Statistical Case Studies — Project 1 part 1 — 2021: NHANES 2005-2006

due: Friday 28. 10. 2022 16:00h

In this first piece of coursework you will work with the easySHARE data. For information on data and variables, see the easySHARE data guide.

Task

You are asked to do some data exploration and analysis regarding risk factors of obesity. The analysis is for an official report which aims to answer the following question:

- Which of the variables are associated with bmi (an indicator for obesity)? Describe the strength and the nature of the association.
- In particular, we are interested in whether the associations between risk factors and obesity are the same for males and females.

Instructions

1. Explore the data. That is read in the data, do some pre-processing e.g. factor levels might need to be defined or missing values might need to be dealt with.
2. Choose 1 -2 waves and pick a maximum of 4 countries .
3. Produce summary statistics and graphical summaries relevant for the analysis.
4. Fit a linear model with bmi as your reponse variable to infer whether and how potential risk factors of obesity are associated with bmi whilst adjusting for potential confounder variables. Check the model assumptions, display and interpret the model results.
5. Write a markdown report which contains the above steps with your comments.
6. In addition the markdown output (pdf) should contain a clearly sign-posted executive summary. That is, a section aimed at members of the general public and politicians presenting and explaining your results (executive summary). This should contain graphs highlighting the most interesting finding of your analysis. This section of the report should be no longer than 500 words (not counting graphs) and should be readable without specialist statistical knowledge.
7. The remaining commentary and text in your report should be technical and aimed at statisticians and health researchers, explaining each step and what the conclusions were. **Your selected model should be presented mathematically.**
8. The work must be completed in your work group of 3, which you must have arranged and registered on Learn. The first paragraph of your markdown must list your names with univeristy user name and give a brief description of what each team member contributed to

the project , and roughly what proportion of work was undertake by each team member. Contributions never end up completly equal, but you should aim for rough equality, with team members each making sure to 'pull their weight', as well as not unfairly dominating.

Report format

Your markdown report should be submitted as a **pdf** file alongside the **.Rmd** file. It must contain no more than 1200 words, not counting graphs, tables and R output. The markdown .Rmd file should run with the project data and be reproducible.

Marking Scheme

There is no single correct analysis for this type of project, so you will not be marked on the basis of how close you get to some particular model answer. The marks are not subdivided, but will be allocated on a combination of statistical approach and justification, interpretation of results in context and presentation.

80 – 100% A report that could be presented to the client or collaborator with little or no revision. Analysis is sound so that conclusions are well-supported statistically. Interpretation is reasonably mature. The project should demonstrate a clear overview of the work, without getting lost in details, and be free of all but minor statistical errors. The work is to a publishable standard.

70-79% A report that could be presented to the client or collaborator with little or no revision. Analysis is sound so that conclusions are well-supported statistically. Interpretation is reasonably mature. The project should demonstrate a clear overview of the work, without getting lost in details, and be free of all but minor statistical errors.

60 – 69% A project that could be presented after a round of revision, but without having to redo much of the actual analysis. Some flaws in the analysis or presentation (or minor flaws in both), but basically sound. A good grasp of the statistics and context, so that interpretation is reasonable.

50 - 59% Major re-working required before the project could be presented, but containing some sound statistics demonstrating understanding of statistical modelling and its application. Reasonable presentation and organisation.

40 – 49% Major flaws in analysis and presentation, but demonstrating some understanding of statistics, and a reasonable attempt to present the results.

Fail (below 40%) Flawed analysis demonstrating little or no understanding of statistics, and/or incomprehensible or very badly organised presentation.