HKS SUP-135 Lab 4: The Tennessee STAR Experiment

Matt Khinda

2/25/2023

Question 1: The Need for Complete Randomization

In the case of the Tennessee STAR Experiment, it is important that teachers be randomly assigned to the control or treatment group (along with students) because not doing so could introduce selection bias. This might look like teachers deciding to teach smaller or larger classes based on a pre-existing preference and knowledge that they teach better under certain classroom conditions. This motivation would act as a confounding variable and obfuscate the validity of the results because in that case it may not be the class size itself that has a causal effect on outcomes, but rather (in part or in whole) the teachers' comfort with the class size.

Question 2: Average class sizes

```
avg_small_class <- mean(star$class_size[star$small == 1])
avg_reg_class <- mean(star$class_size[star$small == 0])

cat("The average class size for small kindergarten classes is", avg_small_class, "
    while for regular kindergarten classes it is", avg_reg_class)
```

```
## The average class size for small kindergarten classes is 15.09576
##     while for regular kindergarten classes it is 22.5232
```

Question 3: SAT Score Index

```


ctrl_grp <- subset(star, small == 0)

math_ctrl_mean <- mean(ctrl_grp$math, na.rm = T)
math_ctrl_sd <- sd(ctrl_grp$math, na.rm = T)

read_ctrl_mean <- mean(ctrl_grp$read, na.rm = T)
read_ctrl_sd <- sd(ctrl_grp$read, na.rm = T)

word_ctrl_mean <- mean(ctrl_grp$wordskill, na.rm = T)
word_ctrl_sd <- sd(ctrl_grp$wordskill, na.rm = T)

listen_ctrl_mean <- mean(ctrl_grp$listen, na.rm = T)
listen_ctrl_sd <- sd(ctrl_grp$listen, na.rm = T)
```

```
star$std_math <- (star$math - math_ctrl_mean)/math_ctrl_sd
star$std_read <- (star$read - read_ctrl_mean)/read_ctrl_sd
star$std_word <- (star$wordskill - word_ctrl_mean)/word_ctrl_sd
star$std_listen <- (star$listen - listen_ctrl_mean)/listen_ctrl_sd
```

3a: Standardized Scores

```
## [1] "Standardized Math Score"
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -3.49539 -0.62800  0.01395  0.05194  0.63451  3.05253
```

```
## [1] "Standardized Reading Score"
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -2.47023 -0.68020 -0.07287  0.05089  0.56643  6.12831
```

```
## [1] "Standardized Wordskill Score"
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -2.79641 -0.76367 -0.15934  0.04772  0.69222  4.40060
```

```
## [1] "Standardized Listening Score"
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -4.22024 -0.61906 -0.01382  0.03368  0.71247  4.07155
```

```
# Calculate composite score based on mean of standardized scores
for (x in 1:nrow(star)){
  star$sat_index[x] <- mean(c(star$std_math[x],
                              star$std_read[x],
                              star$std_word[x],
                              star$std_listen[x]),
                            na.rm = T)}
```

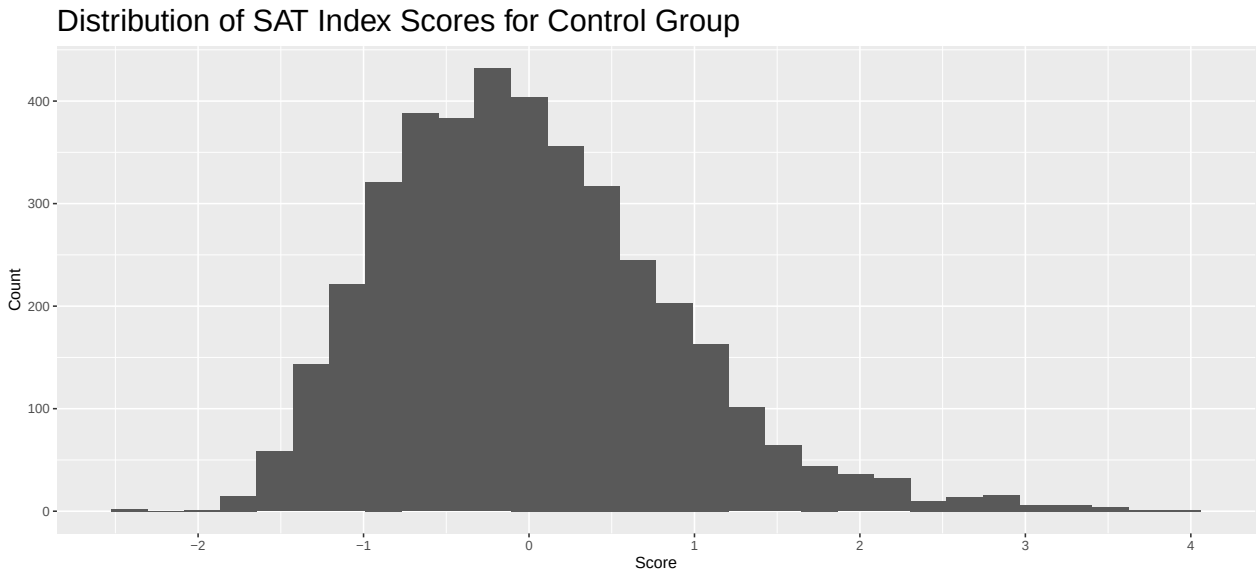
3b: Generating SAT Index

```
## [1] "SAT Index Score"
```

