M358K - Homework 4 (Regression with categoricals)

posted by: Tuesday, October 23rd at 12AM, 2017

due by: Wednesday, November 8th at 11.59PM, 2017

Instructions and grading scheme

Submission instructions. Please submit three files on Canvas with the following names:

- **homework4-writeup**: this is a word file containing your answers. All figures, tables etc must be inside this file! For how to save R figures, see instructions on Canvas.
- **homework4-code-final.R**: this is a text file containing the R codes you used to compute the numbers, tables, and figures CONTAINED IN your report ONLY. This file IS graded, see below.
- **homework4-code-draft.R**: this is a text file containing the R codes you used to explore the data. This file should give an idea on how you explored the dataset, and how you arrived at code-final.R. It can tables/figures which are EXCLUDED from your report, comments, etc. For instance, the R history of your various sessions is enough.

Acceptable word files are: anything that can be opened by LibreOffice, OpenOffice, Microsoft Word, or equivalent software. Example file extensions are .doc, .docx, .odt.

Grading scheme. For the write up: on each question you can earn 0/1/2 points.

- 2 = correct answer

- 1 = partially correct answer, or correct answer but with muddy or missing justification
- 0 = incorrect answer, unreasonable answer

For code-final.R: you can earn 0/5 points.

- 5 = code-final.R runs on the given dataset, gives all the tables and figures included in your report.
- 0 = no code file OR code file does not run at all OR code file does not produce the reports' tables/figures OR code file raises a plagiarism red flag, etc

For code-draft.R: you can earn 0/5 points.

- 5 = The grader is sufficiently convinced that you explored this dataset on your own.
- 0 = no code file OR code file raises a plagiarism red flag OR code file is irrelvant, does not give an indication on how you arrived at code-final.R

Bonus points for presentation:
Write-up: +2 points for nice report(grammatically correct sentences, no rambling discussions, discussions exceed expectations, extra analysis of the dataset)
R code: +2 points for neat layout. (code adequately commented, clearly laid out, easy to understand).

Bonus questions
These are extra challenging questions, and are additional opportunities to score points. On each bonus question you can get 0/4 points.

Questions

This homework concerns a study on obesity and qualification. Reference: Hebl, M. R., & Mannix, L. M. (2003). The weight of obesity in evaluating others: A mere proximity effect. Personality and Social Psychology Bulletin, 29, 28-38.

People who are obese face a great deal of prejudice and discrimination. But what about people who are somehow associated with an obese person? In the study above, participants had to rate how qualified a particular job applicant was. This applicant was sitting by a woman. The researchers manipulated the following two variables: the weight of the woman and the relationship between the woman and the applicant. The woman was either obese or of average weight. This woman was also portrayed as being the applicant's girlfriend or a woman simply waiting to participate in a different experiment.

Data: `weight.txt` on Canvas. You can read this table in with the `read.table` command (same syntax as `read.csv`)

Variables descriptions:

- Weight: The weight of the woman sitting next to the job applicant.
  1 = obese, 2 = average weight
- Relate: type of relationship between the job application and the woman seated next to him
  1 = girlfriend, 2 = acquaintance (waiting for another experiment)
- Qualified: a numerical. Larger numbers represent higher professional qualification ratings

1. Do a plot of qualified vs weight, qualified vs relate, and qualified vs weight:relate. For each plot, briefly describe what you see. Based on these plots alone, do you think sitting next to an obese woman has an adverse effect on the applicant's qualification ratings?
2. Run a linear regression model, called `weights.model1`, of qualified vs weight and relate. Clearly show the R command that you use, and include the R's model summary. Write down the equation that R gives you.

3. Interpret all the coefficients and the p-values associated with the coefficients.

4. Report the R^2 and adjusted R^2 of your model. What are the meaning of these values?

5. Run another linear regression model, called `weights.model2`, of qualified vs weight:relate alone. Write down the equation that R gives you. Interpret all the coefficients and the p-values associated with the coefficients. Report the R^2 and adjusted R^2.

6. Do a diagnostic plot for your two models. Say which, if any, of the (a) independence (no mean trend) (b) normal distribution and (c) constant variance assumptions are violated.

7. Between the two models you have run, which one is better? Why?

8. Run another linear regression model, called `weights.model3`, which includes an interaction term between weight and relate. Write down the equation that R gives you. Interpret all the coefficients and the p-values associated with the coefficients.

9. What effect did the interaction term have on the model, and why do you think this is the case?

10. Bonus question: In what way is `weights.model2` equivalent to `weights.model3`? In what way is it different? Between these two models, which one is better, and why?