Truett Bloxsom, UTID: tsb962

Professor Tom Vogle

TA Bokyung Kim

ECO 348K

Project 1

1. The data set consists of a large sample size of descrete kindergarten variables. These variables include: gender, race, social-economic status, test scores, class size, and teacher characteristics. In the gender variable, males account for 51.4%, and females account for 48.6% of the sample. In the race variable, blacks account for 32.5%, whites account for 66.9%, and the rest, .6%, is other races. In the social-economic status variable, kindergarteners who get free lunch account for 48.3%, and kindergarteners who do not get free lunch account for 51.7% of the sample. Using free or not free lunch at school as the social-economic status factor is concerning since social-economic status should be on a spectrum instead of poor (free lunch) or rich (no free lunch). Minorities excluding blacks, account for 30 individuals so there is little information on these minorities. The sample is also only from Tennessee schools under Project STAR and the data could be effected by state variables not included in the data set. Tabulating race and socio-economic status, there is a disproportionate number of black students with poor socio-economic status compared to white students.

Table 0 below shows descriptive statistics of the data set

Variable | Obs Mean Std. Dev. Min Max

-------------+---------------------------------------------------------

ssex | 6,325 1.486166 .4998481 1 2

srace | 6,322 1.349257 .5676566 1 6

sesk | 6,301 1.515632 .4997952 1 2

tcombssk | 5,786 922.3287 73.7466 635 1253

treadssk | 5,789 436.7253 31.70626 315 627

-------------+---------------------------------------------------------

tmathssk | 5,871 485.3771 47.69788 288 626

cltypek | 6,325 2.052332 .8065276 1 3

classid | 6,325 11800.78 6784.35 100 23600

hdegk | 6,304 2.385787 .5735962 2 5

totexpk | 6,304 9.258249 5.808783 0 27

-------------+---------------------------------------------------------

tracek | 6,263 1.164618 .3708647 1 2

schtypek | 6,325 2.41502 .9349133 1 4

1. Initial assignment might differ from actual enrollment since parents will try to enroll their kids in the best class possible. This could mean that a kindergartener was assigned to a larger class size classroom and their parents moved them to a smaller class size classroom. The parents who moved their kids into smaller classes, might also help their kids with homework and push them to do well in school. This would lead the kids in smaller classrooms to achieve better test results not solely from being in smaller classes but because of their parent’s efforts.

Table 1 below shows the Chi-Squared test between student sex and classroom type

student | classroom type (1s2r3ra) in

sex | kindergarten

(1m2f) | small cla regular c regular + | Total

-----------+---------------------------------+----------

male | 30.06 34.43 35.51 | 100.00

female | 30.02 34.96 35.02 | 100.00

-----------+---------------------------------+----------

Total | 30.04 34.69 35.27 | 100.00

Pearson chi2(2) = 0.2330 Pr = 0.890

Table 2 below shows the Chi-Squared test between student race and classroom type

student | classroom type (1s2r3ra) in

race | kindergarten

(1w2b6o) | small cla regular c regular + | Total

-----------+---------------------------------+----------

white | 30.56 34.77 34.67 | 100.00

black | 28.81 34.50 36.69 | 100.00

other | 40.00 33.33 26.67 | 100.00

-----------+---------------------------------+----------

Total | 30.04 34.67 35.29 | 100.00

Pearson chi2(4) = 4.6757 Pr = 0.322

Table 3 below shows the Chi-Squared test between socio-economic status and classroom type

socio-economic | classroom type (1s2r3ra) in

status in | kindergarten

kindergarten | small cla regular c regular + | Total

---------------+---------------------------------+----------

free lunch | 29.19 34.21 36.60 | 100.00

non-free lunch | 30.81 35.18 34.01 | 100.00

---------------+---------------------------------+----------

Total | 30.03 34.71 35.26 | 100.00

Pearson chi2(2) = 4.7871 Pr = 0.091

From tables 1 through 3, none of the test statistics are significant at the 5% level, se we cannot reject the null that classroom type and baseline student characteristics are independent of each other. Table 3 is the only test statistic, Pr = 0.091, that is significant at the 10% level. Table 2 shows that classroom type was not totally evenly distributed based on racial groups. Proportionally more white and other minorities where enrolled in small classes compared to blacks. Table 1 shows that boys and girls were distributed evenly between the classroom sizes. Kids from poor households were more likely to be in larger classroom sizes than kids from wealthier households.