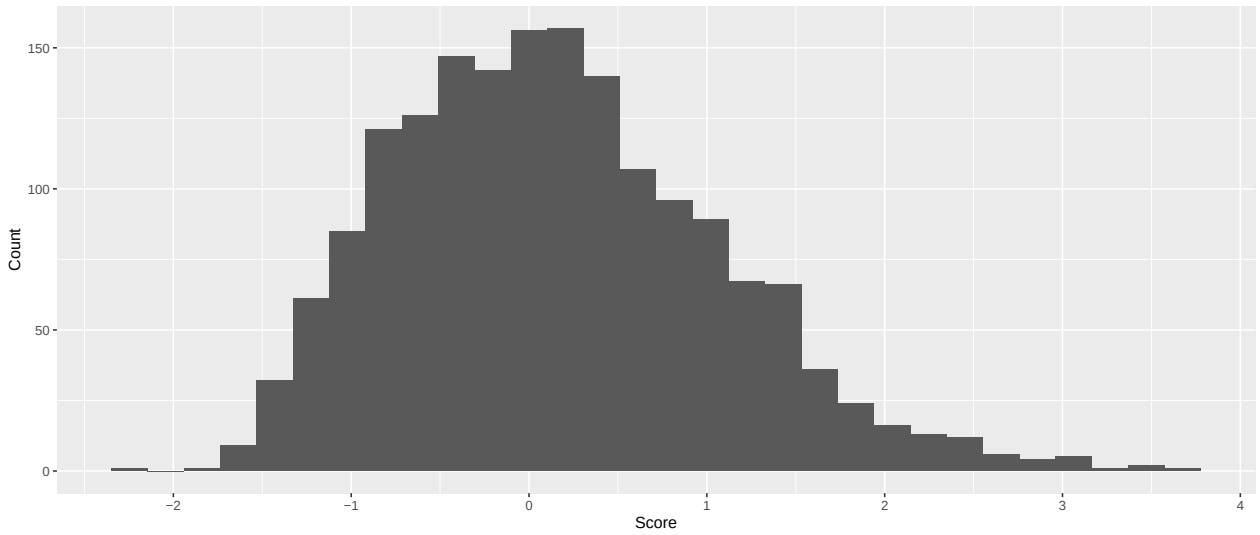
```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -2.43578 -0.59900 -0.04005  0.04606  0.56693  3.92905
```

```
ggplot() +
  geom_histogram(data = filter(star, small == 0), aes(x = sat_index), bins = 30) +
  labs(title="Distribution of SAT Index Scores for Control Group", y = "Count", x = "Score") +
  scale_x_continuous(breaks = seq(-3,5, by =1)) +
  theme(plot.title = element_text(size=21))
```

3c: Histograms




```
ggplot() +
  geom_histogram(data = filter(star, small == 1), aes(x = sat_index), bins = 30) +
  labs(title="Distribution of SAT Index Scores for Treatment Group", y = "Count", x = "Score") +
  scale_x_continuous(breaks = seq(-3,5, by =1)) +
  theme(plot.title = element_text(size=21))
```



As we can see in the histograms above, the distribution of SAT Index scores for both the control and treatment groups is fairly normal. That said, in the treatment group we observe a slightly stronger rightward skew which represents a higher number of overachievers (people who scored 0.5 or more standard deviations above the control group mean). When considering this, it is also worth keeping in mind that there is a notable difference in sample sizes between the control group (n = 3987) and the treatment group (n = 1723).

