RHW3

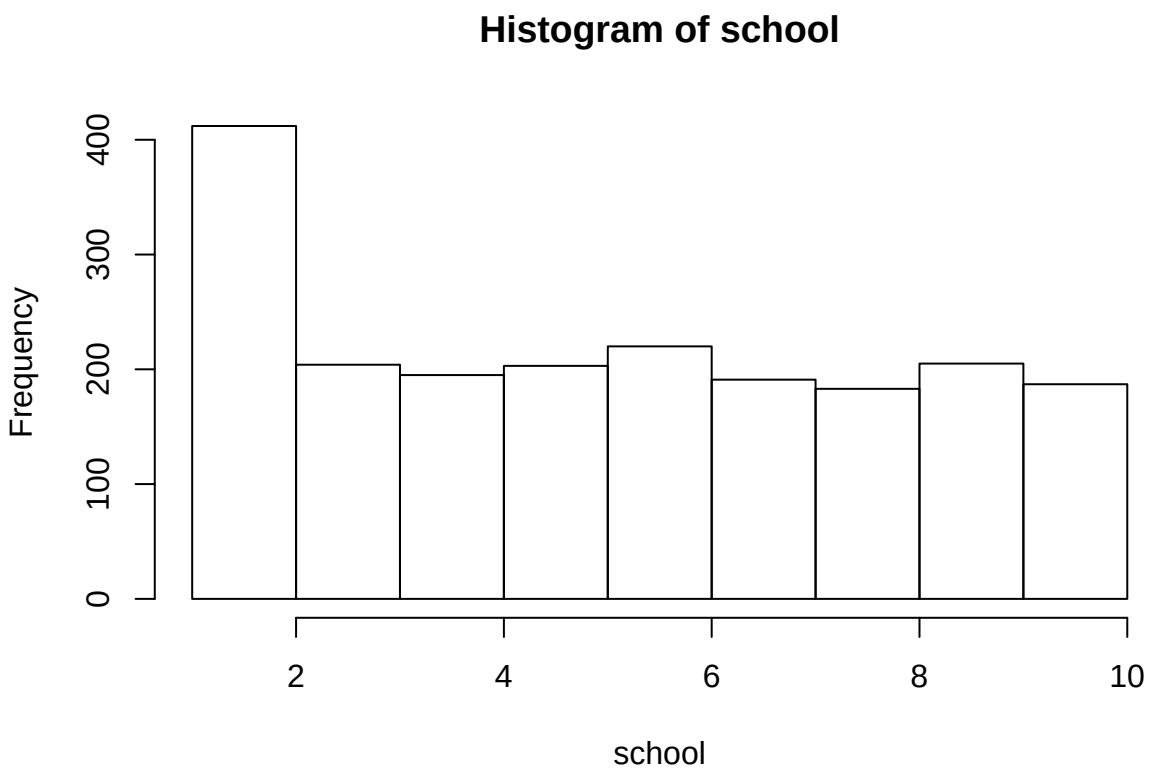
Rabia Suleiman

8/23/2019

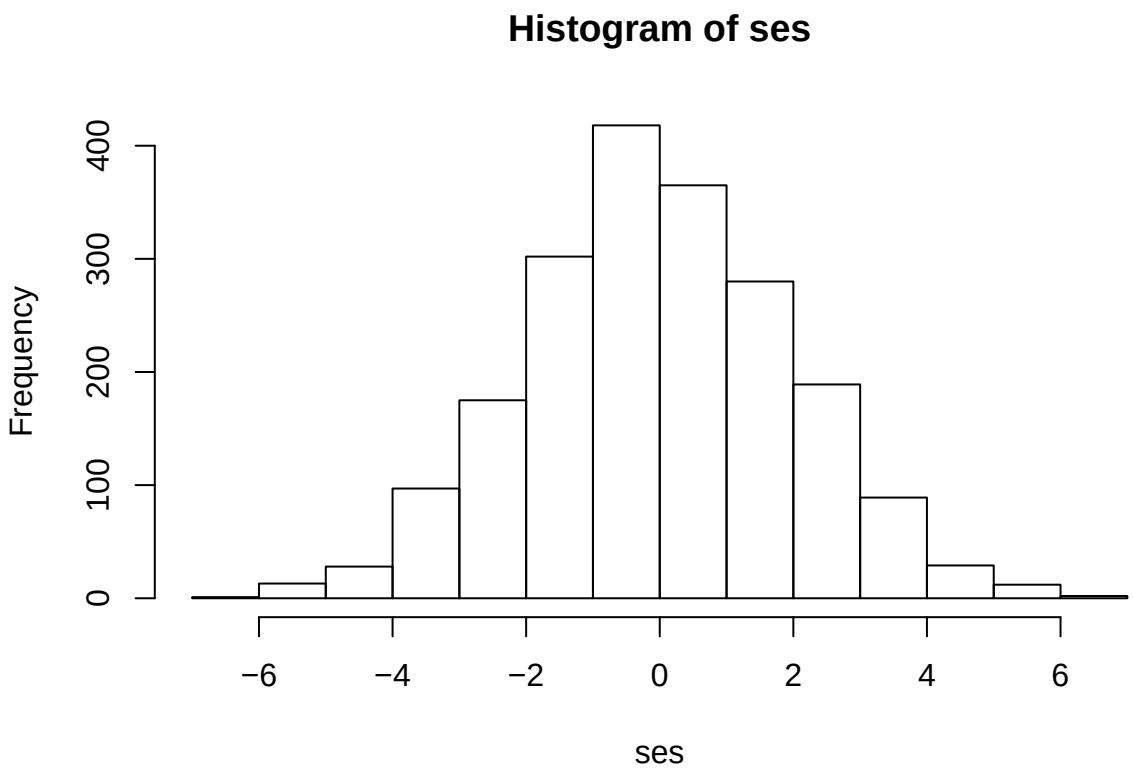
R Markdown

Schools are a stabilizing feature in the unsettled lives of immigrant youth. They provide a safe space, learning opportunities, and diverse interactions. Extensive research indicares that educators and the school environment are important components in promoting socialization and acculturation of immigrant youth. The question I have stems from a current research project which seeks to determine the effectiveness of a specific intervention method. My population consists of school aged immigrants and whether the educational intervention led to better grade outcomes in their first year of schooling in Houston. The dependent variable is a cumulative grade (0<y<100) the independent variable is their enrollment into the program. control variables include school choice, English language ability and socioeconomic status. First we create Variables

```
rm(list=ls())
set.seed(648057)
N = 2000
N_samp <- 50
#covariates
school=sample(1:10, N, replace=TRUE) # School ID (10 total)
hist(school)
```




```
ela <- rbinom(n=N, 1, 0.2)              # English Language ability
ses <- rnorm(n=N, mean=0, sd=2)         # Socio economic status
hist (ses)
```




```
a = 55.5
b = 5
b2 = 3.5
b3= .4
tau = 6
e = rnorm(n=N, mean=0, sd=1)

#no intervention possible outcomes
Y_0 <- a + tau*0 + (b*ela) + (b2*ses) + (b3*school) + e

#intervention outcomes
Y_1 <- a + tau*1 + (b*ela) + (b2*ses) + (b3*school) + e  # intervention possible outcomes
```

Including Plots

You can also embed plots, for example:

```
pop_dat <- data.frame(Y_0, Y_1, ela, ses, school)
head(pop_dat)
```

```
