

Univariate Analysis

Our research question for the whole of this section is: ***What influences the total GCSE scores of young people in the final year of secondary school?***

We'll be using the Youth Cohort Study of England and Wales 2004-2007 (or YCS) dataset to try to answer this question, because this dataset includes detailed information on the lives and goals of young people over a four year period as they make their way out of secondary school, and could help us identify relationships between early and later choices and accomplishments. Because we have information from four survey years within the same dataset, we can use quantitative analysis to make comparisons between years.

We'll begin our quantitative analysis by running some initial univariate analyses. Univariate analysis refers to the quantitative data exploration we do at the beginning of any analysis. These analyses provide us with descriptions of single variables we are interested in using in more advanced tests and help us narrow down exactly what types of bivariate and multivariate analyses we should carry out.

We should always start our univariate analysis by operationalizing our research question. This is done by determining which variables most suit our question of educational achievement, and how best to use them.

In this section, we'll decide what we want our dependent variable to be. Because we are interested in GCSE scores in the last year of secondary school, we should select a variable that codifies educational achievement in some way. There are several variables in the YCS dataset that measure educational or employment outcomes in all four sweeps of the survey. For our purposes in this section, we'll choose **s1gcseptsnew**, which concerns the total GCSE score earned by each young person surveyed in the first sweep. (We can tell this is a Sweep 1 variable because it begins with "s1.")

We'll start our analysis of **s1gcseptsnew** by looking into the frequencies of data points, and then by calculating some means and learning how to build a histogram.