High School Math Data Analysis Report

Carol Cong Chen

October 22, 2015

This report includes two sections: 1) data cleaning and 2) data analysis. All the process for merging, cleaning as well as analyzing data was documented in the R file (see Appendix A).

Section 1: Data Merge and Cleaning

1) Data Merge

Three data files (DistrictA.xlsx, DistrictB.csv, and DistrictC.docx) were merged as one data file (MathData_all.csv). This file includes 119 observations of 10 variables. The structure of the data file is as follow:

Variable	Data Type	Description	Value	Missing data
tc	Numeric	treatment status	0=control;1=treatment	
id_s	Numeric	Student ID		
gen	Numeric	Gender	0=Male; 1=Female	
grade	Numeric	GRADE	0=grade 10; 1=grade 11	9999
eth	Numeric	race and ethnicity	0=other; 1= Hispanic; 2= White	
dist	String	District	A, B, C	
frl	Numeric	FRL(Yes=1/No=0)	0=no; 1=yes	9999
math_name15	String	Name of math standardized test_2015	SBAC	9999
math_score15	Numeric	Math standardized test scale score_2015	>= 0	9999
math_band15	Numeric	Math standardized test band_2015	0,1,2,3,4	9999

Of these 10 variables, *dist* is a new added variable, which contains the district information of each student. Students from district C had missing values on four variables: *frl*, *math_name15*, *math_score15*, and *math_band15*.

2) Data Cleaning

First, the validity of the data was checked by going through each variable to ensure that they had the appropriate data type and valid values. If invalid variable values were identified, I need to recover the correct values by looking up the true values from the actual data records.

However, for this case, I do not have the actual data records, so the invalid values were either corrected by common sense or assigned as missing values. All the missing values in this data were recoded as 9999. The data cleaning process for each variable is as follow:

Variable	Description	Data Cleaning Notes
tc	Treatment status	Recode 2 to 0 for comparison group
id_s	Student ID	Check duplicate ID
gen	Gender	Recode "M" as 0, "F" as 1
grade	GRADE	Recode "10" as 0, "11" as 1; there are 2 invalid values "1", which was recoded as missing values.
eth	Race and ethnicity	Recode "Hispanic" and "Latino" as 1, "White" as 2, all other values as "0"
dist	District	Check frequency
frl	FRL(Yes=1/No=0)	Check frequency
math_name15	Name of math standardized test_2015	Recode " " as missing values
math_score15	Math standardized test scale score_2015	Check range, mean, outliers; there are 4 outliers, 0.
math_band15	Math standardized test band_2015	Check frequency

Section 2: Data Analysis

1) In a table, provide an average and standard deviation for 2015 math standardized test scale scores by treatment status.

Table 1: Mean math scores by treatment status

Treatment Status	N	Mean	SD
Treatment	41	2492.56	68.26
Comparison	60	2377.15	649.83

Table 1 shows that among the 101 students who had their math scores, 41 (41%) students were in the treatment group, and 60 (59%) students were in the comparison group. The mean math score of the treatment group is 2492.56 (SD=68.26), which is 115.41 points higher than the mean math score of the comparison group 2377.15 (SD=649.83).

2) In a table, provide the frequency for gender, grade, race_ethnicity, qualified for free and reduced lunch, and 2015 math standardized test band by treatment status

Table 2: Frequency table of gender, grade, race_ethnicity, qualified for free and reduced lunch, and 2015 math standardized test band by treatment status

		Comparison Group (0)	**Treatment Group (1)**
Gender	Male (0)	34 (60%)	30 (48%)
	Female (1)	23 (40%)	32 (52%)
Grade	Grade 10 (0)	49 (54%)	11 (44%)
	Grade 11 (1)	41 (46%)	14 (56%)
Race_ethnicity	Other (0)	18 (28%)	22 (40%)
	Hispanic or Latino (1)	23 (36%)	21 (38%)
	White (2)	23 (36%)	12 (22%)
Qualified for free and reduced lunch	Not qualified (0)	45 (69%)	16 (44%)
	Qualified (1)	20 (31%)	20 (56%)
2015 math standardized test band	0	4 (7%)	0 (0%)
	1	15 (25%)	16 (39%)
	2	14 (23%)	22 (54%)
	3	16 (27%)	2 (5%)
	4	11 (18%)	1 (2%)

