**Question 1: Summary Statistics**

1. **Calculate the share of students receiving free lunch, the share of white/Asian students, the**

**average age in 1985, the attrition rate, average class size, and gender by the enter-variables**

**generated above and by treatment status (small class, regular class, and regular class with**

**aide). Present your summary statistics for STAR participants in Table 1, structured as in**

**Table I in Krueger (1999). You do not need to provide standard deviations.**

*Please see question 2c for Table 1.*

1. **Calculate the average of the percentile ranks of the math and reading tests for every**

**individual in each year (name the variables testk etc.). (Hint: If one subtest score is missing,**

**take the percentile score corresponding to the only available test as in Krueger (1999),**

**fn.11.) Add the average values by the enter-variables and by treatment status to Table 1.**

*Please see question 2c for Table 1.*

1. **Comment on the characteristics of students assigned to the “small class” treatment, who entered STAR in kindergarten.**

*Some of the most notable characteristics of students who entered the Tennessee’s STAR program when looking at Table1 are that of the 1900 students who entered the program in kindergarten, about 47% of them participated in the free school lunch program.*

*The demographic characteristics of the students participating in the STAR program show that the distribution is fairly homogeneous between class sizes using the mean as a proxy. For example, the whiteasian variable mean has a ranges from 65.9% for reg/aide class to 68.2% for small class sizes. A relatively similar magnitude difference in means is seen in the age and girl variables. The variables that do not have this similarity about the means between class sizes is attrition, class size, and test performance. On average, students that entered the experiment in Kindergarten and that are in a small class types perform better than their counterparts in regular or regular/aide class types. We should also note that though the student demographic means are reasonably close together, their standard deviations are large (standard deviations not included in Table 1*

**Question 2: Random assignment**

**The first question to ask about a randomized experiment is whether the randomization**

**successfully balanced subjects’ characteristics across the different treatment groups.**

1. **The STAR data does not include any pre-treatment test scores. Do you think that this is a**

**problem? Explain briefly.**

*There is a natural challenge with being able to include a pre-treatment test score. It would be useful to have pre-treatment test scores so that, aside from random class selection of students on the basis of demographic characteristics, we could also capture random selection based on innate ability, the proxy for which would be pre-treatment test scores. This baseline test score would help evaluate whether there is an actual effect of learning in a small class or whether the effect is primarily because students are learning in a structured educational environment. Although this does not completely falsify the results of the experiment, it does shed light on the fall backs of a field experiment.*

*The practical problem with this is the age of the students if they were to take a test of ability prior to kindergarten. School attendance is not mandatory and students may have been exposed to different levels of learning from their parents. It might be hard to say that the test score proxy is valid rather than just a reflection of a greater exposure to educational learning content.*

1. **Compare the student characteristics collected in Table 1 across treatments. Formally test the**

**null hypothesis of no difference across treatment groups using an F-test. Add both**

**F-statistics (rounded to 2 digits after the decimal point) and p-values to Table 1. Do you**

**think that randomization was successful, and why/why not? (Hint: Use the regress and test**

**commands.)**

*Please see question 2c for Table 1.*

*To test the randomization of the experiment, we regress the variable of interest (i.e. freelunchk, whiteasian, age…etc.) on smallk (small1, small2, and small3) and regulark (regular1, regular2, and regular3). After running the regression, we run the command “test smallk (small1, small2, small3) regulark (regular1, regular2, regular3). In columns 4 and 5, we observe the F-test statistic and p-values of each individual test. The p-value is for the F-test of equality of all three groups of different class types. In many of the cases, we observe statistically significant p-values which does not support the idea that the experiment was completely random. However, “girl” is statistically insignificant in all cases which means that the distribution of “girls” was similar in all three groups.*

1. **In fact, the treatment was randomly assigned to students and teachers within schools. For**

