis. The resulting association between the two variables was not due to chance occurrence. If the study *r* value does not exceed the critical r value, then the finding is not statistically significant and the null hypothesis is accepted.
6) **Interpret the results**. Besides reporting whether the resulting correlation is significant or not, interpreting results should also include the direction of the finding

(or confirmation of the null hypothesis). For example, stating that UFO belief and trust in authority have a significant association is only a partial interpretation. An appropriate interpretation would include whether the association is positive or negative, and in a "statement format" indicating what the association means.

An alternative way to assess the significance of the Pearson correlation is to use a statistical program such as IBM SPSS or SAS to calculate the exact probability of the Pearson correlation given the null hypothesis is true. If the exact probability is less than .05, then we have a significant finding; the probability that the results are due to chance is less than 5%.

Effect strength

How strong is the resulting association? Once a Pearson correlation is derived and statistically assessed, the correlation may be further interpreted as a measure of effect strength. As covered in Chapter 5 on effect sizes, correlations of +/- .10 are small effects, meaning the variables have minimal influence on each other. Correlations of +/- .30 indicate a medium effect, and values +/- .50 indicate a large effect.

Along with the significance value of the Pearson correlation, noting the effect size is important because statistical significance is partially dependent on study sample size. With an extremely large sample (say $N = 5000$), a Pearson correlation of +/- .10 would be significant, but the effect strength would be small. For some studies, small but significant effects can be important, but for others a small effect may be considered unimportant. For example, a small effect size in a drug vaccine study is important because the drug prevents some individuals from contracting a virus. However, a small effect size in a study assessing an expensive program to increase the attention spans of college students might be deemed unimportant since such a small influence would not justify the cost of the program.

Applying the Pearson Correlation to Case Study 7.1: Association Between UFO Belief and Trust in Authority

Let's formally apply the Pearson Correlation to assess the association between UFO Belief and Trust in Authority. Data are presented in Table 7.3 with Formula 7.1 calculation components for the Pearson correlation.

Table 7.3: UFO Belief (X) and Trust in Authority (Y) sample data: Summary statistics and correlation calculations

Case #	**UFO Belief (X)**	$X - M_X$	$(X - M_X)^2$	**Trust in Authority (Y)**	$Y - M_Y$	$(Y - M_Y)^2$	**Cross Product of Mean Deviations**
1	**5**	0.90	0.81	**2**	-1.50	2.25	**-1.35**
2	**5**	0.90	0.81	**3**	-0.50	0.25	**-0.45**
3	**5**	0.90	0.81	**4**	0.50	0.25	**0.45**
4	**4**	-0.10	0.01	**4**	0.50	0.25	**-0.05**
5	**4**	-0.10	0.01	**4**	0.50	0.25	**-0.05**
6	**3**	-1.10	1.21	**3**	-0.50	0.25	**0.55**
7	**3**	-1.10	1.21	**4**	0.50	0.25	**-0.55**
8	**5**	0.90	0.81	**2**	-1.50	2.25	**-1.35**
9	**4**	-0.10	0.01	**5**	1.50	2.25	**-0.15**
10	**2**	-2.10	4.41	**4**	0.50	0.25	**-1.05**
11	**4**	-0.10	0.01	**2**	-1.50	2.25	**0.15**
12	**4**	-0.10	0.01	**4**	0.50	0.25	**-0.05**
13	**7**	2.90	8.41	**2**	-1.50	2.25	**-4.35**
14	**3**	-1.10	1.21	**5**	1.50	2.25	**-1.65**
15	**2**	-2.10	4.41	**4**	0.50	0.25	**-1.05**
16	**4**	-0.10	0.01	**3**	-0.50	0.25	**0.05**
17	**1**	-3.10	9.61	**5**	1.50	2.25	**-4.65**
18	**5**	0.90	0.81	**4**	0.50	0.25	**0.45**
19	**6**	1.90	3.61	**3**	-0.50	0.25	**-0.95**
20	**6**	1.90	3.61	**3**	-0.50	0.25	**-0.95**
Sum	82.00		41.80	70		19	**-17**
Mean (M)	4.10			3.50			
Sample Variance	2.20			1.00			
Sample Standard Deviation	1.48			1.00			

<u>Covariance</u> is the sum of the mean deviation cross products: (-1.35+-.045+….+-0.95+-0.95) = -17.
<u>Correlation</u> is the standardized covariance: -17/ sqrt(41.80*19) = -17/28.18 = -0.60.

Numerator

For the numerator of the Pearson correlation for Formula 7.1, we follow these steps:

Steps in calculating the Numerator of the Pearson Correlation

1) **Take the difference of each case from its mean value on X and Y separately (Mean Deviations).** These deviations are noted in the table. For Case #1, the mean deviation for UFO Belief is 0.90, and for Trust in Authority the mean deviation -1.50. For Case #2, the mean deviation for UFO Belief is 0.90, and for Trust in Authority the deviation is -0.50.

2) **Multiply these deviations, creating a cross-product.** Table 7.3 has these values. For example, the cross-product for Case #1 is -1.35: 0.90 x -1.50 = -1.35. For Case #2, the cross-product is -0.45.

3) **Sum the cross-products, which becomes the covariance.** Summing all the cross-products between UFO Belief and Trust in Authority gives us a covariance value of - 17. This is the numerator for the Pearson correlation formula.

Denominator

Next, calculate the denominator of the Pearson correlation table.

Steps in calculating the Denominator of the Pearson Correlation

1) Calculate mean deviations of each case for the variables. This is noted in Table 7.3 and was just illustrated for the numerator calculations.

2) Square the mean deviations for each variable to remove the negative sign. Squaring the deviations removes the negative signs. For Case #1, the square of the deviation for UFO Belief (0.90) is 0.81: $0.90^2 = 0.81$. For Trust in Authority, the square of the deviation (-1.50) is 2.25: $-1.50^2 = 2.25$.

3) **Sum the squared deviations for each variable.** Summing the squared deviations for UFO Belief yields a value of 41.80. For Trust in Authority, the sum of the squared deviations is 19.

4) **Multiply the sum of the squared deviations for each variable, then take the square root of the product.** Multiplying the sum of the squared deviations together for UFO Belief and Trust in Authority (41.80 x 19), we get a value of 794.2. Next, we take the square root of 794.2, which gives us a denominator for the correlation formula of 28.18: $\sqrt{794.2} = 28.18$.

The Pearson correlation can now be calculated between UFO Belief and Trust in Authority. The numerator of -17 is divided by the denominator of 28.18. This produces an *r* value of -0.60: -17/28.18 = -0.60. The correlation between UFO Belief and Trust in Authority is -0.60.

Steps *Pearson Correlation Formula 7.1 for the Association between UFO Belief and Trust in Authority*

1 $$r = \frac{\sum (X_i - M_X)(Y_i - M_Y)}{\sqrt{\sum (X_i - M_X)^2 \sum (Y_i - M_Y)^2}}$$

2 $$r = \frac{-17}{\sqrt{(41.8)(19)}}$$

3 $$r = \frac{-17}{\sqrt{794.2}}$$

4 $$r = \frac{-17}{28.18} = -0.60$$

The Pearson correlation is -0.60. This indicates a negative association between UFO Belief and Trust in Authority; higher levels of UFO belief is associated with lower levels of trust in authority  .

Assessing Statistical Significance of Case Study 7.1

Now that we know the correlation value is -0.60, we need to find a critical *r* value for comparison to determine statistical significance. This is done by applying the 6 steps for assessing significance:

Steps for Assessing the Significance of the Pearson Correlation

1) **Adopt a statistical significance level**
   - For the current example, a .05 level of statistical significance will be used. We are willing to accept a less than 5% chance that our test results could have occurred by chance.
2) **Choose a one- or two-tailed test based on the null hypothesis**
   - The null hypothesis states there will be no association between the two variables (*H_0: Rho* = 0*)*. A two-tailed test will be adopted since we want to insure that any association between the two variables, whether positive or negative, will be properly evaluated.
3) **Calculate the degrees of freedom using *N* – 2**
   - For the current example, the degrees of freedom (N – 2) is 18: 20 – 2 = 18
4) **Use Appendix R to find the *r* critical value**
   - Using a two-tail test with a .05 level of statistical significance and *df* = 18, Appendix R shows a critical *r* value of -/+ 0.444. The calculated *r* value will have to be greater than +/- 0.444 to be significant.

