

Chapter 8

Correlation:

Understanding Bivariate Relationships Between Continuous Variables

8.1 Introduction to the Pearson Correlation Coefficient: r

In Chapter 7 we demonstrated how to use the **Crosstabs** procedure to examine the relationship between pairs of categorical variables. As part of this procedure, we also discussed how we could use the statistical measure of association, **Chi square**, to determine whether or not a relationship between two categorical variables is statistically significant.

Recall from Chapter 5 that we created these categorical variables by transforming variables that were originally continuous. For example, scores on the continuous variable, **masctot**, range from 5 to 35. These scores were transformed into a categorical variable, **masc**, with only two levels (1= low masculinity; 2 = high masculinity). While such transformations permitted the construction of contingency tables for crosstabulation, there are some disadvantages to converting continuous data to categorical data.

Such transformations result in a loss of detail and precision that might affect measures of association between variables. Statistically, the reduction in variability in the transformed categorical variables can decrease our ability to assess relationships. There are other ways to evaluate the relationship between continuous variables, and one such procedure involves the calculation of the **Pearson correlation coefficient**, known as r .

In general, the **Pearson correlation coefficient** is a statistic used to determine the **degree** and **direction** of relatedness between two continuous variables. The possible values of the correlation coefficient range from -1.00 to +1.00, and the closer the number is to an absolute value of 1.00, the greater the degree of relatedness. As with **Chi square**, the **Pearson correlation coefficient** can be tested for statistical significance (using the conventional probability criterion of **.05**).

Two variables may be strongly related, as is the case for age and body weight before puberty (10 year olds invariably weigh more than 5 year olds). Or the two variables may be weakly related, as is the case for these same two variables, age and body weight, between ages 30 and 50 (45 year olds are not likely to weigh more than 35 year olds).

Sometimes changes in one variable are associated with no predictable change in the second variable. For example, in adults height and intelligence are unrelated: tall people can be either smart or dull, and the same holds for short people. In such cases the correlation between the variables centers around zero; that is, there is no correlation.

The *direction* of the relationship, either positive or negative, is indicated by the sign (+ or -) of the correlation value (i.e., whether the coefficient is a positive or negative number). High school SAT scores and college GPA's are positively correlated: high values on one variable (SAT score) are associated with high values on the second (college GPA). When two variables are negatively correlated, an increase in the first is associated with a decrease in the second. For example, we typically find that the higher one's alcohol intake, the lower that person's motor coordination becomes.

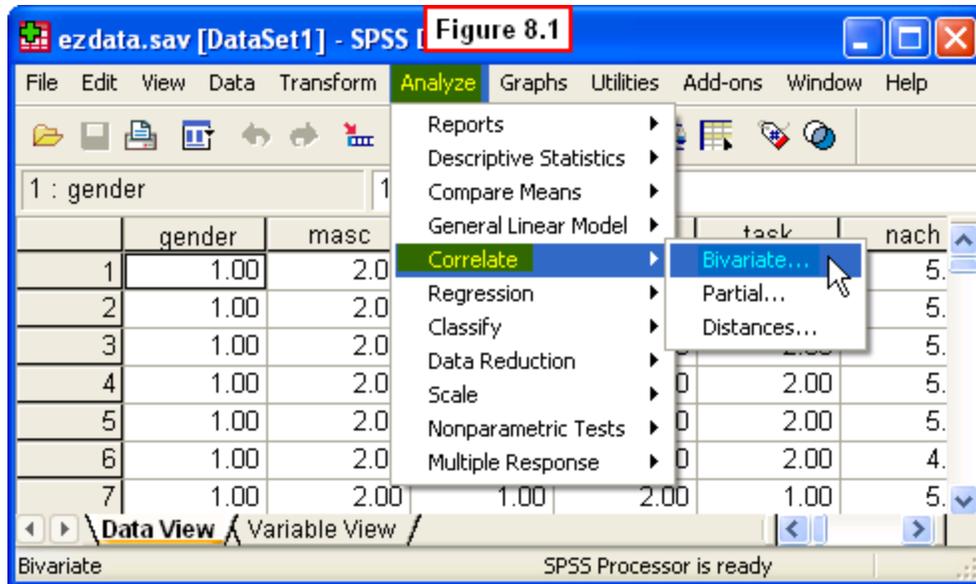
The Pearson correlation coefficient (r) is used specifically to describe relationships when the variables to be correlated are continuous (measured on at least an interval scale). This procedure also assumes that the correlated variables are normally distributed, and that the relationship between the two variables approximates a linear one. Curvilinear relationships are best described with other correlational procedures, most of which are beyond the scope of this book.

