DEPARTMENT OF EDUCATION GRADUATION ANALYSIS

University of Arizona - BNAN 277
May 05, 2024

GROUP 307:

Lyhuong Peou, Sameer Shah, Claire Shambro, Cami Stribling, Ava Weihs

Table of Contents

Introduction

We have prepared a thorough report of graduation rates from approximately 1500 colleges and universities across the United States using data provided by the United States Department of Education. The purpose of our study is to identify significant factors influencing the likelihood of students graduating with a bachelor's degree within four years. Throughout this report, we focus on variables such as institution type, urbanization, tuition fees, and financial aid, we aim to uncover trends and patterns that can inform policy decisions and educational strategies. This report synthesizes our findings and offers recommendations to enhance graduation outcomes, supporting the USDE's efforts to improve educational achievements nationwide.

Data Exploration

Visualizations

Description: The visualization will include data that we used for the equation and there are different types of graphs including, line, map, bubble, and bar. There will be a recommendation that was made by the interpretation of the visual.

- click here – for the Tableau Dashboard

Data Overview and Statistical Analysis

- click here – to look at the glossary for any words that have the « * » before then

We started with raw data from 1500 universities and colleges across the United States. There was General Information, Admissions and Enrollment, Degrees and Graduation, and Financial Aid information for each school. For categorical variables like institution type, type of urbanization, and highest degree offered, *dummy variables were created. All of the data incorporated for each school was looked through, and specific variables were selected to help explain the graduation rate for a student getting a bachelor's degree in 4 years. The data was moved over to a new sheet and for the schools that had blanks in their data, they were filtered out. Hundreds of schools were filtered out due to missing information for each explanatory variable used. Once the data was modified, 6 regressions were ran to come up with our final regression output using 8 explanatory variables to determine the predictor variable of graduation rate (bachelor's degree in 4 years).

Data Variables Used

- Institution type using a *dummy variable for Public
- Total Applicants
- Type of Urbanization that narrowed down to a *dummy variable for City (with a base case of Suburb)
- Tuition and fees for the year 2013-2014
- Total price for in-state students living on campus from 2013-2014
- Total price for out-of-state students living on campus from 2013-2014
- Highest degree offered using *dummy variables for Doctor's degree and Master's degree
- Percent of freshman receiving any financial aid

Overall, 6 regressions were run to end up with these explanatory variables for the final regression. The first regression used the variables Public$_{(d)}$, Applicants total, City$_{(d)}$, Town$_{(d)}$, Rural$_{(d)}$, Estimated undergraduate enrollment total, Tuition and fees for 2013-2014, Total price for in-state students living on campus 2013-2014, Total price for out-of-state students living on campus 2013-2014, Total enrollment, Doctor's degree$_{(d)}$, Master's degree$_{(d)}$, and Percent of undergraduate enrollment that are White. It was found that Rural$_{(d)}$ and Total enrollment were not statistically significant at the significance level of 10%. As we started the second regression, we took out the non-significant variables and Estimated undergraduate enrollment total because it was close to being over the significance level. Variables about Percent of total enrollment that are two or more races and percent of total enrollment that are women were added, but they did not end up being statistically significant. As we started the third regression, we wanted to see if the percent of various races enrolled in the college/university, held a significant impact on graduation rate (bachelor's degree in 4 years). The third regression concluded that some variables according to race were shown to be significant, while others were shown not to be. Therefore, as we started the fourth regression, we decided to take out all variables that related to race. Additionally, the variable for Rural$_{(d)}$'s p-value was close to the significance level, so we decided to take this variable out as well. As we began the fourth regression, we included one variable for Admissions total, but its p-value was relatively close to the significance level, so it wasn't included. The fifth regression didn't have any new variables added to it, just the Admissions total variable taken out. Finally, the sixth and final regression, we decided to include the variable of Percent of freshman receiving any financial aid. This was the processes that were gone through to bring us to our final variables and regression output.

