Chapter 5

Prediction and Association

Section 5.1 Pearson Correlation Coefficient

Description

The Pearson correlation coefficient (sometimes called the Pearson product-moment correlation coefficient or simply the Pearson *r*) determines the strength of the linear relationship between two variables.

Assumptions

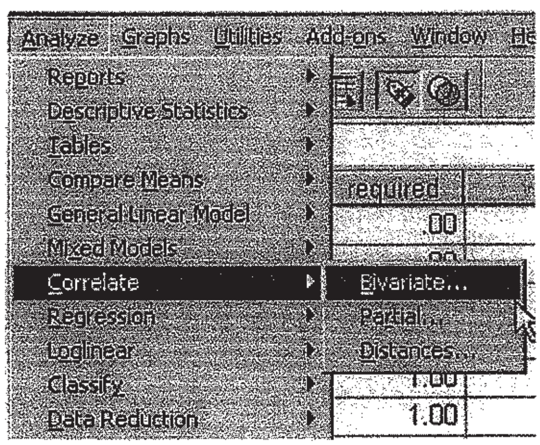
Both variables should be measured on **interval** or **ratio scales**. If a relationship exists between them, that relationship should be linear. Because the Pearson correlation coefficient is computed using *z*-scores, both variables should also be normally distributed. If your data do not meet these assumptions, consider using the Spearman *rho* correlation coefficient instead.

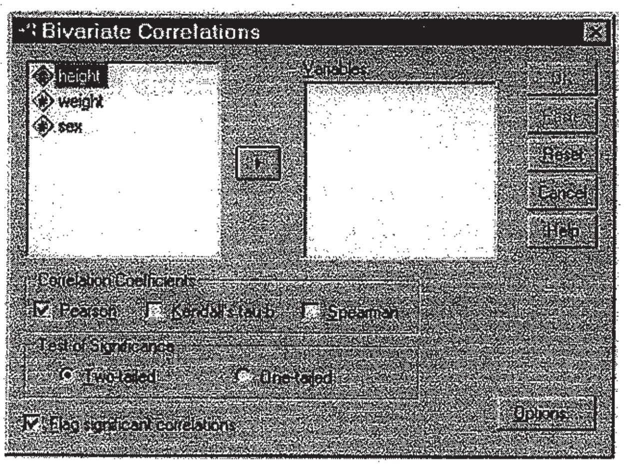
SPSS Data Format

Two variables are required in your SPSS data file. Each subject must have data for both variables.

Running the Command

To select the Pearson correlation coefficient, click *Analyze*, then *Correlate*, then *Bivariate* (bivariate refers to two variables). This will bring up the main dialog box for Bivariate Correlations. This example uses the HEIGHT.SAV data file entered at the start of Chapter 4.





Move at least two variables from the box on the left to the box on the right by using the transfer arrow (or by double-clicking each variable). Make sure that a check is in the *Pearson* box under *Correlation Coefficients*. It is acceptable to move more than two variables. For our example, let's move all three variables over and click *OK*.

Reading the Output

The output consists of a **correlation matrix**. Every variable you entered in the command is represented as both a row and a column. We entered three variables in our command. Therefore, we have a 3 × 3 table. There are also three rows in each cell—the correlation, the **significance** level, and the *N*.

Correlations

Correlations

		height	weight	sex
height	Pearson Correlation	1	.806**	-.644**
	Sig. (2-tailed)		.000	.007
	N	16	16	16
weight	Pearson Correlation	.806**	1	-.968**
	Sig. (2-tailed)	.000		.000
	N	16	16	16
sex	Pearson Correlation	-.644**	-.968**	1
	Sig. (2-tailed)	.007	.000	
	N	16	16	16

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

The correlations are read by selecting a row and a column. For example, the correlation between height and weight is determined by selecting the WEIGHT row and the HEIGHT column (.806). We get the same answer by selecting the HEIGHT row and the WEIGHT column. The correlation between a variable and itself is always 1, so there is a diagonal set of 1s.

