ECON 8080
Homework 1

Derek P. Dyal
University of Georgia
Department of Economics

Problem 1: Project Star Data

Project Star is a very well-known experiment in Tennessee in the 1980s where students were randomly assigned to be in a small class, a regular class, or in a class with a teaching aid. The study was focused on whether or not reducing class sizes improved student's test scores.

P1 (a.)

Exploiting that the treatment is randomly assigned, calculate an estimate of the average treatment effect on the treated (ATT) by comparing the mean of total math score scaled (tmathssk) for boys in small class sizes relative to boys in regular classes.

Recall that the average treatment effect of the treated is:

$$ATT = E[Y(1) - Y(0)|D = 1]$$

where D is a binary random variable that represents treatment ($D = 1$) and no treatment ($D = 0$); $Y(1) = Y|D = 1$ represents the random variable of potential outcome for treated and $Y(0) = Y|D = 0$ for the untreated. Given random assignment, the potential outcomes, Y(1) and Y(0), are independent of treatment D. This implies that ATT becomes:

$$\begin{aligned} ATT &= E[Y(1) - Y(0)|D = 1] \\ &= E[Y(1)|D = 1] - E[Y(0)|D = 1] \\ &= E[Y|D = 1] - E[Y(0)|D = 0] \\ &= E[Y|D = 1] - E[Y|D = 0] \end{aligned}$$

which is the difference between the average outcome for the treated group and the average outcome of the untreated group.

As for an estimator for ATT, by the law of large numbers and continuous mapping theorem, define the difference in the plug-in estimators be an estimator for ATT:

$$\widehat{ATT} := \frac{1}{n}\sum_{i=1}^{n} Y_{i|D=1} - \frac{1}{n}\sum_{j=1}^{n} Y_{j|D=0} \xrightarrow{p} E[Y|D = 1] - E[Y|D = 0]$$

Thus, the estimator we provide is a consistent estimator for ATT. Since we are probably more interested in the effect of small class size on boys, it makes sense to define the indicator variable D in this context as the "small class size" treatment. From the data, we obtain the estimate of small class size on boys using our consistent estimator ($\widehat{ATT}$):

```
# ------------------------------------------------------------------------------
ATT <- mean(tmathssk[classk=="small.class" & sex=="boy"]) -
  mean(tmathssk[classk=="regular" & sex=="boy"])
print(ATT)

13.67522
# ------------------------------------------------------------------------------
```

Note that $\widehat{ATT} > 0$. Hence, the estimated average treatment effect of small class size on the treated scores of boys from the data is **13.67522**.

P1 (b.)

Now, use a regression to calculate the same ATT as we were interested in from part (a). [For this part, you can compare your answer to results that use 'lm', but I'd like for you to use matrix algebra for your main result.] How do the results compare to the ones from part (a)?

Denote the estimator of ATT from regression as $\widehat{ATT}_{SR}$. The linear projection coefficients by "hand"/linear algebra for the linear model is:

```
# ------------------------------------------------------------------------------
X <- as.matrix(cbind(1,Star2$classk)) # design matrix (1919x2).
y <- as.matrix(Star2$tmathssk) # response vector (1919x1).
alpha <- solve(t(X)%*%X) %*% t(X)%*%y # linear projection coefficient (2x1).
print(alpha)

(463.01544, 13.67522)
# ------------------------------------------------------------------------------
```

Hence, our estimate of ATT from regression is $\widehat{ATT}_{SR} =$ **13.67522**, which is identical to our results in (a.). That is, $\widehat{ATT} = \widehat{ATT}_{SR}$.

P1 (c.)

Now, run a regression that additionally includes the teacher's experience (totexpk) and free lunch status (freelunk) as additional regressors. [As before, you can compare your answer to the ones from lm, but please report your final answer using matrix algebra.] How do these results compare to the ones from parts (a) and (b)?

Denote the estimator of ATT from regression that includes the covariates as $\widehat{ATT}_{MR}$. The linear projection coefficients by "hand"/linear algebra for the linear model is:

```
# ------------------------------------------------------------------------------
X2 <- as.matrix(cbind(1,Star2$classk,Star2$totexpk,Star2$freelunk)) # design matrix (1919x4).
y <- as.matrix(Star2$tmathssk) # response vector (1919x1).
alpha2 <- solve(t(X2)%*%X2) %*% t(X2)%*%y # linear projection coefficient (4x1).
print(alpha2)

(491.7956695, 13.4233260, 0.5397799, -22.4616584)
# ------------------------------------------------------------------------------
```

Hence, our estimate of ATT from regression with covariates is $\widehat{ATT}_{MR} =$ **13.4233260**. If we compare our results from (a.) & (b.) to (c.), we find that:

```
# ------------------------------------------------------------------------------
alpha[2]-alpha2[2]

0.2518936
# ------------------------------------------------------------------------------
```

We observe that the difference is positive and that our estimates in (a.) & (b.) provide a larger estimate of ATT than (c.). That is, $\widehat{ATT}_{MR} < \widehat{ATT}_{SR} = \widehat{ATT}$.