**each of the variables in Table 1, test the null hypothesis that, conditional on school of**

**attendance, there are no significant differences across treatment groups. Once again, give**

**both F-statistics and p-values and add them to Table 1. (Hint: Use school dummy**

**variables.) Are the results consistent with random assignment conditional on school attendance? Explain.**

*Similar to the question above, we test whether the treatment was randomly assigned to students and teachers within schools. To test this hypothesis, we regress the variable of interest (i.e. freelunck, whiteasian, age…etc) on smallk (small1, small2, and small3), regulark (regular1, regular2, regular3) and dummy variables for the different schools. In columns 6 and 7, we observe the F-test statistic and p-values of each individual test. The p-value is for the F-test of equality of all three groups of different class types within a school. In most cases, the background variables are statistically insignificant, especially in Kindergarten. This supports the idea that within school treatments were randomly assigned to students. However, it is also important to note that these results are not consistent in all situations like for freelunch and age.*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | (1) | (2) | (3) | No Within School Effects | | Within School Effects | |
|  | Small | Regular | Regular/Aide | F-test statistic | p-value | F-test statistic | p-value |
| 1. Kindergarten | | | | | | | |
| 1. freelunchk | .4709302 | .4773663 | .5027003 | 2.39 | 0.09 | 0.79 | 0.46 |
| 2. whiteasian | .6826316 | .674567 | .6593456 | 1.33 | 0.26 | 0.42 | 0.66 |
| 3. age | 5.255667 | 5.239269 | 5.241256 | 0.80 | 0.45 | 0.44 | 0.64 |
| 4. attritionk | .4868421 | .5177758 | .5284626 | 3.76 | 0.02 | 4.51 | 0.01 |
| 5. csizek | 15.40316 | 22.38286 | 23.20708 | 4171.2 | 0.00 | 7389.2 | 0.00 |
| 6. girl | .4857895 | .4899727 | .4827432 | 0.12 | 0.89 | 0.10 | 0.91 |
| 7. testk | 53.92286 | 49.15445 | 48.9591 | 20.00 | 0.00 | 31.35 | 0.00 |
| *N* | 1900 | 2194 | 2231 |  |  |  |  |
|  | (1) | (2) | (3) | No Within School Effects | | Within School Effects | |
|  | Small | Regular | Regular/Aide | F-test statistic | p-value | F-test statistic | p-value |
| 1. First Grade | | | | | | | |
| 1. freelunch1 | .4878178 | .5296209 | .523532 | 4.20 | 0.01 | 4.43 | 0.01 |
| 2. whiteasian | .6820779 | .621904 | .7025862 | 19.50 | 0.00 | 0.90 | 0.41 |
| 3. age | 5.332465 | 5.419844 | 5.422492 | 16.75 | 0.00 | 16.08 | 0.00 |
| 4. attrition1 | .5286458 | .5126706 | .4684385 | 2.74 | 0.06 | 1.04 | 0.35 |
| 5. csize1 | 15.69714 | 22.69659 | 23.43621 | 8033.6 | 0.00 | 24315 | 0.00 |
| 6. girl | .485447 | .4860465 | .4701557 | 0.75 | 0.47 | 0.50 | 0.60 |
| 7. test1 | 56.22053 | 47.70852 | 51.15209 | 54.97 | 0.00 | 67.48 | 0.00 |
| *N* | 1925 | 2584 | 2320 |  |  |  |  |
|  | (1) | (2) | (3) | No Within School Effects | | Within School Effects | |
|  | Small | Regular | Regular/Aide | F-test statistic | p-value | F-test statistic | p-value |
| 1. Second Grade | | | | | | | |
| 1. freelunch2 | .4864865 | .5276888 | .5143839 | 3.58 | 0.03 | 2.41 | 0.09 |
| 2. whiteasian | .668 | .6321041 | .6430881 | 3.13 | 0.04 | 0.39 | 0.67 |
| 3. age | 5.4215 | 5.512386 | 5.51032 | 14.73 | 0.00 | 15.96 | 0.00 |
| 4. attrition2 | .3653846 | .3353941 | .3533951 | 0.51 | 0.60 | 0.14 | 0.87 |
| 5. csize2 | 15.296 | 23.46855 | 23.47777 | 14274 | 0.00 | 30046 | 0.00 |
| 6. girl | .489 | .48651 | .4769604 | 0.37 | 0.69 | 0.30 | 0.74 |
| 7. test2 | 54.72013 | 48.80313 | 50.23703 | 25.44 | 0.00 | 36.63 | 0.00 |
| *N* | 2000 | 2305 | 2474 |  |  |  |  |
|  | (1) | (2) | (3) | No Within School Effects | | Within School Effects | |
|  | Small | Regular | Regular/Aide | F-test statistic | p-value | F-test statistic | p-value |
| 1. Third Grade | | | | | | | |
| 1. freelunch3 | .4822967 | .5115562 | .5131471 | 2.58 | 0.07 | 5.09 | 0.01 |
| 2. whiteasian | .6835327 | .6718447 | .6515693 | 2.78 | 0.06 | 0.02 | 0.98 |
| 3. age | 5.479521 | 5.553119 | 5.553693 | 9.35 | 0.00 | 12.80 | 0.00 |
| 4. attrition3 | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| 5. csize3 | 15.71113 | 23.63883 | 24.43107 | 6439.9 | 0.00 | 13544 | 0.00 |
| 6. girl | .4894204 | .484466 | .470004 | 0.97 | 0.38 | 1.36 | 0.25 |
| 7. test3 | 55.2845 | 49.96544 | 49.4941 | 28.81 | 0.00 | 32.23 | 0.00 |
| *N* | 2174 | 2060 | 2517 |  |  |  |  |

