Homework #1

Steven Lawrence

September 17, 2019

Contents

Loading Libraries

```
library(tidyverse)
library(arsenal)
require(knitr)
require(survival)
```

Loading in data set

```
#rawE<- readxl::read_xlsx("data/Exercise.xlsx", col_names = TRUE)
exercise = readxl::read_xlsx("data/Exercise.xlsx",col_names = TRUE, skip = 1)
exercise = janitor::clean_names(exercise)
#skimr::skim(exercise)
#str(exercise)
#rawE
```

Problem 1 i

Here I renamed the variables.

```
exercise$group<-as.factor(exercise$group)
exercise$gender<-as.factor(exercise$gender)
exercise$race<-as.numeric(exercise$race)
exercise$htn<-as.factor(exercise$htn)
exercise$depression<-as.factor(exercise$depression)
exercise$smokes<-as.factor(exercise$smokes)
exercise$t2dm<-as.factor(exercise$t2dm)

attach(exercise)
exercise$groupf[group ==1]<- "Intervention Group"
exercise$groupf[group ==0]<- "Control"
```

```r
exercise$genderf[gender == 1]<- "Male"
exercise$genderf[gender == 2]<- "Female"

exercise$racef[race == 1] <- "African American"
exercise$racef[is.na(exercise$racef)]<- 0
exercise$racef[race == 2] <- "Hispanic"
exercise$racef[race == 3] <- "African American"
exercise$racef[race == 4] <- "Caucasian"
exercise$racef[race == 5] <- "Other"
exercise$racef[race==6 ]<- "Other"

exercise$htnf[htn == 1]<- "Yes"
exercise$htnf[htn == 0]<- "No"
exercise$t2dmf[t2dm == 1]<- "Yes"
exercise$t2dmf[t2dm == 0]<- "No"
exercise$depressionf[depression == 1]<- "Yes"
exercise$depressionf[depression == 0]<- "No"
exercise$smokesf[smokes == 1]<- "Yes"
exercise$smokesf[smokes == 0]<- "No"
detach(exercise)
```

Discriptive Statistics

Demographic Table

```r
mylabels <- list(groupf = "Treatment Group", age = "Age", genderf = "Gender
Male n(%)", htnf = "Hypertention Yes n(%)", t2dmf = "Type 2 Diabetes Yes
n(%)", depressionf = "Depression Yes n(%)", racef = "Race Hispanic n(%)",
smokesf ="Smoking Status Yes n(%)")

mycontrols  <- tableby.control(test=T, total=F, numeric.simplify = TRUE,
cat.simplify = T,
                               numeric.test="kwt", cat.test="chisq",
                               numeric.stats=c( "meansd","medianq1q3"),
                               cat.stats=c("countpct"),
                               stats.labels=list(N='Count', meansd='Mean',
countpct = "Count", medianq1q3 = "Median[IQR]"),
                               digits = 2,
                               digits.count = 2,
                               digits.p = 3
                               )

tab1<-tableby(groupf ~ age + genderf + racef + depressionf + smokesf + htnf +
t2dmf, control = mycontrols, data = exercise)

summary(tab1, text = TRUE, labelTranslations = mylabels)
```

	Control (N=36)	Intervention Group (N=36)	p value
Age			0.488

- Mean	51.50 (10.81)	53.58 (9.58)	
- Median[IQR]	51.00 (44.75, 60.25)	55.50 (47.50, 59.25)	
Gender Male n(%)	16.00 (44.4%)	16.00 (44.4%)	1.000
Race Hispanic n(%)	14.00 (38.9%)	5.00 (13.9%)	0.016
Depression Yes n(%)	13.00 (36.1%)	10.00 (27.8%)	0.448
Smoking Status Yes n(%)	5.00 (13.9%)	5.00 (13.9%)	1.000
Hypertention Yes n(%)	20.00 (55.6%)	22.00 (61.1%)	0.633
Type 2 Diabetes Yes n(%)	19.00 (52.8%)	13.00 (36.1%)	0.155

Metabolic parameters

Here is how I renamed the variables

```
exercise<-rename(exercise, pre_LDL =pre_17 , pre_BMI = pre_13 , post_BMI
=post_14  ,post_LDL =post_18, pre_systolic = pre_9, post_systolic = post_10,
pre_diastolic= pre_11, post_diastolic= post_12, pre_glucose = pre_19,
post_glucose= post_20, pre_HDL = pre_15, post_HDL = post_16)
```

