Group Work 03

```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)
library(broom)
```

> **Note**
>
> The Monday problems are a selection from BR Exercises 8.15-8.17, packaged in a different format

> **Note**
>
> Make sure to remove `#| eval: false` from any code chunks if you copy the whole document using the "code" button

1 An exact analysis

This problem uses the `pulse_of_the_nation` data from slides13

```r
data("pulse_of_the_nation")
pulse_of_the_nation %>%
  count(climate_change)
```

```
# A tibble: 3 x 2
  climate_change                   n
  <fct>                        <int>
1 Not Real At All                150
2 Real and Caused by People      655
3 Real but not Caused by People  195
```

(a) Bayesians think of the entire Beta(151,852) posterior pdf as an estimate for π. But for communication purposes, it can be useful to report what values of π are typical. Identify and calculate two possible posterior point estimates.

(b) The posterior estimates above merely capture the typical posterior π value, thus miss the bigger picture. It's important to supplement these estimates with a posterior credible interval. (Bayesians use "credible" instead of "confidence") Calculate a 95% posterior credible interval for π. Revisiting a plot of the posterior might spark some ideas, and you'll need some R code of the `rXXX`, `qXXX`, `dXXX`, `pXXX` variety.

(c) How can we interpret this interval, (a,b)?

(d) What would a strict frequentist say if you asked them "What's the probability that π lies within the Bayesian credible interval?"

(e) A researcher claims that more than 13% of people don't believe in climate change. Using your interval from (b), what do you think about this claim?

(f) Calculate and interpret a posterior probability that helps you test this claim. You'll need some R code of the `rXXX`, `qXXX`, `dXXX`, `pXXX` variety.

(g) There's no common cut-off / threshold (eg: 0.05) for interpreting Bayesian posterior probabilities, hence no binary conclusion. Better yet, Bayesian conclusions are more holistic and nuanced. With this in mind, summarize your conclusions about our hypothesis.

2 Bayes Factor

Posterior Odds

The **posterior odds** for a hypothesis test H_0 against H_a after observing data $Y = y$ is

$$\text{posterior odds} = \frac{P(H_a|Y=y)}{P(H_0|Y=y)}$$

Posterior Odds

The **prior odds** for a hypothesis test H_0 against H_a is

$$\text{posterior odds} = \frac{P(H_a)}{P(H_0)}$$

Bayes Factor

In a hypothesis test of two competing hypotheses, H_a vs H_0, the **Bayes Factor** is an odds ratio for H_a:

$$\text{Bayes Factor} = \frac{\text{posterior odds}}{\text{prior odds}} = \frac{P(H_a|Y)/P(H_0|Y)}{P(H_a)/P(H_0)}$$

Calculate the **prior odds**, **posterior odds**, and **Bayes Factor** for the researcher's claim that "more than 13% of people don't believe in climate change". What do these numbers tell you?

3 BR Exercise 8.21

Now, let's explore how we can do estimation and hypothesis testing with an approximate posterior sample from MCMC.

(a) Load the following packages, and run the following R/stan code to fit an MCMC approximation for the same problem. Make sure you understand all of the pieces of the model fitting code.

```
library(rstan)
library(bayesrules)
library(bayesplot)
library(broom)
```

```
# Define the Beta-Binomial model in rstan notation
climate_model <- "
  data {
    real<lower=0> alpha;
    real<lower=0> beta;
    int<lower=1> n;
    int<lower=0, upper=n> Y;
  }

  parameters {
    real<lower=0, upper=1> pi;
  }

  model {
    Y ~ binomial(n, pi);
    pi ~ beta(alpha, beta);
  }
```

```
"

# Set the random number seed
set.seed(84735)

# SIMULATE the posterior
climate_sim <- stan(
  model_code = climate_model,
  data = list(alpha = 1, beta = 2, Y = 150, n = 1000),
  chains = 4, iter = 5000*2)
```

(b) Give the following plots a quick peak. Do you see any red flags?

```
mcmc_trace(climate_sim, pars = "pi")
mcmc_dens_overlay(climate_sim, pars = "pi")
mcmc_dens(climate_sim, pars = "pi")
```

(c) The four chains in `climate_sim` are currently stored as an array. Use the code below to store all the chains in a single data frame.

```
# Store the array of 4 chains in 1 data frame
climate_chains <- as.data.frame(
  climate_sim,
  pars = "lp__", include = FALSE)

# Check out the results
dim(climate_chains)
head(climate_chains)
```

(d) Recall your exact posterior point estimates of π from earlier. Estimate these posterior features using your MCMC simulation. (How accurate are these estimates?) NOTE: The `sample_mode()` function in {bayesrules} calculates the mode of a sample.

```
climate_chains %>%
  summarize(___)
```

(e) Recall your exact 95% posterior credible interval for π from earlier. Estimate this interval using your MCMC simulation. (How accurate is this estimate?)

(f) Recall your exact analysis of the claim that more than 13% of people don't believe in climate change, i.e. $\pi > 0.13$. Estimate the posterior probability of this claim using your MCMC simulation. (How accurate is this estimate?)

(g) Play around with the following shortcut functions that can address some, but not all, of our posterior questions. These are applied directly to `climate_sim`, not `climate_chains`. Take notes what they do (leaving comments in your .qmd would be sufficient!)

```r
# What is the estimate? The posterior mean, median, or mode?
tidy(climate_sim, conf.int = TRUE, conf.level = 0.95)

#
mcmc_areas(climate_sim, pars = "pi", prob = 0.95)
```

4 TBA Wed

5 TBA Wed