NJCU EDTC 810

Statistics for Educational Research Dr. Glazer

Spring 2019

Assignment 4

Emily Vandalovsky

Introduction

High school competition and graduation are important contributors to further success. It is the first milestone connected to academic and professional accomplishments of students. Ability to earn a high school diploma effects individual people, their families, local communities and districts.

On a macro-level, graduation rates are indicators of the well-being of the economy. Educators, school leaders, district officials, local governing bodies and legislators use graduation rates as indicators of the state of the educational system and its emerging needs.

To create common criteria for accounting and allow all school districts to conform to the universal calculation procedure, the National Governors Association (NGA) established the 4year adjusted cohort graduation rate (NGA, 2005). Once the federal government adopted the formula for the Adjusted Cohort calculation, it was disseminated to states and school districts for adapting. Establishing commonly shared calculation tool and reporting mechanism enabled individual constituencies, such as schools, districts, and counties to approach graduation rates from multiple points of view and track the status of their students' progress from onboarding to graduation.

Body

The data set used in the current report is has been collected by the New Jersey Department of Education and its education data system New Jersey Standards Measurement and Resource for Teaching (NJ SMART). I find this data particularly interesting because 1) it provides the opportunity for the descriptive analysis and central tendency markers; 2) it allows to identify the emerging trends and, possibly, effective interventions through inferential analysis; 3)

1

it challenges college faculty (particularly of community college level) to investigate the relationship between the high school graduation rates and the rates of the college population enrolled in remedial courses (outside of the scope of this paper).

The report uses cohort-based data, where each cohort is defined by their graduation year in a given entity (NJ DOE, 2018). The formula for the graduation rate is calculated as a ratio between the number of graduates (students receiving a High School diploma) and the number of students who entered high school four years ago as 9th graders. It also takes into account students verified movements within and outside the district. The formula for the Graduation Rate in Year Y is calculated as:

Year Y Cohort Graduates [1st time 9th Graders in Year Y-4]+[Transfers in]-[Transerfs Out]-[Excluded from Cohort]

Descriptive Statistics. Descriptive Statistics. The table Cohort 2015-2018 4-Year Graduation Rates (see sample in Appendix A) states graduation rates for years 2015, 2016, 2017, and 2018 on school and district. In total, the table contains 725 records, with the distinct records reported for the school levels and the district levels. In cases when the district contains only one high school, two identical records are reported, one for the school and one for the district. In cases of districts consisting of multiple reporting schools, the district level value is an aggregate of the data from two or more high schools.

To effectively perform descriptive analysis on the homogeneous data, it was necessary to split the original table into two parts by the level of reporting using SPSS tool Split Dataset into Separate Files. As a result of the split, two new data sets were formed: Cohort 2015-2018 4-Year Graduation Rates District Level with 308 records (Appendix B) and Cohort 2015-2018 4-Year Graduation Rates School Level with 417 records.

2

Values and File Names for Split Files Written										
	Values or Labels	Directory	Data File							
1		C:\ED- D\Spring2019 \EDTC810- Stat-for-Educ- Research\As signment 4\Data\Split	levelsav							
2	District	C:\ED- D\Spring2019 \EDTC810- Stat-for-Educ- Research\As signment 4\Data\Split	level_District. sav							
3	School	C:\ED- D\Spring2019 \EDTC810- Stat-for-Educ- Research\As signment 4\Data\Split	level_School. sav							
Based	on Variables: Leve	el								
Figure	Figure 1: Split original data set by level									

The descriptive statistics analysis is applied to each file individually, by district level and by school level. The frequencies table representing the data on the school level shows some missing data for the 2017, 2016 and 2015 graduation rates, therefore the population sizes for the four variables are different: N $_{2018} = 417$, N $_{2017} = 409$, N $_{2016} = 400$, and N $_{2015} = 394$.