##        Y_0      Y_1 ela        ses school
## 1 62.97775 68.97775   1  0.3646740      2
## 2 45.19290 51.19290   0 -3.7789375      6
## 3 66.79468 72.79468   0  2.4982980      6
## 4 51.73052 57.73052   0 -1.3000532      1
## 5 56.96641 62.96641   1 -1.1379328      1
## 6 62.23956 68.23956   0  0.3860879      8
```

```
# Now, we consider our sample from here #GIVEN this vector I want to randomly
sample_ind <- sample(1:nrow(pop_dat), size=N_samp)     # Sampling index
```

```r
sample_dat <- pop_dat[sample_ind, ]     #the sample is the population data # Only keep obs that match t
head(sample_dat)
```

```
##           Y_0      Y_1 ela        ses school
## 189  58.19481 64.19481   0  0.2186004      3
## 1405 62.32110 68.32110   1  0.3956277      4
## 810  57.30187 63.30187   0 -0.2899689      3
## 727  53.90615 59.90615   0 -0.8675956      6
## 1198 62.00015 68.00015   0  2.0371102      2
## 525  42.13054 48.13054   0 -4.2900712      2
```

```r
d <- ifelse(runif(N_samp)<=0.2, 1, 0)

# Treatment assignment indicator (1=Treated, 0=Controlled) the .2 is the indicator
sample_dat$Y_obs <- d*sample_dat$Y_1 + (1-d)*sample_dat$Y_0  # Observed outcomes
                                                             # Y_obs = d*Y_1 + (1-d)*Y_0
sample_dat$Status <- ifelse(d==1, "Treated", "Control")      # Copying the treatment status into sample

head(sample_dat)
```

```
##           Y_0      Y_1 ela        ses school    Y_obs  Status
## 189  58.19481 64.19481   0  0.2186004      3 58.19481 Control
## 1405 62.32110 68.32110   1  0.3956277      4 62.32110 Control
## 810  57.30187 63.30187   0 -0.2899689      3 57.30187 Control
## 727  53.90615 59.90615   0 -0.8675956      6 53.90615 Control
## 1198 62.00015 68.00015   0  2.0371102      2 62.00015 Control
## 525  42.13054 48.13054   0 -4.2900712      2 42.13054 Control
```

```r
t.test(Y_obs ~ Status, data=sample_dat)
```

