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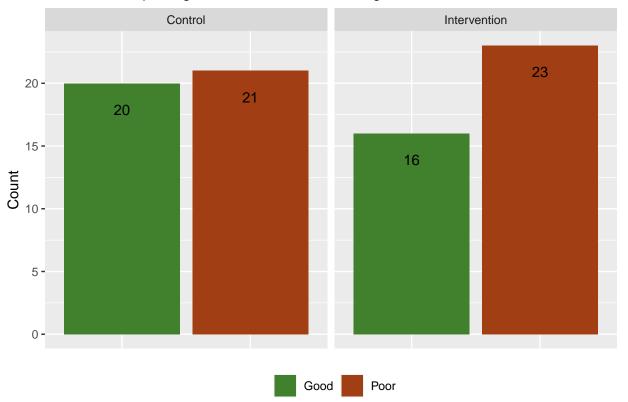
a) **Interpret and discuss the bivariate, cross-sectional relationship between group assignment and health self-rating.**

Samples that were given the control treatment (no educational intervention) had a more even-split health responses, where as lower proportion of samples in the intervention treatment reported good health. By count, 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 2 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)
  )
```

Group Assignment and Health Self–rating at Time of Randomization

b) **Interpret health status over time using a GEE model** Perform a longitudinal data analysis across all study follow-up visits (but not at randomization) to describe the relationship of the participants self-ratings as a function of the effects of health self-rating at the baseline, treatment group, month post randomization, and age group as predictors. Fit a GEE with unstructured correlation structure. Interpret your results.

```
# Create a new column showing baseline health rating, and a new column representing good health as 1, p

data_bl =
  data %>%
  pivot_wider(
    names_from = time,
    values_from = health
  ) %>%
  pivot_longer(
    `2`:`4`,
    names_to = "time",
    values_to = "health"
  ) %>%
  mutate(
```

```r
    time = factor(time),
    nhealth = as.numeric(health == "Good")
) %>%
rename("baseline" = `1`)
```