Question 4: Teacher Randomization

```
star_teachers <- star %>%
  group_by(teacher_id) %>%
  summarise(school_id = mean(school_id),
            small = mean(small),
            teacher_masters = mean(teacher_masters),
            teacher_white = mean(teacher_white),
            teacher_black = mean(teacher_black),
            teacher_experience = mean(teacher_experience),
            sat_index = mean(sat_index),
            class_size = mean(class_size))
```

4a: Aggregate Data by Teacher ID

```
## [1] "Teacher Means in Control Group"
```

```
##      small   teacher_masters  teacher_white    teacher_black
##  Min.   :0   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0   1st Qu.:0.0000   1st Qu.:1.0000   1st Qu.:0.0000
##  Median :0   Median :0.0000   Median :1.0000   Median :0.0000
##  Mean   :0   Mean   :0.3655   Mean   :0.8226   Mean   :0.1723
##  3rd Qu.:0   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:0.0000
##  Max.   :0   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##  teacher_experience   sat_index           class_size
##  Min.   : 0.000     Min.   :-1.326662   Min.   :15.00
##  1st Qu.: 5.000     1st Qu.:-0.333702   1st Qu.:21.00
##  Median :10.000     Median : 0.026161   Median :22.00
##  Mean   : 9.381     Mean   : 0.005253   Mean   :22.36
##  3rd Qu.:13.000     3rd Qu.: 0.299250   3rd Qu.:24.00
##  Max.   :27.000     Max.   : 1.702996   Max.   :28.00
```

```
## [1] "Teacher Means in Treatment Group"
```

```
##      small   teacher_masters  teacher_white    teacher_black
##  Min.   :1   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:1   1st Qu.:0.0000   1st Qu.:1.0000   1st Qu.:0.0000
##  Median :1   Median :0.0000   Median :1.0000   Median :0.0000
##  Mean   :1   Mean   :0.3228   Mean   :0.8661   Mean   :0.1339
##  3rd Qu.:1   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:0.0000
##  Max.   :1   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##  teacher_experience   sat_index           class_size
##  Min.   : 0.000     Min.   :-0.93613   Min.   :12.00
##  1st Qu.: 4.500     1st Qu.:-0.21783   1st Qu.:14.00
```

```
##  Median : 8.000     Median : 0.05043   Median :15.00
##  Mean   : 9.024     Mean   : 0.16308   Mean   :14.96
##  3rd Qu.:13.000     3rd Qu.: 0.54551   3rd Qu.:16.00
##  Max.   :27.000     Max.   : 1.51173   Max.   :17.00
```

```
# Create linear model to test relationship between characteristic and group status
model_exp <- lm(teacher_experience ~ small, data = star_teachers)

# Get lower and upper bounds for 95% confidence interval
bounds_exp <- c(-0.35709 - 1.96*0.66372, -0.35709 + 1.96*0.66372)
cat("The estimated difference in experience between teachers in the control
    and treatment groups is", model_exp$coefficients["small"])
```

4b & c: Measuring Differences in Predetermined Characteristics

```
## The estimated difference in experience between teachers in the control
##     and treatment groups is -0.3570886
```

```
model_edu <- lm(teacher_masters ~ small, data = star_teachers)
bounds_edu <- c(-0.042648  - 1.96*0.054005, -0.042648  + 1.96*0.054005)
cat("The estimated difference in the fraction of teachers with masters degrees
     between the control and treatment groups is", model_edu$coefficients["small"])
```

```
## The estimated difference in the fraction of teachers with masters degrees
##      between the control and treatment groups is -0.04264759
```

```
model_white <- lm(teacher_white ~ small, data = star_teachers)
bounds_white <- c(0.043565  - 1.96*0.040778, 0.043565 + 1.96*0.040778)
cat("The estimated difference in the fraction of white teachers
     between the control and treatment groups is", model_white$coefficients["small"])
```

```
## The estimated difference in the fraction of white teachers
##      between the control and treatment groups is 0.04356499
```

```
model_blk <- lm(teacher_black ~ small, data = star_teachers)
bounds_blk <- c(-0.038489  - 1.96*0.040571, -0.038489 + 1.96*0.040571)
cat("The estimated difference in the fraction of white teachers
     between the control and treatment groups is", model_blk$coefficients["small"])
```

```
## The estimated difference in the fraction of white teachers
##      between the control and treatment groups is -0.03848884
```

4d: Reflecting on Differences in Predetermined Characteristics From the linear models above we can see that none of the differences between the control and treatment groups across any of the predetermined characteristics are statistically significant because all of their 95% confidence intervals include 0. Practically, we can also assess that none of the differences reported are meaningful in a real world sense. The difference in teacher experience is less that a year, the difference in the fraction of masters degree holders and the difference in the number of Black and White teachers are all less than 1 percentage point. Based on these determinations, we can say that the STAR experiment was successful in randomly assigning teachers to the control and treatment groups.