Breakout Box 7.1

What if the resulting correlation between UFO Belief and Trust in Authority was -0.40? We would accept the null hypothesis. A correlation of -0.40 is not beyond what would be expected by chance occurrence. At a .05 level with a sample size of 20, there is a 95% probability that r will be between -.444 and .444 if the null hypothesis is true.

What if *r* from our study fell exactly on -0.444? In other words, say the calculated *r* value between UFO Belief and Trust in Authority was -0.444. Would this still be considered significant at the $p < .05$ level? The answer is no. The value of -0.444 would be considered "at" the .05 level ($p = .05$), but not smaller that .05. Therefore this result would be considered nonsignificant if we adopt a strict $p < .05$ cutoff.

5) Ask whether the Pearson correlation value calculated for the study exceeds the *r* critical value.

- Does the calculated Pearson correlation value of -0.60 exceed the critical value of +/-0.444? It does, and we reject the null hypothesis and conclude that the correlation value was not due to chance occurrence.

6) Interpret the results.

- UFO Belief has a significant association with Trust in Authority at the $p < .05$ level. The association is negative; as UFO belief increases, trust in authority decreases.

A second way to proceed is to analyze the data using a computer program such as IBM SPSS or SAS. These programs calculate the correlation value from the data and provide an exact probability. IBM SPSS shows that the exact probability of getting our obtained correlation is .005 if the null hypothesis is true (SAS provides a probability of .0049). Our probability cutoff is .05 and the exact probability based on the data is lower than .05.

Effect Strength

The correlation value of -0.60, which we now know is significant at $p < .05$, may also be evaluated in terms of effect strength. The correlation value of -0.60 is considered a large effect, indicating the two variables have a considerable amount of influence on each other.

Example Write-up

A study was conducted to examine the association between UFO belief and trust in authority. The research hypothesis is there will be a negative association between UFO belief and trust in authority, with those reporting a stronger belief in UFOs also reporting lower levels of trust in authority figures. The null hypothesis (H_0) is that there will be no association between the variables; Rho = 0. A sample of 20 Introductory Psychology students was provided two items to measure UFO belief and trust in authority. Higher scores on the items indicate stronger belief in UFOs and greater trust in authority. A Pearson correlation coefficient was used to assess the association. Using a .05 level of significance, the Pearson correlation coefficient was significant, $r(20) = -0.60$, $p < .05$, a large effect. The two variables are negatively correlated indicating stronger UFO belief is associated with lower levels of authority trust; as UFO belief increases, trust in authority decreases. We reject the null hypothesis and conclude the two variables are associated.

Case Study 7.2: Association between Runs Scored and Season Progression for a Little League Baseball Team

The second example is taken from actual data from one of our son's Little League baseball teams, and is an interesting application of the Pearson correlation. The book *Money Ball* (and recent Brad Pitt movie of the same name) underscores the use of statistics in baseball, and so why not apply statistical analyses to Little League! As with the first example, we first

present descriptive statistics and plots, then follow with calculating the Pearson correlation test statistic and assessing its statistical significance.

The focus is on a team named the "Cubs" consisting of boys and girls 6-9 years of age who played Little League during the Spring 2011 season. At these ages, run production typically tapers off through a season. This is due to a number of factors, including season fatigue ("When does soccer start? I'm bored!"), and that kids on the other teams simply get better at fielding their positions. To see if there is a significant decline in run production over the season, we can use a Pearson correlation. Our research hypothesis is that there will be a negative association between run production over time; as the season progresses, fewer runs will be scored. *The null hypothesis is H_0: Rho = 0, stating that there is no association between runs scored and season progression.*

Data are taken from a Little League baseball website documenting 14 games the Cubs played and their runs scored (see Table 7.4). Season progression was operationalized using Game Number – higher values indicate further season progression.

Table 7.4: Game Number (X) and Runs Scored by the Cubs (Y)

Game Number (X)	Runs by Cubs Each Game (Y)
1	8
2	18
3	19
4	11
5	14
6	11
7	10
8	10
9	13
10	6
11	12
12	11
13	4
14	8

Exploration Task 7.4: Variables that Correlate with Age

In the current case study, we operationalized season progression using game number, with higher values indicating the game took place later in the season. In essence, we are correlating runs scored with "time". Researchers often perform correlations using time as a variable, whether its monthly time markers, yearly, or some other marker of time. For example, age of a study participant is a widely used variable in many studies! What correlates with age? Weight gain! Income! Even the likelihood of being pulled over by the Police when driving also correlates with age as we saw in Chapter 1! Can you think of other variables that might correlate with age? Make a short list and discuss these with your classmates. By doing this task, you will get a sense of how researchers think about investigating the correlates of age and time.

Summary Statistics: Means, Variances, and Standard Deviations

We start with summary statistics to explore the variables. Components for the summary statistics are noted in Table 7.5.

Mean. The average number of runs scored for the Cubs is 11.07; the Cubs score about 11 runs per game. The mean for game number is 7.5.

Variance. The variance for runs scored by the Cubs is 16.99. For game number, the variance is 17.5.

Standard deviation. The square root of the variance is taken to produce a standard deviation of 4.12 for runs scored by the Cubs, and a standard deviation of 4.18 for game number.

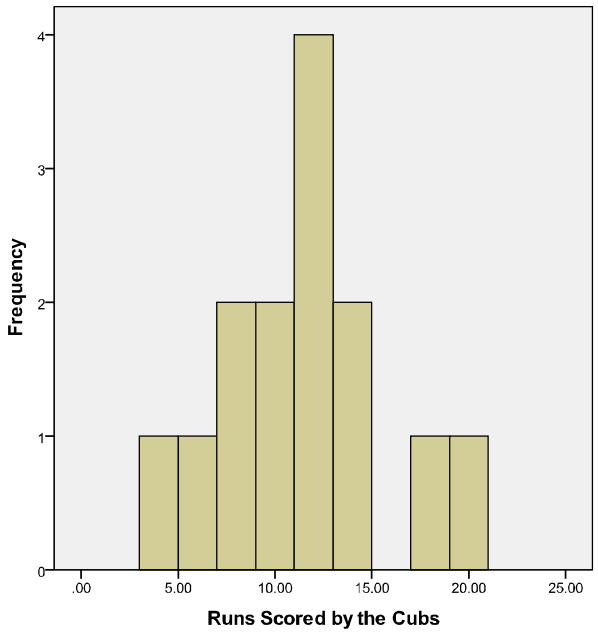
Table 7.5: Game Number (X) and Runs Scored by the Cubs (Y) data: Summary statistics

Case #	**Game Number (X)**	$X - M_X$	$(X - M_X)^2$	**Runs by Cubs Each Game (Y)**	$Y - M_Y$	$(Y - M_Y)^2$
1	**1**	-6.50	42.25	**8**	-3.07	9.43
2	**2**	-5.50	30.25	**18**	6.93	48.01
3	**3**	-4.50	20.25	**19**	7.93	62.86
4	**4**	-3.50	12.25	**11**	-0.07	0.01
5	**5**	-2.50	6.25	**14**	2.93	8.58
6	**6**	-1.50	2.25	**11**	-0.07	0.01
7	**7**	-0.50	0.25	**10**	-1.07	1.15
8	**8**	0.50	0.25	**10**	-1.07	1.15
9	**9**	1.50	2.25	**13**	1.93	3.72
10	**10**	2.50	6.25	**6**	-5.07	25.72
11	**11**	3.50	12.25	**12**	0.93	0.86
12	**12**	4.50	20.25	**11**	-0.07	0.01
13	**13**	5.50	30.25	**4**	-7.07	50.01
14	**14**	6.50	42.25	**8**	-3.07	9.43
Sum	105.00		227.50	155.00		220.93
Mean (M)	7.50			11.07		
Sample Variance	17.50			16.99		
Sample Standard Deviation	4.18			4.12		

