Q-STEP SPSS 'HOW TO' GUIDES:

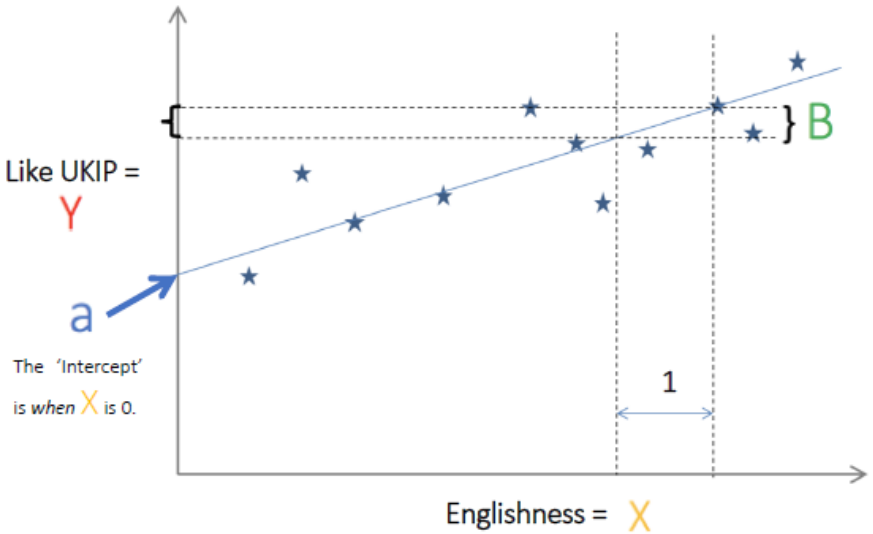
RUNNING LINEAR REGRESSION IN SPSS

Creator: Joshua Townsley

Regression is a widely-used statistical analysis. Knowing how to run and interpret linear regression is a very useful skill that can help your research. This guide will give a brief introduction to the principles of regression analysis, walk you through how to run a linear regression in SPSS, and offer tips on how to interpret the results. The guide is split into the following sections:

1. What is regression?

Imagine we have a correlation between two variables – how 'English' an individual feels (on a scale of 0-10) and how much individuals like UKIP (also on a 0-10 scale), like this:



Q-Step
A step-change in quantitative social science skills
Funded by the Nuffield Foundation, ESRC and HEFCE
Nuffield Foundation
Sheffield Methods Institute.
The University Of Sheffield.

As we can see, people who say they feel more 'English' tend to like UKIP more – and vice versa. What regression does is take us a step further by providing more detailed information on this relationship. Regression essentially plots a 'line of best fit' through the data we provide. This 'regression line' (the blue line above) summarises the relationship between the variables.

The slope of this line is called the 'regression coefficient'. This coefficient shows by how much a one unit increase in the Independent Variable (X) changes the Dependent Variable (Y). E.g. if the coefficient is 2.5, then a one unit increase in X is associated with a 2.5 unit increase in Y.

The regression line is represented by the formula:

$$Y = a + B X$$

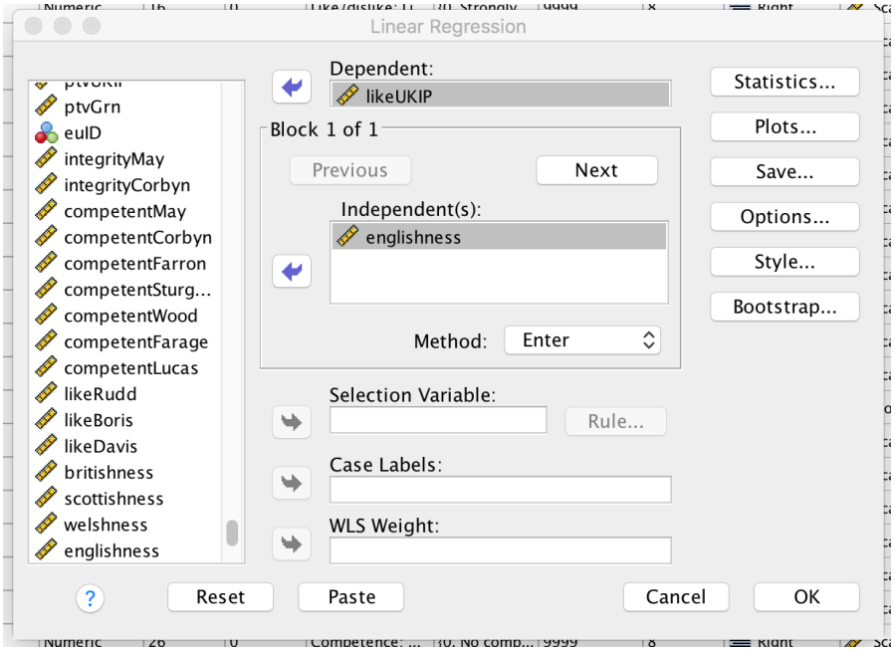
Dependent Variable — constant — coefficient — Independent Variable

