# Instruction

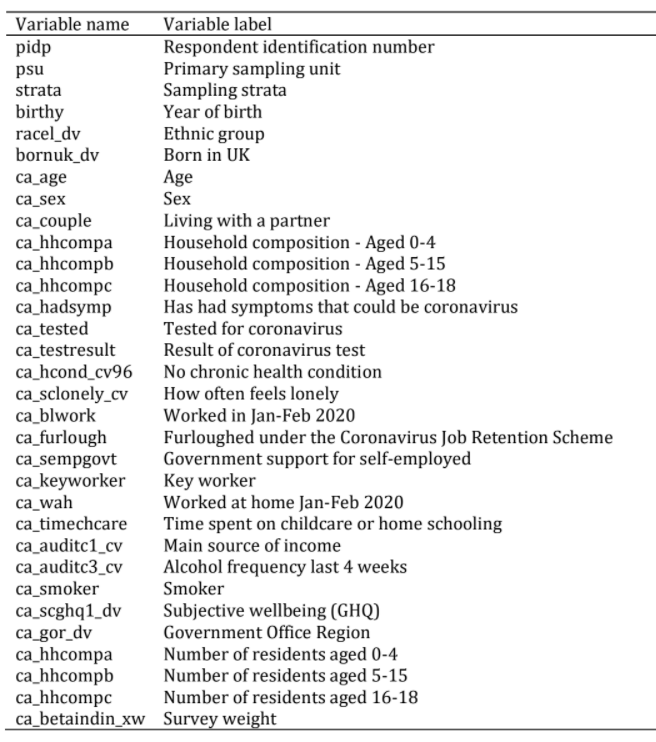
The data for Part A come from the Understanding Society COVID-19 study, April 2020. The study includes all members of the main Understanding Society samples

The data are clustered and stratified probability samples of postal addresses in the UK. A monthly COVID-10 survey was implemented as a web survey between April and July 2020. You have been provided with the April 2020 data. In some months there were additional telephone surveys for households without internet access.

A subset of variables from the COVID-19 study are included in your dataset. The data are collected about the individuals (e.g. age and sex). These are real research data from the UK Data Service. The respondents agreed for their data to be used for research and learning purposes. Before you can access this data, you need to agree to some important term of use. Once you have agreed to the conditions of use, you may open up the dataset and work on the essay questions anytime up until the submission date. There is no limit on the number of times you may open the data files. Be sure to save your data file and R script file.

The essay questions comprise two sections; you must complete each part of each section.

The data file is ca\_indresp\_w\_POLS0010.dta. You can copy this file in the usual way from Moodle once you have agreed to the conditions of use. The variables are:



For more information about these data can be found from following link:

<https://www.understandingsociety.ac.uk/documentation/covid-19>

# DOs and DON’Ts

- DON’T include raw variable names in the text or tables

- DON’T use too many decimal places, but be consistent

- DON’T include unedited R output in the main text of your essay or you will lose marks

- DO make sure tables and figures have titles and referenced in the text

- DO make sure your tables and figures can be understood without reading the text

- DO make sure you have given a clear enough description of what you have done so that the reader can reproduce any numbers/results that you present

- DO be careful how you use the terms ‘significant’ and ‘correlation’ because they have specific meanings in social statistics.

# **PART A: Multiple Linear Regression (60 Points)**

This question uses the ca\_indresp\_w\_POLS0010.dta dataset. You have been asked to write a short report on the relationship between the amount of time spent on childcare or home-schooling during the UK’s coronavirus lockdown and psychological distress. You should create a categorical variable of time spent on childcare or home-schooling with up to four levels and use this variable to fit a multiple linear regression model(s) predicting psychological distress. You may choose to add additional explanatory variables to your model that may explain the relationship between time spent on childcare and home schooling and psychological distress. You may choose to recode or transform some variables in your data. You should report any decisions you take to adjust for individual non-response in your data and to take account of any complex survey design. These decisions should form an introduction that also includes a description of your dataset, subsetting of respondents, descriptive statistics and your research hypothesis. Briefly explain any limitations to your analysis in a concluding section that also summarises your main substantive finding.