This code produced a table which is attached further along.

```
tab2<-tableby(groupf ~ pre_BMI + post_BMI + pre_LDL + post_LDL + pre_systolic
+ post_systolic + pre_diastolic + post_diastolic + pre_glucose + post_glucose
+ pre_HDL + post_HDL, control = mycontrols, data = exercise)

summary(tab2, text = TRUE, labelTranslations = mylabels)
```

This code is the change in metabolic measures by treatment group where the values are in the table below.

```
exercise<- exercise %>%  mutate(
  bmi_change = post_BMI- pre_BMI,
  ldl_change = post_LDL -pre_LDL,
  hdl_change = post_HDL - pre_HDL,
  systolic_change = post_systolic - pre_systolic,
  diastolic_change= post_diastolic - pre_diastolic,
  glucose_change = post_glucose - pre_glucose

)

tab3<- tableby(groupf ~ bmi_change +ldl_change + hdl_change +systolic_change+
diastolic_change+ glucose_change, control = mycontrols, data = exercise)

summary(tab3, text = T, labelTranslations =  mylabels)
```

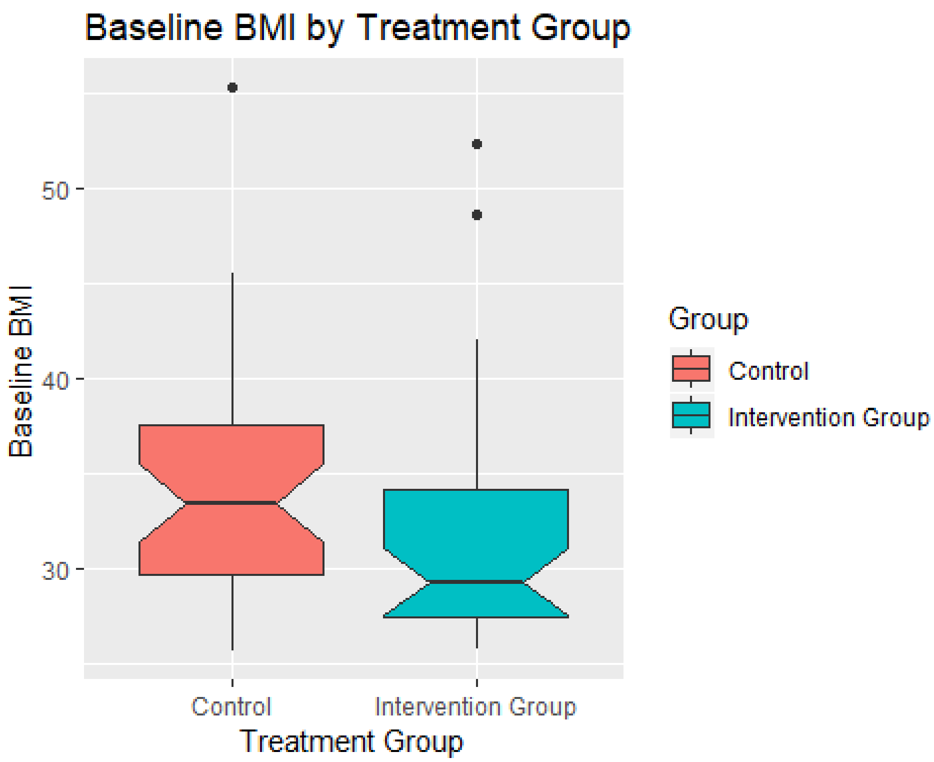
	Intervention Group		Control	
	N = 36		N = 36	
	Basline	6 month	Basline	6 month
BMI	31.97 ± 6.58	31.21 ± 6.13	34.23 ± 6.16	34.51 ± 5.97
	29.25 (27.37 - 34.10)	29.15 (26.80 - 32.88)	33.40 (29.60 - 37.58)	33.05 (30.43 - 37.55)
Δ	-0.76 ± 1.44		0.28 ± 0.97	
Systolic	133.64 ± 15.11	125.06 ± 15.44	133.47 ± 15.94	130.14 ± 14.35
	134 (121.50 - 144.00)	124 (116.75 - 135)	131.00 (122.50 - 143.50)	127.50 (120 - 140)
Δ	-8.58 ± 17.17		-3.33 ± 14.81	
Diastolic	75.44 ± 9.10	74.58 ± 12.37	77.14 ± 9.66	75.69 ± 7.54
	74.50 (69 - 81)	74.00 (65.00 - 80.50)	76.00 (68.75 - 85.00)	76.50 (69.00 - 82.00)
Δ	-0.86 ± 8.30		-1.44 ± 10.11	
Glucose	116.64 ± 74.91	107.14 ± 38.65	128.97 ± 73.86	126.61 ± 63.96
	94.00 (83.75 -116.50)	95.50 (85.25 -129.00)	98.00 (81.75 - 139.00)	106.50 (85.00 - 145.75)
Δ	-9.50 ± 57.36		-2.36 ± 51.22	
HDL	50.17 ± 11.85	50.17 ± 13.07	48.33 ± 13.70	45.19 ± 10.78
	47.50 (40.00 - 60.00)	48.50 (43.00 - 60.25)	43.50 (39.00 - 54.25)	43.50 (38.00 - 52.00)
Δ	0.00 ± 8.09		-3.14 ± 6.91	
LDL	102.94 ± 33.84	100.50 ± 30.39	99.83 ± 29.06	93.61 ± 27.47
	109.00 (75.25 - 124.50)	95.00 (76.50 -120.50)	104.00 (88.25 - 112.25)	96.50 (77.50 - 110.25)
Δ	-2.44 ± 21.27		-6.22 ± 23.12	