Table 1: Summary Statistics

Question 3: OLS estimates of class size effects

1. **Run OLS regressions with the average percentile test score (testk etc.) constructed in**

**Question 1b as a dependent variable (assuming that the errors are iid). Produce regression**

**results similar to those given in columns (1) to (3) of Table V in Krueger (1999) and present**

**them in Table 2 A-D.**

Table 2: OLS: Actual Class Size

|  |  |  |  |
| --- | --- | --- | --- |
|  | (1) | (2) | (3) |
| 1. **Kindergarten** | | | |
| smallk | 4.768\*\*\* | 5.415\*\*\* | 5.449\*\*\* |
|  | (0.880) | (0.783) | (0.751) |
|  |  |  |  |
| regular\_aidek | -0.195 | 0.0173 | 0.258 |
|  | (0.844) | (0.751) | (0.721) |
|  |  |  |  |
| whiteasian |  |  | 9.416\*\*\* |
|  |  |  | (1.264) |
|  |  |  |  |
| girl |  |  | 4.639\*\*\* |
|  |  |  | (0.597) |
|  |  |  |  |
| freelunchk |  |  | -13.29\*\*\* |
|  |  |  | (0.724) |
|  |  |  |  |
| \_cons | 49.15\*\*\* | 56.38\*\*\* | 49.83\*\*\* |
|  | (0.600) | (2.932) | (3.040) |
|  |  |  |  |
| School Fixed Effects | No | Yes | Yes |
| *N* | 5874 | 5874 | 5857 |
| r2 | 0.00677 | 0.249 | 0.311 |
| F | 20.01 | 23.99 | 31.39 |
|  | (1) | (2) | (3) |
| 1. **First Grade** | | | |
| small1 | 8.512\*\*\* | 8.230\*\*\* | 7.646\*\*\* |
|  | (0.812) | (0.724) | (0.702) |
|  |  |  |  |
| regular\_aide1 | 3.444\*\*\* | 2.071\*\*\* | 1.967\*\*\* |
|  | (0.774) | (0.700) | (0.684) |
|  |  |  |  |
| whiteasian |  |  | 8.432\*\*\* |
|  |  |  | (1.084) |
|  |  |  |  |
| girl |  |  | 3.133\*\*\* |
|  |  |  | (0.562) |
|  |  |  |  |
| freelunch1 |  |  | -12.89\*\*\* |
|  |  |  | (0.673) |
|  |  |  |  |
| \_cons | 47.71\*\*\* | 52.77\*\*\* | 48.15\*\*\* |
|  | (0.531) | (2.392) | (2.513) |
| School Fixed Effects | No | Yes | Yes |
| *N* | 6617 | 6617 | 6451 |
| r2 | 0.0164 | 0.248 | 0.309 |
| F | 54.97 | 28.05 | 35.61 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | (1) | (2) | (3) |
| 1. **Second Grade** | | | |
| small2 | 5.917\*\*\* | 6.498\*\*\* | 6.016\*\*\* |
|  | (0.859) | (0.781) | (0.763) |
|  |  |  |  |
| regular\_aide2 | 1.434\* | 1.767\*\* | 1.970\*\*\* |
|  | (0.813) | (0.737) | (0.720) |
|  |  |  |  |
| whiteasian |  |  | 7.906\*\*\* |
|  |  |  | (1.186) |
|  |  |  |  |
| girl |  |  | 3.242\*\*\* |
|  |  |  | (0.596) |
|  |  |  |  |
| freelunch2 |  |  | -13.22\*\*\* |
|  |  |  | (0.727) |
|  |  |  |  |
| \_cons | 48.80\*\*\* | 62.82\*\*\* | 57.16\*\*\* |
|  | (0.586) | (2.714) | (2.860) |
| School Fixed Effects | No | Yes | Yes |
| *N* | 6047 | 6047 | 5794 |
| r2 | 0.00835 | 0.230 | 0.295 |
| F | 25.44 | 23.79 | 31.48 |