Descriptive Data Plots

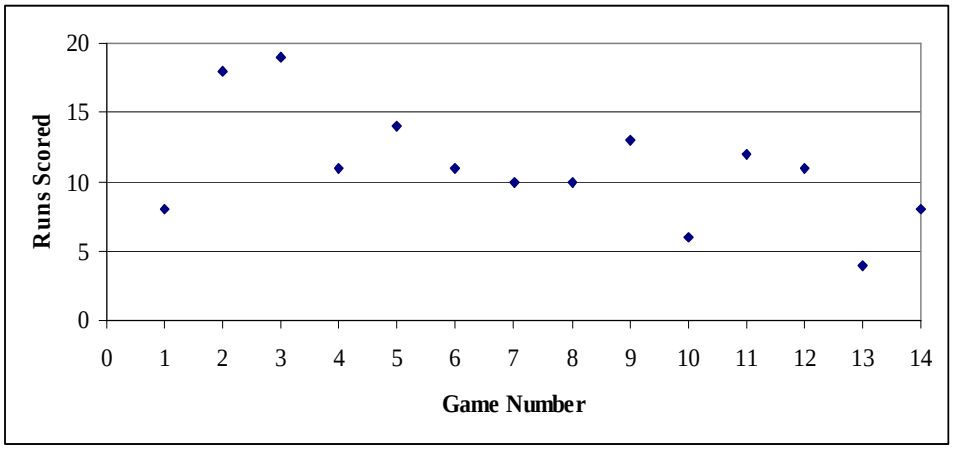
Histograms. A histogram for runs scored by the Cubs is provided in Figure 7.5. The histogram shows a bell-shaped distribution for runs scored.

Figure 7.5: Histogram of Runs Scored by the Cubs (Draft Graph)



Bivariate scatterplot. Figure 7.6 is a bivariate scatterplot of runs scored by the Cubs across the 14 games. Notice a negative downward trend in runs scored moving across the season.

Figure 7.6: Bivariate Scatterplot of Runs Scored by the Cubs across 14 games (Draft Graph)



Applying the Pearson Correlation to Case Study 7.2: Association between Runs Scored and Season Progression for a Little League Baseball Team

Let's apply the Pearson correlation to assess the association between runs scored by the Cubs and game number. Data for the calculations are presented in Table 7.6, which has the calculation components for the Pearson correlation.

Table 7.6: Game Number (X) and Runs Scored by the Cubs (Y) data: Summary statistics and correlation calculations

	Games	$X - M_X$	$(X - M_X)^2$	**Runs Scored**	$Y - M_Y$	$(Y - M_Y)^2$	**Cross Product of Mean Deviations**
	1	-6.50	42.25	**8**	-3.07	9.43	19.96
	2	-5.50	30.25	**18**	6.93	48.01	-38.11
	3	-4.50	20.25	**19**	7.93	62.86	-35.68
	4	-3.50	12.25	**11**	-0.07	0.01	0.25
	5	-2.50	6.25	**14**	2.93	8.58	-7.32
	6	-1.50	2.25	**11**	-0.07	0.01	0.11
	7	-0.50	0.25	**10**	-1.07	1.15	0.54
	8	0.50	0.25	**10**	-1.07	1.15	-0.54
	9	1.50	2.25	**13**	1.93	3.72	2.89
	10	2.50	6.25	**6**	-5.07	25.72	-12.68
	11	3.50	12.25	**12**	0.93	0.86	3.25
	12	4.50	20.25	**11**	-0.07	0.01	-0.32
	13	5.50	30.25	**4**	-7.07	50.01	-38.89
	14	6.50	42.25	**8**	-3.07	9.43	-19.96
Sum	105		227.50	155		220.93	-126.50
Mean (M)	7.50			11.07			
Sample Variance	17.50			16.99			
Sample Standard Deviation	4.18			4.12			

<u>Covariance</u> is the sum of the mean deviation cross products: (19.96+ -38.11+….+ -38.89+ -19.96) = -126.50.
<u>Correlation</u> is the standardized covariance: -126.50/sqrt(227.5*220.93) = -126.50/224.19 = -0.56.

Numerator

For the numerator of the Pearson correlation, we follow these steps:

Steps in calculating the Numerator of the Pearson Correlation

1) **Take the difference of each case from its mean value on X and Y separately (Mean Deviations).** These deviations are noted in the table. For Case #1, the mean deviation for game number is -6.50, and for runs scored the mean deviation is -3.07. For Case #2, the mean deviation for game number is -5.50, and for runs scored the deviation is 6.93.

2) **Multiply these deviations, creating a cross-product.** Table 7.4 has these values. For example, the cross-product for Case #1 is 19.96: -6.50 x -3.07 = 19.96. For Case #2, the cross-product is -38.11.

2) **Sum the cross-products, which is the covariance.** Summing the cross-products between Text Messaging and Extroversion gives a covariance value of -126.50. This is the numerator for the Pearson correlation formula.

Denominator

We next calculate the denominator of the Pearson correlation.

Steps in calculating the Denominator of the Pearson Correlation

1) **Calculate mean deviations of each case for the variables.** This is noted in Table 7.4 and was just illustrated for the numerator calculations.

2) **Square the mean deviations for each variable to remove the negative sign.** Squaring the deviations removes the negative signs. For Case #1 on game number, the square of the deviation (-6.50) is 42.25: $-6.50^2 = 42.25$. For Case #2, the square of the deviation (-5.50) is 30.25: $-5.50^2 = 30.25$.

3) **Sum the squared deviations for each variable.** Summing the squared deviations for game number yields a value of 227.50. For runs scored, the sum of the squared deviations is 220.93.

4) **Multiply the sum of the squared deviations for each variable, then take the square root of the product.** Multiplying the sum of the squared deviations together for game number and runs scored (227.50 x 220.93), a value of 50261.58 is produced. Next, the square root of 50261.58 is taken, which produces a denominator for the correlation formula of 224.19: $\sqrt{50261.58} = 224.19$.

Now we can calculate the Pearson correlation for Season Progression and Runs Scored by the Cubs. We take the numerator of -126.50 and divide by the denominator of 224.19. This produces an *r* value of -0.56: -126.50/224.19 = -0.56. The correlation between Season Progression and Runs Scored by the Cubs is -0.56.

Steps *Pearson Correlation Formula 7.1*

1 $$r = \frac{\sum (X_i - M_X)(Y_i - M_Y)}{\sqrt{\sum (X_i - M_X)^2 \sum (Y_i - M_Y)^2}}$$

2 $$r = \frac{-126.50}{\sqrt{(227.50)(220.93)}}$$

3 $$r = \frac{-126.50}{\sqrt{50261.58}}$$

4 $$r = \frac{-126.50}{224.19} = -0.56$$

The Pearson correlation is -0.56. This indicates there is a negative association between Season Progression and Runs Scored by the Cubs; as the season progresses, the Cubs score fewer runs per game  .

Assessing the Statistical Significance of Case Study 7.2

Now that we know the correlation value is -0.56, we need to assess whether it is statistically significant.

Steps for Determining the Statistical Significance of the Pearson Correlation

1) **Adopt a statistical significance level**

   - For the current example, a .05 level of significant is chosen, indicating we are willing to accept a less than 5% chance that our test results could have occurred by chance.

2) **Choose a one- or two-tailed test based on how the null hypothesis is stated**

   - Here, the null hypothesis is that there is no association between the variables. Thus, a two-tailed test will be adopted.

3) **Calculate the degrees of freedom using *N* – 2**

   - For the current example, the degrees of freedom (N – 2) is 12: 14 – 2 = 12

4) **Use Appendix R to find the *r* critical value**

   - Using a two-tail test with a .05 level of statistical significance, Appendix R shows a critical *r* value of -/+ 0.532.

5) **Ask whether the Pearson correlation value calculated for the study exceeds the critical *r* critical value.**

   - Does the calculated Pearson correlation value of -0.56 exceed the critical value of +/- 0.532? It does, and we therefore reject the null hypothesis and conclude that the correlation value was not due to chance occurrence.

