Homework 3

Bridget Antwiwaa

2/10/2022

#Question 1a

```
x1= read.csv(choose.files(), header = T)
stargazer::stargazer(x1, type = 'text')
```

```
##
## ===============================================================
## Statistic  N   Mean   St. Dev.   Min   Pctl(25) Pctl(75)   Max
## ---------------------------------------------------------------
## meal_pct  420 44.705   27.123   0.000   23.282   66.865  100.000
## testscr   420 654.157  19.053  605.550 640.050  666.663  706.750
## comp_stu  420  0.136   0.065    0.000   0.094    0.164    0.421
## str       420 19.640   1.892   14.000   18.582   20.872  25.800
## avginc    420 15.317   7.226    5.335   10.639   17.629  55.328
## ---------------------------------------------------------------
```

#Question 1b

```
reg1=lm(testscr ~ str, x1)
reg1
```

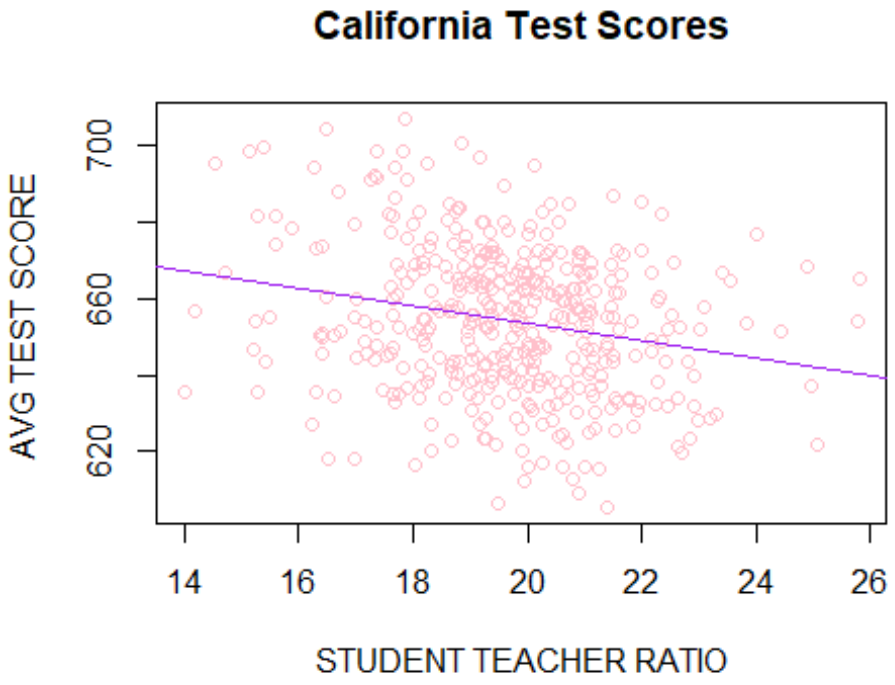
```
##
## Call:
## lm(formula = testscr ~ str, data = x1)
##
## Coefficients:
## (Intercept)          str
##      698.93        -2.28
```

```
jtools::summ(reg1)
```

```
## MODEL INFO:
## Observations: 420
## Dependent Variable: testscr
## Type: OLS linear regression
##
## MODEL FIT:
## F(1,418) = 22.58, p = 0.00
## R² = 0.05
## Adj. R² = 0.05
##
## Standard errors: OLS
## ------------------------------------------------
```

```
##                       Est.   S.E.   t val.      p
## ---------------- -------- ------ -------- ------
## (Intercept)         698.93   9.47    73.82   0.00
## str                  -2.28   0.48    -4.75   0.00
## ------------------------------------------------
```

```
plot(x=x1$str, y=x1$testscr, xlab = "STUDENT TEACHER RATIO", ylab = "AVG TEST
SCORE ", main = "California Test Scores", col="pink")
abline(reg1, col="purple")
```




```
stargazer::stargazer(reg1, type = 'text')
```

```
##
## ===============================================
##                         Dependent variable:
##                     ---------------------------
##                                testscr
## -----------------------------------------------
## str                           -2.280***
##                                (0.480)
##
## Constant                     698.933***
##                                (9.467)
##
## -----------------------------------------------
## Observations                    420
## R2                             0.051
```

```
## Adjusted R2                     0.049
## Residual Std. Error      18.581 (df = 418)
## F Statistic           22.575*** (df = 1; 418)
## ================================================
## Note:               *p<0.1; **p<0.05; ***p<0.01
```

```
cor(x1$str,x1$testscr)
```

```
## [1] -0.2263628
```

The descriptive statistic for text scores gives you the mean close to the constant, a standard deviation close to the residual standard error.

looking at the t-value which is -4.75, the model appears that class size has a statistically significant effect on student learning (test score).