Table 2: OLS: Actual Class Size (Continued)

|  |  |  |  |
| --- | --- | --- | --- |
|  | (1) | (2) | (3) |
| 1. **Third Grade** | | | |
| small3 | 5.319\*\*\* | 5.477\*\*\* | 4.670\*\*\* |
|  | (0.874) | (0.818) | (0.810) |
|  |  |  |  |
| regular\_aide3 | -0.471 | -0.0521 | -0.364 |
|  | (0.842) | (0.786) | (0.777) |
|  |  |  |  |
| whiteasian |  |  | 7.000\*\*\* |
|  |  |  | (1.272) |
|  |  |  |  |
| girl |  |  | 3.123\*\*\* |
|  |  |  | (0.622) |
|  |  |  |  |
| freelunch3 |  |  | -12.35\*\*\* |
|  |  |  | (0.753) |
|  |  |  |  |
| \_cons | 49.97\*\*\* | 43.74\*\*\* | 41.08\*\*\* |
|  | (0.628) | (2.605) | (2.812) |
| School Fixed Effects | No | Yes | Yes |
| *N* | 6061 | 6061 | 5887 |
| r2 | 0.00942 | 0.190 | 0.237 |
| F | 28.81 | 18.42 | 23.18 |

1. **Interpret the coefficients on the small class and the regular/aide class indicators for**

**kindergarten children in column (1) of Table 2.**

*The coefficient for smallk from our regression output is 4.768 with a standard error of .88 and is significant at the 99% level. This coefficient indicates that a kindergarten student who is in the small class treatment group, corresponds with a test score that is a about 4.8 percentile points higher than their cohorts in other class sizes.*

*The coefficient for the regular with aide coefficient is -0.195 with a standard error of 0.844 making it insignificant, so we cannot say that having a teacher’s aide in a regular sized classroom is of any benefit to kindergarten test scores.*

1. **Do the coefficients on the small class indicator change if additional explanatory variables**

**are added to the model? What does this tell you about selection on observables?**

*In Columns (2) and (3) of Table 2, we observe the regression with additional control variables. The coefficient on smallk increases in columns (2) and (3) by approximately 0.647 and 0.681, respectively. Furthermore, we notice that the addition of more explanatory variables reduces standard errors on the coefficient. The fact that we are reducing standard errors but observing an increase in the coefficient estimate may be indicative of our ability to capture characteristics in the error term. The additional variables that we control for are all statistically significant at the 99% level. The selection on observables evaluates whether the treatment is as good as randomly assigned or in other words it is evaluating the strength of the treatment. The fact that our standard error is decreasing, and the coefficient is increasing we can argue that the small classroom size is an effective treatment on student performance.*

