HW 9 Template

Laura Fenton

2022-04-18

```
library(papaja)
library(ggplot2)  # for plots
library(tinytex)
library(here)
library(readxl)  # for reading excel files
library(modelsummary)  # for summarizing data
library(rstan)
rstan_options(auto_write = TRUE)  # save compiled STAN object
options(mc.cores = 2)  # use two cores
library(posterior)
library(bayesplot)
library(dplyr)
library(kableExtra)
theme_set(theme_classic() +
    theme(panel.grid.major.y = element_line(color = "grey92")))
```

Research Question

> Is there a difference in amyloid accumulation between individuals in the control condition and the exercise intervention condition?

Variables

- `amyloid_accum`: accumulation of amyloid from pre (baseline) to post (wk52) intervention
- `apx_group`: AExSupport = aerobic exercise condition, NoSupport = control condition

Import Data

```
# load the data
exercise <- read_excel(here("APEx_FinalData_20201021.xlsx"))

# subset only individuals who have baseline and follow up data
tt <- table(exercise$study_id)
exercise2 <- subset(exercise, study_id %in% names(tt[tt == 3]))

# create a new variable (amyloid_accum) for amyloid accumulation from baseline
```

Table 1: Summary Statistics

		AExSupport	NoSupport
amyloid_accum	Mean	0.01	0.00
	SD	0.06	0.04
	Min	−0.15	−0.09
	Max	0.15	0.08
	Histogram		

SD = standard deviation

```
# to week 52 (smaller values are better)
exercise3 <-
exercise2 %>%
  group_by(study_id)  %>%
  mutate(amyloid_accum = amyloid_Mean6_CB[timept=="WK52"] -
           amyloid_Mean6_CB[timept=="BL"])

# subset only the first entry of each participant
exercise.subset <-
  exercise3 %>%
  group_by(study_id) %>%
  filter(row_number()==1)

# change the condition variable to a factor
exercise.subset$apx_group <- as.factor(exercise.subset$apx_group)

# remove NAs so that STAN will run later
exercise.subset <- na.omit(exercise.subset)
```

Variable Summary

```
## can't get N to work, not sure why
datasummary(amyloid_accum *
                (Mean + SD + Min + Max + Histogram) ~
                factor(apx_group),
            data = exercise.subset,
caption = "Summary Statistics") %>%
    # add table note
    add_footnote("SD = standard deviation", notation = "none")
```

Model

Let Y = change in amyloid from baseline to week 52 (week 52 - baseline; smaller values indicate less accumulation), G = treatment condition Model:

$$Y_{i,G=0} \sim N(\mu_1, \sigma_1)$$
$$Y_{i,G=1} \sim N(\mu_2, \sigma_2)$$

Prior:

$$\mu_1 \sim N(3, 2)$$
$$\mu_2 \sim N(3, 2)$$
$$\sigma_1 \sim N^+(0, 2)$$
$$\sigma_2 \sim N^+(0, 2)$$

Running Stan

We used 4 chains, each with 4,000 iterations (first 2,000 as warm-ups).