Problem 1 ii

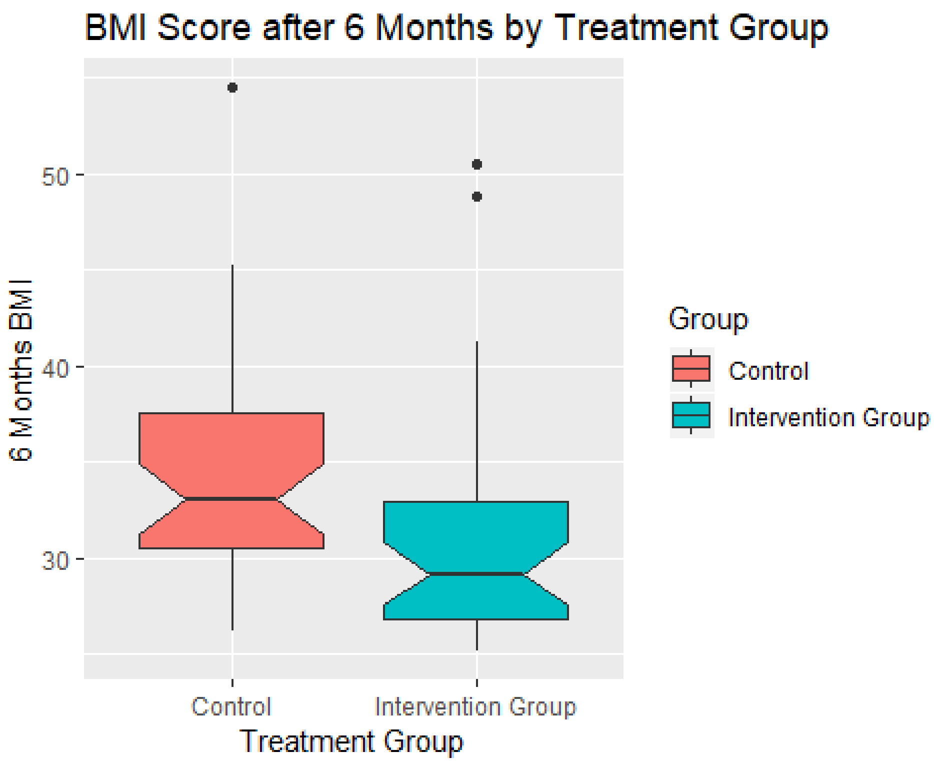
Box Plots

Here I plot BMI and LDL at baseline and after 6 months for both Treatment groups individually

