U.S. EDUCATIONAL ACHIEVEMENT GAP IN THE SALINAS CITY ELEMENTARY SCHOOL DISTRICT

By

JESSICA MEAN AND BIANCA OROZCO

CAPSTONE PROJECT

Submitted in partial satisfaction of the requirements for the degree of

BACHELOR OF SCIENCE

in

MATHEMATICS

in the

COLLEGE OF SCIENCE, MEDIA ARTS, AND TECHNOLOGY

at

CALIFORNIA STATE UNIVERSITY,
MONTEREY BAY

Approved:

(Advisor)
Michael B. Scott
Department of Mathematics and Statistics

Spring 2016

1 Acknowledgements

Firstly, we would like to express our sincere gratitude to our advisor Dr. Michael Scott for the continuous support of our Bachelor's study, for his patience, motivation, and immense knowledge. His guidance helped us all throughout our research and statistical analysis of this capstone paper. We could not have imagined having a better advisor and mentor for our undergraduate study.

Besides our advisor, we would like to thank Professor Steven Kim and Professor Jon Detka for their insightful clarifications and encouragement, but also for the hard questions which led us to widen our statistical analysis from numerous perspectives.

2 Abstract

We investigated if there exists an achievement gap between students' economic status conditioning on grade level in the Salinas City School District. In the United States, there are certain barriers and inequalities (i.e. race, class status, financial standpoint, gender, language, poverty, etc.) within its educational systems. Thus, we focused on how students' economic status and grade level affect their math test performance (pass/fail). We ran a multiple logistic regression test and a mixed effect logistic regression test to see if there is an association between economic status and the probability of failure. Our descriptive statistics shows how a large number of students are economically disadvantaged and the majority of students fail the California Assessment of Student Performance and Progress math test. Our multiple logistic regression analysis explains there is an association between economic status and the probability of failure. Our mixed effect logistic regression expresses how there is an association between economic status and probability of failure where the variability among schools is accounted for, which is a more reasonable model than the multiple logistic model. Our results conclude that grade level and economic status can affect students' math performance. The educational achievement gap is a real issue in our country. We hope to bring awareness and motivate the goal of reaching for quality education for students within the United States as a whole.

Contents

3 Introduction

The achievement gap within the educational system affects students throughout the United States. The purpose of our project is to investigate if there exists an achievement gap between students' economic status conditioning on grade level in the Salinas Elementary City School District. We analyzed data collected from the California Assessment of Student Performance and Progress on math test scores taken in 2015. The analysis test we chose are multiple logistic regression and mixed effect logistic model in order to see if different variables affect students' test score. As college students, to seek raw data was challenging due to access limitations. Once we found legitimate data, we discovered that the details of the data collecting process and the definitions of each variable were not provided by the CAASPP. Despite our limitations, analysis on these standardized math test scores can still be done. Further statistical analysis and problem solving can minimize the issue.

3.1 Definitions

- categorical variable: A variable that belongs to one set of categories
- quantitative variable: A variable that has numerical values that represent different magnitudes of the variable
- dependent variable (response variable): A variable that measures the outcome of interest
- independent variable (explanatory variable): A variable on which the outcome depends on
- multiple logistical regression: A model that applies when the response variable y is categorical and there are several explanatory variables
- probability: The likelihood of a randomly selected subject has a successful outcome
- parameter: Describes a population (true population value)
- statistic: Describes a sample
- point estimate: A single number that is our "best guess" for the parameter
- test statistic: Describes how far our point estimate falls from the parameter value
- p-value: The probability or likelihood of getting the test statistic
- β: The population parameter of the positive or negative association between our variables and the probability of success
- $\hat{\beta}$: The statistic representation of the positive or negative association between our variables and the probability of success

4 Background

Our issue stems from the achievement gap in the U.S. educational system. The achievement gap was not an issue discovered, rather, an issue that was observed over time. According to the Teachers College Columbia University, "by age three, children of professionals have vocabularies that are nearly 50 percent greater than those of working class children, and twice as large as those of children whose families are on welfare." [8] There are many elements that contribute to the educational achievement gap. For example, within the United States, there exists an inequality on the accessibility on education. We know that not all students from diverse backgrounds do not have the same opportunity of a quality education. According to [6], lower-income families and people living in poverty are more prone to under perform in a school environment than fortunate groups due to the financial differences. In addition, from [5], the U.S. families that qualify for free-reduced lunch are more likely to score lower on national test scores than U.S. families that do not qualify. Class status and financial differences affect students' performance in education in the U.S. Similarly, there exist an achievement gap between races and ethnic groups due to various culture dynamics. We seek to reveal trends between students' financial differences. We will analyze the achievement gap and show how these inequalities influence students' performances.