6) **Interpret the results**

- Season Progression has a significant association at the $p < .05$ level with Runs Scored by the Cubs. The association is negative, indicating that as the season progressed, the Cubs scored fewer runs.

Based on the steps above, we can conclude that there is a significant negative association between season progression and Cubs' run production. As the season progressed, there was a significant decline in runs scored by the Cubs.

Both IBM SPSS and SAS show the probability of getting the obtained correlation if the null hypothesis is true is exactly .036. Since the probability cutoff is .05, we can conclude that a correlation of -0.56 is unlikely if the null hypothesis is true. We reject the null hypothesis.

Effect Strength

The correlation value of -0.56 indicates a large effect, beyond the +/-.50 cutoff used for effect sizes based on *r* for large effects.

Example Write-up

A study was conducted to examine the association between runs scored by a Little League baseball team (the Cubs) and season progression. The research hypothesis is that there will be a negative association between number of runs scored by the Cubs and season progression; as the season progresses, the Cubs will score fewer runs per game. The null hypothesis (H_0) is that there will be no association between the two variables, Rho = 0. Data were taken from a Little League website, focusing on a AA Minor team named the Cubs consisting of boys and girls age 6-9. Runs scored across 14 games were used for the analysis, and season progression was operationalized as game number. A Pearson correlation was used to assess association. A negative association beyond the .05 significance level was noted between the two variables, r(14) = -0.56, a large effect.

We can conclude that as the season progressed, the Cubs scored fewer runs per game.

We reject the null hypothesis and conclude the two variables are associated.

Assumptions and Considerations of the Pearson Correlation

To insure the Pearson correlation is valid and reliable, data must meet a number of formal assumptions and considerations about the underlying distributions the data form. Such assumptions and considerations are assessed prior to calculating the Pearson correlation measure. Covered here will be the assumptions of linearity and normality, and additional issues including range restriction, absence of extreme bivariate values, and heterogeneous subsamples.

Linearity

The first assumption is *linearity*, which asks whether there is a straight line association between the two variables. Deviations from linearity can lead the Pearson correlation to be a misleading measure of association. It is assessed by looking at bivariate scatterplots, such as the one shown in Figure 7.3 for UFO Belief and Trust in Authority, and evaluating the shape of the scatterplot. If the scatterplot follows an approximate straight line trajectory, then the assumption is most likely met. In Figure 7.3 there is an implied straight-line association between the two variables. The type of association we wish to avoid is called curvilinear. A curvilinear association suggests a curved association between the variables. Curved associations are valid, but cannot be measured accurately using the Pearson correlation.

Normality

Normality refers to the variables having a normal or bell-shaped curve distribution  , and was covered in Chapter 2. The assumption is assessed by looking at histograms of the variables and seeing if they appear normal. If one or both of the variables are non-normal, the

resulting correlation can result in either an underestimate or overestimate of the actual association. This can lead to either a Type I or Type II error, especially in situations with a correlation that is border-line significant. In such cases, the resulting correlation should be interpreted with caution. Returning to Case Study 7.1, the histogram for UFO Belief (Figure 7.1) is most likely normal, as is the histogram for Trust in Authority (Figure 7.2).

Absence of range restriction

Range restriction refers to the actual range of values in the variables. When one of the variables has a limited range, the resulting correlation can be smaller than it should be. For example, say we wanted to correlate UFO belief with age, and we only had 18 and 19 year old students as study participants. Having such a limited range could lead to a smaller Pearson correlation than if a broader age range was available (say those 18-30 years of age). The smaller Pearson correlation is solely due to the restricted age range.

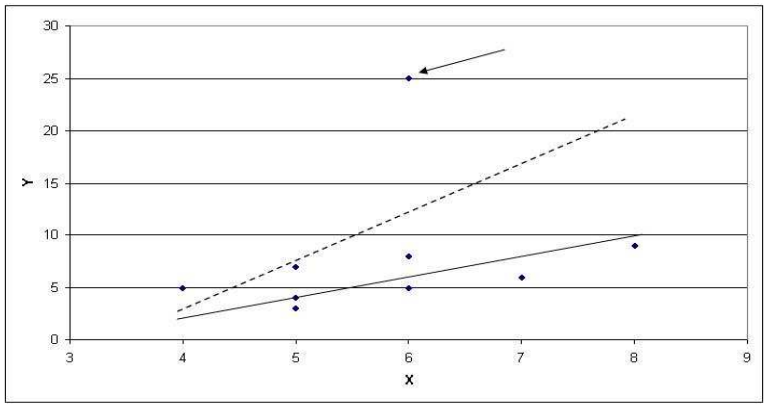
Absence of extreme bivariate values

This consideration is also called absence of bivariate outliers. Extreme bivariate values or bivariate outliers are extreme scores on both of the variables of interest. If bivariate outliers are present, the resulting correlation can overestimate or underestimate the true association depending on the location of the extreme values.

To assess this concern, create a bivariate scatterplot and look for any extreme values that stray from the other plotted values. To illustrate, take a look at the hypothetical bivariate scatterplot in Figure 7.7. An extreme value in the scatterplot has been noted with an arrow. Do you see how far this case is from the other cases? The dotted line represents the straight line or linear association between the two variables when this extreme case is included. The solid line represents the linear association between the two variables if the case is removed. If a correlation

was calculated for all the cases, the correlation would be rather high. However, if the extreme case was removed and a correlation was calculated, the correlation value would be lower.

Figure 7.7: Example of Extreme Values (arrow indicates the extreme value) – Draft Graph



Absence of heterogeneous subsamples

This addresses the possibility that the sample may really consist of two subsamples of individuals who are different. The word heterogeneous is a statistical term which means dissimilar. Since the correlation is calculated on the entire sample, the dissimilar nature of the underlying data may lead to invalid findings. The assumption is assessed by knowing the make-up of the sample and also knowing the research literature. If there are dissimilar subsamples, then separate correlation analyses should be considered for each subsample. For example, regarding UFO belief, there is research evidence suggesting that males are more likely to believe in UFOs than females. In Case Study 7.1 addressing UFO Belief and Trust in Authority, it

would make sense to perform two separate Pearson correlation analyses -- one for males and one for females -- to insure the aggregate findings are valid.

Case Study 9.3: Association between Text Messaging Behavior and Extroverted Personality Style

This third example examines the association between extroversion and text messaging. For this example, we will first present descriptive statistics and plots, and will then address the assumptions and concerns just covered for the Pearson correlation coefficient. The Pearson correlation value will then be calculated and assessed.

We discussed text messaging and extroversion briefly in Chapter 4 in regard to hypothesis testing. Individuals with an extroverted personality style tend to seek social connections, are gregarious and outgoing, and seek excitement. Those higher in extroversion are more likely to exhibit behaviors that lead to a greater set of interactions and connections between themselves and friends and family. Text messaging is a fairly new phenomenon where short messages are sent using a cellphone or smartphone.

We might expect those higher in extroversion will be more likely to send text messages as a way to exhibit their personality. Thus, our research hypothesis is that there will be a positive association between extroversion and text messaging; the more extroversion an individual reports, the more text messages they will send. *The null hypothesis is H_0: Rho = 0, stating that there is no association between the two variables.*

We'll collect data including the number of text messages college students send to their friends and family during a typical day, and their level of extroversion can be assessed using a measure of extroversion. Students are asked to note the number of text messages they had sent in the past 24 hours. They were also asked to complete a brief 10-item extroversion measure.

Individuals were asked how much they disagreed or agreed with 10 statements reflecting extroversion. An example item is "I enjoy being the life of the party." Items were scored on a scale ranging from 1 ("Disagree") to 8 ("Agree"). The items were then averaged to derive an extroversion score, with higher values indicating greater extroversion. Data from 18 college students randomly assessed from the college library are presented in Table 7.7.

Table 7.7: Extroversion (X) and Text Messaging (Y) data for 18 Students

Case #	Extroversion (X)	Text Messages (Y)
1	1	31
2	3	32
3	4	15
4	4	18
5	4	25
6	4	25
7	5	44
8	5	45
9	5	40
10	5	40
11	5	45
12	6	50
13	6	40
14	6	55
15	6	50
16	7	43
17	7	65
18	7	75

Summary Statistics: Means, Variances, and Standard Deviations

We start with summary statistics to explore the variables. Components for the summary statistics are noted in Table 7.8.