Appendix: Salient Code

```
################################################################################
################################################################################
## ---------------------------------- ---------------------------------- ----- ##
## ECON 8080: Homework 1                                                      ##
## ---------------------------------- ---------------------------------- ----- ##
## Author: Derek Payton Dyal                                                  ##
## Contact: ddyal@uga.edu                                                     ##
## Date: 01/31/2023                                                           ##
## ---------------------------------- ---------------------------------- ----- ##
################################################################################
################################################################################

# Setup:
install.packages("Ecdat")
library('Ecdat')

# ----------------------------------------------------------------------------
#  * * * * * * * * * * * * * * * * * * * Q1 * * * * * * * * * * * * * * * * * *  |
# ----------------------------------------------------------------------------
# Q1: Project Star is a very well-known experiment in Tennessee in the 1980s where
#     students were randomly assigned to be in a small class, a regular class, or
#     in a class with a teaching aid. The study was focused on whether or not
#     reducing class sizes improved student's test scores.
# ----------------------------------------------------------------------------
# ----------------------------------------------------------------------------

# Import data:
data(Star, package="Ecdat")
attach(Star)

# ----------------------------------------------------------------------------
# (a.) Exploiting that the treatment is randomly assigned, calculate an estimate of
#      the ATT by comparing the mean of 'tmathssk' for boys in small class sizes
#      relative to boys in regular classes.
# ----------------------------------------------------------------------------

# Estimate ATT of small class size on boys using consistent estimator:
ATT <- mean(tmathssk[classk=="small.class" & sex=="boy"]) -
  mean(tmathssk[classk=="regular" & sex=="boy"])
print(ATT)

# ----------------------------------------------------------------------------



# ----------------------------------------------------------------------------
# (b.) Now, use a regression to calculate the same ATT as we were interested in
#      from part (a). [For this part, you can compare your answer to results that
#      use 'lm', but I'd like for you to use matrix algebra for your main result.]
#      How do the results compare to the ones from part (a)?
# ----------------------------------------------------------------------------

# Subset the data to only contain boys treated with regular and small class
# sizes:
Star2 <- subset(Star, (classk!="regular.with.aide" & sex=="boy"),
                select = c(tmathssk,sex,classk,totexpk,freelunk))
```

```
# Coefficients by "hand"/linear algebra:
X <- as.matrix(cbind(1,Star2$classk)) # design matrix (1919x2).
y <- as.matrix(Star2$tmathssk) # response vector (1919x1).
alpha <- solve(t(X)%*%X) %*% t(X)%*%y # linear projection coefficient (2x1).
print(alpha) # (463.01544, 13.67522)

# ---------------------------------------------------------------------------------


# ---------------------------------------------------------------------------------
# (c.) Now, run a regression that additionally includes the teacher's experience
#      (totexpk) and free lunch status (freelunk) as additional regressors. [As
#      before, you can compare your answer to the ones from lm, but please report
#      your final answer using matrix algebra.] How do these results compare to the
#      ones from parts (a) and (b)?
# ---------------------------------------------------------------------------------

# Coefficients by "hand"/linear algebra:
X2 <- as.matrix(cbind(1,Star2$classk,Star2$totexpk,Star2$freelunk)) # design matrix (1919x4).
y <- as.matrix(Star2$tmathssk) # response vector (1919x1).
alpha2 <- solve(t(X2)%*%X2) %*% t(X2)%*%y # linear projection coefficient (4x1).
print(alpha2) # (491.7956695, 13.4233260, 0.5397799, -22.4616584)

# Comparison of results:
alpha[2]-alpha2[2] # >0.

# ---------------------------------------------------------------------------------

detach(Star)
# ---------------------------------------------------------------------------------
#  * * * * * * * * * * * * * * * * * * * END Q1 * * * * * * * * * * * * * * * * * *  |
# ---------------------------------------------------------------------------------
```