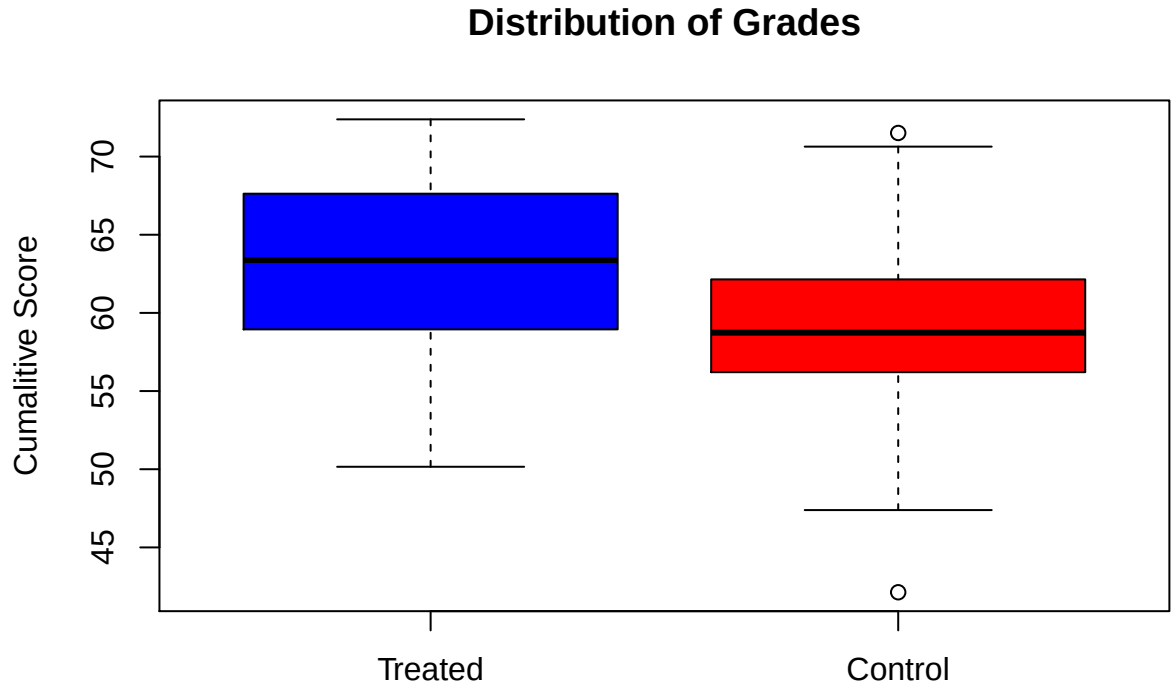
```
##
##  Welch Two Sample t-test
##
## data:  Y_obs by Status
## t = -1.1534, df = 7.2415, p-value = 0.2854
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -10.580430   3.611118
## sample estimates:
## mean in group Control mean in group Treated
##              59.23857              62.72322
```

```r
# Estimating the average treatment effect
tau_ate <- mean(sample_dat$Y_obs[sample_dat$Status=="Treated"]) -
        mean(sample_dat$Y_obs[sample_dat$Status=="Control"])
tau_ate
```