Although the population sizes for all four years are different, the calculated values for mean are very close: $\mu_{2018} = 91.1$, $\mu_{2017} = 91.2$, $\mu_{2016} = 90.1$, $\mu_{2015} = 90.6$. The calculated values for medians are also very close: $M_{2018} = 94.2$, $M_{2017} = 94$, $M_{2016} = 93.8$, $M_{2015} = 93.2$. The calculated values for modes are the same for all four years and equal to 100, the maximum value of the dataset. This explains the negative skewness of data. As depicted in Figure 2, all

skewness coefficients are much less than -1 (-3.2, -2.4, -2.8, and -2.5 respectively), therefore the rates are considered highly skewed.

	Statistics										
		Class of 2018 4-Year Graduation Rate	Class of 2017 4-Year Graduation Rate	Class of 2016 4-Year Graduation Rate	Class of 2015 4-Year Graduation Rate						
N	Valid	417	409	400	394						
	Missing	0	8	17	23						
Mean		91.1431	91.2271	90.8926	90.6266						
Median		94.1500	94.0000	93.7500	93.2200						
Mode		100.00	100.00	100.00	100.00						
Std. Devia	ition	10.20757	9.05737	9.44894	9.23823						
Variance		104.194	82.036	89.282	85.345						
Skewness	S	-3.169	-2.448	-2.805	-2.464						
Std. Error	of Skewness	.120	.121	.122	.123						
Kurtosis		13.948	8.260	12.678	9.492						
Std. Error	of Kurtosis	.238	.241	.243	.245						
Range		79.31	65.57	76.92	70.90						
Minimum		20.69	34.43	23.08	29.10						
Maximum		100.00	100.00	100.00	100.00						
Sum		38006.67	37311.90	36357.04	35706.87						
Figure 2	Figure 2: Descriptive statistics for the school level data for years 2015-2018.										

Kurtosis, which shows the measure of the outliers or heavy tails in the data set (see Figure 2), appears to be high in all cases. As seen in histogram charts below (see Figure 3), every year depicts a leptokurtic distribution, with the highest value in the year 2018, which has no missing data points, as compared with the other three years.



Inferential Statistics. To further analyze the graduation rates reported by the schools and compare it with the reported value by the State of New Jersey (NJ DOE, 2018). *One-sample T*-test is used to evaluate whether a sample mean for each reported year is significantly different from the provided test value. An independent version of *one-sample T*-test is performed on each of the cohort years: 2018, 2017, 2016, and 2015 respectively. The following test values are used: Test Value $_{2018} = 90.91$, Test Value $_{2017} = 90.5$, Test Value $_{2016} = 90.06$, Test Value $_{2015} = 89.67$.

Since there are four independent samples being studied, four null hypothesis and four alternative hypotheses are issued. All versions of null hypothesis state that there is no statistical difference between the graduation rates reported by schools and the test value reported by the state. All versions of the alternative hypothesis will state that the difference between the two (the school-reported graduation rate and the state test value) is statistically significant. The four null hypotheses written in the statistical format are:

H_{0 2018}: $\mu_{2018} = 91.1$ is not statistically different from $T_{2018} = 90.91$; H_{0 2017}: $\mu_{2017} = 91.2$ is not statistically different from $T_{2017} = 90.5$; H_{0 2016}: $\mu_{2016} = 90.1$ is not statistically different from $T_{2016} = 90.06$; H_{0 2015}: $\mu_{2015} = 90.6$ is not statistically different from $T_{2015} = 89.67$.

The four alternative hypotheses written in statistical format are:

H_{A 2018}: $\mu_{2018} = 91.1$ is statistically different from $T_{2018} = 90.91$; H_{A 2017}: $\mu_{2017} = 91.2$ is statistically different from $T_{2017} = 90.5$; H_{A 2016}: $\mu_{2016} = 90.1$ is statistically different from $T_{2016} = 90.06$; H_{A 2015}: $\mu_{2015} = 90.6$ is statistically different from $T_{2015} = 89.67$.

