



University of North Texas at Dallas

Strategic Analysis & Reporting

Analysis on 9-12 Grade Dropout Rate in Texas High Schools

Contents

ABSTRACT.....	2
INTRODUCTION.....	2
METHODOLOGY	2
DISCUSSIONS.....	5
CONCLUSION.....	7

ABSTRACT

With a purpose of better understanding high school dropout rate, we designed a research project which allows us to compare various dimensions and to identify variables that significantly impact dropout rate. In this article, the variables we will test include race, gender, and other geographic information. This study uses big data analysis technology to provide data-based evidence for educators to make policy planning and decisions.

INTRODUCTION

In 2014-15, 30,854 dropped out of Grades 9-12, which is 2.1 percent of all students. In the following part, we will determine the factors which influence the dropout rate and the importance of these variables in the equation.

The data was collected from Texas Education Agency (TEA). The dropout rate is for students who attended Texas public schools in 2014-2015. The data we downloaded is grouped by Region, County then by Campus. Within each campus, the counts of all students, and dropout students were provided. Then these counts were broken down by gender, race, gift, immigrant, migrant, overage. Our goal is to test if all the variables are contributing equally on the dropout rate across Texas public school. Therefore, we have the hypotheses as follows.

- H1: The percentages of major races (African American, Asian, Hispanic, White) will not influence the dropout rates across campus.
- H2: The percentages of gifted students will negatively influence the dropout rate across campus.
- H3: The percentages of immigrant, migrant, economic disability, ESL, and overage students will not influence the dropout rate across campus.
- H4: The percentages of female and male will not influence the dropout rate across campus.

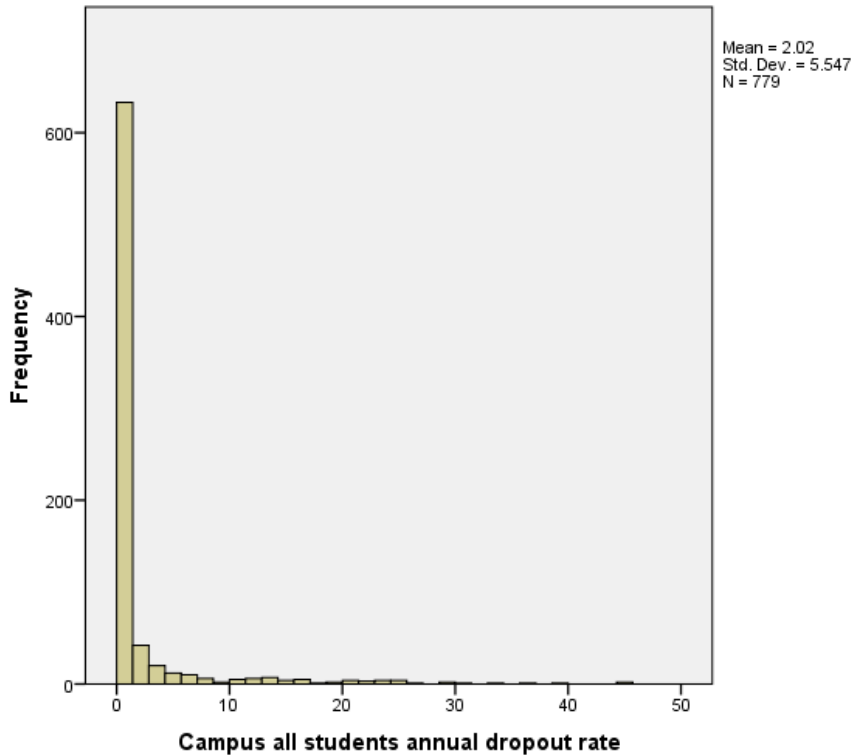
METHODOLOGY

We will test these twelve variables horizontally to see their influence on the performance of the dropout rate across campus. We quantified the variables by percentages (of all students on campus).

- 1) African American
- 2) Asian
- 3) Hispanic
- 4) White
- 5) ESL student
- 6) Economically disadvantage
- 7) Gifted
- 8) Immigrants
- 9) Migrant
- 10) Overage
- 11) Female

