

## Actual Vs. Predicted Graduation Rates: Technical Appendix

By Lauren Taylor

SPSS was used to conduct a linear regression on cohort graduation rates and predictive covariates. Linear regression is used to model the relationship between a dependent variable (graduation rate) and one or more predictive covariates (district demographics). From this model, researchers can extrapolate the possible outcome for an individual with a similar set of covariates. A linear regression model equation can be expressed as such:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_i X_i + \varepsilon$$

The current analysis constructed the above model using observed cohort graduation rates and demographic variables for California Public Schools in the 2009-10 school year. Data was obtained from the California Department of Education website<sup>1</sup>. All variables were reported at the school district level for Unified ( $n = 210$ ) and High School ( $n = 63$ ) districts with an enrollment at or above 1000 9<sup>th</sup> through 12<sup>th</sup> grade students. This sample represented 94% of all 9<sup>th</sup> through 12<sup>th</sup> grade students within the California Public School System.

Cohort graduation rates are recorded as the percentages of students who first entered 9<sup>th</sup> grade in 2006/2007 and graduated within a four-year period. Cohort graduation rates ranged from 49.84% to 99.36% with a mean rate of 83.68% ( $sd = 9.02$ ).

Demographic variables were taken from the California State API dataset. The Department of Education uses these variables to calculate the California Similar Schools rank. In this rank, schools are compared, on their academic performance, to other schools with similar demographic characteristics.<sup>2</sup> Demographic variables are defined and recorded as followed (see Table 1):

---

<sup>1</sup> <http://www.cde.ca.gov/index.asp>

<sup>2</sup> See the 2010-11 APR Glossary at <http://www.cde.ca.gov/ta/ac/ap/glossary11b.asp>

*Racial Composition:* Eight variables were used for this covariate and were presented as the percentage of students within each racial group for a given district. Racial groups included African American, American Indian, Asian American, Filipino, Hispanic, Pacific Islander, White, and Multiple Race.

*Social Economic Status:* Two variables were used for this covariate. These included percentage of students on free and reduced lunch and average parent education. Average parent education was calculated on a 5-point scale ranging from no high school diploma to some graduate school. The parent with the highest education was recorded for students and averaged across the district.

*English Language Learners:* Two variables were used in this covariate: percentage of student classified as English Learners and percentage of student reclassified as English proficient.

*Mobility:* This covariate was gathered at the school and district level. Mobility is defined as the percentage of students who were enrolled on enrollment day during the previous school year and who continue to be enrolled within the same school or district the following year. Both the school and district percentages were used in for this study.

*Other Covariates:* Three other covariates were used in the analysis. These included percentage of Student in Migrant Education Program, percentage of Students in Gifted and Talented Education (GATE) Program, and percentage of Student with Disabilities.

Unstandardized predicted values and residuals were saved to the dataset. The displayed scatter plot was constructed with observed graduation rates and predicted graduation rates. Residuals are highlighted for districts whose observed cohort graduation rates are 10 or more percentage points above or below their predicted graduation rates.

Table 1.

*Percentage Means, Standard Deviations, and Ranges for Demographic Variables.*

Variable	M	SD	Range	
<b>Racial Composition</b>				
<i>African American</i>	5.19	6.17	0.00	38.00
<i>American Indian</i>	0.84	1.65	0.00	14.00
<i>Asian American</i>	8.45	11.68	0.00	66.00
<i>Filipino</i>	2.46	3.39	0.00	30.00
<i>Hispanic</i>	44.49	25.07	5.00	98.00
<i>Pacific Islander</i>	0.48	0.76	0.00	4.00
<i>White</i>	34.52	23.03	0.00	84.00
<i>Multiple Race</i>	2.15	2.54	0.00	23.00
<b>Social Economic Status</b>				
<i>Free and Reduced lunch</i>	46.87	24.19	0.00	100.00
<i>Parent Education</i> (scaled average)	2.91	0.63	2.00	5.00
<b>English Language Learners</b>				
<i>English Learners</i>	18.00	12.00	0.00	58.00
<i>Reclassified English Proficient</i>	13.04	8.04	0.00	47.00
<b>Mobility</b>				
<i>School Mobility</i>	93.33	3.10	75.00	99.00
<i>District Mobility</i>	95.15	2.48	80.00	99.00
<b>Other</b>				
<i>Migrant Education</i>	2.21	4.74	0.00	33.00
<i>GATE</i>	10.05	6.82	0.00	43.00
<i>Disability</i>	10.16	2.14	1.00	16.00

\* These percentages may not sum to 100 due to responses of: other, multiple, declined to state, or non-response.

The following equation was developed using unstandardized coefficients:

$$\begin{aligned}
 Y_{\text{Predicted Graduation}} &= 14.97 - .46X_{AA}^* + .24X_{AI} + .21X_{AS}^* + .35X_F^* + .22X_H^* - .61X_{PI} + .14X_W + .03X_{MR} \\
 &- .04X_{Free} + 5.63X_{Parent}^* - .15X_{EL}^* - .08X_{REP} + .47X_{School}^* - .05X_{District} \\
 &- .03X_{Migrant} - .07X_{GATE} + .46X_{Disability}^*; R^2 = .68
 \end{aligned}$$

\* Unstandardized coefficients significant at 0.05 or greater

After the model equation was constructed using on the observed cohort graduation rates and demographic variables, predictive graduation rates were determined for each school district

within the sample. These predictive rates tell the researcher what a district's graduation rate should be, given their set demographic variables and the model equation. Predicted graduation rates ranged from 58.44% to 102.55% with a mean rate of 83.68% ( $sd = 7.41$ ). Residual values were then calculated to examine the difference between a district's observed graduation rate and predicted graduation rate. Residual values ranged from -18.24 to 18.74.

Table 2.

*Cohort and Predicted Graduation Rates, Residuals, and Enrollment for 273 California School Districts.*

Parameter	M	SD	Range	
Cohort Graduation Rates	83.68	9.02	49.84	99.36
Predicted Graduation Rates	83.68	7.41	58.44	102.55
Residuals	0.00	5.14	-18.24	18.74
Enrollment	6785.47	13614.63	1009.00	208384.00

Similar analysis has been done in the past, when comparing districts throughout the US and School within California. In an analysis done by Education Week, the 150 largest school districts throughout the United States were compared in a similar fashion. Expected graduations rates and residual differences were calculated to determine the top performing districts in the US. The demographics used in Education Week's analysis were slightly different from those used in this analysis.

More recently, California Comprehensive Center (CCC) compared California schools using linear regression. In their analysis, the same sets of demographics, as used in this analysis, were used to calculate predicted rates and residuals. The difference with CCC study was that they used a four-year averaged graduation rate, verse the cohort graduation rate, as their outcome variable.