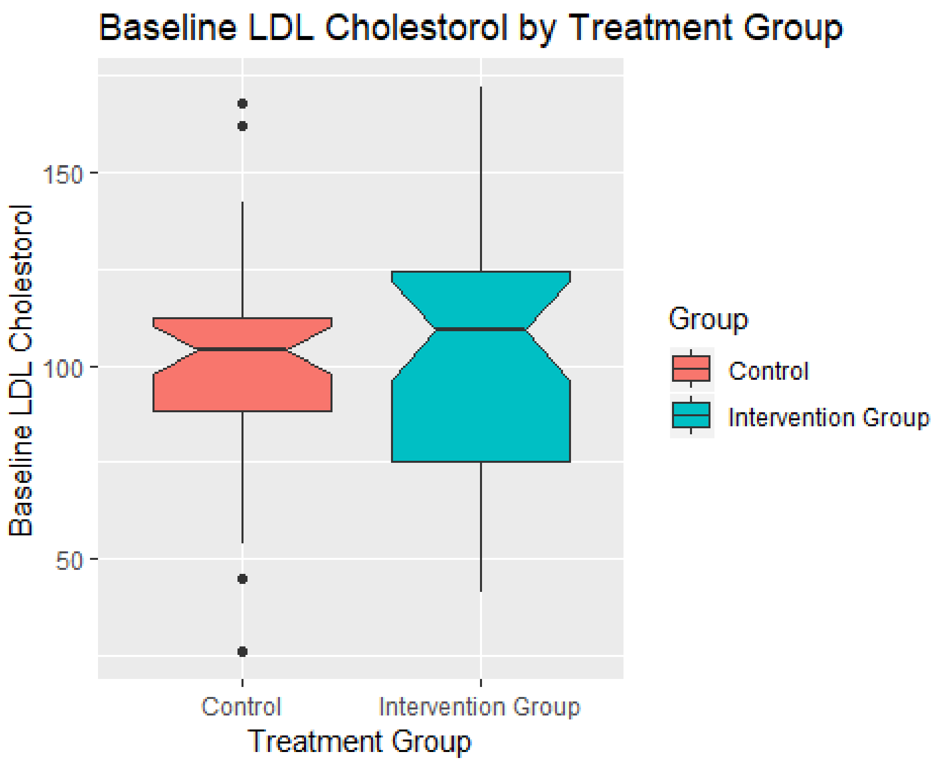
```
exercise %>%
  ggplot(aes(
              groupf,
              pre_BMI,
              fill = groupf
              ))+
  geom_boxplot(notch = T)+
      labs(x = "Treatment Group",
          y = "Baseline BMI",
          title = "Baseline BMI by Treatment Group",
          fill = "Group")
```