12) Male

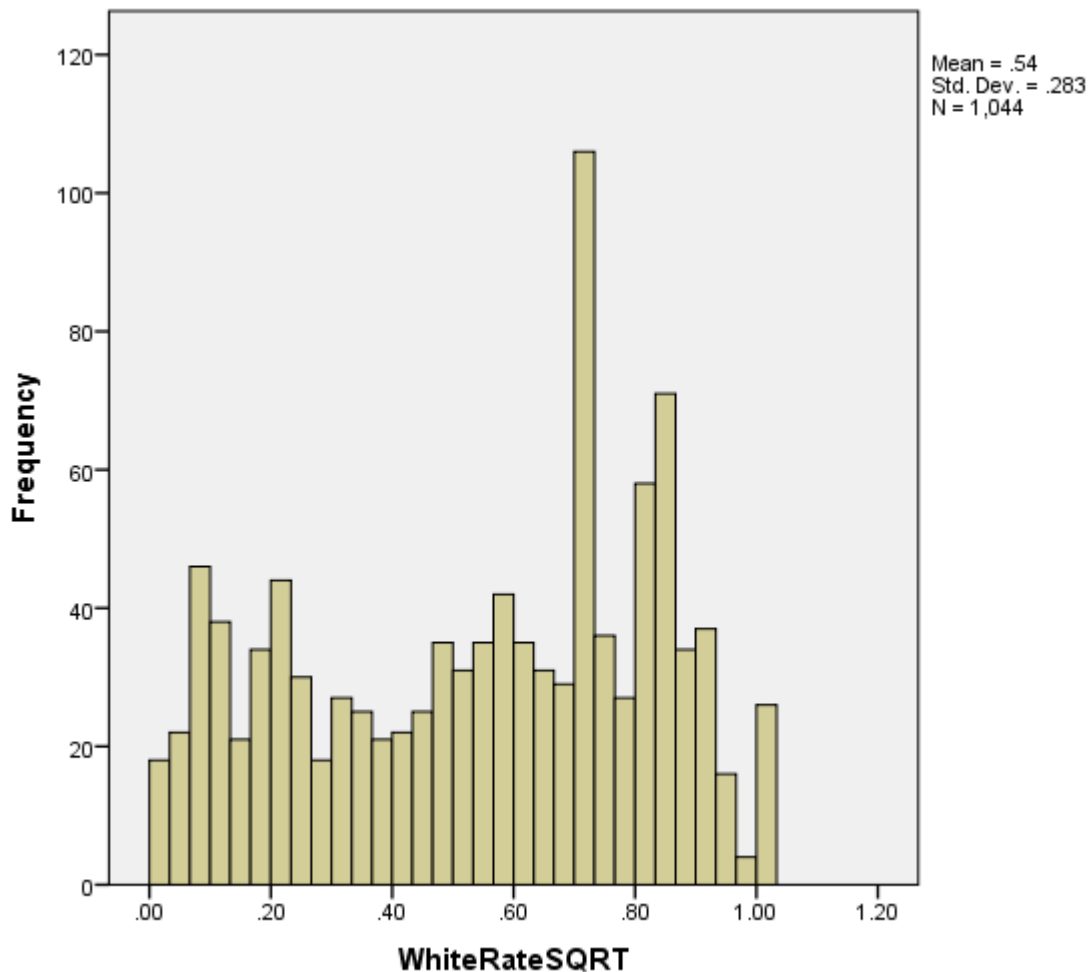
Then we quantified the dropout rate by separating all campuses into two groups, higher than the average and lower than the average. In our study, the average dropout rate is 2.02%. 675 campuses have dropout rate lower than the average, compared to 104 higher than the average. The distribution of the dropout rates is as follows, nearly exponentially decreasing.



Graph 1, distribution of dropout rate across Texas Public schools.

In cleaning the data, many invalid observations were eliminated. The data cleaning was processed by SPSS 24 application.

1. Removed the observations that contain 0% of either female or male. Then removed the ones that the total percentage of female and male was not 100%.
2. Removed the campuses which contain students exceed 1000 due to the small sample size.
3. Due to the high skewed independent variables, a square root transformation had been done on Xs.



Graph 2, White rate after SQRT transform

Data had been analyzed through SPSS Modeler 18 application. Several algorithms had been tested and eventually Logistic Regression was selected due to the advantage of explanation capability this algorithm provides. Not only does Logistic Regression generates coefficients of the variables, which were marked as positive or negative to indicate the directions of the variables, but also it horizontally compared different variables so that the analysts know the weights of variables. The basic interpretation rules are as follows.

- 1) Omnibus Tests of Model Coefficient, if Sig < 0.05, the model itself has ability to distinguish the dependent variables (higher or lower than average dropout rate).
- 2) Cox & Snell R Square shows how good Xs fit Y.
- 3) Classification Table will create accuracy rate.
- 4) Variables in Equation, B is the coefficient of the variable, positive means with the increasing of this variable, the chance of dropout rate higher than average increase.
- 5) Sig. shows if a variable is statistically significant in separating the dependent variable.
- 6) Exp(B) means when other situation stay unchanged, the increase of the variable by one unit, the chance the dependent variable will change.

DISCUSSIONS

		Chi-square	df	Sig.
Step 1	Step	503.950	11	.000
	Block	503.950	11	.000
	Model	503.950	11	.000

Table 1, Omnibus Tests of Model Coefficient

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	108.336 ^a	.476	.875

a. Estimation terminated at iteration number 9 because parameter estimates changed by less than .001.

