HW 2

Cory Costello

October 9, 2017

Part 1: Load Data

In addition to loading the data, I'm going to factorize treatment, setting, and sex to make the output a little prettier later.

```
library(rio)
library(tidyverse)

## Loading tidyverse: ggplot2
## Loading tidyverse: tibble
## Loading tidyverse: tidyr
## Loading tidyverse: readr
## Loading tidyverse: purrr
## Loading tidyverse: dplyr

## Conflicts with tidy packages ----------------------------------------------

## filter(): dplyr, stats
## lag():    dplyr, stats

library(knitr)
sesame13 <- import("Sesame13.sav")
sesame13$treatmen <- factorize(sesame13$treatmen)
sesame13$setting <- factorize(sesame13$setting)
sesame13$sex <- factorize (sesame13$sex)
```

Part 2: Difference between treatments at pretest

```
pretest_differences <- sesame13 %>%
  select(starts_with("pre"), treatmen) %>%
  group_by (treatmen) %>%
  summarize(m_prebody  = mean(prebody, na.rm = TRUE),
            m_prelet   = mean(prelet, na.rm = TRUE),
            m_preform  = mean(preform, na.rm = TRUE),
            m_prenumb  = mean(prenumb, na.rm = TRUE),
            m_prerelat = mean(prerelat, na.rm = TRUE),
            m_preclass = mean(preclass, na.rm = TRUE))

kable(pretest_differences, digits = 2)
```

treatmen	m_prebody	m_prelet	m_preform	m_prenumb	m_prerelat	m_preclass
No Encouragement	20.80	17.05	9.65	20.64	10.09	11.81
Encouraged Viewing	21.75	15.30	10.08	21.05	9.85	12.49

Part 3

Part 3a: Average number gain

```
number_gain <- sesame13 %>%
  select(postnumb, prenumb) %>%
  mutate(numb_gain  = postnumb - prenumb) %>%
  summarize(M_numb_gain = mean(numb_gain, na.rm = TRUE))

kable(number_gain, digits = 2)
```

M_numb_gain
9.16

Looks like the average gain in the numbers outcome was 9.16.

Part 3b: Average number gain by treatment

```
number_gain_by_treat <- sesame13 %>%
  select(treatmen, postnumb, prenumb) %>%
  mutate(numb_gain  = postnumb - prenumb) %>%
  group_by(treatmen) %>%
  summarize(m_numb_gain = mean(numb_gain, na.rm = TRUE))

kable(number_gain_by_treat, digits = 2)
```

treatmen	m_numb_gain
No Encouragement	7.97
Encouraged Viewing	9.85

Looks like those people who received the Encouraged Viewing treatment had a higher gain in number ability (*Mean number gain* = 9.85) than those who received no encouragment(*Mean number gain* = 7.97).

Part 3c: Average number gain by setting

```
number_gain_by_setting <- sesame13 %>%
  select(setting, postnumb, prenumb) %>%
  mutate(numb_gain  = postnumb - prenumb) %>%
  group_by(setting) %>%
  summarize(m_numb_gain = mean(numb_gain, na.rm = TRUE))

kable(number_gain_by_setting, digits = 2)
```

setting	m_numb_gain
Home	9.09
School	9.26

Looks like people in different settings had similar gains in number ability. People in the home setting had an average gain of 9.09 and people in the school setting had an average gain of 9.26.

Part 3d: Average number gain by setting and treatment

I'm going to try to get this by just using both setting and treatment in the group by function.

```
number_gain_by_treatxsetting <- sesame13 %>%
  select(treatmen, setting, postnumb, prenumb) %>%
  mutate(numb_gain  = postnumb - prenumb) %>%
  group_by(setting, treatmen) %>%
  summarize(m_numb_gain = mean(numb_gain, na.rm = TRUE))

kable(number_gain_by_treatxsetting, digits = 2)
```

setting	treatmen	m__numb__gain
Home	No Encouragement	7.86
Home	Encouraged Viewing	9.49
School	No Encouragement	8.04
School	Encouraged Viewing	10.73

Huh, looks like that worked. I'm always amazed at how much sense functions fromt the tidyverse make.

Part 4

Part 4a: Gain in letter outcome by sex

```
letter_gain_by_sex <- sesame13 %>%
  select(sex, prelet, postlet) %>%
  mutate(letter_gain = postlet - prelet) %>%
  group_by(sex) %>%
  summarize (m_letter_gain = mean(letter_gain, na.rm = TRUE))

kable(letter_gain_by_sex, digits = 2)
```

sex	m__letter__gain
Male	10.40
Female	11.09

It appears that average letter gain was lower for men (*mean letter gain* $= 10.4$) than it was for women (*mean letter gain* $= 11.09$).

Part 4b: Most variable gain score

```
most_variable_gain_score <- sesame13 %>%
  select(setting, starts_with("pre"), starts_with("post")) %>%
  mutate(body_gain  = postbody - prebody,
```

```
        let_gain   = postlet - prelet,
        form_gain  = postform - preform,
        numb_gain  = postnumb - prenumb,
        relat_gain = postrel - prerelat,
        class_gain = postclas - preclass) %>%
  group_by(setting) %>%
  summarize(sd_body_gain  = sd(body_gain, na.rm = TRUE),
            sd_let_gain   = sd(let_gain, na.rm = TRUE),
            sd_form_gain  = sd(form_gain, na.rm = TRUE),
            sd_numb_gain  = sd(numb_gain, na.rm = TRUE),
            sd_relat_gain = sd(relat_gain, na.rm = TRUE),
            sd_class_gain = sd(class_gain, na.rm = TRUE)) %>%
  gather(key = "gain_score", value = "sd", -setting) %>%
  arrange(desc(sd))
```

```
## Warning: attributes are not identical across measure variables;
## they will be dropped
```

```
kable(most_variable_gain_score, digits =2)
```

setting	gain_score	sd
School	sd_let_gain	11.57
Home	sd_let_gain	10.92
Home	sd_numb_gain	10.10
School	sd_numb_gain	9.08
Home	sd_body_gain	5.48
Home	sd_class_gain	4.70
School	sd_class_gain	4.17
School	sd_body_gain	4.13
Home	sd_form_gain	3.83
School	sd_form_gain	3.61
Home	sd_relat_gain	3.46
School	sd_relat_gain	3.00

As you can see, it looks like letter gain is the most variable in the School setting ($SD = 11.57$) and the Home setting ($SD = 10.92$).

Part 4c: Average ppvt by treatment within site

```
ppvt_by_treatment_wi_site <- sesame13 %>%
  select(treatmen, site, ppvt) %>%
  group_by (site, treatmen) %>%
  summarize (m_ppvt_score = mean(ppvt, na.rm = TRUE))
```

```
kable(ppvt_by_treatment_wi_site, digits = 2)
```

site	treatmen	m_ppvt_score
1	No Encouragement	44.04
1	Encouraged Viewing	44.19
2	No Encouragement	61.11
2	Encouraged Viewing	61.67

site	treatmen	m__ppvt__score
3	No Encouragement	35.86
3	Encouraged Viewing	38.98
4	No Encouragement	45.22
4	Encouraged Viewing	39.25
5	No Encouragement	49.25
5	Encouraged Viewing	51.79

It looks like there ppvt (Peabody Picture-Vocabulary Test) scores were pretty similar within site across treatment conditions. Although I did not formally test this, it suggests that a decent amount of the variability in ppvt scores is due to site rather than treatment. It further suggests that one should use a modeling approach that accounts for the Site (probably an MLM, with students nested in site).