

Linear Models: Homework 2

2024-2025

The homework is individual work. On Blackboard you find the data file and a R Markdown template that you should use for this homework. You have to submit both the pdf and the Rmd file. Please submit your report via BB Assignments and give your files names of the following format:

LastName_FirstName_HW2.pdf
LastName_FirstName_HW2.Rmd

You really have to submit both files!

Important notes about the use of the R Markdown file:

- Replace “Your name” with your name
- There are two major sections: “Answers to the questions” and “Appendix with R code”:
 - Answers to the questions: here you write the answers to the questions (concisely, to-the-point and in neat English).
This section may NOT contain any R code.
 - Appendix with R code: this part contains your R code to support your answers.

The data that you receive is a random sample of students enrolled in the Linear Models course over the last 5 years. For each student you receive data on

- **exam:** result on the final exam (in percentage). This score excludes the scores on the homeworks and the exam question related to a simulation study.
- **homeworks:** the total score on the 3 homeworks, with a maximum score of 3.
- **simulation:** the score on the exam question related to a simulation study, with a maximum score of 3

- **specialisation:** the specialisation of the student (BS=biostatistics, BI=bioinformatics, D=data science, E=quantitative epidemiology)

First look at the R code given in the appendix of the template R markdown file that is given to you. In this R code several dummy variables are defined, and a linear model is fitted to the data. You may assume that all model assumptions hold true.

Question 1a

Give an interpretation to all parameter estimates of model `m1`.

Question 1b

Look at the results on the line of “homeworks”. Based on this information, calculate a 95% confidence interval (and give an interpretation).

Question 1c

Look at the results on the line of “homeworks”. Based on this information, what do you conclude based on the p -value?

Question 1d

Add the main effect of simulation to the model. What is now the estimate of the effect of the homework score on the average exam score? Give an interpretation of this estimate.

Question 1e

Add the interaction effect of simulation and homework to the previous model. How do you interpret the interaction effect estimate? Also give a 95% confidence interval of this interaction effect, and interpret this interval.

Question 1f

Is there a problem with multicollinearity in the model of the previous question?

Question 1g

Formulate your final conclusion about the relationship between the score on the homeworks and the score on the exam questions (exclusive the question on simulations). Formulate your conclusion in layman's terms.