In-class exercise: Single-factor ANOVA with two levels

Names: (signatures only please, printed names will not be counted)

1.) 4.)

2.) 5.)

3.) 6.)

Overview

Methyl tert-butyl ether (MBTE) is a gasoline additive that has been found in groundwater. In this exercise we suppose we have a series of measurements of the MBTE concentration in groundwater taken in the viscinity of two old gas stations, measured in parts per billion.

We are interested in whether the MBTE concentrations of the groundwater at the two sites are the same or not.

We are also interested in whether the variability in MBTE concentration, as measured by the standard deviation, is the same for both sites.

The assumption of equal standard deviations is required for the classical t-test, which is the frequentist approach to this problem.

Because we are using a Bayesian approach, we do not have to make this assumption.

Instructions

As usual, start by bringing your copy of the `MTH225_Fall2016` archive up to date.

Open a command prompt or terminal window, and use the `cd` command to change to the `MTH225_Fall2016` subdirectory. Then type the command:

```
git pull origin master
```

The pull operation should download the following files:

- The R-knitr code: `Single_factor_two_levels.Rnw`
- The data in Rdata format: `Single_factor_two_levels.csv`

- The STAN model file `Single_factor_two_levels.stan`

In this exercise, the data file is in Rdata format, which you read with a `load` command. The .Rnw file is set up to do this, you should not have to modify it or the .stan file.

Questions

Use the *Compile PDF* button to run the model, and use the output to answer the following questions:

1) What are the point estmates of the mu values for the two sites?

2) What are the upper and lower 95% credible intervals for mu[1] and mu[2]?

3) One way to decide if there is a significant difference in the means for the two sites is to compute diff_mu, the difference between mu[1] and mu[2], from the posterior draw, and see if the 95% credible interval for it contains zero. If it does, we say that based on this data there is no strong evidence that the MBTE concentration at the two sites differ. If it does not contain zero, we say that the data provides evidence that the sites differ on MBTE concentration. Which conclusion do we draw from our output?

4) We also want to decide if there is a significant difference in variablility for the two sites. For this we compute diff_sigma, the difference between sigma[1] and sigma[2], from the posterior draw, and see if the 95% credible interval for it contains zero. If it does, we say that based on this data there is no strong evidence that the variation in MBTE content is different for the two sites. If it does not, we say that the data provides evidence that variability of the MBTE content differs for the two sites. What is the conclusion in this case.