1. **Suppose that test scores are measures of true skills that are noisy (i.e., subject to**

**measurement error). How do you expect this to affect your estimates of the coefficient?**

**Explain briefly.**

*If test scores are a measure of true skills but containing measurement error, we would expect there to be a downward bias from attenuation bias created in the measurement process.*

*The power of the coefficients is reduced because of the noise introduced by measurement error onto the test score variables.*

*If a student’s “true skills” are not of the academic proficiency type, but more creative or “street smart”, it would not be well captured by the test scores implemented over the experiment period. This would indicate that for such individuals, their true skills are yet hidden in the noise. One way to bring out the true skills of such individuals would be employing alternate ascertainment of true skills other than standardized testing.*

Question 4: Instrumental Variable Regression

1. **Reproduce Table IV in Krueger (1999) showing transitions between class types in adjacent grades. Display your results in Table 3.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 3: Transitions Between Class-Size in Adjacent Grades** | | | | |
| **Number of Students in Each Type of Class** | | | | |
| **A. Kindergarten to first grade** | | |  |  |
|  | **First Grade** | | | |
| **Kindergarten** | **Small** | **Regular** | **Reg/Aide** | **All** |
| **Small** | **1293** | **60** | **48** | **1401** |
| **Regular** | **126** | **737** | **663** | **1526** |
| **Aide** | **122** | **761** | **706** | **1589** |
| **All** | **1541** | **1558** | **1417** | **4516** |
|  |  |  |  |  |
|  |  |  |  |  |
| **B. First to second Grade** | |  |  |  |
|  | **Second Grade** | | | |
| **First Grade** | **Small** | **Regular** | **Reg/Aide** | **All** |
| **Small** | **1421** | **23** | **24** | **1468** |
| **Regular** | **152** | **1482** | **194** | **1828** |
| **Aide** | **40** | **115** | **1558** | **1713** |
| **All** | **1613** | **1620** | **1776** | **5009** |
|  |  |  |  |  |
|  |  |  |  |  |
| **C. Second to third grade** | |  |  |  |
|  | **Third Grade** | | | |
| **Second grade** | **Small** | **Regular** | **Reg/Aide** | **All** |
| **Small** | **1552** | **37** | **34** | **1623** |
| **Regular** | **167** | **1461** | **148** | **1776** |
| **Aide** | **36** | **67** | **1837** | **1940** |
| **All** | **1755** | **1565** | **2019** | **5339** |

1. **Generate a variable that contains each student’s initial assignment to a class type, i.e., assigned class type in kindergarten if a student entered STAR in kindergarten, assigned class type in 1st grade if a student entered the program in 1st grade etc. (name this variable ctype\_assigned).**

* *See code for variable creation*

1. **Krueger (1999) argues that initial class assignment is highly correlated with actual class assignment in later years. Show that initial class assignment is a good predictor of actual class treatment in each grade.**

*We ran several sets of regressions in which we look at some grade t’s future state of class size on the left-hand side and the t-1’s prior state(s) on the right-hand side. After this we ran join tests of significance. When progressing from year to year, the number of prior states possible increased, so the number of regressions in subsequent years increased to accommodate the increasing number of prior states. In all of the cases of regressions that were ran, we observed very high F test statistics with small p-values. This indicated the random initial assignment is highly correlated to the actual class assignment in subsequent years.*

1. **Consider using initial class assignment as an instrumental variable for actual class treatment. Name the two conditions that a valid instrument needs to fulfill. Do you think that the requirements are met by initial class assignment, and why/why not?**

*There are two important conditions that a valid instrument must fulfill. The first restriction is the exclusion restriction which states that there is no correlation between the instrument and any unobserved determinant of the dependent variable. In this case, “initial class assignment” is exogenous because it is randomly assigned by the researchers and not influenced by any external factors such as parental influence. The second requirement is relevancy. According to relevancy, IV must be correlated with the endogenous variable. If we assume that the external factor that is influencing actual class assignment is parental influence, then initial class assignment should be highly correlated with actual class assignment. That is, parents would want their student to be assigned to a small class type in order to gain access to more resources. So, if a student is initially in a small class type, we would not expect change in actual class type. However, if a student is initially in a regular/regular\_aide class type, there might be change in the actual class type in the future. Based on this story, we can see that initial class type assigned randomly assigned by researchers can influence actual class type.*

