

SPSS: Simple Linear Regression

Richard Buxton, 2008.

The aim of this handout is to explain how to use SPSS to build a simple linear regression model. It should be read in conjunction with the related handout ‘Statistics: Regression’, which gives an introduction to the statistical ideas involved in regression analysis.

1. Set up the data file. You need two columns of data - one for the response values (e.g. Ice cream sales) and one for the predictor (e.g. Temperature). Both of these variables should be ‘Scale’ variables.
2. Create a scatter plot showing the relationship between the response and predictor. To do this, select **Graphs... Chart Builder** and choose ‘Scatter/Dot plot’. Use a ‘Simple Scatter’ plot, with the response on the vertical axis and the predictor on the horizontal axis. In examining the plot, we’re looking for the following
 - Does the relationship between the response and predictor appear to be roughly linear? If it doesn’t, we’ll need to think about using a more flexible model, such as a quadratic.
You may find it helpful to superimpose a fitted line on to you plot. You can do this by double clicking on the plot to start the Chart Editor and then selecting **Elements... Fit Line at Total**.
 - Are there any obvious outliers - i.e. observations that don’t tie in with the bulk of the data? With a small set of data, it can often be difficult to distinguish between the presence of outliers and the need for a more flexible model.
 - Is there any evidence of non-constant variance, or of a skew error distribution? We’ll be able to check the regression assumptions more carefully once we’ve fitted a model and obtained residual plots - see below.
3. Fit a linear model and obtain diagnostic plots of residuals.
 - (a) Select **Analyze... Regression... Linear**
 - (b) Put the response variable in the ‘Dependent’ box and put the predictor in the ‘Independent(s)’ box.
 - (c) Click ‘Plots’. Put the residuals (*ZRESID) in the ‘Y:’ box and put the predicted values (*ZPRED) in the ‘X:’ box. Select ‘Normal probability plot’. Click ‘Continue’.
 - (d) Click ‘OK’

The fitted model is given in a table headed ‘Coefficients’. The estimates of the intercept and slope are the figures in the column headed ‘Unstandardized coefficients’.

4. Look at the plot of residuals against fitted values. Use the Chart Editor to put a horizontal line at ‘0’ on the vertical axis - this makes the plot much easier to read. In examining the plot, we’re looking for the following
 - Is there any evidence of non-linearity in the relationship between response and predictor? If there is, it will show up as a ‘U’ shape, or inverted ‘U’ shape, on the residual plot.
 - Is there any evidence of non-constant variance? For example, does the amount of variation in the response seem to increase as the predicted value gets larger? If it does, we may need to try a transformation of the response.
5. Look at the Normal probability plot. If the distribution of the errors is roughly Normal, this plot should be roughly linear. Any pronounced curvature suggests that the assumption of Normality may not be reasonable. We can sometimes deal with this problem by transforming the response.
6. Make any required predictions. We often want to predict the value of the response at some specified value of the predictor - e.g. we may wish to predict the fuel consumption of a car with an engine capacity of 1600 cc. To make predictions, proceed as follows
 - (a) Go to the SPSS Data Editor. Add the new predictor values (i.e. the values at which you wish to make predictions) to the bottom of the column containing the predictor.
 - (b) Select **Analyze... Regression... Linear**
 - (c) Click ‘Save’. Under ‘Predicted Values’, click ‘Unstandardized’. Under ‘Prediction Intervals’, click ‘Mean’ or ‘Individual’, depending on whether you’re making a prediction for a single case or for the mean over a large number of cases. Click ‘Continue’ and then click ‘OK’.
 - (d) Go to the SPSS Data Editor. You’ll see five new column containing the predicted values and confidence intervals. These are labelled as follows.

PRE_1	Predicted values
LMCL_1	Lower 95% Confidence Interval for Mean
UMCL_1	Upper 95% Confidence Interval for Mean
LICL_1	Lower 95% Confidence Interval for Individual case
UICL_1	Upper 95% Confidence Interval for Individual case

SPSS makes predictions for the values of the predictor used in fitting the model, as well as for the new predictor values that you added to the bottom of your data file.