Statistics is under the branch of applied mathematics. Its power lies in its' ability to make predictions and assumptions on a population based on a given sample. Statistics is used in our modern daily lives from sports analysis to policy making. Statistics has been around since ancient times. In 3050 B.C., Egyptians used statistics to determine population and wealth. China was found to have used statistics as early as 2300 B.C. In 594 B.C., ancient Greece collected data for levying taxes. Athens would take a census during 309 B.C. to gather information about their population. But the most advanced group of people were the Romans, who collected data and conducted surveys throughout their country and provinces.

We begin to analyze the achievement gap by first understanding simple logistic regression. This type of regression model allows us to study how quantitative/qualitative explanatory variables affect a quantitative response variable. The model provides a proportion that represents a probability that a randomly selected subject has a successful outcome. Regression models are powerful tools used to predict the probability of a specific result. Our topic focuses on if students' pass or fail a standardized math test given their grade level and economic status. To analyze this problem, we will conduct analysis using a multiple logistic regression model and a mixed effect logistic model since our response variable is the test outcomes (pass or fail) and our explanatory variables include economic status and grade level. Multiple logistic regression explains how a probable outcome of a specific categorical response variable depends on the values of multiple explanatory variables.

5 Data

The data used in this study was gathered by the California Assessment of Student Performance and Progress. The CAASPP is a government system that evaluates students' testing progress on mathematics and English literacy. The data is focused on the Salinas City Elementary School District which has 13 elementary school campuses (K-6). We separated the data by grade level, focusing on third, fourth, and fifth grade. The variables collected by this assessment are the test outcomes (pass / fail), economic status (economically disadvantage / not economically disadvantage), and grade level. To define pass and for each grade level, we refer to Figure 1 from [7].

Mathematics

Grade	Minimum Scale Score	Maximum Scale Score	Achievement Level Scale Score Range for Standard Not Met	Achievement Level Scale Score Range for Standard Nearly Met	Achievement Level Scale Score Range for Standard Met	Achievement Level Scale Score Range for Standard Exceeded
3	2189	2621	2189–2380	2381–2435	2436–2500	2501–2621
4	2204	2659	2204–2410	2411–2484	2485–2548	2549–2659
5	2219	2700	2219–2454	2455–2527	2528–2578	2579–2700
6	2235	2748	2235–2472	2473–2551	2552–2609	2610–2748
7	2250	2778	2250–2483	2484–2566	2567–2634	2635–2778
8	2265	2802	2265–2503	2504–2585	2586–2652	2653–2802
11	2280	2862	2280–2542	2543–2627	2628–2717	2718–2862

Figure 1: Score Ranges

We define fail by using Standard Not Met and Standard Nearly Met and pass by using Standard Met and Standard Exceeded. In addition, we categorized economically disadvantaged as low income families. Based on the National Center of Children in Poverty, "families with incomes below this level are referred to as low income $44,700 for a family of four." [1]. Each student is categorized by an economic status in a specific grade level within the Salinas City Elementary School District.

6 Method

In this section, we will introduce multiple logistic regression and mixed effect logistic regression.

6.1 Multiple Logistic Regression

Multiple logistic regression can explain how the probability of failure (standard not met) depends on economic status and grade level. Multiple logistic regression is used when the response variables are categorical with two levels, such as pass or fail. For this model, we will disregard the variability among schools and look at the Salinas City Elementary School District as a whole. For example, students within the same school may perform similarly, and in this section, we disregard this tendency. Then we will account for the variability among school in a later section.

In our model, X represents the economic status and Y represents the performance outcome, and Z represents grade level. We define the three variables as follows:

Let

$$\begin{cases} X = 1 & \text{if not disadvantaged} \\ X = 0 & \text{otherwise (disadvantaged)} \end{cases}$$

Let

$$\begin{cases} Y = 1 & \text{if student failed} \\ Y = 0 & \text{otherwise (passed)} \end{cases}$$

Let P(Y=1) = π which is the probability of failure

Let

$$\begin{cases} Z_1 = 1 & \text{if fourth grade} \\ Z_1 = 0 & \text{otherwise} \end{cases}$$

Let

$$\begin{cases} Z_2 = 1 & \text{if fifth grade} \\ Z_2 = 0 & \text{otherwise} \end{cases}$$

If $Z_1 = Z_2 = 0$ then third grade.