8.2 Running Correlations

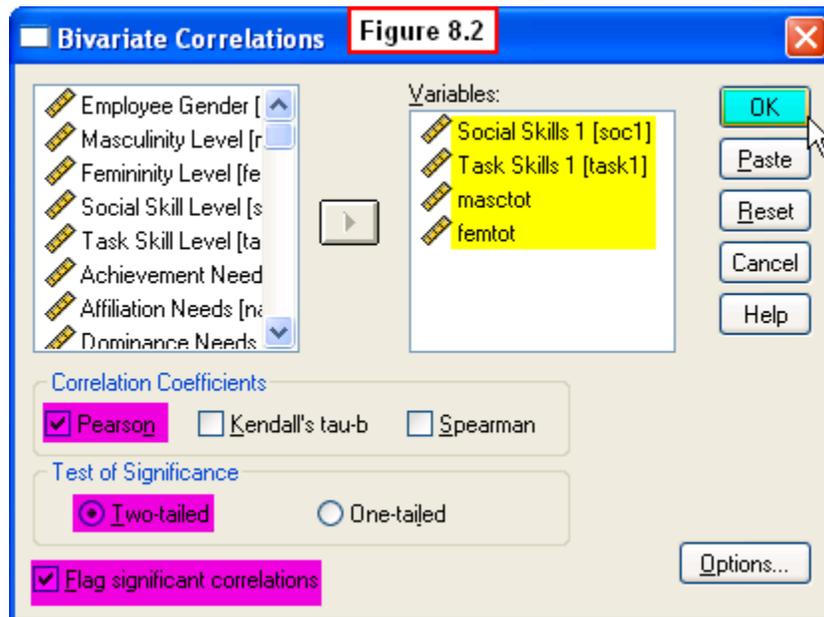
There are numerous bivariate relationships that could be explored using the correlation procedure. For this example we will examine relationships between **masctot** and **femtot** and two other variables, **task1** and **soc1**. In the exercise at the end of the chapter you will be asked to run correlations between **masctot** and **femtot** and two other variables, **naff** and **ndom**. Before we run the analyses, take a moment to think about what kinds of relationships we might expect to obtain here.

For example, given social stereotypes, we might expect a significant positive correlation between degree of masculinity and degree of task skills (i.e., greater masculinity might be related to higher degree of task skills). For the same reasons we might expect a significant positive correlation between degree of femininity and social skills. However, we might find that masculinity is either unrelated or negatively correlated with social skills, and that femininity might be unrelated or negatively correlated with task skills. Actually, the procedure we will use will generate all possible intercorrelations among the included variables.

To run the correlations, open **ezdata.sav** and select **Analyze, Correlate, Bivariate...** from the Data Editor window (Figure 8.1)



In the **Bivariate Correlations** dialog window (Figure 8.2), select **Social Skills 1(soc1)**, **Task Skills1 (task1)**, **masctot** and **femtot** from the variables in the left pane, then click the right-arrow key to move these four variables to the **Variables:** pane on the right. Note that you will have to scroll down the left pane to locate these variables.



Note that by default **Pearson** is checked for the desired correlation, **Two-tailed** is selected for the test of significance, and the **Flag significant correlations** option is checked. Leave these settings as they are. Then click the **OK** button in the upper right corner. The requested correlations will appear in an Output Viewer window.

