Multivariate Analysis: Linear Regression

What factors influence confidence in the police?

Remember that we are using the Crime Survey for England and Wales (CSEW), which is a large-scale survey conducted each year in an effort to better understand public experiences with crime and the police. We are interested in determining how much confidence the survey respondents have in the police, and we can use data from the CSEW to help us investigate this.

We will still be using the variable **policeconf1** as our measure of confidence in the police. Because we are interested in seeing what factors influence **policeconf1** scores, we'll run statistical analyses that illustrate the impact various independent variables may have on the police confidence scores of respondents.

Linear regression is a statistical analysis that allows us to model the relationship between two (or more) variables and predict the values in dependent variables given the influence of certain independent variables. The models created by linear regression require a continuous dependent variable, which we have in the form of **policeconf1**. (There are no restrictions on types of independent variables in linear regression.)

There are a number of individual characteristics that may impact survey respondent confidence in the police, but in this section, we're going to focus on the following explanatory (or independent) variables:

sex: Sex of respondent (binary)
ethngrp2: Respondent's ethnic group
relig2a: Respondent's religion
remploy: Employment status of respondent
educat3: Highest level of education of respondent
health2: Health status of respondent
neighpol: Respondent awareness of neighbourhood policing (binary)

Two of these variables are binary, meaning that they only have two categories. Those that are not binary have more than two categories. For example, the variable **relig2a**, concerning a respondent's religion, has several categories, as there are more than two possible religious affiliations. The variable **sex**, concerning the sex of the respondent, is binary, as there are only two possible responses to this question (i.e. Male or Female).

As social scientists, we may want to investigate if **sex** has a relationship with our outcome variable, **policeconf1**. To do so, we can use linear regression to look at the relationship between the confidence in the police (our dependent or outcome variable Y) and sex (our independent variable X). We'll perform this linear regression first, and then move on to more complex linear regression models further into the section.