In our multiple logistic model, the relationship between X, Y, Z_1, and Z_2 is as follows:

$$P(Y=1) = \pi = \frac{e^{\beta_0+\beta_1 X+\beta_2 Z_1+\beta_3 Z_2}}{1+e^{\beta_0+\beta_1 x+\beta_2 Z_1+\beta_3 Z_2}} \tag{1}$$

We will consider β_0, β_1, β_2, β_3 as our parameters. This equation shows an S-shaped curve for the probability of failure π, because π is always between 0 and 1.

We begin our analysis by assuming,

$H_0 : \beta_1 = 0$, economic status and π are not associated
$H_1 : \beta_1 \neq 0$, economic status and π are associated

6.2 Mixed Effect Logistic Regression

For our second analysis, we will be using a mixed effect logistic model which takes into account the natural variability among schools. We will once again be focusing on the supposition between economic status and performance. We will define X, Y, Z as the same variables as in our previous model. We similarly define the relationship among X, Y, Z as:

$$P(Y=1) = \pi = \frac{e^{\alpha_i+\beta_0+\beta_1 X+\beta_2 Z_1+\beta_3 Z_2}}{1+e^{\alpha_i+\beta_0+\beta_1 x+\beta_2 Z_1+\beta_3 Z_2}}, \tag{2}$$

where α_i is a school level parameter which accounts for the similarity of students belonging to the i^{th} group, and we consider β_0, β_1, β_2, β_3 as population-level parameters.

We begin our analysis by assuming,

$H_0 : \beta_1 = 0$, economic status and π are not associated
$H_1 : \beta_1 \neq 0$, economic status and π are associated

7 Results

In this chapter, we present the graphics, multiple logistic regression results, and mixed effect logistic regression results from our data.

7.1 Descriptive Statisics

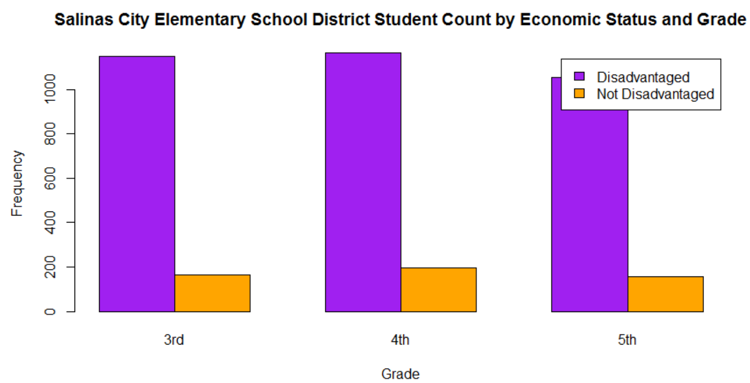

Figure 2: Economic Status by Grade Level

Figure 2 separates economic status between each grade level. The graph shows there are less than $1,200$ students in each grade that are categorized as economically disadvantaged while less than 200 students in each grade are not economically disadvantaged. The majority of students in the Salinas City Elementary School District are economically disadvantaged.

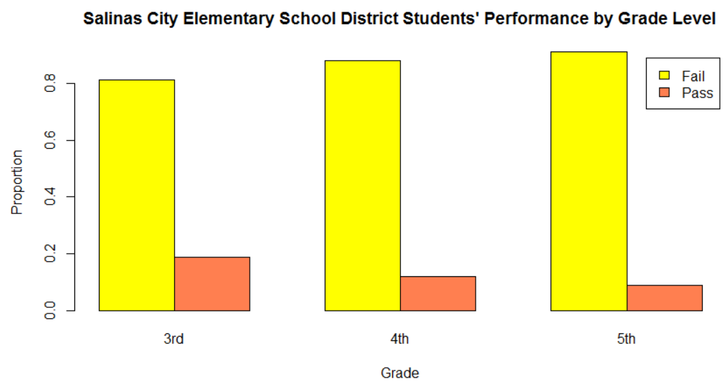

Figure 3: Performance by Grade Level

The total number of students from the 13 schools in the Salinas City Elementary School District is $3,887$. Figure 3 shows the students' grade level split by test performance (i.e. pass/fail). For instance, the proportion of students in the third grade who failed is 0.8130 whereas only 0.1870 proportion of students passed.