Utilizing the analytical tool for *one-sample T*-test from SPSS, the following findings for the four years of graduation rates are recorded. In each scenario, the value of significance (Sig. 2-tailed) is carefully evaluated.

For the graduation year 2018, the significance value Sig. = 0.641 which is greater than 0.05. Therefore, the null hypothesis for the year 2018, H_{0 2018} fails to be rejected. Based on the significance value over 64% of data could be attributed to other causes or chance.

		Or	e-Sample Te	st			
Test Value = 90.91							
				Mean	95% Confidence Interval of Difference		
	t	df	Sig. (2-tailed)	Difference	Lower	Upper	
Class of 2018 4-Year Graduation Rate	.466	416	.641	.23309	7495	1.2157	

In other words, there is not enough evidence to reject that $\mu_{2018} = 91.1$ is not statistically

different from $T_{2018} = 90.91$.

For the graduation year 2017, the significance value Sig. = 0.105 which is greater than 0.05. Therefore, the null hypothesis for the year 2017, H_{0 2017} fails to be rejected. Based on the significance value, over 10% of data could be attributed to other causes or chance.

		Or	e-Sample Te	st				
Test Value = 90.5								
				Mean	95% Confidence Differ	e Interval of the ence		
	t	df	Sig. (2-tailed)	Difference	Lower	Upper		
Class of 2017 4-Year Graduation Rate	1.624	408	.105	.72714	1533	1.6075		

In other words, there is not enough evidence to reject that $\mu_{2017} = 91.2$ is not statistically

different from $T_{2017} = 90.5$.

For the graduation year 2016, the significance value Sig. = 0.709 which is greater than 0.05. Therefore, the null hypothesis for the year 2016, H $_{0\,2016}$ fails to be rejected. Based on the significance value, almost 8% of data could be attributed to other causes or chance.

One-Sample Test									
Test Value = 90.06									
				Mean	95% Confidence Differ	e Interval of the ence			
	t	df	Sig. (2-tailed)	Difference	Lower	Upper			
Class of 2016 4-Year Graduation Rate	1.762	399	.079	.83260	0962	1.7614			

In other words, there is not enough evidence to reject that $\mu_{2016} = 90.1$ is not statistically

different from $T_{2016} = 90.06$.

For the graduation year 2015, the significance value Sig. = 0.041 which is less than 0.05. Therefore, the null hypothesis for the year 2015, H_{0 2015} can be rejected. Based on the significance value, only 4% of data could be attributed to other causes or chance.

One-Sample Test								
Test Value = 89.67								
				Mean	95% Confidence Differ	e Interval of the ence		
	t	df	Sig. (2-tailed)	Difference	Lower	Upper		
Class of 2015 4-Year Graduation Rate	2.055	393	.041	.95657	.0416	1.8716		

Since there is enough evidence to reject that $\mu_{2015} = 90.6$ is not statistically different from $T_{2015} = 89.67$, the alternative hypothesis is supported. Therefore, the graduation rates reported by schools in the year 2015 $\mu_{2015} = 90.6$ are significantly different from the state reported test score $T_{2015} = 89.67$. This piece of information could be interpreted in a variety of ways and provoke further investigation of data and its variances.

To evaluate the variances of data within the dataset and to compare their differences, within groups, one-way ANOVA (or the Analysis of Variances) is used. This report analyzes variances between the means of the graduation rates of four years 2015-2018 between the

schools, as they are grouped by their state counties. The F statistic value confirms whether there is a significant difference between at least the two group means.

Using ANOVA analytical tool from SPSS, the groups are formed based on the schools' county jurisdiction. The county code represents an independent variable in the test. The graduation rates of four years represent the dependent variables in the test. The research needs to be conducted whether there is a significant difference in the graduation rates by the schools from different counties. The null hypothesis states there is no statistical difference between the graduation rates by the county.

