

Lab Worksheet IX

Multiple regression analysis

In this lab worksheet we are going to do some multiple regression analysis on educational attainment using the Scottish School Leavers Survey. You can find the dataset **ssls.dta** on LEARN. Remember three key Stata commands widely used to explore new datasets:

```
describe  
summarize  
codebook, compact
```

We should be paying attention to things like the number of variables in the dataset, what they measure, how they are coded, whether we have missing cases and how these are coded. Note that the variable “female” has no value labels; however, since it is a binary variable and it is named “female” it is rather safe to assume that “1 = female” (although we should always double-check the dataset’s codebook in cases like this one). The variable “cohort90” shows respondents’ ages centred 1990, meaning that a value of 0 corresponds to a respondent who was born in 1990; a value of -2 corresponds to someone who was born in 1988; a value of 6 corresponds to someone born in 1996 and so on and so forth. The social class variable measures respondents’ parents’ social class.