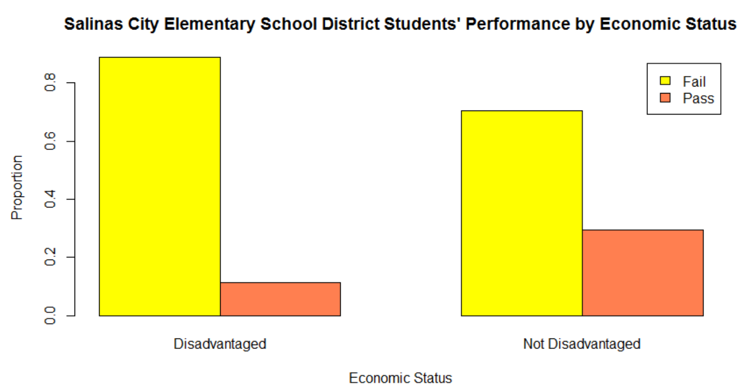

Figure 4: Performance by Economic Status

Figure 4 demonstrates the affects between economic status and students' test performance. The proportion of students who are categorized as economically disadvantaged and failed is 0.8872 while only 0.1128 proportion of students passed. In contrast, only 0.7053 proportion of students who are categorized as not economically disadvantaged and failed while 0.2947 proportion of students passed.

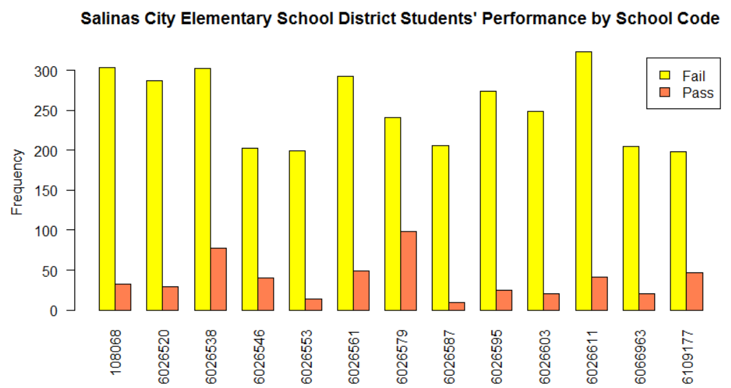

Figure 5: Performance by School

Figure 5 displays how each school performed on the test.

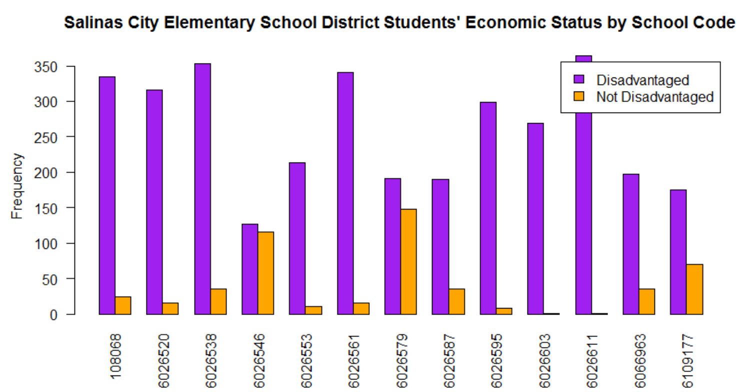

Figure 6: Economic Status by School

Figure 6 shows the number of students from each school who are either disadvantaged or not disadvantaged.