Table 7.8: Extroversion (X) and Text Messaging (Y) data: Summary statistics and calculations

Case #	**Extroversion (X)**	$X - M_X$	$(X - M_X)^2$	**Text Messaging (Y)**	$X - M_X$	$(X - M_X)^2$
1	**1**	-4	16	**31**	-10	100
2	**3**	-2	4	**32**	-9	81
3	**4**	-1	1	**15**	-26	676
4	**4**	-1	1	**18**	-23	529
5	**4**	-1	1	**25**	-16	256
6	**4**	-1	1	**25**	-16	256
7	**5**	0	0	**44**	3	9
8	**5**	0	0	**45**	4	16
9	**5**	0	0	**40**	-1	1
10	**5**	0	0	**40**	-1	1
11	**5**	0	0	**45**	4	16
12	**6**	1	1	**50**	9	81
13	**6**	1	1	**40**	-1	1
14	**6**	1	1	**55**	14	196
15	**6**	1	1	**50**	9	81
16	**7**	2	4	**43**	2	4
17	**7**	2	4	**65**	24	576
18	**7**	2	4	**75**	34	1156
Sum	90		40	738	90	4036
Mean (M)	5.00			41.00		
Sample Variance	2.35			237.41		
Sample Standard Deviation	1.53			15.41		

Mean. The average number of text messages sent in a 24-hour period is 41, and the average Extroversion score is 5.

Variance. The variance for Text Messaging is 237.41: 4036/17 = 237.41. For Extroversion, the variance is 2.35: 40/17 = 2.35.

Standard deviation. The square root of the variance is taken to produce a standard deviation of 15.41 for Text Messaging, and a standard deviation of 1.53 for Extroversion.

Descriptive Data Plots

Histograms. Individual plots can be made of both Text Messaging and Extroversion. Figures 7.8 and 7.9 illustrate these variables using histograms. The histograms show normal distributions for both Text Messaging and Extroversion. In addition, the scores on Extroversion suggest a majority of people have moderate levels of Extroversion.

Figure 7.8: Histogram of Text Messaging (Draft Graph)

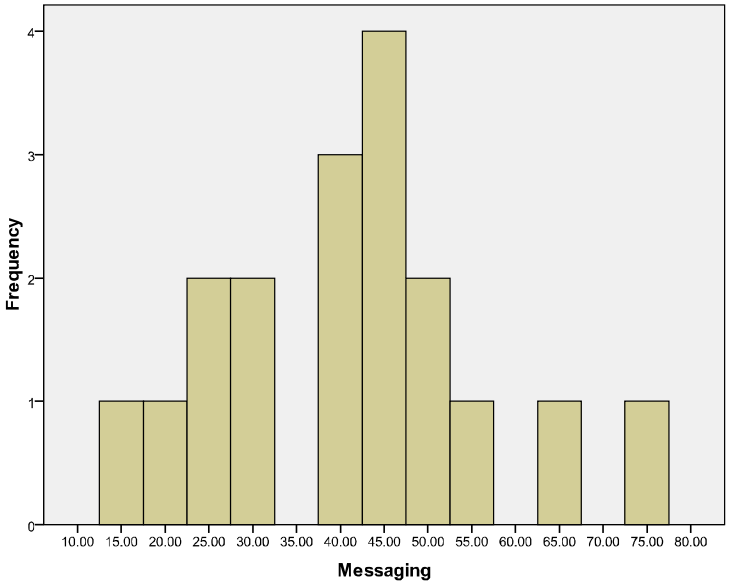
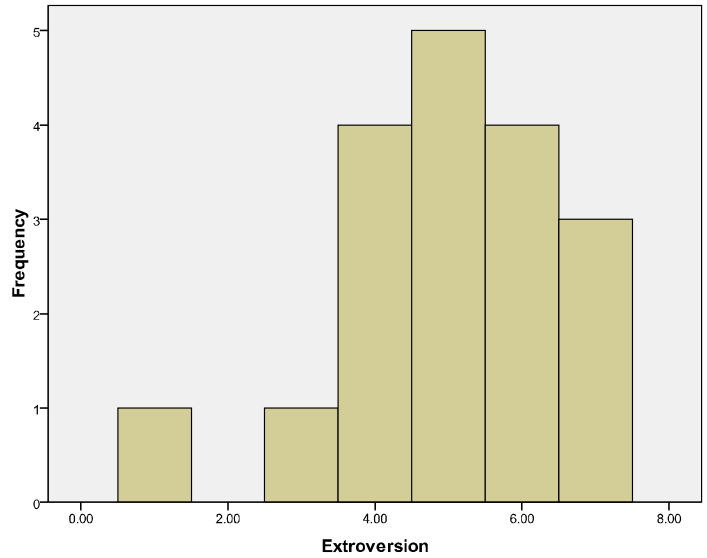
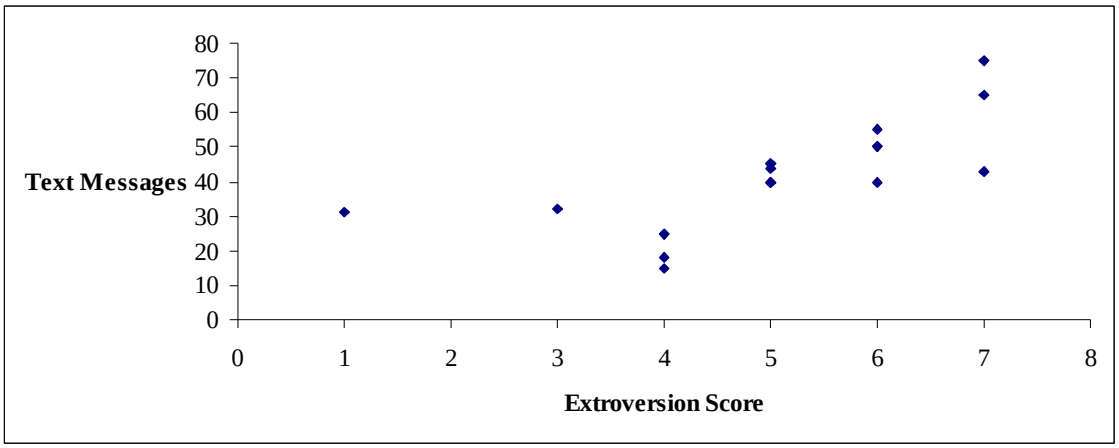


Figure 7.9: Histogram of Extroversion Scores (Draft Graph)



Bivariate scatterplot. Figure 7.10 is a bivariate scatterplot of Text Messaging and Extroversion. There is a positive association between these two variables; as extroversion increases so does text messaging.

Figure 7.10: Bivariate Scatterplot of Text Messaging and Extroversion Data (Draft Graph)



Assumptions and Considerations

Before we calculate the Pearson correlation for this example, the assumptions of linearity and normality should be assessed. To assess linearity, the bivariate scatterplot in Figure 7.10 is evaluated. Here, we ask whether the plotted values for extroversion and text messaging behavior follow a straight line association. From Figure 7.10, it may be inferred that the plotted cases follow a straight trajectory. For normality, the histograms noted in Figures 7.8 and 7.9 are evaluated. We ask whether each of these histograms resemble a normal distribution. Both appear to be bell-shaped, and therefore we can conclude that the variables are normal.

We should also be aware of any range restrictions present within the variables, and whether there are any bivariate outliers. Looking at the data ranges within each variable, range restrictions are not evident. The scatterplot from Figure 7.10 is evaluated for bivariate outliers or extreme values. There is one case that seems a bit removed from the trajectory formed by the other cases. Case #1 with an Extroversion value of “1” and a Text Messaging value of “31”

stands out a little from the other cases. Even though this case is a bit removed, we do not believe the case is "extreme," nor does it have an overt influence on the association between the two variables. Visually, an extreme case will clearly stand out from the remaining cases.

