Assignment 1

Mayaluna Bierlich

February 7, 2019

Question 1

```
b_data <- read.csv("boston.csv")
income.control.tab <- table(income = b_data$income, control = b_data$treatment)
```

```
##Here is the mean and the median for income in the control group

mean(b_data$income[b_data$treatment == 0])
```

```
## [1] 151154.4
```

```
median(b_data$income[b_data$treatment == 0])
```

```
## [1] 135000
```

```
##Here is the mean and the median for income in the treatment group

mean(b_data$income[b_data$treatment == 1])
```

```
## [1] 135181.8
```

```
median(b_data$income[b_data$treatment == 1])
```

```
## [1] 135000
```

The median here is exactly the same, while the mean differs a bit. However, this difference is small enough to conclude that income is controlled for

Next we have the proportion of men in the control group and the treatment group.

```
males.both <- table(treatment = b_data$treatment, males = b_data$male)
```

```
##           males
## treatment  0  1
##          0 28 40
##          1 26 29
```

```
## control group proportion of men ##

males.both[2,1]/sum(males.both[,1])
```

```
## [1] 0.4814815
```

```
## treatment group proportion of men ##

males.both[2,2]/sum(males.both[,2])
```

```
## [1] 0.4202899
```

As one can see, the proportion of men in the control group is ~48%, while the treatment group has ~42% proportion of men. Gender can have an effect on political views, such as on immigration, so it is vital that one ensures the `covariate balance`

Question 2

```
control_dif_means <- (mean(b_data$numberim.post[b_data$treatment == 0], na.rm = T)
                      -
                        mean(b_data$numberim.pre[b_data$treatment == 0], na.rm = T))


treatment_dif_means <- (mean(b_data$numberim.post[b_data$treatment == 1], na.rm = T)
                        -
                          mean(b_data$numberim.pre[b_data$treatment == 1], na.rm = T))
```

```
## [1] -0.1873708
```

For the control group, we can see that the difference between the mean of answers before and after results in a negative number. This means that support for more immigration actually increased.

```
## [1] 0.08128342
```

The opposite is true for the treatment group, as the positive difference between the means that the treatment group actually showed a decrease in support for more immigartion after receiving the treatment.

However, the difference in means for the treatment group is ~0.08, which is not a statistically significant number, and thus the treatment made no difference.

Question 3

```
b_data$type <- NA
b_data$type[b_data$college == 1 & b_data$treatment == 0] <- "EduControl"
b_data$type[b_data$college == 0 & b_data$treatment == 0] <- "UneduControl"
b_data$type[b_data$college == 1 & b_data$treatment == 1] <- "EduTreatment"
b_data$type[b_data$college == 0 & b_data$treatment == 1] <- "UneduTreatment"

b_data$type <- as.factor(b_data$type)


tapply(b_data$numberim.pre, b_data$type, mean, na.rm = T)
```

```
##     EduControl   EduTreatment   UneduControl UneduTreatment
##       2.833333       3.021277       3.571429       3.125000
```

```r
tapply(b_data$numberim.post, b_data$type, mean, na.rm = T)
```

```
##     EduControl   EduTreatment   UneduControl UneduTreatment
##       2.620690       3.113636       3.571429       3.142857
```

Considering the values above, one can observe that there is not much significant change in the means when one considers education. Both educated and uneducated control groups either increase in support or stay at the same level of support. Both treatment groups actually decreased in support for more immigration, with educated showing a sharper decrease.

That being said, all of these differences are way too low to be statistically significant.