7.2 Results from Multiple Logistic Regression

```
Call:
glm(formula = prop.fail ~ econ + grade, family = "binomial",
    data = data0, weights = total.tested)

Deviance Residuals:
       1          2          3          4          5          6
-0.78432   0.03129   1.09988   1.57863  -0.05072  -1.76419

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.64792    0.07539  21.858  < 2e-16 ***
econnot     -1.24161    0.11175 -11.111  < 2e-16 ***
grade4th     0.54623    0.11056   4.941 7.79e-07 ***
grade5th     0.80888    0.12157   6.653 2.86e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Based on our data, we estimated that $\hat{\beta_1} = -1.2416$ where the p-value ≈ 0. We reject the null hypothesis that there is no association between economic status and the probability of failure.

Based on these estimates, we can estimate π (the probability of failure) for each sub-population coded by X, Z_1, and Z_2.
Our multiple logistic model summary states that our:
$\hat{\beta_0} = 1.6479$
$\hat{\beta_1} = -1.2416$
$\hat{\beta_2} = 0.562$
$\hat{\beta_3} = 0.8089$

The probability of failure when X = 1 (not disadvantaged) and $Z_1 = Z_2 = 0$ (third grade):

$$P(Y = 1) = \pi = \frac{e^{\hat{\beta_0}+\hat{\beta_1}}}{1 + e^{\hat{\beta_0}+\hat{\beta_1}}} = 0.6010879 \tag{3}$$

The percentage of failing students among not disadvantaged third graders is 60% in the Salinas City Elementary School District.

The probability of failure when X = 0 (disadvantaged) and $Z_1 = Z_2 = 0$ (third grade):

$$P(Y = 1) = \pi = \frac{e^{\hat{\beta_0}}}{1 + e^{\hat{\beta_0}}} = 0.8388911 \tag{4}$$

The percentage of failing students among disadvantaged third graders is 84%.

The probability of failure when X = 1 (not disadvantaged) and $Z_1 = 1$ and Z_2=0 (fourth grade):

$$P(Y = 1) = \pi = \frac{e^{\hat{\beta}_0+\hat{\beta}_1+\hat{\beta}_2}}{1 + e^{\hat{\beta}_0+\hat{\beta}_1+\hat{\beta}_2}} = 0.72331218 \tag{5}$$

The percentage of failing students among not disadvantaged fourth graders is 72% in the Salinas City Elementary School District.

The probability of failure when X = 0 (disadvantaged) and $Z_1 = 1$ and $Z_2 = 0$ (fourth grade):

$$P(Y = 1) = \pi = \frac{e^{\hat{\beta}_0+\hat{\beta}_2}}{1 + e^{\hat{\beta}_0+\hat{\beta}_2}} = 0.9002495 \tag{6}$$

The percentage of failing students among disadvantaged fourth graders is 90%.

The probability of failure when X = 1 (not disadvantaged) and $Z_1 = 0$ and Z_2=1 (fifth grade):

$$P(Y = 1) = \pi = \frac{e^{\hat{\beta}_0+\hat{\beta}_1+\hat{\beta}_3}}{1 + e^{\hat{\beta}_0+\hat{\beta}_1+\hat{\beta}_3}} = 0.7720635 \tag{7}$$

The percentage of failing students among not disadvantaged fifth graders is 77% in the Salinas City Elementary School District.

The probability of failure when X = 0 (disadvantaged) and $Z_1 = 0$ and $Z_2 = 1$ (fifth grade):

$$P(Y = 1) = \pi = \frac{e^{\hat{\beta}_0+\hat{\beta}_3}}{1 + e^{\hat{\beta}_0+\hat{\beta}_3}} = 0.9212897 \tag{8}$$

The percentage of failing students among disadvantaged fifth graders is 92%.

7.3 Results from Mixed Effect Logistic Regression

```
Formula: y ~ econ + grade + (1 | school)
   Data: better

     AIC      BIC   logLik deviance df.resid
  2782.9   2814.1  -1386.5   2772.9     3779

Scaled residuals:
    Min      1Q  Median      3Q     Max
-5.3306  0.2557  0.3128  0.4057  0.9785

Random effects:
 Groups Name        Variance Std.Dev.
 school (Intercept) 0.1882   0.4339
Number of obs: 3784, groups:  school, 13

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   1.7418     0.1448  12.031  < 2e-16 ***
econnot      -0.9884     0.1498  -6.598 4.16e-11 ***
grade4th      0.5539     0.1142   4.849 1.24e-06 ***
grade5th      0.8716     0.1278   6.822 8.97e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Based on our data, we estimated that $\hat{\beta}_1 = -0.9884$ where the p-value ≈ 0. We reject the null hypothesis that there is no association between economic status and the probability of failure. Compared to $\hat{\beta}_1$ = -1.2416 from the previous model, the estimate $\hat{\beta}_1$ = -0.9884 is closer to 0 (i.e. weaker association) but we would treat as a more reasonable estimate because this model takes into account the natural variability among schools.

