Individual HW06

Your Name Here

```r
library(bayesrules) # R package for our textbook
library(tidyverse) # Collection of packages for tidying and plotting data
library(janitor) # Helper functions like tidy and tabyl
library(rstan) # for MCMC
library(bayesplot) # for plotting
library(broom.mixed) # for tidy() for mcmc
```

1 BR Exercise 9.9

Throughout this chapter, we explored how bike ridership fluctuates with temperature. But what about humidity? In the next exercises, you will explore the Normal regression model of rides (Y) by humidity (X) using the `bikes` dataset. Based on past bikeshare analyses, suppose we have the following prior understanding of this relationship: - On an average humidity day, there are typically around 5000 riders, though this average could be somewhere between 1000 and 9000. - Ridership tends to decrease as humidity increases. Specifically, for every one percentage point increase in humidity level, ridership tends to decrease by 10 rides, though this average decrease could be anywhere between 0 and 20. - Ridership is only weakly related to humidity. At any given humidity, ridership will tend to vary with a large standard deviation of 2000 rides.

(a) Tune the Normal regression model (Eq. 9.6) to match our prior understanding. Use careful notation to write out the complete Bayesian structure of this model.
(b) To explore our combined prior understanding of the model parameters, simulate the Normal regression **prior** model with 5 chains run for 8000 iterations each. HINT: You can use the same `stan_glm()` syntax that you would use to simulate the posterior, but include `prior_PD = TRUE`.
(c) Plot 100 prior plausible model lines $(\beta_0 + \beta_1 X)$ and 4 datasets simulated under the priors.
(d) Describe our overall **prior** understanding of the relationship between ridership and humidity.

2 BR Exercise 9.10

With the priors in place, let's examine the data. (a) Plot and discuss the observed relationship between ridership and humidity in the bikes data. (b) Does simple Normal regression seem to be a reasonable approach to modeling this relationship? Explain.

3 BR Exercise 9.11

We can now simulate our posterior model of the relationship between ridership and humidity, a balance between our prior understanding and the data. (a) Use `stan_glm()` to simulate the Normal regression posterior model. Do so with 5 chains run for 8000 iterations each. *HINT:* You can either do this from scratch or `update()` your prior simulation from Exercise 9.9 using `prior_PD = FALSE`. (b) Perform and discuss some MCMC diagnostics to determine whether or not we can "trust" these simulation results. (c) Plot 100 posterior model lines for the relationship between ridership and humidity. Compare and contrast these to the prior model lines from Exercise 9.9.

4 BR Exercise 9.13

Tomorrow is supposed to be 90% humidity in Washington, D.C. What levels of ridership should we expect?

(a) Without using the `posterior_predict()` shortcut function, simulate two posterior models:

- the posterior model for the typical number of riders on 90% humidity days
- the posterior predictive model for the number of riders tomorrow.

(b) Construct, discuss, and compare density plot visualizations for the two separate posterior models in part a.
(c) Calculate and interpret an 80% posterior prediction interval for the number of riders tomorrow.
(d) Use `posterior_predict()` to confirm the results from your posterior predictive model of tomorrow's ridership.

5 BR Exercise 10.15 (adapted to use `bikes` data)

Before putting too much stock into your regression analysis, step back and consider whether it's wrong.

(a) Your posterior simulation contains multiple sets of posterior plausible parameter sets, $(\beta_0, \beta_1, \sigma)$. Use the first of these to simulate a sample of 500 new daily bike rides from the observed humidities.

(b) Construct a density plot of your simulated sample and superimpose this with a density plot of the actual observed `bikes` data. What do you notice?

(c) Think bigger. Use `pp_check()` to implement a more complete posterior predictive check.

(d) Putting this together, do you think that the assumptions of the Normal regression model are reasonable? Explain.