

Session 2 homework task

To complete this homework task, you will need to download the `language_exams` data file from our Moodle page into your working directory.

In the file, you will find the (made-up) scores and ages of 475 students who took an intermediate Portuguese language course at university. Students were tested three times: first in September to check their Portuguese proficiency at the beginning of the course, then again in January as part of their mid-term examination, and finally in June as part of their final examination. On each occasion, students had to complete three subtests to respectively assess their Portuguese vocabulary, grammar and pronunciation. The scores for exams 1, 2 and 3 are composite scores, i.e. each combines the results of the three subtests.

Your task is to run a basic analysis of the exam data using an R script. In your script, please include all the steps, including the command that loaded the data. Please also include sections to make your script very clear, as well as comments.

1. How many observations and columns does the data file contain?
2. Run commands to display the first and the last six lines of the table.
3. What is the average age of participants?
4. What type of variable is `student_id`?
5. What is the rounded mean score on exam 3?
6. What is the difference between the mean scores on exams 1 and 2?

Please save the script to discuss at the next session.