A final concern is whether heterogeneous subsamples underlie the data. This is assessed by knowing your data and also knowing the research literature. Here, we have no reason to believe that there may be underlying group differences in the data, and an evaluation of the research literature on text messaging and extroversion does not support separate analyses for any groups that may underlie the data.

Applying the Pearson Correlation to Case Study 7.3: Association between Text Messaging and Extroverted Personality Style

Now that assumptions and considerations have been evaluated, let's apply the Pearson correlation to assess the association between Text Messaging and Extroversion. Data for the calculations are presented in Table 7.9, which has the calculation components for the Pearson correlation.

Table 7.9: Extroversion (X) and Text Messaging (Y) data: Summary statistics and correlation calculations

Case #	**Extroversion (X)**	$X - M_X$	$(X - M_X)^2$	**Text Messaging (Y)**	$X - M_X$	$(X - M_X)^2$	**Cross Product of Mean Deviations**
1	**1**	-4	16	**31**	-10	100	**40**
2	**3**	-2	4	**32**	-9	81	**18**
3	**4**	-1	1	**15**	-26	676	**26**
4	**4**	-1	1	**18**	-23	529	**23**
5	**4**	-1	1	**25**	-16	256	**16**
6	**4**	-1	1	**25**	-16	256	**16**
7	**5**	0	0	**44**	3	9	**0**
8	**5**	0	0	**45**	4	16	**0**
9	**5**	0	0	**40**	-1	1	**0**
10	**5**	0	0	**40**	-1	1	**0**
11	**5**	0	0	**45**	4	16	**0**
12	**6**	1	1	**50**	9	81	**9**
13	**6**	1	1	**40**	-1	1	**-1**
14	**6**	1	1	**55**	14	196	**14**
15	**6**	1	1	**50**	9	81	**9**
16	**7**	2	4	**43**	2	4	**4**
17	**7**	2	4	**65**	24	576	**48**
18	**7**	2	4	**75**	34	1156	**68**
Sum	90		40	738	90	4036	**290**
Mean (M)	5.00			41.00			
Sample Variance	2.35			237.41			
Sample Standard Deviation	1.53			15.41			

<u>Covariance</u> is the sum of the mean deviation cross products: (40+18+….+48+68) = 290.
<u>Correlation</u> is the standardized covariance: 290/sqrt(4036*40) = 290/401.796 = 0.72.

Numerator

For the numerator of the Pearson correlation, we follow these steps:

Steps in calculating the Numerator of the Pearson Correlation

1) **Take the difference of each case from its mean value on X and Y separately (Mean Deviations).** These deviations are noted in the table. For Case #1, the mean deviation for Text Messaging is -10, and for Extroversion the mean deviation is -4. For Case #2, the mean deviation for Text Messaging is -9, and for Extroversion the deviation is -2.

2) **Multiply these deviations, creating a cross-product.** Here again, Table 7.4 has these values. For example, the cross-product for Case #1 is 40: -10 x -4 = 40. For Case #2, the cross-product is 18.

3) **Sum the cross-products, which is the covariance.** Summing the cross-products between Text Messaging and Extroversion gives a covariance value of 290. This is the numerator for the Pearson correlation formula.

Denominator

We next calculate the denominator of the Pearson correlation.

Steps in calculating the Denominator of the Pearson Correlation

1) **Calculate mean deviations of each case for the variables.** This is noted in Table 7.4 and was just illustrated for the numerator calculations.

2) **Square the mean deviations for each variable to remove the negative sign.** Squaring the deviations removes the negative signs. For Case #1 on Text Messaging, the square of the deviation (-10) is 100: $-10^2 = 100$. For Case #2, the square of the deviation (-9) is 81: $-9^2 = 81$.

3) **Sum the squared deviations for each variable.** Summing the squared deviations for Text Messaging yields a value of 4036. For Extroversion, the sum of the squared deviations is 40.

4) **Multiply the sum of the squared deviations for each variable, then take the square root of the product.** Multiplying the sum of the squared deviations together for Extroversion and Text Messaging (4036 x 40), a value of 161440 is produced. Next, the square root of 161440 is taken, which produces a denominator for the correlation formula of 401.796: $\sqrt{161440} = 401.796$.

Now we can calculate the Pearson correlation for Extroversion and Text Messaging. We take the numerator of 290 and divide by the denominator of 401.796. This produces an *r* value of 0.72: 290/401.796 = 0.72. The correlation between Text Messaging and Extroversion is 0.72.

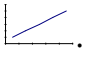
Steps *Pearson Correlation Using Formula 7.1*

1 $$r = \frac{\sum (X_i - M_X)(Y_i - M_Y)}{\sqrt{\sum (X_i - M_X)^2 \sum (Y_i - M_Y)^2}}$$

2 $$r = \frac{290}{\sqrt{(4036)(40)}}$$

3 $$r = \frac{290}{\sqrt{161440}}$$

4 $$r = \frac{290}{401.796} = 0.72$$

The Pearson correlation is 0.72. This indicates there is a positive association between Text Messaging and Extroversion; as Extroversion increases, Text Messaging increases  .

If curious, calculations for the Pearson correlation using Formula 7.2 (with population standard deviations) are presented in Table 7.10. Note that the same Pearson correlation of 0.72 is derived. Remember though that if you use this formula, you have to calculate the population variance that uses N instead of N-1 in the denominator $\frac{\sum (X_i - M\ )^2}{n}$ for each variable, then take the square root to derive the population standard deviation.

Table 7.10: Extroversion (X) and Text Messaging (Y) correlation using Formula 7.2

Case #	**Extroversion (X)**	**Text Messaging (Y)**	X*Y
1	**1**	**31**	-10
2	**3**	**32**	-9
3	**4**	**15**	-26
4	**4**	**18**	-23
5	**4**	**25**	-16
6	**4**	**25**	-16
7	**5**	**44**	3
8	**5**	**45**	4
9	**5**	**40**	-1
10	**5**	**40**	-1
11	**5**	**45**	4
12	**6**	**50**	9
13	**6**	**40**	-1
14	**6**	**55**	14
15	**6**	**50**	9
16	**7**	**43**	2
17	**7**	**65**	24
18	**7**	**75**	34
Sum	90	738	90
Mean (M)	5.00	41.00	
Population Variance	224.22	2.22	
Population Standard Deviation	14.97	1.49	

Steps *Pearson Correlation Using Formula 7.2*

1 $$r = \frac{\frac{\sum XY}{N} - (M_X)(M_Y)}{(\sigma_X)(\sigma_Y)}$$

2 $$r = \frac{\frac{90}{18} - (50)(41)}{(14.97)(1.49)}$$

3 $$r = \frac{16.11}{22.32} = 0.72$$

Assessing the Statistical Significance of Case Study 7.3

Now that we know the correlation value is 0.72, we need to assess whether it is statistically significant.

Steps for Determining the Statistical Significance of the Pearson Correlation

1) Adopt a statistical significance level

- For the current example, a .05 level of significant is chosen, indicating we are willing to accept a less than 5% chance that our test results could have occurred by chance.

2) Choose a one- or two-tailed test based on how the null hypothesis is stated

- Here, the null hypothesis is that there is no association between the variables. Thus, a two-tailed test will be adopted.

3) Calculate the degrees of freedom using *N* – 2

- For the current example, the degrees of freedom (N – 2) is 16: 18 – 2 = 16

4) Use Appendix R to find the *r* critical value

- Using a two-tail test with a .05 level of statistical significance, Appendix R shows a critical *r* value of -/+ 0.468.

5) Ask whether the Pearson correlation value calculated for the study exceeds the critical *r* critical value.

- Does the calculated Pearson correlation value of 0.72 exceed the critical value of +/- 0.468? It does, and we therefore reject the null hypothesis and conclude that the correlation value was not due to chance occurrence.

6) Interpret the results

- Text messaging behavior has a significant association at the $p < .05$ level with Extroversion. The association is positive, indicating that as extroversion increases, so does text messaging.

Based on the steps above, we can conclude that Extroversion has a significant association with Text Messaging. The association is positive, indicating that higher levels of Extroversion are associated with more text messaging.

