Do Small Classes in Elementary School Improve Student Outcomes? (with Solutions)

(Based on Frederick Mosteller, "The Tennessee Study of Class Size in the Early School Grades," *Future of Children* 5, no.2 (1995): 113–27.)

Let's answer this question by analyzing the data from Project STAR, where students in Tennessee were randomly assigned to attend either a small class or a regular-size class from kindergarten until third grade. (See Chapter 2 of DSS for further details.)

The dataset is in a file called *STAR.csv*. Table 1 shows the names and descriptions of the variables in this dataset, where the unit of observation is students.

variable	description
classtype	class size the student attended: "small" or "regular"
reading	student's third-grade reading test scores (in points)
math	student's third-grade math test scores (in points)
graduated	identifies whether the student graduated from high school: 1=graduated or 0=did not graduate

Table 1: Variables in "STAR.csv"

In this problem set, we practice estimating average treatment effects using data from a randomized experiment and a linear model, and determining whether the estimated average treatment effect is statistically significant at the 5% level.

As always, let's start by loading and looking at the data:

```
## load and look at the data
star <- read.csv("STAR.csv") # reads and stores data
head(star) # shows first six observations
##   classtype reading math graduated
## 1     small     578  610         1
## 2   regular     612  612         1
## 3   regular     583  606         1
## 4     small     661  648         1
## 5     small     614  636         1
## 6   regular     610  603         0
```

Then, we run the following code to create our treatment variable and confirm that it is created correctly:

```
## create variable pressure inside dataframe voting
star$small <- # stores return values in new variable
         ifelse (star$classtype=="small", # logical test
          1, # return value if logical test is true
          0) # return value if logical test is false
```

```
## look at the first observations again to ensure small was created correctly
head(star, n=3) # shows first three observations
##     classtype reading math graduated small
## 1       small     578  610         1     1
## 2     regular     612  612         1     0
## 3     regular     583  606         1     0
```

1. First, let's estimate the average causal effect of attending a small class on the three different measures of student outcomes we have in the dataset: (i) third-grade reading test scores, (ii) third-grade math test scores, and (iii) probability of graduating from high school.

   a. To estimate each of the three average treatment effects, fit a linear model to the data in such a way that the estimated slope coefficient is equivalent to the difference-in-means estimator you are interested in and store the fitted model in an object. Call the three objects *fit_reading*, *fit_math*, and *fit_graduated*, respectively. Then, run the names of the objects, *fit_reading*, *fit_math*, and *fit_graduated*, so that R will provide you with the contents of each object. Do you arrive at the same estimates as we arrived in Chapter 2 when we computed the three difference-in-means estimators directly? (See page 43 of DSS.) A yes/no answer will suffice. (10 points)

   <u>R code</u>:

```
## fit and store linear models that estimate the three causal effects of small
fit_reading <- lm(star$reading ~ star$small) # on reading
fit_math <- lm(star$math ~star$small) # on math
fit_graduated <- lm(star$graduated ~ star$small) # on graduated
```

```
fit_reading # provides contents of object fit_reading
##
## Call:
## lm(formula = star$reading ~ star$small)
##
## Coefficients:
## (Intercept)   star$small
##     625.492        7.211
```

```
fit_math # provides contents of object fit_reading
##
## Call:
## lm(formula = star$math ~ star$small)
##
## Coefficients:
## (Intercept)   star$small
##      628.84         5.99
```

```
fit_graduated # provides contents of object fit_reading
##
## Call:
## lm(formula = star$graduated ~ star$small)
##
## Coefficients:
## (Intercept)   star$small
##    0.866473     0.007031
```