[Show Me Video!](#)

8.3 Interpreting the Output

The output file contains a matrix of all possible correlations among the variables (Figure 8.3).

Figure 8.3
Correlations

		Social Skills 1	Task Skills 1	masctot	femtot
Social Skills 1	Pearson Correlation	1	.301**	.124	.622**
	Sig. (2-tailed)		.000	.061	.000
	N	228	228	228	228
Task Skills 1	Pearson Correlation	.301**	1	.488**	.283**
	Sig. (2-tailed)	.000		.000	.000
	N	228	228	228	228
masctot	Pearson Correlation	.124	.488**	1	-.063
	Sig. (2-tailed)	.061	.000		.346
	N	228	228	228	228
femtot	Pearson Correlation	.622**	.283**	-.063	1
	Sig. (2-tailed)	.000	.000	.346	
	N	228	228	228	228

** . Correlation is significant at the 0.01 level (2-tailed).

The four variables (highlighted blue) are listed in rows as well as in the columns (thereby creating the *matrix* of all possible correlations). In each cell (or "box") of this matrix are three rows.

- The first row contains the **Pearson r** value (highlighted yellow)
- The second row contains the two-tailed probability (**Sig.**, highlighted purple)
- The third row shows the number of pairs of scores on which each r value is computed.

Next note that the r value for the correlation between each variable and itself (e.g., **Social Skills 1** with **Social Skills 1**) equals 1. This is because the correlation of any variable with itself is perfect - something of no interest, so these boxes can be ignored.

Also note that the 1's form a diagonal cutting from the upper left to the lower right of this matrix. *Note also that all of the correlations listed below the diagonal are duplicates of the correlations shown above the diagonal.* So we need only pay attention to the six correlations shown above the diagonal to interpret our output, since the information below the diagonal is redundant. Last, note that asterisks are placed next to the r values for which the probability is less than or equal to .01 to flag these as statistically significant correlations.

Now we can begin interpreting the **degree** and **direction** of the relationships indicated by the six correlation coefficients of interest. Recall that the **direction** is easily seen by the

sign of the r value (+ or -). A positive correlation indicates that high scores on one variable are associated with high scores on the second variable, while a negative correlation indicates that high scores on one variable are related to low scores on the second variable. All of our significant correlations are positive in this example, so we will discuss the more complex issue of interpreting degree of relationship.

There are two ways to interpret the degree of relationship. The first is an all or none judgment of whether or not the variables are significantly related. The null hypothesis (**H₀**) is that $r = 0.0$, meaning that the variables are unrelated (hence, the **degree** of relationship is zero). Recall the convention is:

- If the **Sig.**, or probability (p), associated with the r value **.05** or less, then we reject **H₀**, and conclude that there is a statistically significant relationship between pair of variables.
- If $p > .05$, then we retain **H₀**, and conclude that the variables are unrelated.

Thus, if the **Sig.** value listed for a correlation is .05 or less, we can assume that the correlation is not the result of chance or random sampling error. That is why we would reject **H₀** and conclude that the correlation is a real one, and thus, one that can be generalized from the sample to the overall population in which we are interested.

So to interpret our correlation matrix, we first need to look at the **Sig.** values (p) listed below each r value (recall that SPSS makes this easier by flagging correlations for which $p < .05$, placing asterisks next to the r values). Examining these, we see that four of the six correlations are significant. The nonsignificant correlations are between **Social Skills** and **masctot** ($r = .12, p = .06$), and between **masctot** and **femtot** ($r = -.06, p = .35$). Thus, our first conclusions concern the nonsignificant correlations. We can conclude that employees' social skills are unrelated to their masculinity, and their degree of femininity is unrelated to their masculinity.

The latter conclusion is of theoretical interest, because intuitively one might have expected masculinity and femininity to be significantly correlated in a *negative* direction (high masculinity would be related to low femininity). The fact that the correlation is not significant supports the proposition that these variables are independent, and that it is possible to have all combinations of low/high levels of masculinity and femininity.

