MTH225 Spring2017 Final Problem 11

A study to examine the effect of ambient levels of nitrogen dioxide and fine particulate matter (PM2.5) examined the prevalence of acute bronchitis in metropolitan areas.

The areas were classified according to whether either of these were present at levels thought to be harmful on more than 30% of days.

The data is in `MTH225_Spring2017_Final_Problem11.csv`.

- `CB` Incidence of chronic bronchitis (cases per 100,000 population per year)
- `NO2` Nitrogen dioxide at harmful levels 30% or more of days? (1=no, 2=yes)
- `PM` PM2.5 particulate matter at harmful levels 30% or more of days? (1=no, 2=yes)

Use STAN to the analyze data using a two-factor ANOVA. A question of interest is whether there is an interaction between NO2 and PM, meaning that when present together, whether they have a stronger impact than you would expect from either alone.

You can use the STAN model file `two_factor_anova_with_interaction_general.stan` to fit a cell mean model, which is a general way to analyze models with interactions. With this approach, you have to decide which parameters to compare, based on the structure. The mapping of levels of the two factors in this problem to cell number is as follows:

- `alpha[1]` Cell 1: NO2 = 1 (no), PM = 1 (no)
- `alpha[2]` Cell 2: NO2 = 1 (no), PM = 2 (yes)
- `alpha[3]` Cell 3: NO2 = 2 (yes), PM = 1 (no)
- `alpha[4]` Cell 4: NO2 = 2 (yes), PM = 2 (yes)

If there is no interaction, then the difference between NO2 present and NO2 absent should be the same whether PM is present or not. Likewise, the difference between PM present and PM absent should be the same whether NO2 is present or not.

In the first case, no interaction means

$$\alpha_2 - \alpha_1 = \alpha_4 - \alpha_3$$

In the second case, no interaction means

$$\alpha_3 - \alpha_1 = \alpha_4 - \alpha_2$$

- 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 paramter(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 confidence intervals for the differences listed in the first case of no interaction.
- 1 point: Compute confidence intervals for the differences listed in the second case of no interaction.

(10 points possible)