Our mixed effect logistic model summary states that our:

α's

school code 108068: $\hat{\alpha_1} = 0.0627$
school code 6026520: $\hat{\alpha_2} = 0.1138$
school code 6026538: $\hat{\alpha_3} = -0.6545$
school code 6026546: $\hat{\alpha_4} = 0.0286$
school code 6026553: $\hat{\alpha_5} = 0.3546$
school code 6026561: $\hat{\alpha_6} = -0.3237$
school code 6026579: $\hat{\alpha_7} = -0.7100$
school code 6026587: $\hat{\alpha_8} = 0.7336$
school code 6026595: $\hat{\alpha_9} = 0.1941$
school code 6026603: $\hat{\alpha_{10}} = 0.2664$
school code 6026611: $\hat{\alpha_{11}} = -0.0679$
school code 6066963: $\hat{\alpha_{12}} = 0.2215$
school code 6109177: $\hat{\alpha_{13}} = -0.3430$

β's

$\hat{\beta_0} = 1.7418$
$\hat{\beta_1} = -0.9884$
$\hat{\beta_2} = 0.5539$
$\hat{\beta_3} = 0.8716$

Suppose we have many schools similar to school "6026546" (i.e. same economic level). For unknown reasons (i.e. not accounted under the model), $\hat{\alpha_1}$ = 0.0286 means this school is worse than average when compared to similar schools. On the contrary, schools similar to "6026538" with $\hat{\alpha_3}$ = -0.6545 are better than average when compared to similar schools.

Here are estimated failure probabilities for each economic status and grade level. We let $\hat{\alpha_i} = 0$. In other words, our estimates are for average performance schools.

The probability of failure when X = 1 (not disadvantaged) and $Z_1 = Z_2 = 0$ (third grade):

$$P(Y = 1) = \pi = \frac{e^{\hat{\alpha_i}+\hat{\beta_0}+\hat{\beta_1}}}{1 + e^{\hat{\alpha_i}+\hat{\beta_0}+\hat{\beta_1}}} = 0.6799191 \tag{9}$$

The percentage of failing students among not disadvantaged third graders is 68% in the Salinas City Elementary School District.

The probability of failure when X = 0 (disadvantaged) and $Z_1 = Z_2 = 0$ (third grade):

$$P(Y = 1) = \pi = \frac{e^{\hat{\alpha_i}+\hat{\beta_0}}}{1 + e^{\hat{\alpha_i}+\hat{\beta_0}}} = 0.8509156 \tag{10}$$

The percentage of failing students among disadvantaged third graders is 85%.

The probability of failure when X = 1 (not disadvantaged) and $Z_1 = 1$ and Z_2=0 (fourth grade):

$$P(Y = 1) = \pi = \frac{e^{\hat{\alpha_i}+\hat{\beta_0}+\widehat{\beta_1+\hat{\beta_2}}}}{1 + e^{\hat{\alpha_i}+\hat{\beta_0}+\widehat{\beta_1+\hat{\beta_2}}}} = 0.787061 \tag{11}$$

The percentage of failing students among not disadvantaged fourth graders is 79% in the Salinas City Elementary School District.

The probability of failure when X = 0 (disadvantaged) and $Z_1 = 1$ and $Z_2 = 0$ (fourth grade):

$$P(Y=1) = \pi = \frac{e^{\hat{\alpha_i}+\hat{\beta_0}+\hat{\beta_2}}}{1+e^{\hat{\alpha_i}+\hat{\beta_0}+\hat{\beta_2}}} = 0.9085203 \tag{12}$$

The percentage of failing students among disadvantaged fourth graders is 91%.

The probability of failure when X = 1 (not disadvantaged) and $Z_1 = 0$ and Z_2=1 (fifth grade):

$$P(Y=1) = \pi = \frac{e^{\hat{\alpha_i}+\hat{\beta_0}+\beta_1\hat{+}\hat{\beta_3}}}{1+e^{\hat{\alpha_i}+\hat{\beta_0}+\beta_1\hat{+}\hat{\beta_3}}} = 0.8354835 \tag{13}$$

The percentage of failing students among not disadvantaged fifth graders is 84% in the Salinas City Elementary School District.