1. **For each grade, run an instrumental variable regression of average math/reading test score on actual class type dummies (small class and regular with aide class), using initially assigned class type dummies as instruments (without further control variables). Report your results in Table 4. How do your results compare to those found in Table 2? Do they suggest that non-random transitions between class types were a problem?**

Table 4: Initial Class Size

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | (1) | (2) | (3) | (4) |
|  | testk | test1 | test2 | test3 |
| smallk | 4.768\*\*\* |  |  |  |
|  | (0.880) |  |  |  |
|  |  |  |  |  |
| regular\_aidek | -0.195 |  |  |  |
|  | (0.844) |  |  |  |
|  |  |  |  |  |
| small1 |  | 5.850\*\*\* |  |  |
|  |  | (1.577) |  |  |
|  |  |  |  |  |
| regular\_aide1 |  | 4.644\*\*\* |  |  |
|  |  | (1.197) |  |  |
|  |  |  |  |  |
| small2 |  |  | 1.163 |  |
|  |  |  | (1.940) |  |
|  |  |  |  |  |
| regular\_aide2 |  |  | -3.355\*\* |  |
|  |  |  | (1.592) |  |
|  |  |  |  |  |
| small3 |  |  |  | 2.434 |
|  |  |  |  | (2.083) |
|  |  |  |  |  |
| regular\_aide3 |  |  |  | -3.730\* |
|  |  |  |  | (1.926) |
|  |  |  |  |  |
| \_cons | 49.15\*\*\* | 42.74\*\*\* | 45.44\*\*\* | 44.94\*\*\* |
|  | (0.600) | (0.816) | (1.127) | (1.368) |
| *N* | 5874 | 2180 | 1258 | 1010 |
| r2 | 0.00677 | 0.00970 | 0.00557 | 0.00909 |
| F | 20.01 | 10.66 | 3.515 | 4.617 |

Standard errors in parentheses

\* *p* < 0.10, \*\* *p* < 0.05, \*\*\* *p* < 0.01

*The IV regression results for kindergarten yielded 4.77 for smallk and -.195 for regular\_aidek which is the same as the OLS coefficients which makes sense since there was no difference between the initially assigned classes and the actual class.*

*For first grade, the coefficients of 5.850 and 4.644 differ compared to the OLS figures of 8.512 and 3.444 respectively. This difference and the fact that the IV results are statistically significant suggests that the is upward bias in the OLS regression. While overall, both indicate that there is a positive benefit of smaller class size on grades, the upward bias of the OLS may be due to parents that decided to affect the randomness of class assignment by requesting a class change. It is possible these parents are more involved with educational decisions of their children, which extends more than just class assignment. If they also devote more time, energy, and resources to their children’s education through tutors or other learning materials, it may cause test scores coefficients on small class size to be overstated because of these factors apart from class size. This suggests that nonrandom transition between class types biases our OLS model.*

*For second grade, the coefficients of 1.163 for small2 and -3.355 for regular\_aide2 are significantly different compared to 5.917 and 1.434 respectively in the OLS regression from Table 2. In the IV model, we observe that small2 is not significantly significant. However, in the OLS model it was statistically significant at the 99% level for small2 and 90% for the regular\_aide2. These results continue to support our hypothesis that parental decisions are influencing the randomness of the class sizes in the experiment.*

*For third grade, the IV coefficients of 2.434 for small3 and -3.730 for regular\_aide3 are different from the 5.319 and -0.471in the OLS model. Again, the IV small3 coefficient is not statistically significant and the regular\_aide3 coefficient is significant only at the 90% level. The OLS coefficient for small3 is significant at the 99% level while the regular\_aide3 coefficient is not significant. Consequently, the story for small class sizes remains unchanged.*