Question 5: Evaluating the Experiment

```
model_multi <- lm(sat_index ~ small + factor(school_id), data = star_teachers)
coeftest(model_multi, vcov = vcovHC(model_multi, type="HC1"))
```

5a: Multivariate Regression

```
##
## t test of coefficients:
##
##                           Estimate Std. Error t value  Pr(>|t|)
## (Intercept)             -0.851634   0.135875 -6.2678 1.647e-09 ***
## small                    0.163630   0.042549  3.8457 0.0001534 ***
## factor(school_id)123056  0.742982   0.160901  4.6176 6.282e-06 ***
## factor(school_id)128068  0.591252   0.147261  4.0150 7.916e-05 ***
## factor(school_id)128076  0.530225   0.145817  3.6362 0.0003375 ***
## factor(school_id)128079  0.561778   0.361237  1.5552 0.1212063
## factor(school_id)130085  0.802593   0.156663  5.1231 6.110e-07 ***
## factor(school_id)159171  1.381224   0.227829  6.0625 5.058e-09 ***
## factor(school_id)161176  0.547440   0.140887  3.8857 0.0001315 ***
## factor(school_id)161183  1.244425   0.217357  5.7253 3.021e-08 ***
## factor(school_id)162184  0.654059   0.193163  3.3861 0.0008263 ***
## factor(school_id)164198  1.111463   0.282533  3.9339 0.0001090 ***
## factor(school_id)165199  1.671316   0.291318  5.7371 2.841e-08 ***
## factor(school_id)166203  0.555454   0.217549  2.5532 0.0112823 *
## factor(school_id)168211  0.820108   0.175955  4.6609 5.183e-06 ***
## factor(school_id)168214  1.304527   0.145723  8.9521 < 2.2e-16 ***
## factor(school_id)169219  1.440278   0.149202  9.6532 < 2.2e-16 ***
## factor(school_id)169229  1.270937   0.169761  7.4866 1.285e-12 ***
## factor(school_id)169231  0.236876   0.147926  1.6013 0.1106011
## factor(school_id)169280  1.066253   0.270786  3.9376 0.0001074 ***
## factor(school_id)170295  1.026736   0.382163  2.6866 0.0077135 **
## factor(school_id)173312  1.706037   0.234176  7.2853 4.425e-12 ***
## factor(school_id)176329  1.268590   0.245277  5.1721 4.827e-07 ***
## factor(school_id)180344  0.833024   0.218947  3.8047 0.0001795 ***
## factor(school_id)189378  0.653349   0.161363  4.0489 6.913e-05 ***
## factor(school_id)189382  0.935711   0.196907  4.7520 3.441e-06 ***
## factor(school_id)189396  0.489242   0.170405  2.8711 0.0044504 **
## factor(school_id)191411  0.865285   0.186437  4.6412 5.659e-06 ***
## factor(school_id)193422  1.188009   0.164024  7.2429 5.725e-12 ***
## factor(school_id)193423  0.988013   0.223691  4.4169 1.506e-05 ***
## factor(school_id)201449  1.341854   0.233627  5.7436 2.747e-08 ***
## factor(school_id)203452  0.968086   0.181369  5.3377 2.152e-07 ***
## factor(school_id)203457  1.273071   0.167369  7.6064 6.098e-13 ***
## factor(school_id)205488  1.025869   0.191234  5.3645 1.885e-07 ***
## factor(school_id)205489  1.328150   0.367440  3.6146 0.0003653 ***
## factor(school_id)205490 -0.100167   0.135445 -0.7395 0.4602918
## factor(school_id)205491  0.939741   0.275403  3.4122 0.0007541 ***
## factor(school_id)205492  0.909886   0.281002  3.2380 0.0013708 **
## factor(school_id)208501  0.802863   0.211682  3.7928 0.0001879 ***
## factor(school_id)208503  0.448054   0.146444  3.0596 0.0024642 **
## factor(school_id)209510  0.674459   0.179215  3.7634 0.0002100 ***
```

```
