# Ccc

## Transcript

See video: <https://youtu.be/9h77viejtn4>

Regression is a statistical modelling technique that we use to analyse and predict relationships between variables.

**Multiple Linear Regression**, or MLR allows us to understand the linear relationship between a continuous response variable and two or more explanatory variables.

Note that instead of response you will sometimes see the word dependent or outcome, and instead of explanatory you sometimes see the words independent or predictor. There are small technical differences between these but for most practical purposes they are interchangeable. We use response and explanatory as it is the most general form.

By modelling these relationships, MLR enables us to predict outcomes and, importantly, provides insights into the importance of the predictors we are working with.

In this session, I’ll be guiding you through the application of MLR, drawing on examples from the social sciences to demonstrate its relevance in real-world contexts. A downloadable worksheet and workbook are also available for you.

Multiple Linear Regression builds upon simple linear regression, offering a more sophisticated statistical analysis. With MLR, we can explore the influence of multiple explanatory variables on a single response variable, allowing us to examine more complex relationships.

MLR is an important tool in predictive modelling and data interpretation. For instance, we might be interested in predicting outcomes such as income, health measures, or educational performance, where multiple factors might explain the variability of the outcome.

Let’s briefly revisit simple linear regression.

In simple linear regression, we estimate the relationship between a response variable and a single explanatory variable, given a set of data. This data includes observations for both variables across a particular sample.

For example, we might ask: can we predict exam performance at age 16—the response variable—based on exam results at age 11—the explanatory variable? Essentially, we are examining the relationship between scores at two different ages. Our hypothesis is that the exam score at age 16 can be predicted from the score at age 11.

In Multiple Linear Regression (MLR), we extend the idea of simple linear regression by including more than one explanatory variable. Although we now have multiple variables, the term "linear" remains, because we still assume that the response variable is directly related to a linear combination of these explanatory variables.

MLR can be applied to a wide range of practical problems. For example, we might use it to predict an individual’s income based on several socio-economic characteristics, or we might estimate systolic or diastolic blood pressure, taking into account factors such as occupation, smoking habits, drinking habits, age, and so on.

There are certain assumptions we need to be aware of when using MLR.

Firstly, we assume that the **response variable is continuous**, and the **explanatory variables are either continuous or binary**. The relationship between the response and the explanatory variables should be **linear**. This means that, in theory, a straight line could describe the relationship between the variables.

Next, the **residuals**—which are the differences between the observed and predicted values—should be **homoscedastic**, meaning that they have constant variance across all levels of the explanatory variables. Additionally, the residuals should be **normally distributed**.

Another key assumption is that there should be **no more than limited multicollinearity**. Multicollinearity occurs when two or more explanatory variables are highly correlated with each other, which can make it difficult to assess their individual contributions to the model.

We also assume that there are **no extrinsic variables**—meaning no omitted variables that could strongly influence the response variable after controlling for those included in the model. Lastly, both the **errors** and **observations** should be **independent** of each other.

It is important to note that if any of these assumptions are violated, it does not necessarily mean that we cannot use MLR. Rather, it means we may need to acknowledge certain limitations of our model, or perhaps adjust our interpretation of the results. In some cases, we might need to transform the data to make it more suitable for linear modelling.

National Centre for Research Methods (NCRM)
Social Sciences
Murray Building (Bldg 58)
University of Southampton
Southampton SO17 1BJ
United Kingdom

**Web** www.ncrm.ac.uk
**Email** info@ncrm.ac.uk
**Tel** +44 23 8059 4539
**Twitter** @NCRMUK