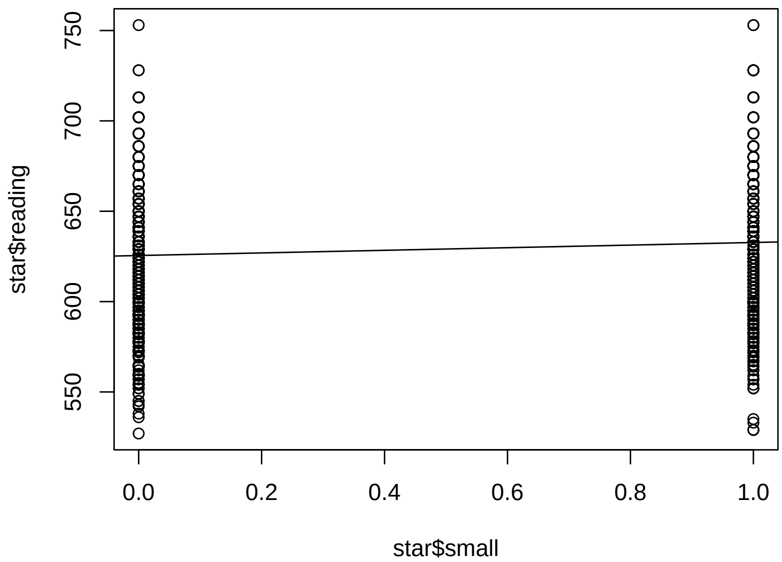
(Note: Remember that to fit a linear model in R we use the lm() function. This function requires an argument of the type Y ~ X. Here, *reading*, *math*, and *graduated* are the outcome variables, Y, and *small* is the treatment variable, X. To specify the dataframe where the variables are stored, we can use either the $ operator (as in the code above) or the optional argument data. If we wanted to use the latter, the code to fit the first linear model above would be lm(reading ~ small, data = star). Now, to store the fitted linear model as an object named *fit_reading*, we use the assignment operator <- and specify *fit_reading* to its left.)

Answer: Yes, we arrive at the same estimates as when we computed the difference-in-means estimator directly.

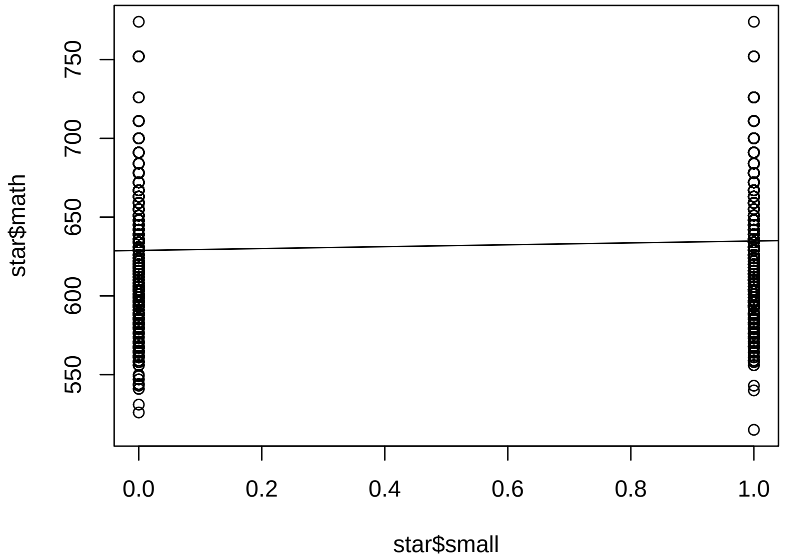
b. For each of the three linear models above, create a visualization of the relationship between X and Y and add the fitted line. (Hint: The functions plot() and abline() might be helpful here.) (R code only.) (5 points)

R code:

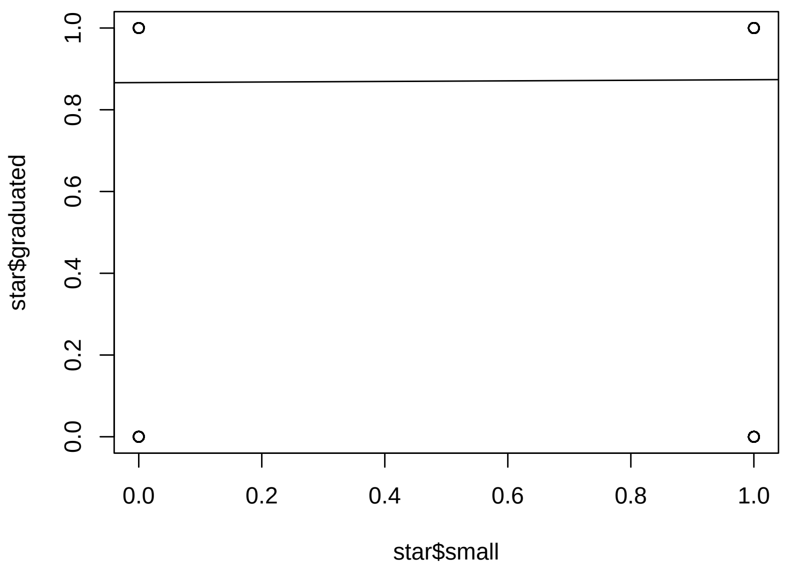
```
plot(star$small, star$reading) # creates scatter plot
abline(fit_reading) # adds fitted line
```



```
plot(star$small, star$math) # creates scatter plot
abline(fit_math) # adds fitted line
```




```
plot(star$small, star$graduated) # creates scatter plot
abline(fit_graduated) # adds fitted line
```



