

SAS – Programming

Assessment

November 2018

Instructions

The data you will be working with involves 1428 independent records of college student attainment in a mathematics exam. There is some additional information for the variables you will work with, contained in `maths_details.pdf`. You should produce the code to answer the questions in this assessment in a SAS editor. The code for all steps from this point forward must be included in the editor that you submit for marking. Your final script should write your log to a text file in your writeable folder (note that you should only submit your script, no logs).

Deadline for submission: **11.55pm on Monday 19th November 2018.**

Setup

The folder `P:\SCIENG\MATHS\DATA\SAS Course\Programming\Assessment` contains the following datasets:

- `maths_rural.sas7bdat`
- `maths_urban_1.csv`
- `maths_urban_2.sas7bdat`

- 1) Set up the folder above as a library in your SAS system. Since you are unable to write to the p-drive, ensure you have your own SAS folder assigned as a writeable library.
- 2) Using SAS code, copy the SAS datasets **`maths_rural`** and **`maths_urban_2`** to your work directory. Import the dataset **`maths_urban_1.csv`**. You may use the menus to do this, but you must provide your resulting code in your editor.
- 3) Write a macro that renames a variable name for a dataset and overwrites the original dataset. For the datasets, **`maths_rural`**, **`maths_urban_1`** and **`maths_urban_2`**, apply your macro to correct any variable names that are different to the variable names in `maths_details.pdf`.

- 4) Datasets **maths_urban_1** and **maths_urban_2** contain different information for the same students. Merge these datasets into a new dataset called **maths_urban** by the variable that is common to both datasets.
- 5) Find (only) the mean, median, minimum and maximum test scores for urban students and rural students (separately), split up by gender, to 2 decimal places. In a comment in your editor, state which gender in what location (urban or rural) has the lowest observed test score and what value this score is.
- 6) Combine datasets **maths_rural** and **maths_urban** into 1 dataset called **maths**. Make sure this dataset is ordered by studentid, from largest value to smallest.
- 7) Produce 2 boxplots (in 1 figure) of the distribution of test scores, for students who took paid maths classes and students who did not take paid maths classes. Ensure test score is plotted on the y-axis, include a reference line at test score 70% and give appropriate names to your y-axis and x-axis.
- 8) In order to assess whether students qualify for having extenuating circumstances, the college wishes to consider students who did not pass the course, had 10 or more absences and have either of the two worst health statuses. Put the records that meet all 3 criteria into a new dataset called **contact**.
- 9) The college has decided to contact people from the dataset **contact**, based on whether or not a test score is less than 50%. The college will contact a student if the student is aged 21 or over and will contact a parent/guardian if the student is aged under 21. Write one data step that does all of the following:
 - a) In the dataset **contact**, create a new variable called decision. This variable should say "Contact" or "Don't contact" if the student's test score was less than 50% or greater than or equal to 50%, respectively.
 - b) In the dataset **contact**, create a new variable called outcome. This variable should say "Contact parent/guardian of student <studentid>" or "Don't contact parent/guardian of student <studentid>" if the student is aged under 21 and their test score was less than 50% or greater than or equal to 50%, respectively. The variable should say "Contact student <studentid>" or "Don't contact student <studentid>" if the student is aged 21 or over and their test score was less than 50% or greater than or equal to 50%, respectively. Ensure in all cases, you replace <studentid> with the student's actual id number.

- 10) Print the dataset **contact** from SAS to an external file, only displaying the variables decision and outcome, where the output is sorted alphabetically by decision, and give appropriate labels to your columns. The external file should be html format.
- 11) Obtain a 99% confidence interval for the variance of score in the dataset **contact**. Write this interval in a comment in your editor.
- 12) Using SQL, output to your writeable library (not your work library) a dataset for the students who need to be contacted by the college (based on your results from question 9)). Do not include the variables p1edu and p1job, but include all other variables.

[End of assessment]