Data

The New York State (NYS) Department of Health collects and releases data on mortality due to heart attacks. Their 2019 report was the second public data release by the NYS Department of Health on risk-adjusted mortality outcomes for heart attack patients at hospitals across the state. We have data on 65 hospitals in all five boroughs of the region of New York City (Manhattan, the Bronx, Brooklyn, Queens, and Staten Island), along with the Mid-Hudson and Nassau-Suffolk regions. Some of these hospitals perform percutaneous coronary intervention (PCI), a minimally invasive procedure to open blocked coronary (heart) arteries.

According to the report, the outcome of mortality is defined as mortality "within 30 days after initial admission or emergency department encounter. All deaths that occur inside or outside the hospital but within 30 days of admission are counted as deaths in the report. Those who expire during the same hospital stay but more than 30 days after admission are not counted as a mortality."

It may also be worth accounting for differences in patient risk. There are many patient characteristics that influence outcomes after a heart attack, including age and pre-existing conditions. It could be that a given hospital has high observed rates of heart-attack mortality due to the demographics (broadly defined) of the patients that go to that hospital, and not necessarily the heart attack itself. Thus, I have also included the risk adjusted mortality ratio (RAMR), which is the best estimate, based on another statistical model, of what the provider's mortality rate would have been if the provider had a mix of patients similar to the statewide mix. In particular, to obtain the provider's RAMR, we take the observed mortality rate and divide it by the expected morality rate (sum of the predicted probabilities of death for all patients divided by the total number of patients), then multiply that fraction by the state-wide overall observed 30-day mortality rate of 7.3%.

Each row of the dataset is a particular hospital. The variables in the data set are:

1. `Hospital`: name of hospital
2. `Region`: region where hospital is located (Mid-Hudson, Nassau-Suffolk, NYC)
3. `Borough`: borough of NYC `Region` in which hospital is located, if applicable
4. `Type`: hospital type (N = Non-PCI hospital; P = PCI hospital)
5. `Cases`: the total number of heart attack cases in 2019
6. `Death`: the number of deaths among the heart attack `Cases` in 2019
7. `RAMR`: risk adjusted mortality ratio in 2019
8. `ccd_beds`: the number of coronary care beds in 2025
9. `total_beds`: the total number of beds, regardless of type in 2025

Task

Carry out a Bayesian analysis that attempts to answer the following question: **How, if at all, do heart attack mortality rates in New York State vary by hospital, borough, region, and/or hospital type? Create both 1) an informative, aesthetic, and professional plot and 2) informative, clean table that support your findings.**

You are required to use a hierarchical model to answer this research question, however, not every component of the data needs to be incorporated 'hierarchically'. *You are also required to consider and implement a simpler model to compare your final model to.* As the modelers, you get to decide what your models looks like. You don't need to use all variables nor cases (i.e. rows) in the data set if you don't wish. Don't be swayed by other groups!

As I expect there to be a lack of (semi)-conjugacy, you should use JAGS to implement the MCMC. This leaves lots of room for flexibility in priors! You will still have to be thoughtful in your prior choice and provide justification, but you will not be limited to priors that play nice for coding your sampler.

Your case study should be completed in a `.Rmd` or `.qmd`. This time, you will not be provided a starter template!

Some questions to consider

1. What is a simple/basic model that might help answer the research question but perhaps ignores some of the complexities of the data? *It's always good to start simple and build complexity!*

2. What is your ultimate hierarchical model that will help answer the research question, and why? What assumptions are you making?

3. What are the unknown parameter(s) in your model(s)? What prior distribution(s) do you use, and why? Solid interpretation of parameters and/or derived quantities and defending/explaining your prior choices in the context of the problem and your beliefs are very important.

4. What is your computation scheme? How many chains will you run, amount of burn-in and samples taken, etc? (Don't forget MCMC diagnostics to make sure your chain has converged.)

5. What posterior summary of parameters or derived quantities will you need to obtain in order to answer the research question? What sort of visualizations and tables will you create that shed light on the research question?

6. What is your conclusion? Are there any limitations/unreasonable assumptions you made? How do your results compare to those obtained from a model that ignores the grouping structure?

Requirements/General rubric

- State your research question/goal of interest in the context of the data, and provide a statistical model in correct statistical notation that can answer this question. (5 pts)
- Identify the unknown parameters of interest in the model, and provide priors with appropriate interpretation in the context of your model. (5 pts)
- Explain how your statistical model (e.g. via the parameters or perhaps derived quantities) will address the research question of interest. (3 pts)
- Provide a simpler model to compare your more complex model against, and briefly explain why your ultimate model is more desirable for this analysis. (2 pts)
- Clearly state how you will implement your models. You should provide enough detail such that someone else could implement your analysis. (2 pts)
- JAGS code that correctly implements both of your models + appropriate diagnostics. (7 pts)
- Provide results for both models *with interpretation*. You should spend the majority of this space on the results from your more complex model, but do provide a result from the simpler model to compare to. (10 pts)
  - Effectiveness, clarity, and beauty of your visualization and table (3 pts)
- Give a conclusion that answers the research question and demonstrates the insights gained from fitting a hierarchical model. (3 pts)
- Your code should be clean. That means only code necessary for implementing your sampler should be included in your final document. If you think some code for a particular result is worth showing, please do include it! (2 pts)
- Your report should be well-written. This means correct spelling, complete sentences, a nice flow to the report, etc. You should only provide plots/results that you actually intend to discuss in words. Think about this as a *professional* report. I estimate that the final document to be somewhere between 10-14 pages, depending on figure/table sizes and length of code, but feel free to write more/less if you think that is what the report needs. (3pts)