H $_{01} = \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = \dots = \mu_n$, where n is the number of counties in the state.

Consequently, the alternative hypothesis states that there is a significant difference between at least two graduation rates by the county.

H A1 \neq $\mu_1 \neq$ $\mu_2 \neq$ $\mu_3 \neq$ $\mu_4 \neq$ $\mu_5 \neq$ $\mu_6 \neq$ μ_n , where n is the number of counties in the state.

ANOVA										
		Sum of Squares	df	Mean Square	F	Sig.				
Class of 2018 4-Year	Between Groups	4419.268	21	210.441	2.135	.003				
Graduation Rate	Within Groups	38925.641	395	98.546						
	Total	43344.909	416							
Class of 2017 4-Year	Between Groups	3839.886	21	182.852	2.388	.001				
Graduation Rate	Within Groups	29630.788	387	76.565						
	Total	33470.675	408							
Class of 2016 4-Year	Between Groups	4033.838	21	192.088	2.298	.001				
Graduation Rate	Within Groups	31589.834	378	83.571						
	Total	35623.672	399							
Class of 2015 4-Year	Between Groups	3437.400	21	163.686	2.023	.005				
Graduation Rate	Within Groups	30103.144	372	80.922						
	Total	33540.545	393							

Based on the analysis performed by SPSS, all four graduation tests were considered statistically significant since the calculated significance values are smaller than 0.05. The reported values vary from 0.01 to 0.03 to 0.005, all of which are minimally small supporting insignificantly small probability for the observed differences occurring due to chance. Therefore, the null hypothesis of no statistical difference in graduation rates by county is rejected and the alternative hypothesis of exitance of statistical difference in graduation rates by county is supported.

Conclusion

The above-stated statistical analysis presents a strong evaluative and predictive tool for measuring the projected outcomes and meeting the benchmarks for the graduation rates.

The central tendency data depicts the normality of distribution, skewness, kurtosis and pictorial representations of the dataset. Some of these characteristics are important for the qualification for the inferential analysis. In some cases of inferential statistics, it is important for the data to be normally distributed, so the descriptive statistics assist with that.

The inferential statistics allows to evaluate data for the relationships and statistically test hypotheses. The conclusion of these hypotheses may have an essential role in decision making. For example, the scenario, where ANOVA was able to identify the statistically significant differences between the graduation rates by the county, could be used by the county and state legislators for implementing reforms, distributing funds, promoting developments. It can also be used by the educators for incorporating new methodologies or interventions to equalize the learning opportunities for all students in the state.

Bibliography

National Governors Association. (2005). Graduation counts. A report of the National Governors Association task force on state high school graduation data. Retrieved from <u>https://classic.nga.org/files/live/sites/NGA/files/pdf/0507GRAD.PDF</u>

NJ DOE. (June, 2018). High School Graduation Guidance. An introduction to the 4-year adjusted cohort graduation rate calculation in New Jersey. Retrieved from <u>https://www.nj.gov/education/njsmart/download/grad/AnIntroductiontothe4YGAdjusted</u> <u>CohortRate.pdf</u>

Appendix A

Cohort 2015-2018 4-Year Graduation Rates (sample)