Population Regression Equation

$$\begin{aligned} Graduation\ Rate\ 4\ year \\ &= \beta_0 + \beta_1 * \left(Public_{(d)}\right) + \beta_2 * (Applicants\ total) + \beta_3 * \left(City_{(d)}\right) + \beta_4 \\ &* (Tuiton\ and\ fees, 2013-14) + \beta_5 \\ &* (Total\ price\ for\ in-state\ students\ living\ on\ campus\ 2013-14) + \beta_6 \\ &* (Total\ price\ for\ out-of-state\ students\ living\ on\ campus\ 2013-14) + \beta_7 \\ &* (Doctor's\ Degree_{(d)} + \beta_8 * \left(Master's\ Degree_{(d)}\right) + B_9 \\ &* (Percent\ of\ freshman\ receiving\ any\ financial\ aid) + \varepsilon \end{aligned}$$

Sample Regression Equation

$$\begin{aligned} \widehat{Graduation\ Rate}\ 4\ years &= 36.235 + 2.637 * \left(Public_{(d)}\right) + 0.0003 * (Applicants\ Total) - 2.813 \\ &* \left(City_{(d)}\right) + 0.002 * (Tuition\ and\ fees, 2013-14) - 0.002 \\ &* (Total\ price\ for\ in-state\ students\ living\ on\ campus\ 2013-14) + 0.001 \\ &* (Total\ price\ for\ out-of-state\ students\ living\ on\ campus\ 2013-14) \\ &- 4.075 * \left(Doctor's\ Degree_{(d)}\right) - 2.666 * \left(Master's\ Degree_{(d)}\right) - 0.284 \\ &* (Percent\ of\ freshman\ recieving\ any\ financial\ aid) \end{aligned}$$

Fit of the Model

The $R^2 = 0.6673$ tells us we are 66.73% of the way toward perfectly predicting Graduation rate (bachelor's degree in 4 years) using this model.

The S_e = 12.343 tells us our predictions of graduation rate (bachelor's degree in 4 years) are off by an average of 12.34 percentage points.

- See figure 1 for a full review of the regression output – click here

Test of Joint Significance

H_0: None of the independent variables significantly impacts graduation rate (bachelor's degree in 4 years).
H_A: At least one of the independent variables significantly impacts Graduation rate (bachelor's degree in 4 years).

Because the p-value (Significance F) = 0 and is less than the significance level of 0.10, we can reject the null hypothesis and can conclude that at least one of the independent variables significantly impacts graduation rate (bachelor's degree in 4 years).

Test of Individual Variable Significance

Institution Type Individual Significance Test
H_0: Institution Type does not significantly impact graduation rate (bachelor's degree in 4 years).
H_A: Institution Type significantly impacts graduation rate (bachelor's degree in 4 years).

Since the p-value = 0.176, is greater than the significance level of 0.10, we can fail to reject the null and can conclude that Institution Type does not significantly impact graduation rate (bachelor's degree in 4 years).

Applicants Total Individual Significance Test
H_0: Applicants total does not significantly impact graduation rate (bachelor's degree in 4 years).
H_A: Applicants total significantly impacts graduation rate (bachelor's degree in 4 years).

Since the p-value = 1.43E-07, is less than the significance level of 0.10, we can reject the null and can conclude that applicants total significantly impacts graduation rate (bachelor's degree in 4 years).

The Other Variables' Significance

- Type of urbanization ($City_{(d)}$) has the p-value of 0.0001 which is less than significance level of 0.10, so this variable is significant.
- Tuition and fees, 2013-14 has p-value of 1.58E-37 which is less than significance level of 0.10, so this variable is significant.
- Total price for in-state students living on campus 2013-14 has a p-value of 5.51E-17, which is less than significance level of 0.10, so this variable is significant.
- Total price for out-of-state students living on campus 2013-14 has a p-value of 2.11E-15, which is less than significance level of 0.10, so this variable is significant.
- Highest degree offered (Doctor's $degree_{(d)}$ and Master's $degree_{(d)}$) have the p-values of 0.0002 and 0.011 accordingly. These p-values are less than the significance level of 0.10, which makes these dummy variables for highest degree offered, significant.

- Percent of freshman receiving any financial aid has a p-value of 4.44E-14, which is less than significance level of 0.10, so this variable is significant.

