P8131-HW8

Brian Jo Hsuan Lee

4/13/2022

Load packages

```
library(tidyverse)
library(readxl)
```

Import data

```
data =
  read_excel("HW8-HEALTH.xlsx") %>%
  janitor::clean_names() %>%
  mutate(
    id = factor(id),
    time = factor(time),
    txt = factor(txt),
    health = factor(health),
    agegroup = factor(agegroup)
  )
```

a) **Interpret and discuss the bivariate, cross-sectional relationship between group assignment and health self-rating.**

Samples that were given the control treatment (no intervention) had a more evenly split health self-rating responses, where as lower proportion of samples in the intervention treatment reported good health. There are more samples who reported poor health in the intervention group than the control group, even when the total sample count in the control group (41) exceeds that of the intervention group (39). While the treatment assignments were random, the baseline status for the groups are not equivalent and the discrepancy could impact the study conclusions, especially when the within-group progression over time is ignored.

```
data %>%
  filter(
    time == "1"
  ) %>%
  group_by(txt, health) %>%
  summarize(count = n()) %>%
  ggplot(aes(x = health, y = count, fill = health)) +
  geom_col() +
  scale_fill_manual(labels = c("Good", "Poor"), values = c("#41802C", "#A23E14")) +
  facet_grid(cols = vars(txt)) +
  geom_text(aes(label = count), vjust = 3) +
  labs(
    title = "Group Assignment and Health Self-rating at Time of Randomization",
    y = "Count"
  ) +
  theme(
    axis.title.x = element_blank(),
```

```
    axis.text.x = element_blank(),
    axis.ticks.x = element_blank(),
    legend.title = element_blank(),
    legend.position = "bottom",
    plot.title = element_text(size = 11, hjust = 0.5)
  )
```