(Note: To visualize the relationship between X and Y, we create a scatter plot with the function plot(). This function requires two arguments and in this order: (1) the X variable, (2) the Y variable. We always plot the X variable along the X axes, and the Y variable along the Y axes. Here, *small* is the X variable and *reading*, *math*, and *graduated* are the Y variables. Alternatively, if we do not want the order of the arguments to matter, we could specify the names of the arguments, x and y, in the code. For example, plot(x=star$small, y=star$reading) and plot(y=star$reading, x=star$small) would produce the same scatter plot as the first one above. Now, to add the fitted line to the scatter plot, we use the function abline(). The only required argument is the name of the object where we have stored the fitted line, fit_reading in this first case.)

c. Now, let's answer the questions: (i) What is the estimated average causal effect of attending a small class on third-grade reading test scores?, (ii) What is the estimated average causal effect of attending a small class on third-grade math test scores?, and (iii) What is the estimated average causal effect of attending a small class on the probability of graduating from high school? Provide a full substantive answer for each (make sure to include the assumption, why the assumption is reasonable, the treatment, the outcome, as well as the direction, size, and unit of measurement of the average treatment effect) (10 points)

Answers:

(i) Assuming that the students who attended a small class were comparable before schooling to those who attended a regular-size class (a reasonable assumption given that the dataset comes from a randomized experiment), we estimate that attending a small class increases student performance on the third-grade reading test by 7 points, on average.

(ii) Assuming that the students who attended a small class were comparable before schooling to those who attended a regular-size class (a reasonable assumption given that the dataset comes from a randomized experiment), we estimate that attending a small class increases student performance on the third-grade math test by 6 points, on average.

(iii) Assuming that the students who attended a small class were comparable before schooling to those who attended a regular-size class (a reasonable assumption given that the dataset comes from a randomized experiment), we estimate that attending a small class increases the probability of graduating from high school by about 1 percentage point, on average.

(Note: These are the exact same conclusion statements we wrote in Chapter 2 of DSS, on page 43.)

2. Second, let's figure out whether the effects are statistically significant at the 5% level?

a. For each of the three treatment effects, specify the null and alternative hypotheses. Please provide both the mathematical notations and their meaning. (5 points)

Answers:

(i) The null and alternative hypotheses for the treatment effect on reading test scores are:

H_0: β=0 (meaning: attending a small class has no average causal effect on third-grade reading test scores at the population level).

H_1: $\beta \neq 0$ (meaning: attending a small class either increases or decreases third-grade reading test scores, on average, at the population level)

(ii) The null and alternative hypotheses for the treatment effect on math test scores are:

H_0: $\beta = 0$ (meaning: attending a small class has no average causal effect on third-grade math test scores at the population level).

H_1: $\beta \neq 0$ (meaning: attending a small class either increases or decreases third-grade math test scores, on average, at the population level)

(iii) The null and alternative hypotheses for the treatment effect on the probability of graduating from high school are:

H_0: $\beta = 0$ (meaning: attending a small class has no average causal effect on the probability of graduating from high school at the population level).