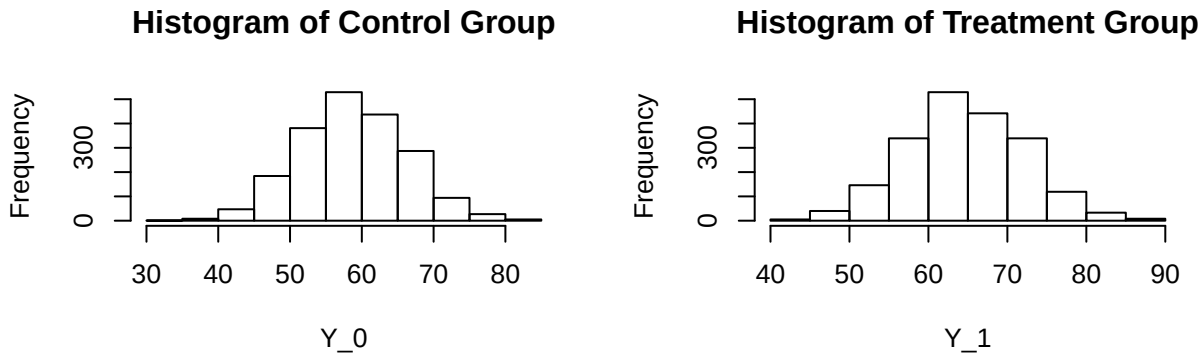
```
## [1] 3.484656
```

Now we visualize the impact of the treatment

```r
boxplot(sample_dat$Y_obs[sample_dat$Status=="Treated"],
        sample_dat$Y_obs[sample_dat$Status=="Control"],
        names=c("Treated", "Control"), ylab="Cumalitive Score",
        main="Distribution of Grades",
        col=c("blue", "red"))
```

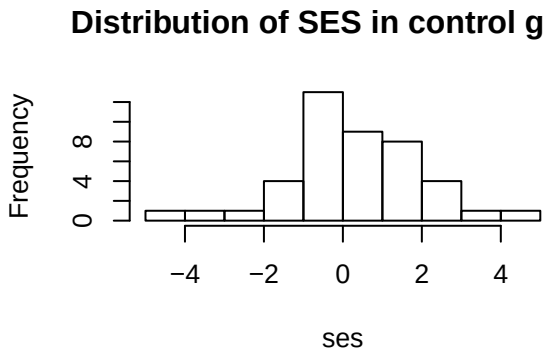



```
  par(mfrow=c(2,2))    # Specify the graphic parameter
  hist(Y_0, xlab=deparse(substitute(Y_0)),
       main=paste0("Histogram of", sep=" Control Group", substitute()))
  hist(Y_1, xlab=deparse(substitute(Y_1)),
       main=paste0("Histogram of", sep=" Treatment Group", substitute()))
```



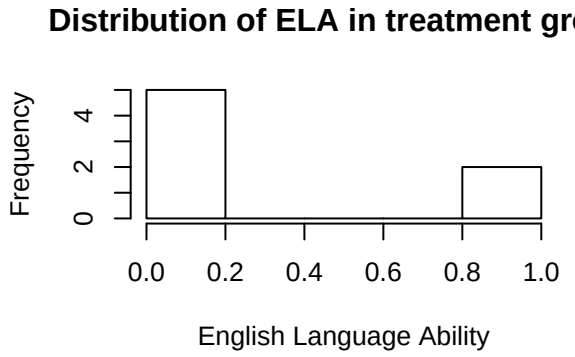
Now we can see the variation in contol variables between the two groups.

