Group Assignment 'Survey analysis'

The final assignment involves doing survey data analysis for statistical inferences, and preparing a presentation where you:

In the final assignment you will have to:

1. Do self-study to figure out how you can deal with a specific challenge, and explain how you used additional literature to tackle this problem.
2. Perform one or more analyses to answer a substantive research question posed in the assignment.
3. Carry out the analyses in R.
4. Write a report documenting what you did, and why you did so. You have to write your R-code and accompanying text, tables and figures in R-Markdown. You will not be graded on how the report looks, so you can keep the layout of the Markdown document to a bare minimum and use all the default settings. Also, do not worry about how your references or formatting of tables look. As long as everything is consistent, and readable it is fine with me. Here are some good videos and tutorials if you want to know more:
   https://rmarkdown.rstudio.com/lesson-1.html
   https://ourcodingclub.github.io/tutorials/rmarkdown/
   https://rstudio.com/resources/webinars/getting-started-with-r-markdown/

5. Apart from the written report, you will also be graded on an online document. Prepare an online document that shows other students what challenge you were given, and how you dealt with the challenge. You may (or in fact should!) be creative here. The document should be self-contained, so it could be a website (.html document), a wiki, a digital poster, a shiny app (a bit of a challenge, but fun!), a discussion/performance that is recorded on video or something else. It is **not allowed** to hand in a document with text (or code) only or use Powerpoint/Prezi slides. Creativity get bonus points (see rubric final assignment on Blackboard).
   The product should be self-contained, and take no more than 6 minutes to watch/read/listen to.
   You have to send the presentation to the course coordinator, and all other students by **December 9, 17:00.** The easiest way is to simply e-mail your document around. On blackboard, you can go to groups -> send e-mail.
6. Every group reviews the work of 1 other group (see the assigned schedule on Blackboard for this), and prepares 3 questions for the group to be asked during the final meeting on **December 12.**
7. You have the opportunity to also hand in a draft of your report by **16 December at 17:00 latest**. You will then receive written feedback on what can be improved before Monday 19 December 17:00. It is not obligatory to hand in a draft version.
8. The deadline for handing in the final report is 16 January at 17:00. You will in principle receive the same grade as a group, but in case I have strong signs that people contributed much more or less than other group members, I may adjust the grade for individuals up- or downwards. This assignment will be graded using the rubric posted on Blackboard and counts for 50% towards your final grade. Three of the rubrics here concern the presentation and discussion of week 15. All students in the group normally get the same grade. However, it may be the case that some members of the group get a higher or lower grade depending on their performance in the group work as judged by the course coordinator.
9. You can ask for help. Simply write an e-mail any time to explain where/how you get stuck or ask me before/after. Be aware that you are dealing with real-life survey data challenges, and that I may not know myself how to deal with particular issues!

Some additional notes about grading

- The challenges you are given are real-life datasets with real-life problems. There will not be one right way to address the challenge; this is for you to try and work out. You will therefore also not be graded on how 'correct' your solution is, but rather on your approach; what steps did you take to get to a solution? This involves reading some literature (I would expect you to read about 5 articles for

background information), discussing potential solutions (found in the literature or solutions you come up with yourself), lining out the pros and cons of these solutions, and using empirical analyses to evaluate these possible solutions. If you come up with a smart and creative solution this will be regarded positively, but finding a simple and efficient solution even more so!

- In the process of finding a solution, there are also aspects where there is a 'right' or 'wrong' way to analyse the data. Depending on the challenge you are addressing, I would expect you to use course materials to:
    - Evaluate one or more components of the TSE framework
    - Assess bias and/or precision of your proposed solution in a good way
    - Specify a correct survey design object reflecting the sampling design of your dataset (if you are using sample data)
    - Deal with item missing data correctly (e.g. item missings, don't knows)
    - Use survey weights (if they are available) appropriately.
    - Carry out regression or other models correctly in R.

For more documentation on what challenge you are given, see the next page.

Assignment: European Social Survey, cross-nationally?

Questions? Stuck? E-mail p.lugtig@uu.nl

The challenge in this assignment is to fix missing data in the variable hinctnta, which measures total net household income for respondents in deciles. You should impute the data, and describe in your report how you build the imputation model and evaluate the quality of the imputation procedure.

You may choose yourself how to tackle this challenge, and also how to structure your report. Below are a few notes that I believe are „complications" or „challenges" that you will encounter and have to deal with. Feel free to structure your report along these challenges, but also feel free to structure it differently. Also, you don't necessarily need to combine all the variations of imputation designs. You can also use arguments (ideally based on reading extra literature!) to „tackle" a challenge. So, for some questions you can use the literature to make a decision (e.g. XX found that approach XX works best in this situation), and sometimes to figure this out in the data (e.g. to understand the effect of outliers, we did X , Y, Z, and found that all procedures led to almost the same results).

1. The ESS round 9 contains different kinds of weights. Read the weighting documentation online to determine which weight you will want to use. If you are unsure evaluate what differences the weights lead to in the results.
2. As the dependent variable is measured across deciles, you have to think about how to evaluate the dependent variable. Will you look at the mean, median, mode, or something else? (i.e. the idea of deciles is that every category contains 10% of all observations).
3. There are two ways you can impute the data.

- First impute the data without using weights, and then weight the data after imputations.
- First weight the data, and then build an imputation model, and then impute the data wile weighting the results. Does this matter, and which method do you prefer?

4. As the dependent variable is measured in deciles at the country level, the cut-offs between the deciles are different too in the different countries. Select 5 countries from the European Social Survey for which you want to do imputations (you can do this any way you like). Investigate whether the nature of the missing data for income is different in these countries. If so, describe what you do to accommodate your imputation model to differences in the nature of missing data in the income variable.

Good luck!