1. Table 4 below shows summarized data from student combined test scores from small classes

Variable | Obs Mean Std. Dev. Min Max

-------------+---------------------------------------------------------

tcombssk | 1,738 931.9419 76.35863 747 1253

Table 5 below shows summarized data from student combined test scores from regular classes

Variable | Obs Mean Std. Dev. Min Max

-------------+---------------------------------------------------------

tcombssk | 2,005 918.0429 73.13799 635 1229

Table 6 below shows the t-test of student combined test scores on small and regular classes

Two-sample t test with unequal variances

------------------------------------------------------------------------------

Group | Obs Mean Std. Err. Std. Dev. [95% Conf. Interval]

---------+--------------------------------------------------------------------

small cl | 1,738 931.9419 1.831611 76.35863 928.3495 935.5343

regular | 2,005 918.0429 1.633375 73.13799 914.8396 921.2462

---------+--------------------------------------------------------------------

combined | 3,743 924.4967 1.225268 74.96196 922.0944 926.8989

---------+--------------------------------------------------------------------

diff | 13.89899 2.454121 9.087394 18.71059

------------------------------------------------------------------------------

diff = mean(small cl) - mean(regular) t = 5.6635

Ho: diff = 0 Satterthwaite's degrees of freedom = 3616.04

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0

Pr(T < t) = 1.0000 Pr(|T| > |t|) = 0.0000 Pr(T > t) = 0.0000

Using Tables 4 and 5, I computed the standard t-test:

(931.94-918.04)/sqrt((76.35^2/1738)+(73.13^2/2005)) = 5.6645722

Since the t-stat is greater than 1.96, we can reject the null hypothesis that the means are equal at the 5% level. Using the actual t-test in stata returned the value 5.6635 which is close to the value I obtained. Since the t-stat is significant, class size positively influences test scores, so a kindergartener in a smaller class is predicted to get a better test score than a kindergartener in a regular class on average. The difference in means between reading and math scores between small and regular sizes was equal to 1.93 points which is of comparable magnitude. Computing the reading and math tests separately, the t-score for reading was t = 5.6039 and the t-score for math was t = 4.8959. The difference of these scores is .708 which is in a comparable magnitude since both t-scores are already highly significant.

1. Table 7 below shows summarized data from student reading scores from classes without an aid

Variable | Obs Mean Std. Dev. Min Max

-------------+---------------------------------------------------------

treadssk | 2,006 434.7323 30.9359 315 627

Table 8 below shows summarized data from student reading scores from classes with an aid

Variable | Obs Mean Std. Dev. Min Max

-------------+---------------------------------------------------------

treadssk | 2,044 435.4295 31.50247 372 627

Computing the t-score using Tables 7 and 8:

(435.429-434.732)/(sqrt(31.502^2/2044+30.936^2/2006)) = .71041363

Table 9 below shows summarized data from student math scores without an aid

Variable | Obs Mean Std. Dev. Min Max

-------------+---------------------------------------------------------

tmathssk | 2,032 483.1993 47.63593 320 626

Table 10 below shows summarized data from student math scores with an aid

Variable | Obs Mean Std. Dev. Min Max

-------------+---------------------------------------------------------

tmathssk | 2,077 482.7959 45.78352 339 626

Computing the t-score using Tables 9 and 10:

(482.79-483.19)/(sqrt(45.78^2/2077+47.63^2/2044)) = -.27478942

Since both t-scores listed above are between – 1.96 and 1.96, we can accept the null that the mean test scores between regular classes with aids and without aids are equal. This means that a kindergartener in a class with an aid is predicted get the same test score on average as a kindergartener in a class without an aid.

1. Table 11 shows the t-test of black students combined test scores on small and regular classes

------------------------------------------------------------------------------

Group | Obs Mean Std. Err. Std. Dev. [95% Conf. Interval]

---------+--------------------------------------------------------------------

small cl | 541 913.7283 3.179171 73.94564 907.4832 919.9733

regular | 2,005 918.0429 1.633375 73.13799 914.8396 921.2462

---------+--------------------------------------------------------------------

combined | 2,546 917.1261 1.453033 73.31702 914.2768 919.9753

---------+--------------------------------------------------------------------

diff | -4.314612 3.551711 -11.27915 2.649927

------------------------------------------------------------------------------

diff = mean(small cl) - mean(regular) t = -1.2148

Ho: diff = 0 degrees of freedom = 2544

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0

Pr(T < t) = 0.1123 Pr(|T| > |t|) = 0.2246 Pr(T > t) = 0.8877

Table 12 shows the t-test of white students combined test scores on small and regular classes