H_1: $\beta \neq 0$ (meaning: attending a small class either increases or decreases the probability of graduating from high school, on average, at the population level)

(Note that the null and alternative hypotheses refer to β, which is the true average treatment effect at the population level, not to $\widehat{\beta}$, which is the estimated average treatment effect at the sample level.)

b. For each of the three treatment effects, what is the value of the observed test statistic, z^{obs}? (Hint: the code summary()$coeff might be helpful here.) (5 points)

R code:

```
summary(fit_reading)$coeff
##                Estimate Std. Error    t value     Pr(>|t|)
## (Intercept) 625.492017   1.393119 448.986670 0.0000000000
## star$small    7.210547   2.055867   3.507301 0.0004684467
```

```
summary(fit_math)$coeff
##                Estimate Std. Error    t value    Pr(>|t|)
## (Intercept) 628.837446   1.475823 426.092746 0.000000000
## star$small    5.989905   2.177916   2.750293 0.006038609
```

```
summary(fit_graduated)$coeff
##                Estimate Std. Error    t value  Pr(>|t|)
## (Intercept) 0.866473149 0.01283401 67.5138061 0.0000000
## star$small  0.007031124 0.01893953  0.3712406 0.7105202
```

Answers:

(i) The value of the observed test statistic for the treatment effect on reading test scores, z^{obs}, is 3.51.

(ii) The value of the observed test statistic for the treatment effect on math test scores, z^{obs}, is 2.75.

(iii) The value of the observed test statistic for the treatment effect on the probability of graduating from high school, z^{obs}, is 0.37.

(Note: The observed test statistic for regression coefficients equals $\widehat{\beta}$ divided by the standard error of $\widehat{\beta}$. For the first analysis above, it equals 7.21055/2.05587= 3.51, which is exactly what R provides as the t-value for the coefficient affecting *small*, that is, the value in the cell in the second row, third column of the first table above.)

c. For each of the three treatment effects, what is the associated p-value? (5 points)

<u>Answers</u>:

(i) The associated p-value is 0.0005. (Note: We can interpret this as indicating that, if the null hypothesis were true, the probability of observing a test statistic equal to or larger than 3.51 (in absolute value) is about 0.05%. This is a tiny probability, well below 5%, so we will reject the null hypothesis.)

(i) The associated p-value is 0.006. (Note: We can interpret this as indicating that, if the null hypothesis were true, the probability of observing a test statistic equal to or larger than 2.75 (in absolute value) is about 0.6%. This is a small probability, well below 5%, so we will reject the null hypothesis.)

(i) The associated p-value is 0.711. (Note: We can interpret this as indicating that, if the null hypothesis were true, the probability of observing a test statistic equal to or larger than 0.37 (in absolute value) is about 71.1%. This is a large probability, well above 5%, so we will fail to reject the null hypothesis.)

d. For each of the three treatment effects, answer the question: Is the effect statistically significant at the 5% level? Please provide your reasoning. (10 points)

<u>Answers</u>:

(i) Yes, the effect of attending a small class on reading test scores is statistically significant at the 5% level. Because (a) the absolute value of the observed test statistic is greater than 1.96 ($|3.51|>1.96$), and/or (b) the p-value is smaller than 0.05 ($0.0005<0.05$), we reject the null hypothesis and conclude that there is likely to be an average treatment effect different than zero at the population level. In other words, we conclude that attending a small class from kindergarten until third grade is likely to have an average effect different than zero on third-grade reading test scores *at the population level*.

(i) Yes, the effect of attending a small class on math test scores is statistically significant at the 5% level. Because (a) the absolute value of the observed test statistic is greater than 1.96 ($|2.75|>1.96$), and/or (b) the p-value is smaller than 0.05 ($0.006<0.05$), we reject the null hypothesis and conclude that there is likely to be an average treatment effect different than zero at the population level. In other words, we conclude that attending a small class from kindergarten until third grade is likely to have an average effect different than zero on third-grade math test scores *at the population level*.

(i) No, the effect of attending a small class on the probability of graduating from high school is not statistically significant at the 5% level. Because (a) the absolute value of the observed test statistic is less than 1.96 ($|0.37|>1.96$), and/or (b) the p-value is larger than 0.05 ($0.711<0.05$), we fail to reject the null hypothesis and conclude that we do **not** have enough evidence to state that the average treatment effect is likely to be non-zero at the population level. In other words, we do **not** have enough evidence to state that attending a small class from kindergarten until third grade is likely to have a non-zero average effect on the probability of graduating from high school *at the population level*.

(Note: In each of these answers, you do not need to provide both reasons, (a) and (b). One of them suffices since both procedures should lead to the same conclusion.)

3. For a discussion about the internal and external validity of this study, see subsection 5.5.4 on page 156 of DSS.