Interpretation of Each Significant Variable

- A public institution type has a Graduation rate (bachelor's degree in 4 years) 2.637 percentage points higher than a private institution type, on average and all else constant.
- As the total number of applicants per college/university increases by 1 applicant, Graduation rate increases by 0.0003 percentage points, on average and all else constant.
- An institution in the city has a Graduation rate 2.81 percentage points lower than an institution in the suburbs, on average and all else constant.
- As tuition and fees in 2013-14 increases by $100, Graduation rate increases by 0.24 percentage points, on average and all else constant.
- As total price for in-state students living on campus in 2013-14 increases by $100, Graduation rate decreases by 0.18 percentage points, on average and all else constant.
- As total price for out-of-state students living on campus in 2013-14 increases by $100, Graduation rate increases by 0.11 percentage points, on average and all else constant.
- An institution offering a doctor's degree as their highest degree has a Graduation rate 4.07 percentage points lower than an institution offering a bachelor's degree as their highest degree, on average and all else constant.
- An institution offering a master's degree as their highest degree has a Graduation rate 2.67 percentage points lower than an institution offering a bachelor's degree as their highest degree, on average and all else constant.
- As the percent of freshmen receiving any financial aid increases by 1 percentage point, Graduation rate decreases by 0.28 percentage points, on average and all else constant.

Residual Plot Interpretations

Applicants total residual plot- This residual plot appears to have a funnel shape which indicates issues with heteroskedasticity. This could cause the standard error for Applicant total to be biased, which would mean the hypothesis testing for Applicants total is also biased.

- See figure 3 for a full review of residual plots – click here

Tuition and fees, 2013-2014 residual plot- This residual plot appears to have a random shape which indicates no issues with heteroskedasticity, which means so far, we can trust the hypothesis associated with Tuition and fees, 2013-14.

- See figure 4 for a full review of residual plots – click here

Total price for in-state students living on campus 2013-2014 residual plot- This residual plot appears to have a random shape which indicates no issues with heteroskedasticity, which means so far, we can trust the hypothesis associated with Total price for in-state students living on campus 2013-14.

- See figure 5 for a full review of residual plots – click here

Total price for out-of-state students living on campus 2013-14 residual plot- This residual plot appears to have a random shape which indicates no issues with heteroskedasticity, which means so far, we can trust the hypothesis associated with Total price for out-of-state students living on campus 2012-14.

- See figure 6 for a full review of residual plots – click here

Percent of freshman receiving any financial aid residual plot- This residual plot appears to have a funnel shape which indicates issues with heteroskedasticity. This could cause the standard error for Percent of freshman receiving any financial aid to be biased, which would mean the hypothesis testing for Percent of freshman receiving any financial aid is also biased.

- See figure 7 for a full review of residual plots – click here

Check for Multicollinearity

Correlations from the matrix that show evidence of Multicollinearity:

- Tuition and fees 2013-14 and $Public_{(d)}$: -0.81
- Total price for in-state students living on campus 2013-14 and Tuition and fees 2013-14: 0.98
- Total price for out-of-state students living on campus 2013-14 and Tuition and fees 2013-14: 0.86
- Total price for in-state students living on campus 2013-14 and Total price for out-of-state students: 0.90
- See figure 2 for a full review of correlation matrix – click here

Since these correlations are larger than 0.8 or less than -0.8, we do have evidence of multicollinearity. This could be the reason why Institution Type ($Public_{(d)}$) is not showing up as significant. To look into the true impact of each variable, they should be looked at through two separate regressions.

Check for Endogeneity/Omitted Variables

Since the correlations between the residuals and explanatory variables are all very close to zero, we don't have evidence of endogeneity. This means that our coefficients are unbiased.

- See figure 2 for a full review of correlation matrix – click here

Relevance of Findings

It's important for the United States Department of Education to know that there is an effect on Graduation rate (bachelor's degree in 4 years) when it comes to the costs associated and living for students of institutions. There is a negative effect on Graduation rate for every increase in the Total price for an in-state student living on campus in 2013-2014. The data from 2013-2014 was used for all financial variables because they were the most recent years in the data set. It was surprising to see that the regression showed an increase in Graduation rate for an increase in Tuition and fees, 2013-14, Total price for out-of-state students living on campus 2013-14, and a decrease in Percent of freshman any financial aid. Since there is a negative effect for the in-state student variable, there should be the same effect on the out-of-state student variable. Additionally receiving more financial aid allows for the student to be incentivized to go to the institution. The variables relating to the other *dummy variables for Type of urbanization ($Town_{(d)}$ and $Rural_{(d)}$) ended up not being significant as we continued running the

regressions. This was the reason why they were not included in the final regression with $City_{(d)}$. Additionally, variables relating to race and gender ended up not being significant either when it came to affecting Graduation rate. The final model includes the variables that we thought were accurate determinates of the rate that a student graduates with a bachelor's degree in 4 years

Recommendation

We recommend that the United States Department of Education increase financial aid programs for students in urban colleges and universities, where our study shows graduation rates are often lower. This financial support should focus on providing more scholarships and grants that can help cover tuition and living expenses, making college more accessible and affordable. Enhanced aid can reduce the financial pressures on students and support their journey to successfully complete their degrees. By concentrating resources in this area, the USDE can make a significant impact on improving graduation outcomes in urban educational settings.

