Seventh Edition

Introduction to SPSS in Psychology

Dennis Howitt and Duncan Cramer

Pearson

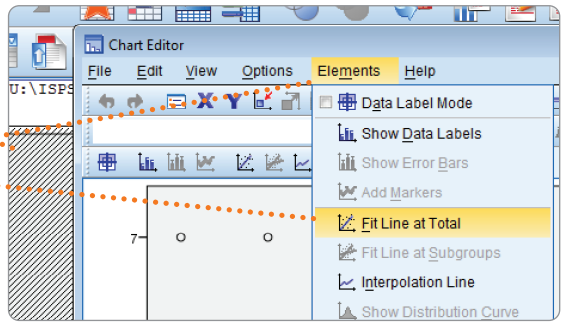
Introduction to SPSS in Psychology

Visit the *Introduction to SPSS in Psychology, seventh edition* Companion Website at www.pearsoned.co.uk/howitt to find valuable student learning material, including:

- Learning objectives for each chapter
- Datasets and syntax
- Annotated links to relevant sites on the web

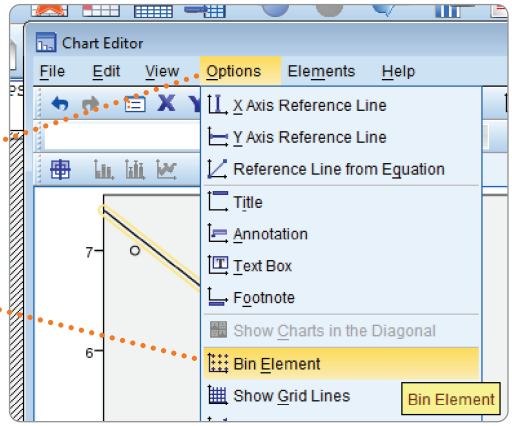
Step 4

To fit a correlation line to the scatterplot, double-click anywhere in it, which opens the 'Chart Editor'.
Select 'Elements' and 'Fit Line at Total'.
Select 'Close' in the 'Properties' box.
This is a regression line (Chapter 10).
To obtain a correlation line the 2 variables have to be standardised (Chapter 5).



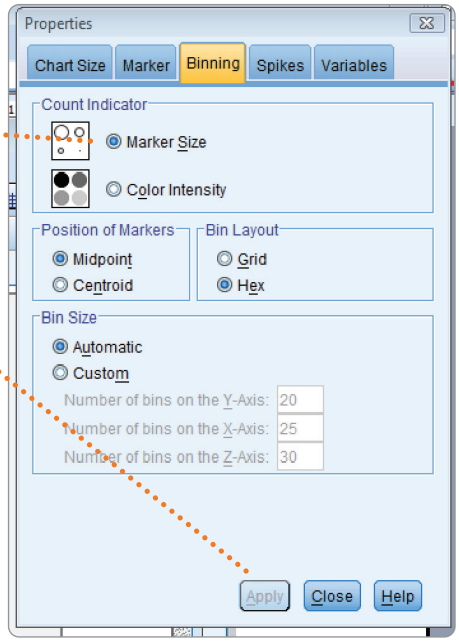
Step 5

A large data set will usually have more than one case with the same two values. The line drawn will depend on how many of these there are, so it is useful to indicate this by selecting 'Options' and then 'Bin Element'. This will give the box in Step 6.



Step 6

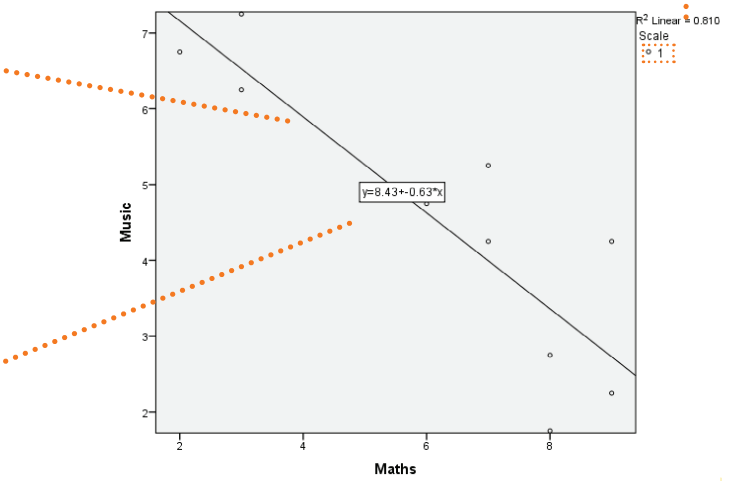
If not already selected, select 'Marker Size' and 'Apply'.
Select 'Close'.
An example of data with more than one case per data point is shown in Section 9.14.