💑 CountyCode	💑 CountyName	🚜 DistrictCode	🚜 DistrictName	💑 SchoolCode	🚜 SchoolName	💑 Level	Classof20184 YearGraduatio	Classof2017 🔗 4YearGradua	Classof2016 🔗 4YearGradua <	Classof2015 🔗 4YearGradua
							nRate	tionRate	tionRate	tionRate
01	ATLANTIC	0110	ATLANTIC CITY	010	Atlantic City High School	School	81.11	80.54	76.14	73.71
01	ATLANTIC	0110	ATLANTIC CITY	999	District Total	District	81.11	80.54	76.14	73.71
01	ATLANTIC	0120	ATLANTIC CO VOCATIONAL	010	Atlantic County Institute of Technology	School	99.11	98.74	99.04	98.08
01	ATLANTIC	0120	ATLANTIC CO VOCATIONAL	999	District Total	District	99.11	98.74	99.04	98.08
01	ATLANTIC	0590	BUENA REGIONAL	025	BUENA REGIONAL HIGH SCHOOL	School	83.02	86.98	91.71	90.56
01	ATLANTIC	0590	BUENA REGIONAL	999	District Total	District	83.02	86.98	91.71	90.56
01	ATLANTIC	1310	EGG HARBOR TWP	005	EGG HARBOR TOWNSHIP HIGH SCHOOL	School	95.51	94.18	92.27	93.85
01	ATLANTIC	1310	EGG HARBOR TWP	999	District Total	District	95.51	94.18	92.27	93.85
01	ATLANTIC	1790	GREATER EGG HARBOR REG	040	Absegami High School	School	91.30	91.83	91.75	90.25
01	ATLANTIC	1790	GREATER EGG HARBOR REG	050	Oakcrest High School	School	91.76	87.94	91.35	91.98
01	ATLANTIC	1790	GREATER EGG HARBOR REG	060	Cedar Creek High School	School	94.38	89.87	94.32	92.02
01	ATLANTIC	1790	GREATER EGG HARBOR REG	999	District Total	District	92.34	90.11	92.27	91.20
01	ATLANTIC	1960	HAMMONTON TOWN	050	Hammonton High School	School	94.06	86.83	91.58	91.17
01	ATLANTIC	1960	HAMMONTON TOWN	999	District Total	District	94.06	86.83	91.58	91.17
01	ATLANTIC	2910	MAINLAND REGIONAL	050	MAINLAND REGIONAL HIGH SCHOOL	School	95.45	95.88	93.66	95.93
01	ATLANTIC	2910	MAINLAND REGIONAL	999	District Total	District	95.45	95.88	93.66	95.93
01	ATLANTIC	4180	PLEASANTVILLE CITY	050	PLEASANTVILLE HIGH SCHOOL	School	88.08	86.55	80.85	81.36
01	ATLANTIC	4180	PLEASANTVILLE CITY	999	District Total	District	88.08	86.55	80.85	81.36
03	BERGEN	0290	BERGEN COUNTY VOCATIONAL	020	Bergen County Academies	School	100.00	99.22	99.60	99.46
03	BERGEN	0290	BERGEN COUNTY VOCATIONAL	070	Bergen County Technical High School - Teterb	School	98.79	100.00	99.35	100.00
03	BERGEN	0290	BERGEN COUNTY VOCATIONAL	999	District Total	District	99.52	99.52	99.50	99.69
03	BERGEN	0300	BERGENFIELD BORO	020	Bergenfield High School	School	97.35	97.32	96.49	97.64
03	BERGEN	0300	BERGENFIELD BORO	999	District Total	District	97.35	97.32	96.49	97.64
03	BERGEN	0440	BOGOTA BORO	020	Bogota Jr./Sr. High School	School	93.98	92.71	92.71	88.00
03	BERGEN	0440	BOGOTA BORO	999	District Total	District	93.98	92.71	92.71	88.00
03	BERGEN	0745	CARLSTADT-EAST RUTHERFORD	050	HENRY P. BECTON REGIONAL HIGH SCHOOL	School	90.97	96.92	97.50	92.81

Appendix B

Cohort 2015-2018 4-Year Graduation Rates by School Level (sample)