## factor(school_id)212522  0.942536   0.169279  5.5680 6.783e-08 ***
## factor(school_id)215533  1.225273   0.220248  5.5631 6.951e-08 ***
## factor(school_id)216536  0.841960   0.184617  4.5606 8.078e-06 ***
## factor(school_id)218562  1.145140   0.239538  4.7806 3.023e-06 ***
## factor(school_id)221571  0.139932   0.193191  0.7243 0.4695635
## factor(school_id)221574  0.448606   0.154486  2.9039 0.0040237 **
## factor(school_id)225585  0.654056   0.184466  3.5457 0.0004694 ***
## factor(school_id)228606  1.097980   0.314331  3.4931 0.0005668 ***
## factor(school_id)230612  1.366720   0.154073  8.8706 < 2.2e-16 ***
## factor(school_id)231616  1.018462   0.170314  5.9799 7.889e-09 ***
## factor(school_id)234628  0.978400   0.170931  5.7239 3.042e-08 ***
## factor(school_id)244697  0.105866   0.213116  0.4968 0.6198104
## factor(school_id)244708  0.050129   0.162932  0.3077 0.7585994
## factor(school_id)244723  0.072909   0.171261  0.4257 0.6706886
## factor(school_id)244727  0.620647   0.188059  3.3003 0.0011103 **
## factor(school_id)244728  0.181251   0.199447  0.9088 0.3643692
## factor(school_id)244736  0.869944   0.190092  4.5764 7.535e-06 ***
## factor(school_id)244745  1.072443   0.153032  7.0080 2.349e-11 ***
## factor(school_id)244746  0.763222   0.286752  2.6616 0.0082932 **
## factor(school_id)244755  0.812430   0.294085  2.7626 0.0061716 **
## factor(school_id)244764  0.629867   0.141236  4.4597 1.253e-05 ***
## factor(school_id)244774  0.420183   0.198555  2.1162 0.0353403 *
## factor(school_id)244776  0.393112   0.197855  1.9869 0.0480553 *
## factor(school_id)244780  1.690340   0.483683  3.4947 0.0005635 ***
## factor(school_id)244796  0.327487   0.155581  2.1049 0.0363205 *
## factor(school_id)244799  0.473004   0.163045  2.9011 0.0040588 **
## factor(school_id)244801  0.648929   0.181216  3.5810 0.0004130 ***
## factor(school_id)244806  1.180155   0.241279  4.8913 1.820e-06 ***
## factor(school_id)244818  0.401005   0.141477  2.8344 0.0049755 **
## factor(school_id)244831  0.547438   0.184310  2.9702 0.0032730 **
## factor(school_id)244839  1.307445   0.373253  3.5028 0.0005474 ***
## factor(school_id)252885  1.014697   0.237851  4.2661 2.847e-05 ***
## factor(school_id)253888  0.864719   0.190502  4.5392 8.873e-06 ***
## factor(school_id)257899  0.715732   0.151235  4.7326 3.757e-06 ***
## factor(school_id)257905  1.229334   0.163073  7.5385 9.308e-13 ***
## factor(school_id)259915  0.710962   0.216715  3.2806 0.0011870 **
## factor(school_id)261927  1.123087   0.166953  6.7270 1.223e-10 ***
## factor(school_id)262937  1.736390   0.219499  7.9107 8.903e-14 ***
## factor(school_id)264945  1.320569   0.151734  8.7032 4.903e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
itt_effect <- 0.163630

bounds_multi <- c(0.16363 - 1.96*0.042549,
                  0.16363 + 1.96*0.042549)

cat("The estimated effect of small class sizes on test scores is", itt_effect, "
    and because the 95% confidence interval does not include zero, we can say
    this finding is statistically significant.")
```

5b: Estimated Effect

```
## The estimated effect of small class sizes on test scores is 0.16363
##     and because the 95% confidence interval does not include zero, we can say
##     this finding is statistically significant.
```

```
# Calculate control group mean score
teachers_ctrl_mean <- mean(star_teachers$sat_index[star_teachers$small == 0])

# Construct dataframe to plot
star_plot_data <- star_teachers %>%
  group_by(small) %>%
  summarise(small = mean(small),
            score = ifelse(small == 0,
                           teachers_ctrl_mean,
                           teachers_ctrl_mean+itt_effect)) %>%
  select(small, score) %>%
  rename (group = small) %>%
    mutate(ub = ifelse(group == 1, bounds_multi[2], NA),
           lb = ifelse(group == 1, bounds_multi[1], NA)) %>%
  mutate(group = ifelse(group == 1, "Treatment", "Control"))

ggplot(data=star_plot_data, aes(x=group, y=score, fill=group)) +
  geom_bar(stat="identity", show.legend = FALSE, width=.6) +
  geom_errorbar(aes(ymin=lb, ymax=ub),
                width=.1, size = 0.7, color="black") +
  scale_fill_manual(values=c("red", "blue")) +
  labs(y = "SAT Index Score", x = "")
```

5c: Bar Chart