Conclusion

In our report of graduation rates from about 1500 colleges and universities across the United States, using data provided by the United States Department of Education, we found several factors influencing the prospect of students obtaining a bachelor's degree within four years. Our study highlighted the significant impact of institution type, urbanization, tuition fees, and financial aid on graduation outcomes. Specifically, our data analysis showed that urban colleges and an increase in total price for in-state students living on campus often correlate with lower graduation rates. Public institutions and those with a higher number of total applicants also tend to have higher graduation rates. We were surprised by the decrease in Graduation rate with an increase in the percent of freshman receiving any financial aid. Even so, based on these findings, we still recommend the enhancement financial aid programs to better support students. Therefore, this relates particularly in urban settings, in order to tackle the financial challenges that may hinder their educational progress. Overall, this strategic focus aims to directly address the financial variables that most significantly impact student success, thereby improving 4 year graduation rate outcomes nationwide. If you have any inquiries about this analysis or recommendations, please contact us ant one of these emails: sameershah@arizona.edu lyhuongpeou@arizona.edu cshambro@arizona.edu avaweihs@arizona.edu cstribling@arizona.edu

Glossary

Dummy Variable: an indicator variable that allows for categorical variables to be changed into a binary value (0 or 1). For example: if the college or university was private for their institution type, the equation would input a “0” in place of the “$\text{Public}_{(d).}$” So on and so forth for the type of urbanization and highest degrees offered variables

*Return to Data Overview and Statistical Analysis

Appendix

Figure 1 - Regression Based on Graduation Rate (Bachelor's Degree Within 4 Years)

Regression Statistics								
Multiple R	0.816883376							
R Square	0.667298449							
Adjusted R Square	0.664991582							
Standard Error	12.34280242							
Observations	1308							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	9	396613.5406	44068.17118	289.2660555	0			
Residual	1298	197743.5137	152.3447717					
Total	1307	594357.0543						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	36.23495181	4.798808908	7.550821986	8.11837E-14	26.82068066	45.64922296	26.82068066	45.64922296
Public (d)	2.636862655	1.946632924	1.354576214	0.175788349	-1.182028766	6.455754077	-1.182028766	6.455754077
Applicants total	0.000284665	5.38089E-05	5.290290693	1.43204E-07	0.000179103	0.000390227	0.000179103	0.000390227
City (d)	-2.813011954	0.722572081	-3.893053756	0.000104022	-4.230549019	-1.395474889	-4.230549019	-1.395474889
Tuition and fees, 2013-14	0.002456869	0.000185839	13.22039018	1.58356E-37	0.002092291	0.002821448	0.002092291	0.002821448
Total price for in-state students living on campus 2013-14	-0.001862336	0.000219321	-8.491389484	5.50925E-17	-0.002292598	-0.001432075	-0.002292598	-0.001432075
Total price for out-of-state students living on campus 2013-14	0.001087634	0.000135393	8.033173801	2.11501E-15	0.000822021	0.001353246	0.000822021	0.001353246
Doctor's degree (d)	-4.074918612	1.100203024	-3.703787867	0.000221322	-6.233289529	-1.916547694	-6.233289529	-1.916547694
Master's degree (d)	-2.665712057	1.048567284	-2.542242256	0.011129782	-4.722784326	-0.608639788	-4.722784326	-0.608639788
Percent of freshmen receiving any financial aid	-0.284257498	0.037243227	-7.632461604	4.44042E-14	-0.357321011	-0.211193986	-0.357321011	-0.211193986
Peou, Shah, Shambro, Stribling, Weihs								