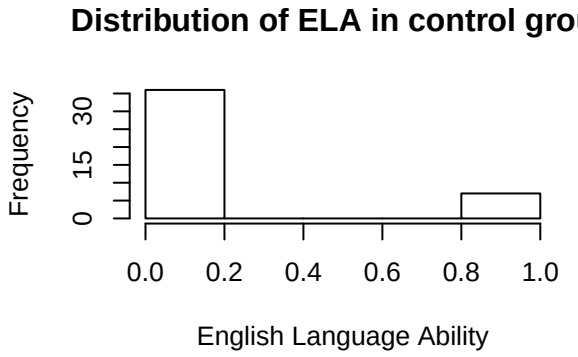
```
 par(mfrow=c(2,2))
hist(sample_dat$ses[sample_dat$Status=="Treated"],
     main= "Distribution of SES in treatment group ", xlab="ses")
hist(sample_dat$ses[sample_dat$Status=="Control"],
    main= "Distribution of SES in control group",
    xlab="ses")
```



“‘

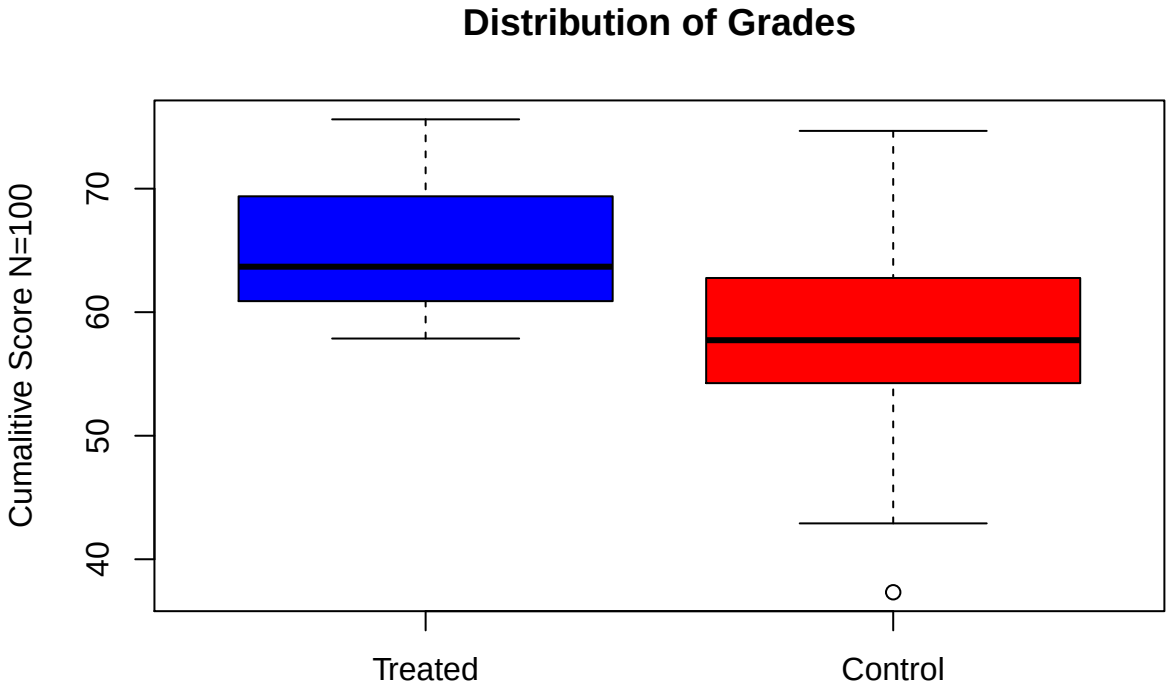
```
par(mfrow=c(2,2))
hist(sample_dat$ela[sample_dat$Status=="Treated"],
     main= "Distribution of ELA in treatment group ", xlab="English Language Ability")
hist(sample_dat$ela[sample_dat$Status=="Control"],
    main= "Distribution of ELA in control group",
    xlab="English Language Ability")
```





Now we replicate with other sample sizes

```
##           Y_0      Y_1 ela        ses school
## 1083 57.70080 63.70080   1 -1.4321008      4
## 261  62.03209 68.03209   0  0.7911216      9
## 80   57.76423 63.76423   0  0.6413876      1
## 467  67.32541 73.32541   0  2.6987369      5
## 1995 64.63977 70.63977   0  1.7390773      8
## 970  66.75029 72.75029   0  2.6359049      7
```




```
##           Y_0      Y_1 ela        ses school
## 1259 70.01849 76.01849   1  2.3790221      4
## 994  53.32917 59.32917   0 -0.3531940      2
## 557  58.17062 64.17062   0  0.5682245      4
## 1429 62.15532 68.15532   0  0.6055224      7
## 595  67.35448 73.35448   0  1.9877506     10
## 241  62.15008 68.15008   0  1.0353184      8
```