9.13 Interpreting the output

Each circle represents 1 case.

In this scattergram the scatter of points is relatively narrow, indicating that the correlation is high.
The slope of the scatter lies in a relatively straight line, indicating it is a linear rather than a curvilinear relationship.
The line moves from the upper left to the lower right, which signifies a negative correlation.
If the relationship is curvilinear, then Pearson's or Spearman's correlations may be misleading.
This is the regression equation (see Chapter 10).



Reporting the output

- You should never report a correlation coefficient without examining the scattergram for problems such as curved relationships or outliers (*SP*, Chapter 7).
- In a student project it should always be possible to include graphs of this sort. Unfortunately, journal articles and books tend to be restricted in the figures they include because of economies of space and cost.
- We would write of the scattergram: 'A scattergram of the relationship between mathematical ability and musical ability was examined. There was no evidence of a curvilinear relationship or the undue influence of outliers.'

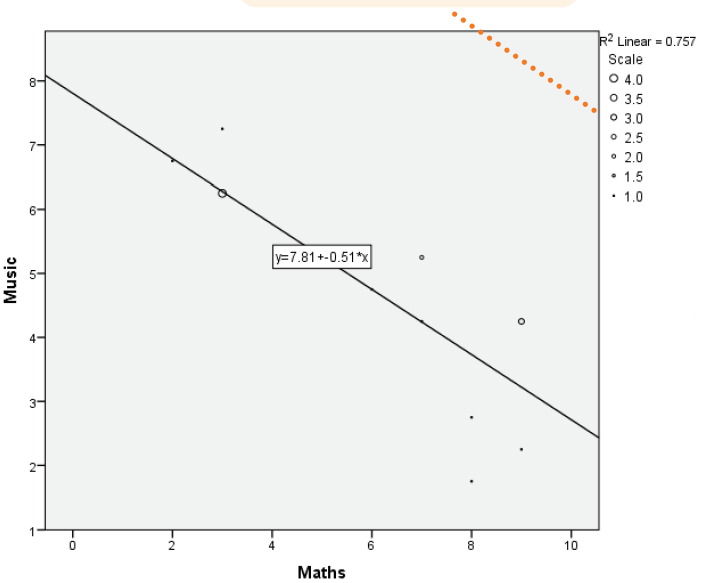
9.14 Scattergram with more than one case with the same two values

To illustrate a scattergram with more than one case with the same two values, we have increased the number of children from 10 in Table 9.2 to 16 in Table 9.3. In Table 9.3, four children have scores of 6 on 'Music' and 3 on 'Maths', three children have scores of 4 on 'Music' and 9 on 'Maths' and two children have scores of 5 on 'Music' and 7 on 'Maths'.

Table 9.3 Scores on musical ability and mathematical ability for 16 children

Music score	Mathematics score
2	8
6	3
6	3
6	3
6	3
4	9
4	9
4	9
5	7
5	7
7	2
7	3
2	9
3	8
5	6
4	7

Bigger circles represent more cases.



Reporting the output

- You should never report a correlation coefficient without examining the scattergram for problems such as curved relationships or outliers (*SP*, Chapter 7).
- In a student project it should always be possible to include graphs of this sort. Unfortunately, journal articles and books tend to be restricted in the figures they include because of economies of space and cost.
- We would write of the scattergram: 'A scattergram of the relationship between mathematical ability and musical ability was examined. There was no evidence of a curvilinear relationship or the undue influence of outliers.'

Summary of SPSS steps for correlation

Data

- In 'Variable View' of the 'Data Editor', 'name' the variables.
- In 'Data View' of the 'Data Editor', enter the data under the appropriate variable names.

Analysis

- For the correlation, select 'Analyze', 'Correlate' and 'Bivariate....'
- Move appropriate variables to the 'Variables:' box.
- Select the appropriate correlation.
- Select 'Options...' if 'Means and standard deviations', 'Zero-order correlations' or 'Exclude cases pairwise' is needed and then 'Continue'.
- Select 'OK'.
- For the scatterplot, select 'Graphs'.
- Then *either* select 'Chart Builder' (this chapter), 'OK', 'Scatter/Dot' and move the 'Simple Scatter' figure to the box above.
- Move appropriate variable names to the vertical and horizontal axes.
- *Or* select 'Legacy Dialogs' (Chapter 10), 'Scatter/Dot...', 'Define', 'Y Axis' variable, 'X Axis' variable and 'OK'.

Output

- The correlation table shows the correlation, its significance level and the sample size.
- For the scattergram, a regression line can be fitted by double-clicking anywhere on the scattergram to bring up the 'Chart Editor'.
- Select 'Elements' and 'Fit Line at Total'.
- To indicate number of cases with the same two values, double-click on the chart and select 'Options', 'Bin Elements' and 'Close' on Bin Elements 'Property' box.