2. Running regression in SPSS

Firstly, we need to remember that linear regression is used when the dependent/outcome variable (i.e. the variable we are interested in explaining) is continuous.

To run regression in SPSS, go through the following menus:

Analyze/Regression/Linear



The Linear Regression box will then appear.

1. Put the likeUKIP variable in the 'Dependent' box.

2. Put englishness in the Independent(s) box.

3. Click OK.

3. Interpreting regression output

Now we have run our regression model, we can interpret the results.

The "SPSS Output" viewer will display a series of boxes that contain general information about our regression model, including the main results we need. The "Coefficients" box (below) contains the main information.

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	.346	.193		1.798	.072
	English identity scale	.421	.036	.319	11.683	.000

a. Dependent Variable: Like/dislike: United Kingdom Independence Party

The *coefficient* (B) for ***englishness*** is 0.421. This is how steep the regression line is. It means that a one unit increase in ***englishness*** leads to a 0.421 unit increase in the dependent variable – i.e. ***likeUKIP***.

The '(Constant)' is showing what ***likeUKIP*** would be if ***englishness*** was 0. This is where the regression line intercepts the Y axis (a in the formula).

The coefficient also has a *p-value* that we need to check for statistical significance.

In this case, the p-value is less than .000, so we can reject the null hypothesis with 99.9% confidence (i.e. we're very confident that this relationship exists in the population).

The only other bit to interpret from the regression output is the **r-squared….**

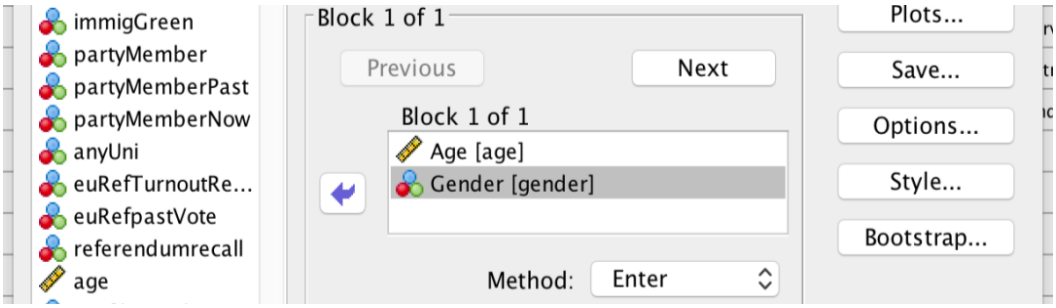
Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.244[a]	.060	.060	3.299

a. Predictors: (Constant), Age

R-squared is basically a measure of how good our regression model is at explaining our dependent variable. We interpret it as the amount of variance in the DV that our model "accounts for". In this case, 6% of the variation in is accounted for.

We could also add other variables to the model. We do this by simply by adding more variables to the Linear Regression box:



Each variable will have its own coefficient (and p-value) in the "Output Viewer". The only difference is that we read each variable's coefficient as being "the change we expect to see in the Dependent Variable from a one unit increase in that Independent Variable, **controlling for the other variable(s) in the model.**

We need to be careful to consider what a "one unit increase in X" actually means – this will depend on what the variable is and how it's coded. For example, if we had a "Gender" variable (where 1=Male and 2=Female), then its coefficient would simply show what the difference is between males and females.