This folder contains the programs necessary for reproducing the results in the paper "Genetics and Socioeconomic Status: Some Preliminary Evidence on Mechanisms"

The main dataset used in the analysis is the UK Biobank (UKB)

See http://www.ukbiobank.ac.uk/register-apply/ for how to gain access to the UKB data

The file "UKB_fields.txt" lists the phenotypes from the UKB data that are used in the analysis

A polygenic index for educational attainment can be obtained from the SSGAC Polygenic Index (PGI) Repository: https://www.thessgac.org/pgi-repository

Aside from the main data, there is one special file, "Linking_Children_to_Their_Parents.dta", that identifies the father and/or mother of a UKB participant when they also participated in the study

- The variable "child_id" corresponds to the ID of the latter
- The variables "father_id" and "mother_id" correspond to the IDs of the parents

The variable "fam" identifies sets of siblings that participated in the study.

In addition to the UKB data, the analysis uses data from the Occupational Information Network (O*NET) and from the 2009 Annual Survey of Hours and Earnings (ASHE). The folder "data" contains the cleaned data used from these two sources. The file "SOC2000_Labels.dta" has the labels of SOC2000 occupations used in Figure 4.

There are four do-files:

- "crdata.dta" cleans the data
- "FIGURES.dta" produces the figures
- "TABLES.dta" produces all tables with the exception of Table 4
- "Table 4.dta" produces Table 4

The last three do-files use secondary datasets created when the do-file "crdata.dta" is run.