Table 2, Model Summary, Correlation

		Predicted			
		DROP_OUT_RATE_STATE_MEAN		Percentage Correct	
		Lower than Mean	Higher than mean		
Observed					
Step 1	DROP_OUT_RATE_STATE_MEAN	Lower than Mean	666	9	98.7
		Higher than mean	13	91	87.5
	Overall Percentage				97.2

Table 3, Classification Table

Table 1 shows the P-value of the Omnibus Tests of the Model Coefficients. In the third row, the Sig. of the Model is .000, which means the Model is highly significant, and statistically speaking, the variables we selected can successfully separate the dependent variable into two groups.

Table 2 suggests a good fitness between all the independent variables and the dependent variable.

Table 3 indicates a good accuracy rate of the model. 97.2% means with the model we created, we can successfully predict whether a campus will have a dropout rate higher or lower than average.

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a						
AARateSQRT_transformed	1.321	.277	22.799	1	.000	3.749
AsianRateSQRT_transformed	.365	.232	2.469	1	.116	1.440
HispanicRateSQRT_transformed	.758	.413	3.375	1	.066	2.134
WhiteRateSQRT_transformed	-.334	.414	.651	1	.420	.716
ESLRateSQRT_transformed	.104	.258	.164	1	.685	1.110
EcoDisRateSQRT_transformed	.111	.356	.098	1	.755	1.118
GiftedRateSQRT_transformed	-.656	.272	5.821	1	.016	.519
ImmigRateSQRT_transformed	.439	.269	2.668	1	.102	1.551
MigrantRateSQRT_transformed	-.175	.238	.544	1	.461	.839
OverageRateSQRT_transformed	4.981	.742	45.051	1	.000	145.566
FemaleRatesqrt_transformed	-.466	.351	1.768	1	.184	.627
Constant	-6.507	.882	54.437	1	.000	.001

a. Variable(s) entered on step 1: AARateSQRT_transformed, AsianRateSQRT_transformed, HispanicRateSQRT_transformed, WhiteRateSQRT_transformed, ESLRateSQRT_transformed, EcoDisRateSQRT_transformed, GiftedRateSQRT_transformed, ImmigRateSQRT_transformed, MigrantRateSQRT_transformed, OverageRateSQRT_transformed, FemaleRatesqrt_transformed.

Table 4, Variables in Equation

Table 4 shows many valuable information. These are the variables that statistically significant (Sig<0.1).

- 1) Rate of African American
- 2) Rate of Hispanic
- 3) Rate of Gifted
- 4) Rate of Overage

The following are the variables that have positive direction, which means with the increase of the percentage of these students, the dropout rate of the campus has larger chance to be higher than average.

- 1) Rate of African American
- 2) Rate of Hispanic
- 3) Rate of Overage

On the opposite, rate of Gifted student has a negative direction, which means with the increase of the percentage of gifted students, the campus has larger chance to have a lower than average dropout rate.

With the information that we collected from table 4, we can finish the hypothesis finding.

- H1: The percentages of major races (African American, Asian, Hispanic, White) will not influence the dropout performance across campuses.

Hypothesis is not accepted, because, according to the result, the percentages of African American and Hispanic are positively significant. Meanwhile, the percentages of Asian and White are not significant. Hence, races do not have equal influence on the dropout rate.

- H2: The percentage of gifted students will negatively influence the dropout rates across campuses.

Hypothesis accepted, as the result shows, the percentage of gifted students has a coefficient of -0.66, and it statistically significant. As a result, the percentage of gifted students on campus will lower the chance the campus has a higher than average dropout rate.

- H3: The percentages of immigrant, migrant, economic disability, ESL and overage students will not influence the dropout rate across campus.

Hypothesis not accepted, with the percentage of overage students' significant P-value. This is the most unexpected found in this study. The rate of overage student has the heaviest weight among the variables. So, with the increasing of the percentage of overage student, a campus has larger chance to have a higher than average dropout rate.

- H4: The percentages of female and male will not influence the dropout rate across campus.

Hypothesis accepted, the P-value of the percentage of female student is not significant. The percentage of male student is not in the test as it has a perfect negative relationship with the percentage of female student (combine equal to 100%). However, the rate of female has a negative direction, which means even though not significant, the increase of the percentage of female on campus lower the chance the campus has a higher than average dropout rate.

CONCLUSION

Many variables influence the dropout rate of a campus. In this study, we demonstrated how the percentage of races, gifted and other factors influence the dropout rates. Educators and decision makers can use the result to better understand these variables. This study proved the increasing percentage of African American, Hispanic and overage student will increase the dropout rate where the percentage of gifted student will decrease the dropout rates. The limitation of this study is also obvious, we cannot take the parents' education background into consideration, which we expect has high impact on the dropout rate. Secondly, we do not have financial data to assist the model building, which could be another crucial factor for dropout.