HW3

Hanyu Lu

October 2020

Problem 1

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0

## -- Conflicts ------------------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(arsenal)
```

```
exercise_df = read_csv("./Exercise.csv")
```

```
## Parsed with column specification:
## cols(
##   Group = col_double(),
##   Age = col_double(),
##   Gender = col_double(),
##   Race = col_double(),
##   HTN = col_double(),
##   T2DM = col_double(),
##   Depression = col_double(),
##   Smokes = col_double(),
##   Systolic_PRE = col_double(),
##   Systolic_POST = col_double()
## )
```

```
exercise_df2 = mutate(exercise_df, Group2 = Group) %>%
  mutate(
    Group = as.factor(Group),
    Group2 = as.factor(Group2),
```

```
  )

levels(exercise_df2$Group2) =
  list(Systolic_POST_intervention = "1", Systolic_POST_control = "0")

levels(exercise_df2$Group) =
  list(Systolic_PRE_intervention = "1", Systolic_PRE_control = "0")

exercise_df2  =
exercise_df2  %>%
pivot_wider(
  names_from = Group,
  values_from = Systolic_PRE,
  ) %>%
  pivot_wider(
  names_from = Group2,
  values_from = Systolic_POST,
  )  %>%
  mutate(
    control_difference = Systolic_POST_control - Systolic_PRE_control,
    intervention_difference = Systolic_POST_intervention - Systolic_PRE_intervention
  )

my_controls <- tableby.control(
                total = F,
                test = F,  # No test p-values yet
                numeric.stats = c("meansd", "medianq1q3"),
                stats.labels =
                list(
                meansd = "Mean (SD)",
                medianq1q3 = "Median (Q1, Q3)")
                )

tab2 <- tableby( ~ Systolic_POST_intervention + Systolic_POST_control  + Systolic_PRE_intervention + Sys

summary(tab2, title = "Descriptive Statistics", text = T, digits = 2)
```

```
##
## Table: Descriptive Statistics
##
## |                           |     Overall (N=72)      |
## |:--------------------------|:-----------------------:|
## |Systolic_POST_intervention |                         |
## |-  Mean (SD)               |     125.06 (15.44)      |
## |-  Median (Q1, Q3)         | 124.00 (116.75, 135.00) |
## |Systolic_POST_control      |                         |
## |-  Mean (SD)               |     130.14 (14.35)      |
## |-  Median (Q1, Q3)         | 127.50 (120.00, 140.00) |
## |Systolic_PRE_intervention  |                         |
## |-  Mean (SD)               |     133.64 (15.11)      |
## |-  Median (Q1, Q3)         | 134.00 (121.50, 144.00) |
## |Systolic_PRE_control       |                         |
## |-  Mean (SD)               |     133.47 (15.94)      |
```

```
## |-  Median (Q1, Q3)         | 131.00 (122.50, 143.50) |
## |control_difference         |                         |
## |-  Mean (SD)               |      -3.33 (14.81)      |
## |-  Median (Q1, Q3)         |  -3.50 (-12.25, 8.25)   |
## |intervention_difference    |                         |
## |-  Mean (SD)               |      -8.58 (17.17)      |
## |-  Median (Q1, Q3)         |  -5.50 (-23.00, 3.00)   |
```

a) Perform appropriate tests to assess if the Systolic BP at 6 months is significantly different from the baseline values for each of the groups:
b) Intervention group (5p)

Since we don't know true population variance. We are going to use paired t-test because we intend to compare scores on two different variables but on the same group. Additionally, we test for the mean of the differences with unknown variance.

H_0: the Systolic BP at 6 months is equal to the baseline values for intervention group H_1: the Systolic BP at 6 months is significantly different from the baseline values for intervention group

$\bar{d} = \sum_{i=1}^{n} d_i/n$ = -8.58 $s_d = \sqrt{\sum_{i=1}^{n}(d_i - \bar{d})^2/(n-1)}$ = 17.17

$t = \frac{\bar{d}-0}{s_d/\sqrt{n}} = \frac{-8.58-0}{17.17/\sqrt{36}}$ = -3

$t_{36-1,0.975}$ = 2.03

Since this t-test is two-sided, |t| = 3 > $t_{36-1,0.975}$ = 2.03.

We can reject H_0. We can conclude that the Systolic BP at 6 months is significantly different from the baseline values for intervention group

```
sd(exercise_df2$intervention_difference, na.rm = TRUE)
```

```
## [1] 17.1687
```

```
-8.58/(17.17/6)
```

```
## [1] -2.998253
```

```
qt(0.975,35)
```

```
## [1] 2.030108
```

ii) Control group (5p)

Since we don't know true population variance. We are going to use paired t-test because we intend to compare scores on two different variables but on the same group. Additionally, we test for the mean of the differences with unknown variance.

H_0: the Systolic BP at 6 months is equal to the baseline values for control group H_1: the Systolic BP at 6 months is significantly different from the baseline values for control group

$\bar{d} = \sum_{i=1}^{n} d_i/n$ = -3.33 $s_d = \sqrt{\sum_{i=1}^{n}(d_i - \bar{d})^2/(n-1)}$ = 14.81

$t = \frac{\bar{d}-0}{s_d/\sqrt{n}} = \frac{-3.33-0}{14.81/\sqrt{36}}$ = -1.35

$t_{36-1,0.975} = 2.03$

Since this t-test is two-sided, $|t| = 1.35 < t_{36-1,0.975} = 2.03$.

We cannot reject H_0. We can conclude that the Systolic BP at 6 months is not significantly different from the baseline values for intervention group

```
sd(exercise_df2$control_difference, na.rm = TRUE)
```

```
## [1] 14.81312
```

```
-3.33/(14.81/6)
```

```
## [1] -1.349088
```

```
qt(0.975,35)
```

```
## [1] 2.030108
```