------------------------------------------------------------------------------

Group | Obs Mean Std. Err. Std. Dev. [95% Conf. Interval]

---------+--------------------------------------------------------------------

small cl | 1,187 940.3218 2.204641 75.95622 935.9964 944.6473

regular | 2,005 918.0429 1.633375 73.13799 914.8396 921.2462

---------+--------------------------------------------------------------------

combined | 3,192 926.3277 1.326852 74.96421 923.7261 928.9293

---------+--------------------------------------------------------------------

diff | 22.27893 2.717332 16.95103 27.60682

------------------------------------------------------------------------------

diff = mean(small cl) - mean(regular) t = 8.1988

Ho: diff = 0 degrees of freedom = 3190

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0

Pr(T < t) = 1.0000 Pr(|T| > |t|) = 0.0000 Pr(T > t) = 0.0000

Table 13 shows the t-test of low socio-economic status students combined test scores on small and regular classes

------------------------------------------------------------------------------

Group | Obs Mean Std. Err. Std. Dev. [95% Conf. Interval]

---------+--------------------------------------------------------------------

small cl | 820 911.1 2.472742 70.80857 906.2463 915.9537

regular | 2,005 918.0429 1.633375 73.13799 914.8396 921.2462

---------+--------------------------------------------------------------------

combined | 2,825 916.0276 1.364527 72.52558 913.352 918.7032

---------+--------------------------------------------------------------------

diff | -6.942893 3.00402 -12.83319 -1.052597

------------------------------------------------------------------------------

diff = mean(small cl) - mean(regular) t = -2.3112

Ho: diff = 0 degrees of freedom = 2823

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0

Pr(T < t) = 0.0104 Pr(|T| > |t|) = 0.0209 Pr(T > t) = 0.9896

Table 14 shows the t-test of high socio-economic status students combined test scores on small and regular classes

------------------------------------------------------------------------------

Group | Obs Mean Std. Err. Std. Dev. [95% Conf. Interval]

---------+--------------------------------------------------------------------

small cl | 913 950.8675 2.529487 76.43069 945.9032 955.8318

regular | 2,005 918.0429 1.633375 73.13799 914.8396 921.2462

---------+--------------------------------------------------------------------

combined | 2,918 928.3132 1.401681 75.71675 925.5648 931.0616

---------+--------------------------------------------------------------------

diff | 32.82458 2.961812 27.01712 38.63203

------------------------------------------------------------------------------

diff = mean(small cl) - mean(regular) t = 11.0826

Ho: diff = 0 degrees of freedom = 2916

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0

Pr(T < t) = 1.0000 Pr(|T| > |t|) = 0.0000 Pr(T > t) = 0.0000

The effect of class size on test scores of black students is insignificant at the 5% level with a t-score of -1.2148. The effect of class size on test scores of white students is highly significant at the 5% level with a t-score of 8.1988. This means that a black student from a regular class is predicted to get the same score as a black student in a small class. A white student in a small class is predicted to get a much higher score on their tests than a white student in a regular class.

sqrt((1.633-2.205)^2-(1.633-3.179)^2) = 2.063 = standard error between white and black students

((918.04-940.32)-(918.04-913.72))/1.436 = -18.523677 = t-stat for testing statistical difference between white and black students test scores in small or regular classes. So the effect of moving a black student compared to a white student is highly statistically different.

From table 13, the average predicted test score of a student from a small class compared to a large class has a significant negative effect if they are from a low socio-economic status. So low socio-economic students in regular classes did better on tests than the same student in a small class. From table 14, the average predicted test score of a student from a small compared to a regular class has a very significant positive effect if they are from a high socio-economic status. So a student form a high socio-economic status is predicted to do better on tests than the same student in a regular class.