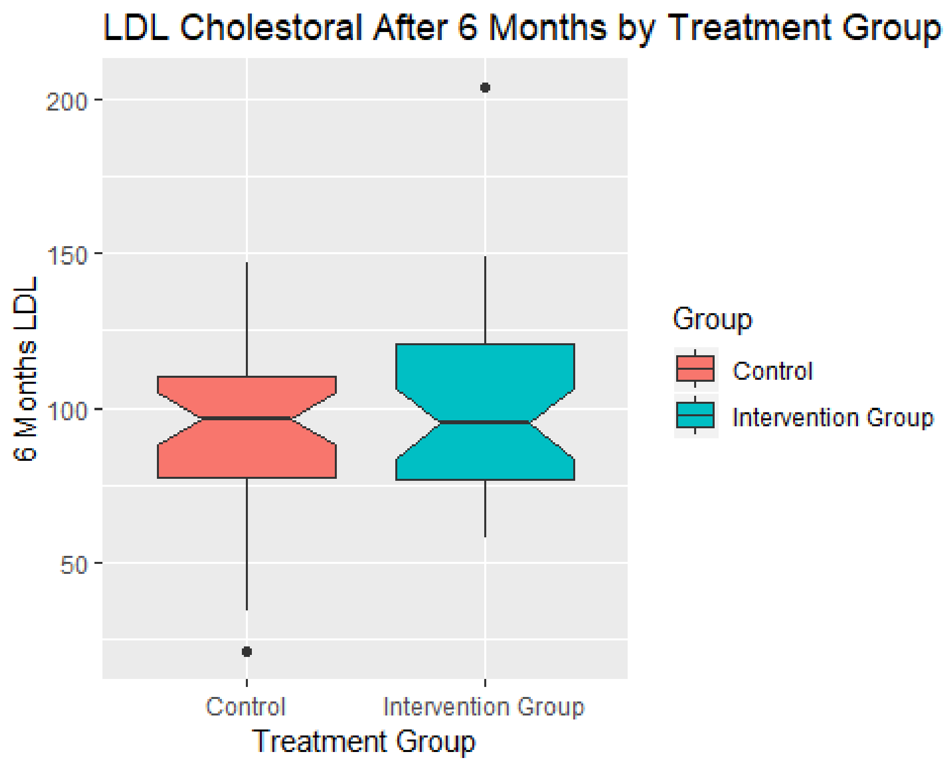

```
exercise %>%
  ggplot(aes(
              groupf,
              post_BMI,
              fill = groupf
              ))+
  geom_boxplot(notch = T)+
      labs(x = "Treatment Group",
          y = "6 Months BMI",
          title = "BMI Score after 6 Months by Treatment Group",
          fill = "Group")
```




```
exercise %>%
  ggplot(aes(
            groupf,
            pre_LDL,
            fill = groupf
            ))+
  geom_boxplot(notch = T)+
      labs(x = "Treatment Group",
          y = "Baseline LDL Cholestorol",
          title = "Baseline LDL Cholestorol by Treatment Group",
          fill = "Group")
```




```
exercise %>%
  ggplot(aes(
              groupf,
              post_LDL,
              fill = groupf
              ))+
  geom_boxplot(notch = T)+
      labs(x = "Treatment Group",
          y = "6 Months LDL",
          title = "LDL Cholestoral After 6 Months by Treatment Group",
          fill = "Group")
```



Problem 1 iii

Discussion of Findings in parts i) and ii).

Demographically there is no differce of distribution between the two treatment groups except for race using non parametric tests to asses differences numerically. This suggest that the findings from this sample are mostly derived from the African Americans who participated in the study.

As for the metabolic measures, the BMI, systolic blood pressure and glucose levels of those who participated in the study decreased over time more than those who did not participate. The Diastolic blood pressure, LDL and HDL choloesterol levels of those who participated in the study differed by an increase in those measures from those who did not participate.

Problem 2

Calculating the PPV of the Triple Test

Information given

In this problem we note the following information that will be used to calculate the positive predictive value.

The prevalecne of Downsyndrome is `0.001` and the more reliable test, that cuases less deaths, has a sensitivity of `0.60` and a specificity of `0.95` given an false postive value of '0.05'. The problem is asking what is the probability that a baby will be born with downsyndrome given that they have a positive test value `p(DS+| T+)`

Calculation of PPV

The positive predictive value (PPV) is 0.012

```
sen <- 0.6
spe <- 1 - 0.05
prev <- 1/1000
signif(sen*prev/( sen*prev + (1 -spe)*(1-prev)), digit = 2)
```

Explanation

The PPV of `1.2%` means that useing this test the probability of a baby having Downsyndrome given a postive test score is `1.2%`. In other words theres a small chance that a baby will actually have downsyndrome at 16 weeks of pregnacy using the triple test.

Problem 3

The Impact of Climate change on Human Health

Johnathan Patz et.al speak on the effects of climate change on human health. There premise for this study was based on the estimates of the World Health Organisation that claims that anthropoligical climate change takes 150,000 per year. TWo of the main issues suggested in this paper, attributied to climate change, is heat waves and crop failure. The authors aim to fill the gap of long term and reliable climate datasets by providing a more detailed means of analysis for the effect of climate change on volunerable populations.

Along with an increase in temperature comes an increase in infectious diseases [Patz et.al]. This leaves countries that are alread prone to have exrtreem temperatures to be worsen. Figure 1 shows the effects of an increase in climate temperature positively correlates with mosquito population size with the fluctuations of temperature. The authors suggest that this will only become more evident with time and that countries that are undeveloped and close to the equator will recieved most of the impact of climate change in the near future.

Reference

Patz, J. A., Campbell-Lendrum, D., Holloway, T., & Foley, J. A. (2005). Impact of regional climate change on human health. Nature, 438(7066), 310.