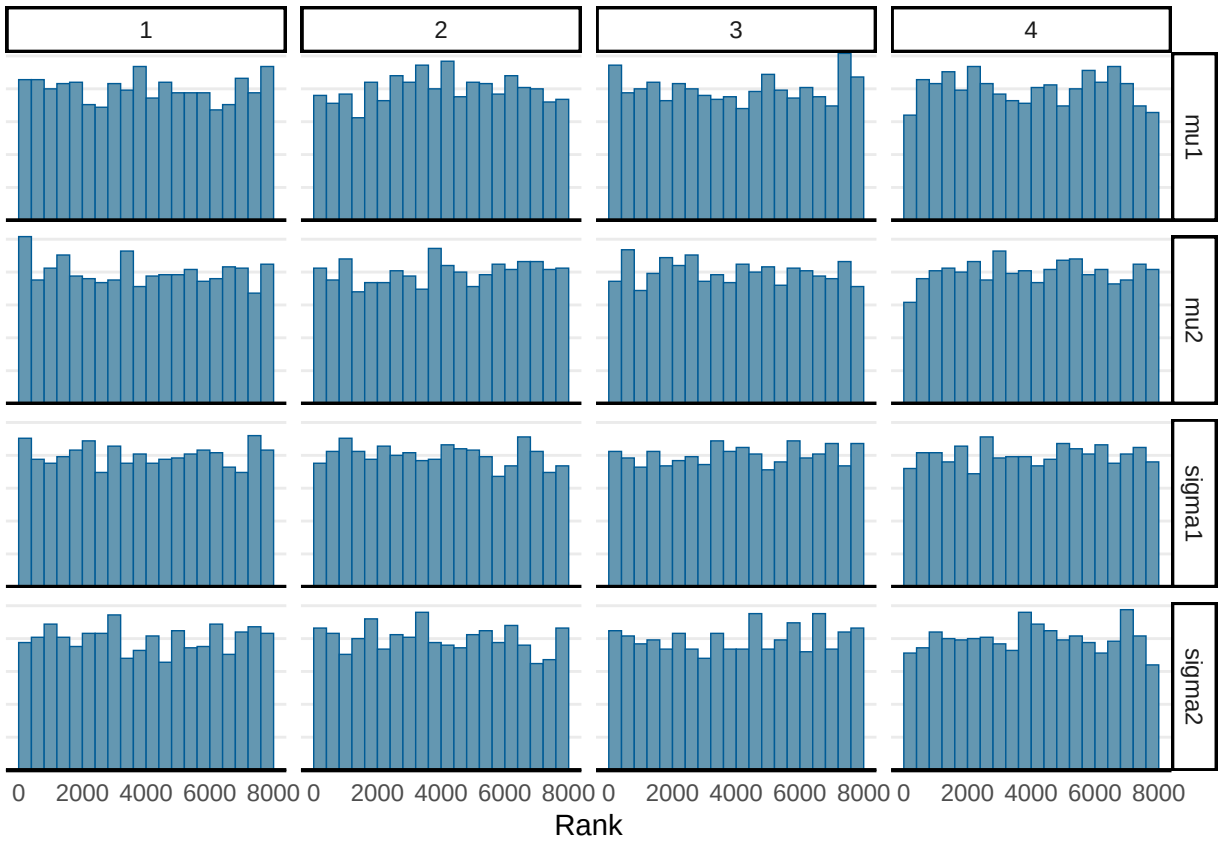
```
# Rescale data so all are greater than 0 and STAN will run (chose .16 based
# on min values from summary data)
exercise.subset$amyloid_accum_rescale <- exercise.subset$amyloid_accum + .16

# 1. form the data list for Stan
stan_dat <- with(exercise.subset,
    list(N1 = sum(apx_group == "NoSupport"),
         N2 = sum(apx_group == "AExSupport"),
         y1 = amyloid_accum_rescale[which(apx_group == "NoSupport")],
         y2 = amyloid_accum_rescale[which(apx_group == "AExSupport")])
)
# 2. Run Stan
m1 <- stan(
    file = here("normal_2group.stan"),
    data = stan_dat,
    seed = 1234,  # for reproducibility
    iter = 4000
)
```

Results

As shown in the graph below, the chains mixed well.

```
mcmc_rank_hist(m1, pars = c("mu1", "mu2", "sigma1", "sigma2"))
```



The following table shows the posterior distributions of μ_1, μ_2, σ_1, σ_2, and $\mu_2 - \mu_1$.

```
summ_m1 <- as_draws_df(m1) %>%
    subset_draws(variable = c("mu1", "mu2", "sigma1", "sigma2")) %>%
    mutate_variables(`mu2 - mu1` = mu2 - mu1) %>%
    summarise_draws()
knitr::kable(summ_m1, digits = 2)
```

variable	mean	median	sd	mad	q5	q95	rhat	ess_bulk	ess_tail
mu1	0.16	0.16	0.01	0.01	0.15	0.18	1	8443.84	5850.82
mu2	0.17	0.17	0.01	0.01	0.15	0.18	1	8022.43	6024.80
sigma1	0.04	0.04	0.01	0.01	0.03	0.05	1	8778.01	6158.22
sigma2	0.06	0.06	0.01	0.01	0.06	0.07	1	8324.20	5875.25
mu2 - mu1	0.00	0.00	0.01	0.01	-0.01	0.02	1	8048.69	5860.81

The analysis showed that on average, there was no difference in amyloid accumulation from baseline to week 52 between individuals in the aerobic exercise condition and individuals in the education control condition. Individuals in the control condition had a posterior mean of 0.16, while individuals in the exercies condition has a posterior mean of 0.17. The 90% CI's were identical for both groups. As a reminder, the actual values are .16 lower than those we see in the table, due to data transformation prior to running our model.