Both IBM SPSS and SAS show the probability of getting the obtained correlation if the null hypothesis is true is exactly .001. Since the probability cutoff is .05, we can conclude that a correlation of 0.72 is very unlikely if the null hypothesis is true. We reject the null hypothesis.

Effect Strength

The correlation value of 0.72 indicates a large effect, well beyond the .50 cutoff used for effect sizes based on *r* for large effects.

Example Write-up

A study was conducted to examine the association between text messaging and extroverted personality style. The research hypothesis is that there will be an association between test messaging and extroversion; the more extroversion an individual reports, the more likely they are to send text messages (a positive association). The null

hypothesis (H_0) is that there will be no association between the two variables, Rho = 0. A sample of 18 college students taken from a college library was asked to report the number of text messages they made during a 24-hour period. In addition, they were given a measure of Extroversion. Higher scores on the Extroversion scale indicate greater Extroversion. A Pearson correlation was used to assess association. Prior to assessment, data were evaluated for the assumptions and considerations of the Pearson correlation, including normality, linearity, extreme bivariate values, range restrictions, and heterogeneous subsamples. Histograms showed both variables to be normal. A bivariate scatterplot was inspected, revealing a linear association between both variables, and no extreme bivariate values were evident. Data for both variables had a sufficient range of values, and prior research did not suggest that our results would differ based on possible underlying groups in the data. A Pearson correlation was next conducted, showing a positive association beyond the .05 significance level between the two variables, r(18) = 0.72, a large effect. We can conclude that as extroversion increases, so does text messaging. We reject the null hypothesis and conclude the two variables are associated.

Breakout Box 7.2: Nonsense Correlations

You now know that correlation assesses whether phenomena move together. Further, given that your research is planned, you should be able to explain or "make sense" of the association between your phenomena of interest. For example, in Case Study #2, the gradual decline in runs scored by the Cubs can be explained through season fatigue. Yes, makes sense! But sometimes it's tempting to correlate all sorts of variables, which is not a good idea. Why? Because sometimes variables that move together cannot be explained. *Google Correlate* http://www.google.com/trends/correlate/ can be used to illustrate variables that correlate yet

they have nothing to do with each other. *Google Correlate* works by providing a correlation between search terms – for example, enter "puppies for sale" into the Search Correlations box, and you will see the search term "boxer puppies for sale" correlates about 0.97 with "puppies for sale". In other words, there is a 0.97 correlation between people searching for "puppies for sale" and searching for "boxer puppies for sale". This makes sense – people who are interested in puppies for sale might also be interested in Boxer puppies. But if you expand the correlation list (click "show more") you will see other terms that correlate with "puppies for sale" that make no sense, with r values over 0.90. This includes "prepaid verizon", "convert flac to mp3", and "princess games". Yes, there is a correlation, but its random occurrence why these terms correlate. The point is that sometimes variables will correlate for no apparent reason. Therefore, we need to be mindful when interpreting correlations; having a good theory and proper explanations for the variables we correlate is a must.

Some Final Notes on Correlation

Causality and the Pearson Correlation

A correlation between two variables tells us that the variables are associated; the size of the correlation is an index of the strength of the association or effect size. However, the correlation does not indicate whether one variable *caused* a change in the other unless the research includes one variable that is manipulated using the experimental method with random assignment of study participants to particular conditions, or the research design is quasi-experimental (review Chapter 4 for an explanation of experimental and quasi-experimental designs).

Both of us have encountered this correlation/causality problem over and over again during our careers. Students produce correlations and then say that one variable caused a change

in the other, yet their study designs cannot support such an argument. Causality is not determined by the test statistic used. Instead, it is determined by the design of the study, the variables assessed, and whether the researcher manipulated the variables.

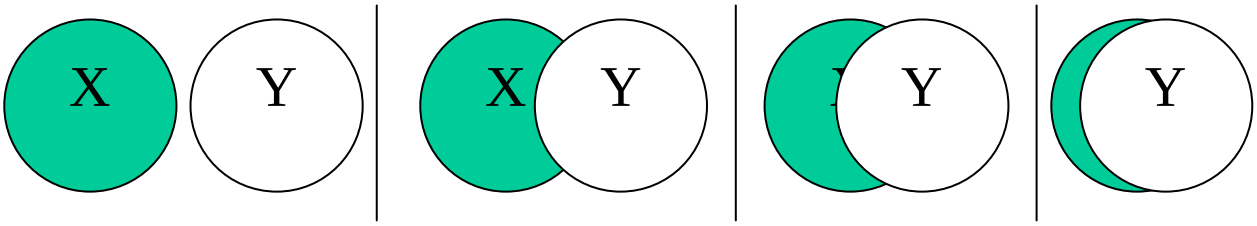
In the example on Relationship Insecurity and Jealousy, all we can say is that higher levels of Relationship Insecurity are associated with lower levels of Jealousy. Does greater relationship insecurity *cause* jealousy to decrease? We really cannot say. If the study was redesigned to manipulate levels of relationship insecurity, then we could infer that higher levels of insecurity caused lower jealousy scores.

Using the Pearson Correlation to Calculate Percent Variance Overlap

Once the Pearson correlation is calculated, we can use the value to calculate how much variance the two variables have in common. Recall that the Pearson correlation is a measure of shared variability between two variables. To assess the amount of shared variability, the Pearson correlation can be squared. Squaring the Pearson correlation converts the correlation coefficient into a measure of shared variance ranging from 0 to 1. An r^2 value close to 0 indicates little or no shared variance across two variables, while an r^2 closer to 1 indicates a great deal of shared variance.

The r^2 value can also be discussed in terms of "percent" variance overlap, with the r^2 value interpreted as a percent. For example, if the resulting Pearson correlation is .50, the amount of variance the two variables have in common is .25 or 25%: $r^2 = .50^2 = .25$. Figure 7.11 shows four different r^2 values and visual representations of variance overlap.

Figure 7.11: Illustration of various squared Pearson correlations for percent variance overlap (Draft Graph)



$r^2 = 0$ (0%) $r^2 = .25$ (25%) $r^2 = .50$ (50%) $r^2 = .79$ (79%)

Let's use the three case studies in this chapter to illustrate. In Case Study 7.1, the overlapping variance between UFO Belief and Trust in Authority is $r^2 = (-0.60)^2 = .36$. Thus, these two variables share 36% of their variance. In other words, 36% of the variation in UFO Belief scores can be accounted for by scores addressing Trust in Authority. This leaves 64% of the variance in both variables to be explained by other factors.

In Case Study 7.2, the overlapping variance between Runs Scored by the Cubs and Season Progression was is $r^2 = (-0.56)^2 = .31$. This indicates that 31% of the variation in Runs Scored by the Cubs can be accounted for by Season Progression (leaving 69% to be explained by other factors).

In Case Study 7.3, the overlapping variance between Extroversion and Text Messaging is $r^2 = (0.72)^2 = .52$, indicating 52% of the variation in Text Messaging behavior can be accounted for by Extroversion scores. With 52% shared variance between the two variables, that leaves 48% of the variance in both variables to be explained by other factors.

Correlation Measures When One or Both Variables are Dichotomous

The Pearson Correlation is designed to be used with continuous or discrete data. When one or both of the variables are dichotomous (having only 2 levels), the Pearson correlation formula is still used to assess association. However, the name used to report the measure changes. When one variable is continuous and the other is dichotomous, the correlation is called a **point-biserial correlation**. When both variables are dichotomous, the correlation is referred to as the **phi coefficient**. The same Pearson correlation formula is used for both sets of variables.

Summary of Important Chapter Points

1. Correlation is used to assess association between two variables that are continuous and/or discrete.
2. Correlation focuses on the linear or straight-line relationship between two variables. If variables have a curvalinear association, correlation is not appropriate to use.
3. The Pearson correlation coefficient is one of the most widely used statistical tests in Psychology and the Social Behavioral Sciences.
4. The possible numeric range of the Pearson correlation is -1.0 to +1.0. Values exceeding +/- 1.0 indicate a calculation error has occurred.
5. The “classic” Pearson correlation formula calculates the covariance between the variables, then standardizes the covariance to form the correlation.
6. Correlation values may be interpreted in terms of an effect size.
7. A Pearson correlation value may be squared to provide the percent variance overlap between the two variables.