Since the R^2 is close to 0, the regressor is not at predicting Y. Therefore we can infer that the model does not appear to fit the data well

#Question 1c

```
reg2=lm(testscr ~ str+avginc, x1)
jtools::summ(reg2)
```

```
## MODEL INFO:
## Observations: 420
## Dependent Variable: testscr
## Type: OLS linear regression
##
## MODEL FIT:
## F(2,417) = 218.30, p = 0.00
## R² = 0.51
## Adj. R² = 0.51
##
## Standard errors: OLS
## ------------------------------------------------
##                        Est.   S.E.   t val.      p
## ----------------- -------- ------ -------- ------
## (Intercept)         638.73   7.45    85.75   0.00
## str                  -0.65   0.35    -1.83   0.07
## avginc                1.84   0.09    19.82   0.00
## ------------------------------------------------
```

The omitted variable that income control for the square of income

The T value and P value is changed

No class size does not continue to be statistically significant

The R^2 increased compared to the first regression. Also, it seem to be a lot closer to 1 but still small, therefore we can infer that it is not a better fit.

#Question 1d

```
reg3=lm(testscr ~ str+avginc+meal_pct, x1)
jtools::summ(reg3)
```

```
## MODEL INFO:
## Observations: 420
## Dependent Variable: testscr
## Type: OLS linear regression
##
## MODEL FIT:
## F(3,416) = 513.28, p = 0.00
## R² = 0.79
## Adj. R² = 0.79
##
## Standard errors: OLS
## ------------------------------------------------
##                        Est.   S.E.   t val.      p
## ----------------- -------- ------ -------- ------
## (Intercept)         685.03   5.31   129.02   0.00
## str                  -0.83   0.23    -3.55   0.00
## avginc                0.53   0.08     6.33   0.00
## meal_pct             -0.51   0.02   -23.23   0.00
## ------------------------------------------------
```

The percent qualifying for reduced-price lunch variable appears to be statistically significant based in the t-value

The model predicts that an increase in percent qualifying for reduced-price lunch is associated with a decrease of 0.51 in test scores.

#Question 1e

```
reg4=lm(testscr ~ str+avginc+meal_pct+comp_stu, x1)
jtools::summ(reg4)
```

```
## MODEL INFO:
## Observations: 420
## Dependent Variable: testscr
## Type: OLS linear regression
##
## MODEL FIT:
## F(4,415) = 393.43, p = 0.00
## R² = 0.79
## Adj. R² = 0.79
##
## Standard errors: OLS
## ------------------------------------------------
##                        Est.   S.E.   t val.      p
## ----------------- -------- ------ -------- ------
## (Intercept)         678.36   5.77   117.55   0.00
```

```
## str                  -0.64   0.24    -2.64   0.01
## avginc                0.52   0.08     6.31   0.00
## meal_pct             -0.50   0.02   -22.96   0.00
## comp_stu             19.84   7.02     2.83   0.00
## ------------------------------------------------
```

The model predicts that an increase in computers per student is associated with an increase of 19.85% in test scores.

I think district should add more capital (computers) in order to improve test scores

```
stargazer::stargazer(reg1,reg2,reg3, type = 'text')

##
##
================================================================================
================
##                                              Dependent variable:
##                     ---------------------------------------------------------
------------------
##                                                      testscr
##                              (1)                      (2)
(3)
## --------------------------------------------------------------------------
------------------
## str                        -2.280***               -0.649*
-0.831***
##                             (0.480)                (0.354)
(0.234)
##
## avginc                                             1.839***
0.527***
##                                                    (0.093)
(0.083)
##
## meal_pct
-0.506***
##
(0.022)
##
## Constant                  698.933***              638.729***
685.031***
##                             (9.467)                (7.449)
(5.310)
##
## --------------------------------------------------------------------------
------------------
## Observations                 420                     420
420
## R2                          0.051                   0.511
0.787
```

```
## Adjusted R2                   0.049                    0.509
0.786
## Residual Std. Error    18.581 (df = 418)       13.349 (df = 417)
8.819 (df = 416)
## F Statistic         22.575*** (df = 1; 418) 218.302*** (df = 2; 417)
513.284*** (df = 3; 416)
##
================================================================================
================
## Note:                                                          *p<0.1;
**p<0.05; ***p<0.01
```