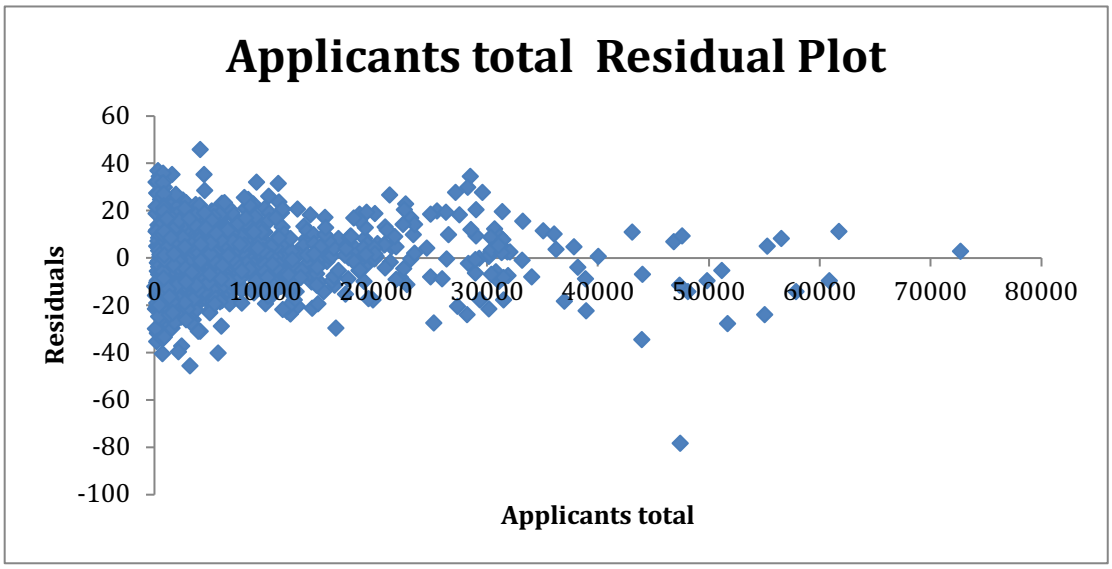
*Return to Fit of Model

Figure 2 - Correlation Matrix

	Residuals	*Public (d)*	*Applicants total*	*City (d)*	*Tuition and fees, 2013-14*	*Total price for in-state students living on campus 2013-14*	*Total price for out-of-state students living on campus 2013-14*	*Doctor's degree (d)*	*Master's degree (d)*
Residuals	1								
Public (d)	-6.70871E-15	1							
Applicants total	-1.11867E-16	0.30240762	1						
City (d)	7.46163E-16	0.02536209	0.224347225	1					
Tuition and fees, 2013-14	8.10659E-15	-0.8098078	-0.049723386	0.02117789	1				
Total price for in-state students living on campus 2013-14	8.38818E-15	-0.7686446	0.021090591	0.05501362	0.984796725	1			
Total price for out-of-state students living on campus 2013-14	7.33511E-15	-0.4769629	0.274973491	0.12216094	0.855513639	0.903295731	1		
Doctor's degree (d)	-2.20925E-17	0.21469666	0.416078582	0.25780498	-0.072705461	-0.026601291	0.124695355	1	
Master's degree (d)	-4.64627E-16	-0.0786964	-0.289108035	-0.1593399	-0.034168784	-0.056409972	-0.154566444	-0.73162596	1
Percent of freshmen receiving any financial aid	-1.21946E-16	-0.306717	-0.526983493	-0.1040867	0.009013314	-0.041838702	-0.249928163	-0.162375521	0.128613775
Peou, Shah, Shambro, Stribling, Weihs									