8. A bivariate scatterplot of the variable one wishes to correlate can be used to illustrate the association between the variables.

9. Correlation does not equate causality. Causality may only be shown given your research is designed to manipulate variables.

10. Mindfulness is important when it comes to evaluating correlations between variables. If you plan your study appropriately, you should be able to explain the correlations derived.

12. In exploratory research, many variables may correlate due to happenstance. Beware not to "interpret" correlation findings that may in fact do not make sense.

Key Terms

Association	Causality	Positive/Negative Relationships
Correlation	Scatterplot	Effect size
Pearson Correlation "*r*"	Descriptive Plots	Variance Overlap
Covariance	Rho	Correlation with Dichotomous Variables

Web Resources

- Dr. John Marden's website has a cool application for guessing correlations (give it a try!). A scatterplot is presented, and you guess the correct correlation -- http://istics.net/stat/node/35

- The *Consortium for the Advancement of Undergraduate Statistics Education* provides an interactive program to animate correlation values with scatterplots -- http://www.causeweb.org/repository/statjava/CorrMovieApplet.html

- *Google Trends* to illustrate Internet search & News trends over time http://www.google.com/trends

- *Google Flu* for U.S. and World-Wide influenza trends - http://www.google.org/flutrends/

- *Google Correlate* for illustrating correlations between search terms

http://www.google.com/trends/correlate/

Applied Examples Using IBM SPSS and SAS

In IBM SPSS the Pearson Correlation may be calculated using the "Bivariate Correlation" procedure. This procedure is located in the pull-down menus under *Analyze → Correlation → Bivariate*. Programming syntax may be written, or the pull-down menus can be used. Here we provide the programming syntax.

In SAS, correlation is performed using the procedure "Proc Corr". The syntax code is provided below.

IBM SPSS Pearson Correlation Example: UFO Belief and Trust in Authority

Data from Table 7.1: UFO Belief and Trust in Authority data for 20 Students

Case	UFO Belief	Trust in Authority
1	5	2
2	5	3
3	5	4
4	4	4
5	4	4
6	3	3
7	3	4
8	5	2
9	4	5
10	2	4
11	4	2
12	4	4
13	7	2
14	3	5
15	2	4
16	4	3
17	1	5
18	5	4
19	6	3
20	6	3

This is the example data we have used in this chapter. The following steps will walk you through how to enter and analyze the data using IBM SPSS syntax. You may also use the pull-

down menus after the data have been entered:

1) Enter this data directly into IBM SPSS "Data View." Only enter the data.

2) Once the 3-columns of data have been entered, go to "Variable View" and enter the three variable names (Participant, UFO, and Trust). The variable labels can also be filled in (e.g., "UFO Belief" for UFO).

3) Next, open a syntax window *(File →New →Syntax)*. Then directly enter the commands we have listed below. Then, highlight all the syntax, and press the "Play" icon which looks like a button from a VCR or DVD player. The output should resemble what is illustrated below. Note that IBM SPSS calculates the Pearson Correlation as -0.603. Our hand calculated value of -0.60 was rounded – if we do not round the hand calculation, the value is -0.603.

```
CORRELATIONS
   /VARIABLES=UFO Trust
   /PRINT=TWOTAIL SIG
   /MISSING=PAIRWISE.
```

Correlations

		UFO	Trust
UFO	Pearson Correlation	**1**	**-.603**
	Sig. (2-tailed)		**.005**
	N	**20**	**20**
Trust	Pearson Correlation	**-.603**	**1**
	Sig. (2-tailed)	**.005**	
	N	**20**	**20**

SAS Pearson Correlation Example: UFO Belief and Trust in Authority

In SAS, we will use the exact same data setup noted in Table 7.1. Unlike the IBM SPSS example, we have provided the full data setup and output. Once SAS is opened, simply go to the *Program Editor* window, and type the following syntax:

```
Data Corr_Example;
Input Case UFO Trust;
Datalines;
1     5     2
2     5     3
3     5     4
4     4     4
5     4     4
6     3     3
7     3     4
8     5     2
9     4     5
10    2     4
11    4     2
12    4     4
13    7     2
14    3     5
15    2     4
16    4     3
17    1     5
18    5     4
19    6     3
20    6     3
;
Proc Corr;
Var UFO Trust;
Run;
```

Next, highlight all of this syntax in the *Program Editor*, then press the "Run" icon which looks like a person running at the top of the screen. The output created will be similar to the output presented on the next page. Notice that all of the relevant information produced by SAS matches the output from IBM SPSS. Also notice that SAS provides *H0: Rho = 0* in the output.

The CORR Procedure

2 Variables: UFO Trust

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
UFO	20	4.10000	1.48324	82.00000	1.00000	7.00000
Trust	20	3.50000	1.00000	70.00000	2.00000	5.00000

Pearson Correlation Coefficients, N = 20
Prob > |r| under H0: Rho=0

	UFO	Trust
UFO	1.00000	-0.60323 0.0049
Trust	-0.60323 0.0049	1.00000

Draft Sample Exercises (additional exercises will be created)

1. You are assessing the association between relationship commitment and how much individuals "like" their relationship partners. Higher values indicate greater relationship commitment, and greater liking. Your data consisting of just 8 cases are below:

Sub	Commitment	Liking
1	6	87
2	3	73
3	7	95
4	7	94
5	7	58
6	6	58
7	7	89
8	5	71

a. Derive the null and alternative hypotheses for this problem.

b. List the assumptions of the Pearson correlation.

c. Make a scatterplot of the data.

d. By hand, calculate the Pearson correlation using Formula 7.1.

e. Is there a statistically significant association between relationship commitment and liking at the .05 level?

f. What percentage of variance is accounted for between the two variables?

g. Interpret your findings.

2. Use Formula 7.2 (the easier correlation formula) and calculate the correlation coefficient. Your results should match the value derived from Formula 7.1.

3. You are interested in attitudes toward music, and are curious about the relationship between the number of years of formal musical training (X) and the depth of that person's interest in classical music as an adult (Y – higher values indicate greater interest). Years in music is measured in whole numbers (0 to 5). Your N is 28, and the data are below.

	Interest in Classical Music (Y)					
	15	22				
	17	16				
	31	25				
	20	17	11		18	
	16	24	29	24	20	25
	15	20	22	26	25	22
	18	14	16	24	22	12
Years Music training (X)	0	1	2	3	4	5

a. Derive the null and alternative hypotheses for this problem.

b. Do the data meet the assumptions of correlation?

c. How did you test the assumptions?

d. Provide a scatterplot of the data.

e. Calculate a Pearson correlation for these data.

f. Is there a statistically significant association between formal musical training and interest in classical music at the .05 level?

g. What percentage of variance is accounted for across the two variables?

h. Interpret your findings.

4. Using both IBM SPSS and SAS, cross-check your hand calculations from #1 and #2 above.

a. What are the exact probabilities generated by these computer programs for #1 and #2.

b. Do your conclusions change at all based on the computer runs vs. your hand calculations?

Appendix R (DRAFT – to be expanded)

df = *n* -2			
		Alpha	
Df	0.100	0.050	0.010
1	0.988	0.997	1.000
2	0.900	0.950	0.990
3	0.805	0.878	0.959
4	0.729	0.811	0.917
5	0.669	0.754	0.874
6	0.622	0.707	0.834
7	0.582	0.666	0.798
8	0.549	0.632	0.765
9	0.521	0.602	0.735
10	0.497	0.576	0.708
11	0.476	0.553	0.684
12	0.458	0.532	0.661
13	0.441	0.514	0.641
14	0.426	0.497	0.623
15	0.412	0.482	0.606
16	0.400	0.468	0.590
17	0.389	0.456	0.575
18	0.378	0.444	0.561
19	0.369	0.433	0.549
20	0.360	0.432	0.537

Note: Draft Table is for Two-tailed test only

References

Patry, A. L., & Pelletier, L. G. (2001). Extraterrestrial beliefs and experiences: An application of the Theory of Reasoned Action. *The Journal of Social Psychology, 141*, 199-217.