For further resources, including data sets and questions, please refer to the website accompanying this book.

CHAPTER 10

Regression

Prediction with precision

Overview

- Where there is a relationship between two variables, it is possible to estimate or predict a person's score on one of the variables from their score on the other variable. The stronger the correlation, the better the prediction. This is known as **simple regression** because there are only two variables involved.
- Regression can be used on much the same data as the correlation coefficient. However, it is far less commonly used, partly because of the problem of comparability between values obtained from different sets of variables. (The standardised beta weight can be used if such comparability is required.)
- The dependent variable in regression is the variable whose value is to be predicted. It is also known as the criterion variable, the predicted variable or the *Y*-variable.
- The independent variable is the variable that is used to make the prediction. It is also known as the predictor variable or the *X*-variable.
- Great care is needed to avoid getting the independent variable and the dependent variable confused. This can easily happen with simple regression. The best way of avoiding problems is to examine the scatterplot or scattergram of the relationship between the two variables. Make sure that the horizontal *x*-axis is the independent variable and the vertical *y*-axis is the dependent variable. One can then check what the cut point is approximately from the scattergram as well, to get an idea of what the slope should be. The cut point is where the slope meets the vertical axis. These estimates may be compared with their calculated values to ensure that an error has not been made. If problems are found, the most likely reason is that the independent and dependent variables have been confused.
- The simple regression technique described in this chapter expresses relationships in terms of the original units of measurement of the variables involved. Thus, if two different studies use slightly different variables, it is difficult to compare the outcomes of the studies using this form of regression.

- In regression, the relationship between two variables is described mathematically by the slope of the best-fitting line through the points of the scattergram, together with the point at which this regression line cuts the (vertical) axis of the scattergram. Therefore, the relationship between two variables requires the value of the slope (usually given the symbol *B* or *b*) and the intercept or cut point in the vertical axis (usually given the symbol *a* or described as the constant).
- Regression becomes a much more important technique when one is using several variables to predict values on another variable. These techniques are known as multiple regression (see Chapters 29–32). When the dependent variable is a nominal category variable, then the appropriate statistical analysis will be a form of logistic regression (see Chapters 42 and 43 of Howitt, D. and Cramer, D. (2017) *Statistics in psychology using SPSS,* Harlow: Pearson).

10.1 What is simple regression?

One of the most difficult statistical techniques for novices to understand is that of simple regression. It is actually one of the earliest statistical methods and historically predated the correlation coefficient. The first thing to understand is that regression is applied to exactly the same data as the Pearson correlation coefficient. But it does something different. Look at the scatterplot in Section 10.11. When we discussed Pearson correlation in Chapter 9, we explained how the correlation coefficient is a measure of how closely the data points fit the straight line through the data points – the bigger the correlation coefficient, the closer the data points tend to be to the straight line. Regression is different in that it describes the characteristics of the straight line itself. In order to describe the best-fitting straight line we need to know two things: (1) where the line cuts the vertical axis of the scatterplot and (2) what the slope of the line is. The point at which the vertical axis is cut by the straight line is known as the cut point, intercept or constant, whereas the slope is the regression weight. The slope is merely the amount that the line goes up (or down) for every unit that one goes along the horizontal axis. The slope can be negative, which indicates that the line goes downwards towards the right of the scatterplot.

The good news is that all of the hard work is done for you in SPSS. The best-fitting straight line is calculated by SPSS, so it is not a matter of trial and error.

However, it needs to be understood that regression is affected by which variable you put on the horizontal (or *X*) axis and which variable you put on the vertical (or *Y*) axis. Quite different figures emerge in regression depending on your choice. Why does this matter? One of the functions of regression is to allow the user to make predictions from the value of one variable to the other variable. This is possible if it is established from a sample of data that there is a good correlation between the two variables. For example, if you measured the heights and weights of a sample of participants, you would find a correlation between the two. On the basis of this information, you would assume that someone who is tall is likely to be heavier than someone who is small. The only problem is that this is not very precise.

Figure 10.1 gives a scattergram for the relationship between musical and mathematical ability using the data in Table 10.1, identical to that in Table 9.2 in the previous chapter. The best-fitting straight line (also known as the regression line) has been drawn in. We have added in some extra things: (a) the cut point of the regression line on the vertical axis (which, by the way, can be negative – i.e. below the horizontal axis) and (b) the words 'independent variable' and 'dependent variable' that refer to the horizontal and vertical axes, respectively. You could predict music ability scores using this scatterplot if you wished simply by drawing a vertical line from the relevant point on the horizontal axis (i.e. a particular individual's maths ability score) to the regression line, then horizontally to the vertical axis. The point at which the vertical axis is cut is the predicted score on musical ability. It would be more