The probability of failure when X = 0 (disadvantaged) and $Z_1 = 0$ and $Z_2 = 1$ (fifth grade):

$$P(Y=1) = \pi = \frac{e^{\hat{\alpha_i}+\hat{\beta_0}+\hat{\beta_3}}}{1+e^{\hat{\alpha_i}+\hat{\beta_0}+\hat{\beta_3}}} = 0.931719 \tag{14}$$

The percentage of failing students among disadvantaged fifth graders is 92%.

8 Conclusion

The purpose of our project is to investigate if there exists an achievement gap between students' economic status conditioning on grade level in the Salinas Elementary City School District. Through our descriptive statistics, a large number of students are disadvantaged. Also, the majority of students fail the CAASPP math test. Note that as students move up in each grade level the proportion of students who fail increases, whereas the proportion of students who pass decreases.

Based on our multiple logistic regression analysis, the data shows there is an association between economic status and the probability of failure. Since this association is proven likely, further analysis is done. Disadvantaged third graders are 1.4 times more likely to fail than not disadvantaged third graders. Disadvantaged fourth graders are 1.2 times more likely to fail than not disadvantaged fourth graders. Disadvantaged fifth graders are 1.2 times more likely to fail than not disadvantaged fifth graders.

Based on our mixed effect logistic regression analysis, the data shows there is an association between economic status and the probability of failure where the variability among schools is accounted for. For this reason, we assume this model is more reasonable than the multiple logistic regression model. Furthermore, we were able to see which schools tend to be above or below average (when compared to hypothetical schools with the same economic status and grade level) under the structure of a logistic regression model. Disadvantaged third graders are 1.3 times more likely to fail than not disadvantaged third graders. Disadvantaged fourth graders are 1.2 times more likely to fail than not disadvantaged fourth graders. Disadvantaged fifth graders are 1.1 times more likely to fail than not disadvantaged fifth graders.

Not every student can have the same quality of education as others within the United States. Through our statistical analysis, we can say there exist an association between economic status and probability of failure in the Salinas City Elementary School District. In other words, the vast majority of elementary school students who fail the CAASPP math test tend to be economically disadvantaged. When 0.8872 proportion of disadvantaged students fail while only 0.1128 proportion of disadvantaged students who passed, this is an issue within our community. We focus on economic status as a factor in the educational achievement gap. However, race, class status, gender, language, parental education, etc. are other contributions to this gap. We must be aware of this issue, because the education achievement gap effects children within our community and generations of children as a whole. We hope to raise awareness on this issue as well as encourage bringing quality education to future young minds.

References

[1] Addy, Sophia, Will Engelhardt, and Curtis Skinner. *"Basic Facts About Low-income Children."* NCCP. N.p., Jan. 2013. Web. 04 Mar. 2016. $< http://www.nccp.org/publications/pub_{1}074.html >$.

[2] Agresti, Alan, and Christine A. Franklin. *Statistics: The Art and Science of Learning from Data.* Upper Saddle River, NJ: Pearson Prentice Hall, 2007. Print.

[3] California Assessment of Student Performance and Progress (CAASPP)

[4] Eves, Howard W.. *"A Very Brief History of Statistics".* The College Mathematics Journal 33.4 (2002): 306–308. Web.

[5] *Reference Tables*. National Center of Education Statistics, 2015.

[6] Richard Rothstein *Whose Problem is Poverty?*. Research Associate at the Economic Policy Institute, 2008.

[7] *Smarter Balanced Scale Score Ranges* California Department of Education

[8] *The Academic Achievement Gap: Facts & Figures.* "TC Media Center from the Office of External Affairs." N.p., 09 June 2005. Web. 18 Mar. 2016.

