STA365: Homework 2

Daniel Simpson

Instructions

This should be prepared using RMarkdown and **all R and Stan code should be included in the document**. A large portion of the marks are for the explanations accompanying the questions, and if these are missing you will get very few marks.

The assignment should be submitted via Quercus as a single pdf file. It is due at 12pm (Midday) on 26 March, 2020.

Estimating the effect of an intervention

A company has approached you to evaluate the efficacy of their intervention for maths anxiety among univeristy students. They have found a group of people to experiment on. They measured their anxiety (on a scale from 10-50) before and after treatment. The participants were randomized into either receiving the treatment or a placebo and their post-intervention anexity was measured using the same scale.

The company wants you to estimate the average treatment effect (ATE)

$$\mathbb{E}(y \mid Z = 1) - \mathbb{E}(y \mid Z = 0),$$

where y is the difference between pre- and post-intervention anxiety scores and Z is the treatment indicator ($Z = 1$ indicates that the participant received the treatment and $Z = 0$ indicates they received the placebo).

The problem is that the group of people is *not* representative of the population.

We have only measured two variables (other than the outcome): gender and major. In the univeristy, they are given by the following table.

```
table(population %>% select(major, gender))
```

```
##       gender
## major   1   2   3
##     1 265 225  14
##     2 268 235  12
##     3 428 531  25
##     4 485 478  23
##     5 714 723  41
```

The overall task of this assignment is to use MRP to estimate the average treatment effect.

Modelling pre- and post-intervention

Use Stan to fit a multilevel model to predict both pre- and post-intervention anxiety scores.

- The pre-intervention score should be predicted using gender and major. -The post-intervention score should be predicted using gender, major, pre-intervention score, and the treatment indicator Z. (The effect of Z should be a fixed effect and not a random effect.)
- You should use a truncated normal distribution to reflect the natural limits of the data.
- You must write Stan code for this task and it must be displayed.

- You can use any priors that are sensible.
- You must perform posterior predictive checks and comment on the fit of the model.

Estimating the ATE in the population

Estimate the ATE in the population. The challenge here is that the pre-intervention score is not known in the population. Therefore you must predict it.

- Use the model in the previous section to predict the pre-intervention score in the population. You should predict multiple samples (20+).
- For each pre-intervention score, predict the post intervention score when $Z = 0$ and when $Z = 1$, and hence the difference between pre- and post-intervention scores under each treatment arm.
- Use this to construct a posterior distribution for the ATE.
- Comment on the effectiveness of the intervention.