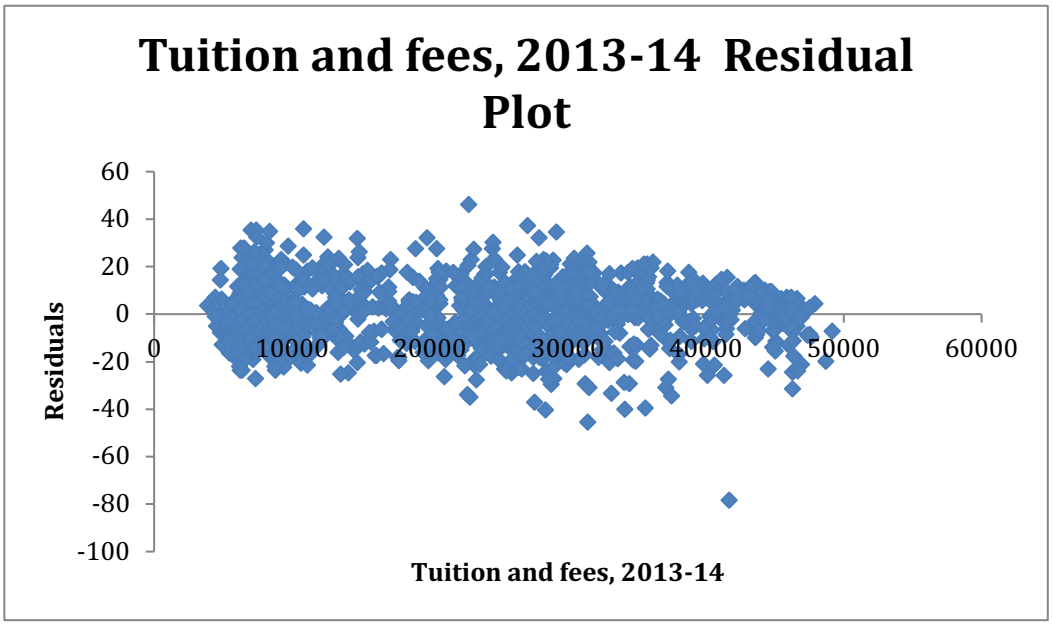
*Return to Multicollinearity and Endogeneity/Omitted Variables

