

MTH225 Spring2017 Final Problem 15

SAT scores from a sample of students at each of two schools are to be analyzed to see if there is a significant difference in the mean SAT score for the two schools. We can assume that the SAT is standardized to have a mean of 500 and a standard deviation of 100.

The data is in `MTH225_Spring2017_Final_Problem15.csv`. The variables are:

- `SAT` SAT score
- `school` School (1 or 2)

Use STAN to model the data as two samples from normal populations, and use the posterior draw to compare the value of the mean parameters to each other, and to estimate the probability that the mean SAT score is higher at school 1.

SAT scores range between 200 and 800, and can be assumed to have a standard deviation of 100. Your prior for the mean score should allow for values in this range.

- 2 points: Write R code to read the data and convert it to an R data frame.
- 1 point: Write the data block of a STAN model file that extracts the data from the R workspace.
- 1 point: Write the parameters block of a STAN model file that declares the parameter(s) of your model.
- 2 points: Write the model block of a STAN model file that specifies the priors and likelihood for your model.
- 1 point: Write R code to apply the `extract` function to the data structure output from the `stan` function.
- 1 point: Use the `extract()` function of the RSTAN package to obtain the values for the parameters from the posterior draw.
- 1 point: Compute 95% confidence intervals for the difference between the two mean SAT scores.
- 1 point: Use the posterior draw for this year's mean to estimate the probability that the mean for school 1 is higher than that for school 2.

(10 points possible)