9 Appendix A: Graphics R Code

```
##Install Package mosaic
library(mosaic)

##Data
data < − read.csv(file.choose())
#choose student_level_data

##Restructuring data
data1 < − subset(data, school != 0)
data1$economic < − ifelse(data1$econ == ”disadv”, ”Disadvantaged”, ”Not Disadvantaged”)
str(data1)
schoolcode < − factor(data1$school)

##Tables
t1 < − table(data1$econ, data1$grade)
t2 < − table(data1$performance, data1$grade)
t3 < − table(data1$performance, data1$economic)
t4 < − table(data1$performance, schoolcode)
t5 < − table(data1$econ, schoolcode)

##Barplots
barplot(t1, main = ”Salinas City Elementary School District Student Count by Economic Status and Grade”, xlab = ”Grade”, ylab = ”Frequency”, col = c(”purple”, ”orange”), legend = c(”Disadvantaged”, ”Not Disadvantaged”), beside= T, axes = T)

barplot(prop.table(t2, 2), main = ”Salinas City Elementary School District Students’ Performance by Grade Level”, xlab = ”Grade”, ylab = ”Proportion”, col = c(”yellow”, ”coral”), legend = c(”Fail”, ”Pass”), beside=T, axes = T)

barplot(prop.table(t3, 2), main = ”Salinas City Elementary School District Students’ Performance by Economic Status”, xlab = ”Economic Status”, ylab = ”Proportion”, col = c(”yellow”, ”coral”), legend = c(”Fail”, ”Pass”), beside=T, axes = T)

barplot(t4, main = ”Salinas City Elementary School District Students’ Performance by School Code”, ylab = ”Frequency”, col = c(”yellow”, ”coral”), legend = c(”Fail”, ”Pass”), beside=T, axes = T, las = 2)

barplot(t5, main = ”Salinas City Elementary School District Students’ Economic Status by School Code”, ylab = ”Frequency”, col = c(”purple”, ”orange”), legend = c(”Disadvantaged”, ”Not Disadvantaged”), beside=T, axes = T, las = 2)
```

10 Appendix B: Multiple Logistic Regression R Code

```
##Install Package mosaic
library(mosaic)

    ##Data
data < − read.csv( file.choose())
#choose school_level_data

    ##Restructuring data
head(data,7)
data0 < − data[1:6,]

    ##Multiple Logistic Model
fit < − glm( prop.fail ∼ econ + grade, family=”binomial”, weights=total.tested,
data=data0 )
summary(fit)

    ##Results
beta0 < − 1.65
beta1 < − (-1.24)
beta2 < − 0.55
beta3 < − 0.81

#When X=1, Z1=Z2=0
pi1 < − (exp(beta0+beta1))/(1+exp(beta0+beta1))

#When X=0, Z1=Z2=0
pi2 < − (exp(beta0))/(1+exp(beta0))

#comparing 3rd graders
pi2/pi1

#When X=1, Z1=1, Z2=0
pi3 < − (exp(beta0+beta1+beta2))/(1+exp(beta0+beta1+beta2))

#When X=0, Z1=1, Z2=0
pi4 < − (exp(beta0+beta2))/(1+exp(beta0+beta2))

#comparing 4th graders
pi4/pi3
```

```
#When X=1, Z1=0, Z2=1
pi5 <- (exp(beta0+beta1+beta3))/(1+exp(beta0+beta1+beta3))

#When X=0, Z1=0, Z2=1
pi6 <- (exp(beta0+beta3))/(1+exp(beta0+beta3))

#comparing 5th graders
pi6/pi5
```

11 Appendix C: Mixed Effect Logistic Regression R Code

```
##Install Package lme4
library(lme4)

##Data
good <- read.csv( file.choose() )
#choose student_level_data

##Restructuring Data
better <- subset( good, school = 0 )
better$y <- ifelse(better$performance == "pass", 0, 1 )
head(better)

##Mixed Effect Logistic Model
fit3 <- glmer( y ~ econ + grade + (1| school), family="binomial", data=better)
summary(fit3)
random.effects(fit3)

##Results
beta0 <- 1.7418
beta1 <- (-0.9884)
beta2 <- 0.5539
beta3 <- 0.8716

#When X=1, Z1=Z2=0
pi1 <- (exp(beta0+beta1))/(1+exp(beta0+beta1))

#When X=0, Z1=Z2=0
pi2 <- (exp(beta0))/(1+exp(beta0))

#comparing 3rd graders
```

```
pi2/pi1

#When X=1, Z1=1, Z2=0
pi3 < − (exp(beta0+beta1+beta2))/(1+exp(beta0+beta1+beta2))

#When X=0, Z1=1, Z2=0
pi4 < − (exp(beta0+beta2))/(1+exp(beta0+beta2))

#comparing 4th graders
pi4/pi3

#When X=1, Z1=0, Z2=1
pi5 < − (exp(beta0+beta1+beta3))/(1+exp(beta0+beta1+beta3))

#When X=0, Z1=0, Z2=1
pi6 < − (exp(beta0+beta3))/(1+exp(beta0+beta3))

#comparing 5th graders
pi6/pi5
```