In-class exercise: Two sample normal

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

1.) 4.)

2.) 5.)

3.) 6.)

Overview

In this exercise we suppose we have measurements of the moisture content in two types of grapes.

We are interested in whether the moisture content of the two types differ.

We are also interested in whether the variability in moisture content, as measured by the standard deviation, is the same for both types.

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: `MTH225-4_normal_two_sample.Rnw`
- The data in csv format: `MTH225-4_normal_two_sample.csv`
- The STAN model file `two_sample_normal.stan`

Modifying the files for this problem

Unlike a rigged textbook problem, in a realistic situation you will need to write the R-knitr code and STAN model file.

Experienced programmers seldom write anything from scratch, instead taking code written for a similar task and tweaking it for the task at hand.

In this exercise, we will assume that you have a working R-knitr and STAN model file from a similar analysis that we will modify for this problem.

The following changes to the R-knitr file are likely to be required:

- The input file name will probably not match the one we have
- The column names might not match the names from the earlier analysis

Modifying the .Rnw file

The first change just amounts to changing the file name in the `.Rnw` file.

The second change requires that we discover the column names in the .csv file. There are a number of ways we could do this, but perhaps the simplest is to print the first part of the file. From a command prompt or terminal window, if you are not already in the `MTH225_Fall2016` subdirectory, use the `cd` command to change to it.

Once in the directory, enter the command:

```
more two_sample_normal.csv
```

The column names appear in the first line of the dataset listing.

You will need to modify the `.Rnw` file to replace the column name(s) originally in the program with the column names from `MTH225-4_normal_two_sample.Rnw`

Modifying the STAN model file

Once the `.Rnw` file has been modified, you will need to modify the STAN model file by changing the names that appear in the `data` block to match the column names in our dataset.

Additionally, we will assume that we want to use a normal(50,20) prior for `mu` and a cauchy(0,10) prior for `sigma`.

These priors appear in the `model` block. If the values in the file to begin with are different, we need to change them.

Questions

Once you have made the necessary changes, 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 groups?

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 types 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 moisture contents of the two types differ. If it does not contain zero, we say that the data provides evidence that the types differ on moisture content. 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 types. 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 moisture content is different for the two types. If it does not, we say that the data provides evidence that variability of the moisture content differs for the two types. What is the conclusion in this case.