Figure 3 - Applicants Total Residual Plot



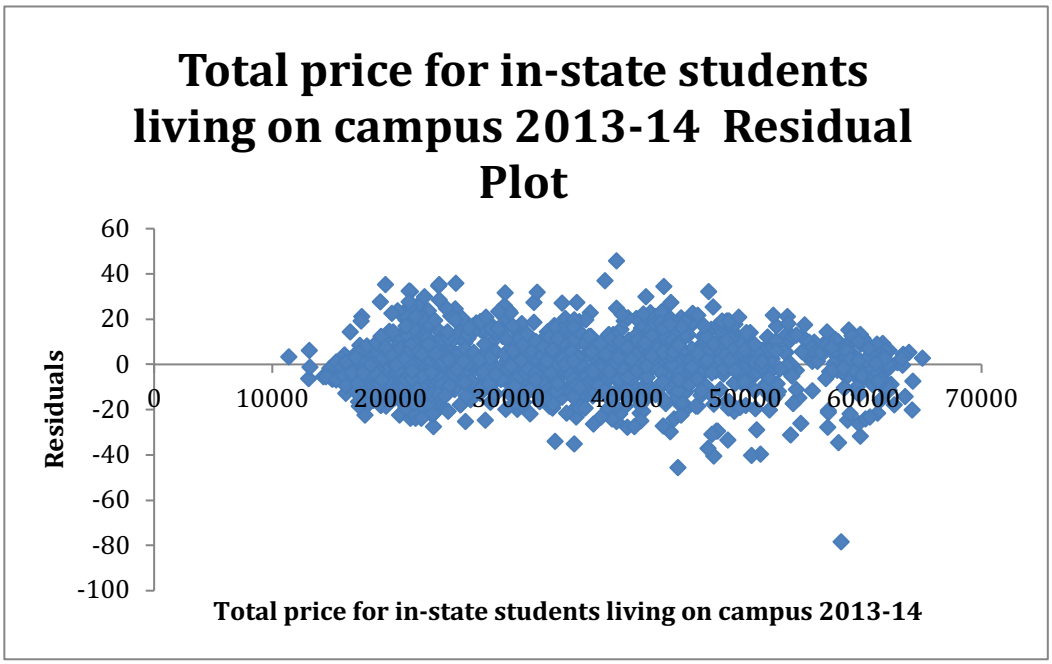
*Return to Interpretation

Figure 4 - Tuition and Fees, 2013-14 Residual Plot



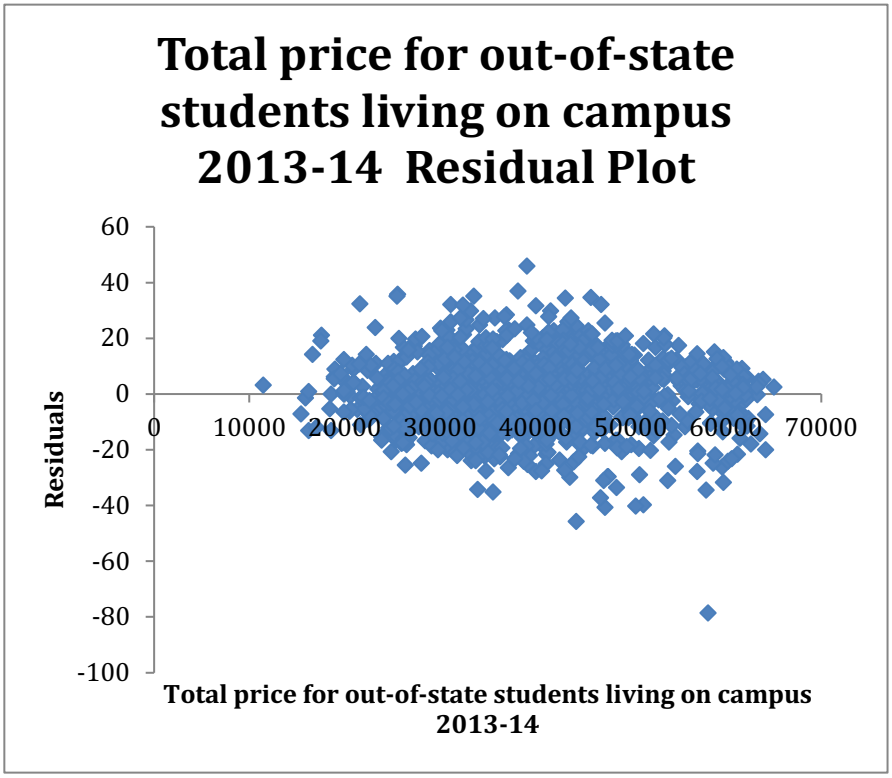
*Return to Interpretation

Figure 5 - Total Price for In-State Students Living on Campus 2013-14 Residual Plot



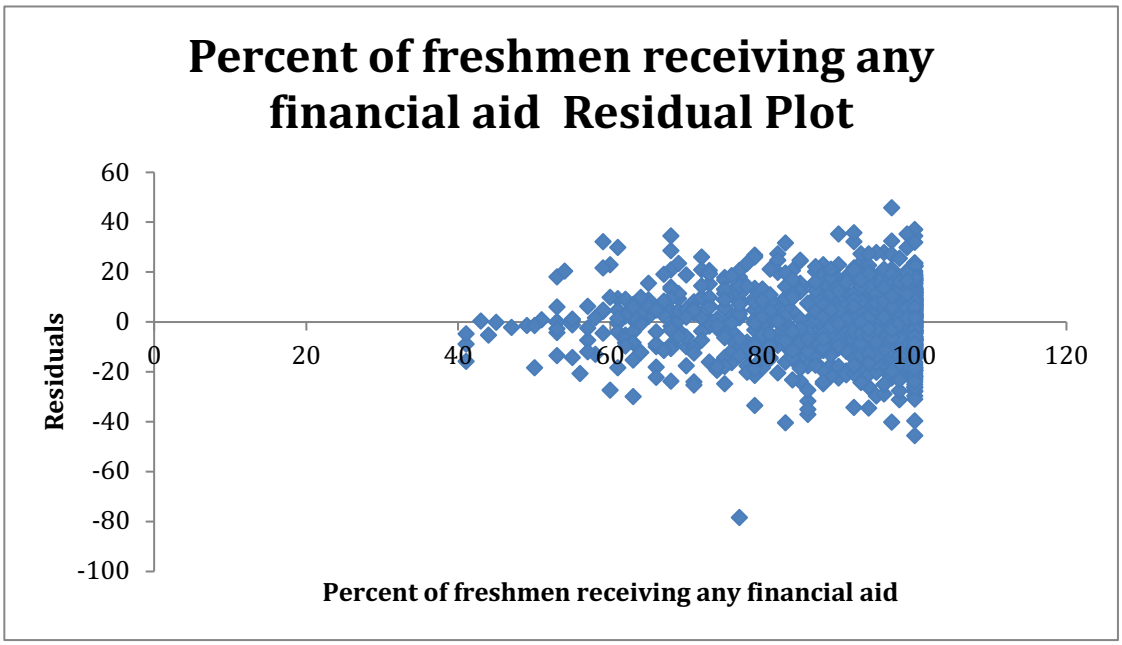
*Return to Interpretation

Figure 6 - Total Price for Out-of-State Students Living on Campus 2013-14 Residual Plot



*Return to Interpretation

Figure 7 - Percent of Freshman Receiving Any Financial Aid Residual Plot



*Return to Interpretation