PS947: Homework 3

Dr. Alasdair D. F. Clarke & Dr Anna E. Hughes

Deadline: 04/03/22, 1pm

Welcome to Homework 3. This is worth 12.5% of your overall mark. Write your answers in an **Rmarkdown script** and submit the file to FASER. Please include code, comments, plots, output and discussions of what you have done. Please ensure that your submission is your own work, in your own words. For further information see the university webpages on academic offences.

In addition to the marks for each question, you will also be awarded up to:

- **10 marks** for submitting a well-formatted answer file (suitable use of headings and figures, captions, well-presented graphs etc.)

- **10 marks** for code clarity (sensible names for variables, white space, commented code, only printing relevant code and output, etc.)

NB: note the higher weightings for these aspects compared to previous homeworks!

1 Distinction material: using Github

Create an online Github repository for this piece of homework, and make multiple commits while you're completing the work. Share the repository with us (*scienceanna* and *Riadsala*) by either making us collaborators, or by making the repository public and sending us the link. NB - you still need to submit via FASER (see above)! [**15 marks**]

2 Stroop, standing up or sitting down

In one study, participants who completed the Stroop task showed a smaller Stroop effect (difference between congruent and incongruent trials) when they performed the task standing than when sitting. Another group of researchers decided to try to replicate this effect. Their data is available in *'stroop_standing_data.csv'*

The column meanings are as follows:

Subject	subject ID
Age	age in years
Gender	preferred gender assignment f = female, m = male, o = other
Condition	standing or sitting trials first 1 = start sitting, 2 = start standing
Phase	block type - standing or sitting
Trial	trial number in each block
Stimulus	the stimulus on each trial
Congruency	incongruent or congruent trial
RT	response time for trial
Correct	correct (1), incorrect (0) or mistrial (999)

Fit an appropriate model to this data. [**15 marks**] Note: this data comes from an actual experiment and is therefore still a bit messy compared to the data we have simulated for teaching purposes! Some preprocessing may be required before analysis. Remember to describe your analysis choices and why you made them.

Using this model and any plots you have made, qualitatively assess whether the replication study showed a smaller Stroop effect when participants performed the task standing v.s. sitting (i.e. whether it was able to replicate the original study). [**15 marks**]

3 The Dimensions of Popular Music

Alasdair thinks that modern pop music is rubbish, and all sounds the same. To support his view, he found an interesting article on the internet that walks through applying PCA to the Top 200 songs from Spotify (from 2018). `https://towardsdatascience.com/all-about-the-music-01-ad1b989260df`. The writer claims that when PCA is applied to this dataset, 99% of the variance can be explained using only two dimensions! (Alasdair feels justified in his biases.)

- Download the dataset used for the article. Can you replicate their PCA results? [**10 marks**]
- How would you interpret their findings? [**10 marks**]
- Improve on their analysis in any way you see fit, visualise your results and discuss your conclusions. [**15 marks**]

4 Rubric

The basic rubric framework we will be using during marking is as follows:

Pass:

- The modelling/statistical tests have been carried out and reported correctly. There may be some minor mistakes.
- Basic plots are included.

Merit:

- R code is mostly clear and well commented.
- Modelling and statistical tests are appropriate. The choice of test has been justified and the results are correctly reported.
- High quality plots are included that outline the key points of the analysis.
- Discussion of results is overall good, but may be lacking in depth in some places.

Distinction:

- Code is extremely clear, concise and well commented.
- Appropriate modelling and statistical tests are carried out, with good explanations of why particular choices have been made.
- Publication-quality formatting, including plots (labels, legends, clarity, etc).
- Discussion of results shows excellent understanding and awareness of limitations.