	🔗 CountyCode 윩 CountyName	Distric tCode	윩 DistrictName	SchoolCode	🖧 SchoolName	👌 Level	Classof20184	Classof2017 4YearGradua	Classof2016 4YearGradua	Classof2015 4YearGradua
۲	1 ATLANTIC	110	ATLANTIC CITY	10	Atlantic City High School	School	81.11	80.54	76.14	73.71
۲	1 ATLANTIC	120	ATLANTIC CO VOCATIONAL	10	Atlantic County Institute of Technology	School	99.11	98.74	99.04	98.08
۲	1 ATLANTIC	590	BUENA REGIONAL	25	BUENA REGIONAL HIGH SCHOOL	School	83.02	86.98	91.71	90.56
۲	1 ATLANTIC	1310	EGG HARBOR TWP	5	EGG HARBOR TOWNSHIP HIGH SCHOOL	School	95.51	94.18	92.27	93.85
۲	1 ATLANTIC	1790	GREATER EGG HARBOR REG	40	Absegami High School	School	91.30	91.83	91.75	90.25
۲	1 ATLANTIC	1790	GREATER EGG HARBOR REG	50	Oakcrest High School	School	91,76	87.94	91.35	91.98
۲	1 ATLANTIC	1790	GREATER EGG HARBOR REG	60	Cedar Creek High School	School	94.38	89.87	94.32	92.02
٦	1 ATLANTIC	1960	HAMMONTON TOWN	50	Hammonton High School	School	94.06	86.83	91.58	91.17
۲	1 ATLANTIC	2910	MAINLAND REGIONAL	50	MAINLAND REGIONAL HIGH SCHOOL	School	95.45	95.88	93.66	95.93
٦	1 ATLANTIC	4180	PLEASANTVILLE CITY	50	PLEASANTVILLE HIGH SCHOOL	School	88.08	86.55	80.85	81.36
٦	3 BERGEN	290	BERGEN COUNTY VOCATIONAL	20	Bergen County Academies	School	100.00	99.22	99.60	99.46
٦	3 BERGEN	290	BERGEN COUNTY VOCATIONAL	70	Bergen County Technical High School - Teterboro	School	98.79	100.00	99.35	100.00
٦	3 BERGEN	300	BERGENFIELD BORO	20	Bergenfield High School	School	97.35	97.32	96.49	97.64
٦	3 BERGEN	440	BOGOTA BORO	20	Bogota Jr./Sr. High School	School	93.98	92.71	92.71	88.00
٦	3 BERGEN	745	CARLSTADT-EAST RUTHERFORD	50	HENRY P. BECTON REGIONAL HIGH SCHOOL	School	90.97	96.92	97.50	92.81
٦	3 BERGEN	890	CLIFFSIDE PARK BORO	30	Cliffside Park High School	School	87.89	82.41	83.61	82.61
٦	3 BERGEN	990	CRESSKILL BORO	40	Cresskill High School	School	98.64	96.80	97.83	96.18
٦	3 BERGEN	1130	DUMONT BORO	40	Dumont High School	School	97.49	97.21	96.62	93.07
٦	3 BERGEN	1345	ELMWOOD PARK	50	Memorial Senior High School	School	89.72	91.71	90.10	89.37
٦	3 BERGEN	1360	EMERSON BORO	50	Emerson Jr Sr High	School	98.97	98.08	96.23	93.48
٦	3 BERGEN	1370	ENGLEWOOD CITY	40	Dwight Morrow High School/Academies@Englewood	School	86.29	87.29	86.53	82.03
Ī	3 BERGEN	1450	FAIR LAWN PUBLIC SCHOOLS	50	FAIR LAWN HIGH SCHOOL	School	97.61	96.62	96.74	94.52
٦	3 BERGEN	1550	FORT LEE BORO	50	Fort Lee High School	School	96.39	97.35	97.22	97.30
	3 BERGEN	1700	GARFIELD CITY	50	GARFIELD HIGH SCHOOL	School	90.43	81.45	90.29	85.71
	3 BERGEN	1760	GLEN ROCK BORO	50	GLEN ROCK HIGH SCHOOL	School	97.55	99.39	99.49	97.95
Ĩ	3 BERGEN	1860	HACKENSACK CITY	50	Hackensack High School	School	89.94	92.07	90.40	90.45
	1					Contract of the local division of the local]