Table 2 shows that 1) the percentage of **male** of the comparison group (60%) was higher than the treatment group (48%); 2) the students were almost evenly split by grade for the both groups; 3) the percentage of *white* students was higher for the comparison group (36%) than the treatment group (22%), while the percentage of students with the *other* race category was higher

for the treatment group (40%) than the comparison group (28%); 4) the percentage of students who were qualified for free or reduced lunch of the comparison group (69%) was higher than the treatment group (44%); 5) Most treatment group students (83%) had math test bands 1 and 2, while most comparison group students (93%) had math test band 1, 2, 3, and 4. To conclude, there are some differences between the treatment group and comparison group in their demographic variables and test band variable.

3) Run analysis of 1) and 2) by district and provide tables by district.

Table 3: Frequency table of gender, grade, race_ethnicity, qualified for free and reduced lunch, and 2015 math standardized test band by district

		District A	**District B**	**District C**
Gender	Male (0)	22 (45%)	29 (49%)	6 (54%)
	Female (1)	27 (55%)	30 (51%)	5 (46%)
Grade	Grade 10 (0)	34 (71%)	48 (86%)	8 (73%)
	Grade 11 (1)	14 (29%)	8 (14%)	3 (27%)
Race_ethnicity	Other (0)	16 (33%)	18 (31%)	6 (55%)
	Hispanic or Latino (1)	32 (65%)	8 (13%)	4 (36%)
	White (2)	1 (2%)	33 (56%)	1 (9%)
Qualified for free and reduced lunch	Not qualified (0)	9 (21%)	56 (97%)	
	Qualified (1)	34 (79%)	2 (3%)	
2015 math standardized test band	0	0 (0%)	4 (7%)	
	1	19 (45%)	12 (20%)	
	2	16 (38%)	20 (34%)	
	3	5 (12%)	13 (22%)	
	4	2 (5%)	10 (17%)	

Table 3 shows that 1) the students were almost evenly split by gender for all the three districts; 2) the percentage of grade 10 students of district B (86%) was higher than that of district A (71%) and C (73%); 3) the majority of students of district A (65%) were Hispanic or Latino, of district B were White (56%), and of district C were other races; 4) the percentage of students who were qualified for free or reduced lunch of district A (79%) was much higher than that of district B(3%); 5) Most students of district A (95%) had math test bands 1, 2 and 3, while most students of district B (93%) had math test band 1, 2, 3, and 4. To conclude, there are obvious differences between the three districts in their demographic variables and test band variable.

4) Do the data suggest that the performance of treatment students is better than the comparison students?

Table 4: An independent sample t-test: mean math scores of treatment group and control group

Group	**N**	**Mean**	**SD**	**Mean difference**	**df**	**t**	**P-value**	**Effect Size**
Treatment	41	2492.56	68.26	115.41	60.9	-1.37	0.177	0.23
Control	60	2377.15	649.83					

To answer this research question, two analyses were conducted on this data: 1) an independent sample t-test on the student math test scores was performed to test whether the performance of treatment students is better than the comparison students. Specifically, Welch two sample t-test was applied to deal with the nonhomogeneity of variance issue. Table 3 shows that the mean score difference between the two groups is 115.41 (t=-1.37, p=0.178) with effect size 0.23, which indicates that the mean math scores of the treatment and comparison groups are not significantly different. 2) One-way ANOVA was also performance to determine whether there is any significant difference between the means of two groups. Table 5 indicates that no significant differences exist between the mean math scores of the treatment and control groups, which is consistent with the results from t-test. To conclude, the performance of treatment

students is not better than the comparison students. And the intervention was not more effective for certain students.

Table 5:One-way ANOVA

Source	**df**	**Sum of square**	**Mean of square**	**F**	**P-value**
Student Group	1	324420	324420	1.28	0.261
Residual	470	25101140	253547		

Summary

This data consisted of student demographic information and math test information for 119 students from 3 districts who participated in an intervention study about high school mathematics education. Out of the 119 students, 55 (46%) students were in the treatment group and 64 (54%) students in the comparison group. Results from the frequency tables show that the demographic information of students are obviously different by student treatment status and by district. Further, both the independent sample t-test and one-way ANOVA show that the mean math scores of the treatment and comparison groups are not significantly different. That means that the intervention had no significant impact.