1. I am assuming in question 5 that individuals in a particular classroom are independent of other classrooms and individuals within a particular classroom are also independent. The assumption that classes are independent of each other is a reasonable assumption but not reasonable between individuals within the same classroom. We can collapse student test scores by class and take the mean so every student has the same test score in each class. This method assumes students in each class are dependent on each other. After collapsing both the reading and math scores by class id and class room type, I recomputed the t-tests for test scores for class room type. The t-score for reading score = 2.3423 and the t-score for the math score = 1.8967. These t-scores are much smaller in significance than what was calculated in question three. This difference is caused by grouping the individuals by which class they are enrolled in and by relaxing the assumption that individuals in each class room are independent of each other.
2. Table 15 shows the regression of student test scores on class size

Source | SS df MS Number of obs = 3,743

-------------+---------------------------------- F(1, 3741) = 32.27

Model | 179850.266 1 179850.266 Prob > F = 0.0000

Residual | 20847551.4 3,741 5572.72158 R-squared = 0.0086

-------------+---------------------------------- Adj R-squared = 0.0083

Total | 21027401.7 3,742 5619.29495 Root MSE = 74.651

------------------------------------------------------------------------------

tcombssk | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

cltypek | -13.89899 2.446592 -5.68 0.000 -18.69578 -9.102211

\_cons | 945.8409 3.950318 239.43 0.000 938.0959 953.5859

Table 16 shows regression of student reading scores on class size with collapsed by class id

Source | SS df MS Number of obs = 225

-------------+---------------------------------- F(1, 223) = 5.49

Model | 1866.77672 1 1866.77672 Prob > F = 0.0200

Residual | 75876.6802 223 340.254171 R-squared = 0.0240

-------------+---------------------------------- Adj R-squared = 0.0196

Total | 77743.4569 224 347.069004 Root MSE = 18.446

------------------------------------------------------------------------------

treadssk | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

cltypek | -5.80276 2.477366 -2.34 0.020 -10.6848 -.9207167

\_cons | 446.6584 3.773411 118.37 0.000 439.2223 454.0945

Table 17 shows regression of student math scores on class size with collapsed by class id

Source | SS df MS Number of obs = 225

-------------+---------------------------------- F(1, 223) = 3.60

Model | 2831.9117 1 2831.9117 Prob > F = 0.0592

Residual | 175537.889 223 787.165422 R-squared = 0.0159

-------------+---------------------------------- Adj R-squared = 0.0115

Total | 178369.801 224 796.293754 Root MSE = 28.056

------------------------------------------------------------------------------

tmathssk | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

cltypek | -7.147073 3.768091 -1.90 0.059 -14.5727 .2785491

\_cons | 498.0498 5.739385 86.78 0.000 486.7395 509.3602

Table 18 shows the regression of student test scores on class size clustering class id

Linear regression Number of obs = 3,743

F(1, 224) = 5.56

Prob > F = 0.0192

R-squared = 0.0086

Root MSE = 74.651

(Std. Err. adjusted for 225 clusters in classid)

------------------------------------------------------------------------------

| Robust

tcombssk | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

cltypek | -13.89899 5.893393 -2.36 0.019 -25.51258 -2.28541

\_cons | 945.8409 9.105881 103.87 0.000 927.8967 963.785

------------------------------------------------------------------------------

As the results show from regressions in tables 15,16, and 17 we calculated the same t-score from questions 3 and 6. Table 15 shows that moving a student from a small class to a large class is predicted to reduce their combined score by 13.9 points. Table 16 shows, by allowing for dependence between students in the same class, that moving a student from a small to a regular class is predicted to reduce their reading test score by 5.8 points. Table 17, with the collapse method, shows that a student moved from a small to a regular class is predicted to reduce their score by 7.2 points.

Table 18 allows for dependence between individuals within the same class by clustering some of them together, which is similar to what we did with the collapse method in question 6. The difference is that the cluster method does not totally assume strict dependence between students in each class rooms but does allow for some covariance between them. This can be shown by collapsing by class id and running regression again and we obtain a t-score of -2.25 which is smaller than the clustered regression t-score of -2.36. By allowing for dependence, the t-scores are smaller but are still significant at the 5%. So even allowing for dependence between students, class room size has a significant effect, so students in a small class are predicted to have higher tests scores than students in regular classes.

1. By using clustered regression or regular regression, the effects on class size are still large, a predicted difference of 14 points. Clustered regression is less significant than regular regression which had a standard error of 2.44 compared to the clustered robust standard error of 5.89. Running regression of test score on race has a more significant effect. A white student is predicted to get more than 22 points on their combined test score compared to a black student. Running regression of test score on socio-economic status had the largest effect on test scores. A student from a wealthy socio-economic background was predicted to get more than 39 points on their combined test score than a student from a poor socio-economic background.