Drawing Conclusions

The correlation coefficient will be between –1.0 and +1.0. Coefficients close to 0.0 represent a weak relationship. Coefficients close to 1.0 or –1.0 represent a strong relationship. Significant correlations are flagged with asterisks. A significant correlation indicates a reliable relationship, not necessarily a strong correlation. With enough subjects, a very small correlation can be significant. Please see Appendix A for a discussion of effect sizes for correlations.

Phrasing a Significant Result

In the example above, we obtained a correlation of .806 between HEIGHT and WEIGHT. A correlation of .806 is a strong positive correlation, and it is significant at the .001 level. Thus, we could state the following in a results section:

> A Pearson correlation coefficient was calculated for the relationship between subjects' height and weight. A strong positive correlation was found ($r(14) = .806$, $p < .001$), indicating a significant linear relationship between the two variables. Taller subjects tend to weigh more.

The conclusion states the direction (positive), strength (strong), value (.806), degrees of freedom (14), and **significance** level (< .001) of the correlation. In addition, a statement of direction is included (taller is heavier).

Note that the degrees of freedom given in parentheses is 14. The output indicates an *N* of 16. While most SPSS procedures give degrees of freedom, the correlation command gives only the *N* (the number of pairs). For a correlation, the degrees of freedom is $N - 2$.

Phrasing Results That Are Not Significant

Using our SAMPLE.SAV data set from the previous chapters, we could calculate a correlation between ID and GRADE. If so, we get the output on the right. The correlation has a **significance** level of .783. Thus, we could write the following in a results section (note that the degrees of freedom is N – 2):

Correlations

		ID	GRADE
ID	Pearson Correlation	1.000	.217
	Sig. (2-tailed)	.	.783
	N	4	4
GRADE	Pearson Correlation	.217	1.000
	Sig. (2-tailed)	.783	.
	N	4	4

A Pearson correlation was calculated examining the relationship between subjects' ID numbers and grades. A weak correlation that was not significant was found (r (2) = .217, p > .05). ID number is not related to grade in the course.

Practice Exercise

Use Practice Data Set 2 in Appendix B. Determine the value of the Pearson correlation coefficient for the relationship between salary and years of education.

Section 5.2 Spearman Correlation Coefficient

Description

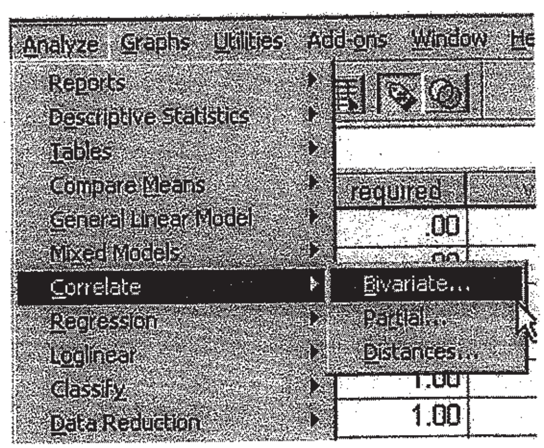
The Spearman correlation coefficient determines the strength of the relationship between two variables. It is a nonparametric procedure. Therefore, it is weaker than the Pearson correlation coefficient, but it can be used in more situations.

Assumptions

Because the Spearman correlation coefficient functions on the basis of the ranks of data, it requires **ordinal** (or **interval** or **ratio**) data for both variables. They do not need to be normally distributed.

SPSS Data Format

Two variables are required in your SPSS data file. Each subject must provide data for both variables.



Running the Command

Click *Analyze*, then *Correlate*, then *Bivariate*. This will bring up the main dialog box for Bivariate Correlations (just like the Pearson correlation). About halfway down the dialog box is a section where you indicate the type of correlation you will compute. You can select as many as you want. For our example, remove the check in the *Pearson* box (by