

Statistical Learning (2022/2023): Assignment 3

The goal of this assignment is that you select a data set of your own choice and that you analyze this data set based on the techniques learned during this course (or related techniques). This assignment is a **group assignment**. A group consists of 3 students (should you be unable to create a group of three, please contact j.d.karch@fsw.leidenuniv.nl)

Data

You can use your own data set or choose a dataset (many datasets are available online). The websites mentioned below could be used as a starting point for finding a dataset:

- <http://archive.ics.uci.edu/ml/>
- <http://www.statsci.org/datasets.html>
- <http://www.europeansocialsurvey.org/>
- <http://www.cs.cmu.edu/~awm/15781/project/data.html>
- http://www.dmoz.org/Computers/Artificial_Intelligence/Machine_Learning/Datasets/
- Open Psychometrics Project: <https://openpsychometrics.org/> (all data sets can be downloaded from https://openpsychometrics.org/_rawdata/)

Analysis

The first step in the analysis consists of exploring your data and formulating some research question(s) and/or problem(s). Next, choose at least three different techniques (more than three are also allowed) that can give some insight into your research question/problem. Apply these techniques to your data, describe and compare their results and indicate what you learned from these analyses regarding your research question/problem.

When you are not sure whether the data set you took and/or the questions/problems you want to tackle are adequate for the assignment (i.e., fit within a course on statistical learning), you can send us a (very) short description of your data and/or question/problem, and we will let you know whether your proposal suits the assignment or whether you should change your proposal a bit (no other feedback on the proposal will be given). Note that you can do this only one time but that you are not at all required to make use of this opportunity.

Presentation

The assignment consists of a (group) presentation you give to your fellow students and the instructors. At the end of the presentation, the instructors and your fellow students will ask questions regarding your presentation and performed analyses. Answering these questions is part of the assignment. For this assignment, **no report** has to be prepared or handed in (only the presentation counts).

In the presentation, you should pay attention to the following topics: (1) description of the data set, (2) research question(s) or problem(s) that you want to tackle, (3) methodology (exploration, which techniques, justify your choices), (4) results and discussion of the results. **Do NOT** present your R-code during the presentation.

Some practicalities regarding the presentation:

- The (group) presentation should be in English and should take maximally 20 minutes (we will stop you after 20 minutes).
- You can decide for yourself which team member presents which part of the presentation. Divide the presentation time (more or less) equally.
- It is mandatory to use slides during your presentation.
 - There is no minimal or maximal number of slides (given the time limit, 15-20 slides is probably enough).
 - You can make your slides in PowerPoint (or any other program you want to use). If you do not use PowerPoint, export your presentation to the pdf-format and use this pdf-version for the presentation.
 - **You should hand in your slides via Brightspace before the presentation. You can find the deadline in the course schedule.** We assume you do not want to present if we do not receive your presentation. This version of the slides does not count for your grade. We will only use these slides to prepare ourselves (i.e., knowing already your data set, research questions, analysis pipeline and main results). For the presentation, you can use an upgraded version of the slides (so you can make changes).
- After each presentation, 5/10 minutes are reserved for questions.

The presentation schedule will be announced on Brightspace a few days before the meeting. To prepare the presentation schedule, we need to know who will take the third assignment (and who will not). Therefore, **send an email to j.d.karch@fsw.leidenuniv.nl by 15.12.2022 and indicate whether or not you will take the third assignment (also send an email when you will NOT take the assignment).** Indicate in this email also your group composition. The schedule will also contain information about attendance requirements.

Evaluation

The evaluation will take the following aspects of the presentation into account (with a tentative grading scheme):

- Did you clearly present the data and the research question(s)/problem(s)? [1/10]
- Did you explore the data in a way relevant to your research question/problem? [1/10]
- Did you clearly explain and justify the techniques you want to use in your analysis (and are these techniques relevant to your question/problem)? [2/10]
- Did you clearly present and compare the results of the analyses? [2/10]
- Did you draw correct conclusions from your results regarding the postulated research questions? [1/10]
- How complex are your data, research question/problem and analysis [1/10]
 - Note that more complex data/problem/analysis will only positively affect your grade when you do everything correctly. Including an analysis technique not discussed in the lecture is a bonus. Be sure, however, that you apply it correctly.
- We will also evaluate your presentation style [1/10]
- Did you appropriately answer the questions after your presentation? [1/10]
- **Extra credits can be gained by taking part actively in the discussion after each presentation (max. 1 point)**

In the break after your presentation, you will get some short feedback from us regarding your performance. The grade for this assignment counts for 1/3 of the final grade (i.e., your final grade will be the mean of your grades obtained on the three assignments). The final grades will be posted on Brightspace as soon as possible.

Questions

You can email us if you have remaining questions regarding this assignment. Extra information regarding the assignment that is relevant for all students will be provided through Brightspace (if needed).

Success,

Marjolein, Tom, and Julian