p8130_hw4_jn2761

jiaying Ning

11/6/2020

```
rm(list=ls())
```

```
library(readxl)
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)
```

Problem 2

part a

```
#Import Data
Knee_df=
   read.csv("./data/Knee.csv")
```

```
summary(Knee_df)
```

```
##      Below       Average          Above
##  Min.   :29   Min.   :28.00   Min.   :20.00
##  1st Qu.:36   1st Qu.:30.25   1st Qu.:21.00
##  Median :40   Median :32.00   Median :22.00
##  Mean   :38   Mean   :33.00   Mean   :23.57
##  3rd Qu.:42   3rd Qu.:35.00   3rd Qu.:24.50
##  Max.   :43   Max.   :39.00   Max.   :32.00
##  NA's   :2                    NA's   :3
```

From the current descriptive statistics, it seems like lower physical status are associate with longer time (days) required in physical therapy until successful rehabilitation on average. This trend is observable across all descriptive data (min, median,mean,max) Those who are below average physical status have the higher value in min, median, mean, and max for time (days) required in physical therapy when comparing to the other group.

part b

- Hypothesis
  - H0 = Levels of Physical status are not associated with days required in physical therapy until successful rehabilitation.
  - H1 = Levels of Physical status are associated with days required in physical therapy until successful rehabilitation.
- Anova Table/Test Statistics

```
#Tidy the data frame
Knee_Anova=
Knee_df %>%
  pivot_longer(
    Below:Above,
    names_to = "Physical_Status",
    values_to = "Days_in_Physical_Therapy")

#perform anova test
teststats = aov(Days_in_Physical_Therapy ~ Physical_Status, data = Knee_Anova)
summary(teststats)
```

```
##                 Df Sum Sq Mean Sq F value   Pr(>F)
## Physical_Status  2  795.2   397.6   19.28 1.45e-05 ***
## Residuals       22  453.7    20.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 5 observations deleted due to missingness
```

- Critical Value

```
F_critic=qf(.99, df1=2, df2=22)
F_critic
```

```
## [1] 5.719022
```

F_critic(5.719022) < F_stats(19.28)

- Decision: Since out F-stats is larger than F critical value with alpha=0.01, we thus reject the null and conclude that