Recall that the concept of *androgyny* implies that these variables are independent. That is, it is possible to score high or low on one variable and either high or low on the other. For example, androgynous individuals score high on both, while sex-typed persons score high on one, but low on the other). Thus, this nonsignificant correlation supports the theory of androgyny.

Turning to the four significant correlations, note that although they are all significant, the r values vary widely - the coefficients range from .68 (the strongest) to .28 (the weakest). Beyond the all-or-none decision about statistical significance, researchers also

often elaborate on the degree of relationship. While there are no hard-fast rules, here is a general rule of thumb:

- r values greater than **.50** indicate a **strong** correlation
- r values around **.30** indicates **moderate** correlation
- r values less than **.20** indicate a **weak** correlation

The import of this is that it is possible for two variables to be significantly related (statistically), even if the relationship is a weak one. Thus, researchers typically expect a statement of degree to modify the basic conclusion that a relationship is statistically significant.

Using the above guidelines looking across the first row of the matrix, we can conclude that there is a moderate positive relationship between social skills and task skills ($r = .30, p = .000$), and a strong positive correlation between social skills and femininity ($r = .62, p = .000$). Thus, a higher degree of social skills is related to both a higher degree of task skills and higher femininity, but the latter is a much stronger relationship.

Looking in the second row, we can conclude that there is a moderate positive relationship between task skills and femininity ($r = .28, p = .000$), and a strong positive correlation between task skills and masculinity ($r = .49, p = .000$). Thus, a higher degree of task skills is related to both higher femininity and higher masculinity, but the latter is a much stronger relationship.

So these analyses have confirmed some expectations, but not others. It is perhaps not surprising that the strongest positive correlations were obtained between masculinity and task skills and between femininity and social skills. Task skills in leadership are associated with instrumentality, a stereotypically masculine characteristic. And Social skills in leadership are associated with expressiveness, a stereotypically feminine characteristic.

However, it is interesting to note that although masculinity was unrelated to social skills, the correlations between femininity and task skills was moderate and statistically significant. Thus, masculinity was strongly related to task skills, and masculinity was independent of social skills (as one might have intuited). However, while femininity was strongly related to social skills, it was not unrelated to task skills (as one might have expected based on social stereotypes).

Confused? We are still in the preliminary stages of identifying the pieces of the puzzle, so we don't yet have the larger picture in front of us. The relationships will become clearer as we carry out other kinds of analyses in the succeeding chapters. However, this first run has generated some interesting correlations that suggest what kinds of characteristics seem to go together in EZ employees. You will discover some similar types of relationships in the exercise to follow, and you are encouraged to explore other correlations in the data file on your own!

8.4 Chapter Review Video

[Review Me!](#)

8.5 Try It! Exercises

1. Correlations among Affiliation needs, Dominance needs, Masculinity & Femininity

In the chapter example we examined relationships between masculinity/femininity and task/social skills. Are there also relationships between masculinity/femininity and need for dominance and need for affiliation? Use the **Pearson correlation** procedure described in Section 8.2 to generate a correlation matrix of the variables, **Affiliation Needs (naff)**, **Dominance Needs (ndom)**, **masctot** and **femtot**.

- In the **Bivariate Correlations** dialog window, move the above four variables from the left pane to the **Variables:** pane on the right.
- Make sure Note that the default **Pearson** is checked for the desired correlation, **Two-tailed** is selected for the test of significance, and the **Flag significant correlations** option is checked.
- Click the **OK** button. The requested correlation matrix will appear in an Output Viewer window.
- **Print** your output file to submit to your instructor.
- **Write an interpretation of the results** to submit to your instructor. Follow the example in Section 8.3, stating your conclusions:
 - Interpret the **Pearson r values**; remember to state which variables are significantly related and which ones are unrelated.
 - State the **degree** and **direction** of relationship among those correlations that are significant (p less than or equal to .05).
- **Write a statement explaining** these patterns of relationships. That is, how might these patterns be explained in terms of social stereotypes about femininity, masculinity, need for affiliation and need for dominance?