2 Regression analysis

Regression analysis is a technique that helps us to analyse relationships between variables to better understand the world around us.

Variables are items of data that describe an attribute or characteristic of an object or participant. They can be categorical – placing objects or participants into discrete groups – or continuous – placing an object or participant somewhere on a scale according to a specific attribute.

In data analysis, variables can be assigned as predictors – also known as independent variables – or outcomes – also called dependent variables.

There are many applications for regression analysis. Within longitudinal studies, we can examine how early-life circumstances or characteristics relate to outcomes in adulthood, middle age or later life.

For example, we may want to see whether there is a clear relationship between people's academic performance at school and their life satisfaction in middle age.

Using a scatter diagram, we can plot the life satisfaction score of individuals in our sample against their average academic achievement.

We can see that there is a pattern – but can we say how much more satisfied with life people can expect to be if they averaged 1 grade point higher at school? How about if they averaged half a grade point lower?

Suggested citation: Kaye, N., Hardy, R., & O'Neill, D. (2021). *Analysing longitudinal data*. CLOSER Learning Hub, London, UK: CLOSER

We can summarise the data on the graph by drawing a line roughly through the middle of all the data points. This is the regression line and we can use it to estimate (or model) the relationship between our independent and dependent variables.

From this model, we can predict that someone who had a grade point average of 2.5 would approximate a life satisfaction score of 7.6 out of 10; this is around 0.8 lower, on average, than someone with a GPA of 3.5, whose life satisfaction score would be around 8.4.

It is important to remember that the line is just an estimation and will not predict the outcome perfectly.

One of the advantages of regression analysis is that this simple model can be extended to include more and more independent variables. This is important as it allows us to adjust for variables such as confounders. Confounding variables are related to both the independent and the dependent variable but are not an intermediate factor on the pathway between them

For example, we might suspect that sex confounds the relationship between academic performance at school and life satisfaction in middle age, because girls achieve better academic results than boys and women also report higher life satisfaction than men.

With regression analysis, we can simply add a variable for sex to our model. After controlling for sex, the estimated association between academic performance and life satisfaction in middle age is weaker.

You can continue adding more and more variables to regression models. Regression models are therefore a powerful statistical method for understanding the relationship between variables. To learn more about regression and how it can be performed in different software packages, check out our guidance and suggestions for further reading on the CLOSER website and Learning Hub.

Suggested citation: Kaye, N., Hardy, R., & O'Neill, D. (2021). *Analysing longitudinal data*. CLOSER Learning Hub, London, UK: CLOSER