Layla Burgan, Daniel DeFoe, Sadie Rhen
STAT 324
Professor Walker

STAT 324 Final Project Report

Introduction:

The US News data set that we received hoped to help predict what factors would influence graduation rate of American colleges based on data from the 1993-1994 school year. This data contained problems with violation of some assumptions, a great deal of missing data, and variables that did not appropriately contribute to the regression predictions. However, through the following steps, the data was able to be managed in such a way that it was ultimately able to effectively contribute to the regression predicting the graduation rate from the different colleges.

Actions Taken:

We began our process by creating a region categorical variable by grouping the states. To arrive at out final model, we started by checking the regression assumptions and believed that equal variance was violated(graph 1). We attempted to fix this problem through transforming the dependent variable(models 2 and 3), but this was ultimately unsuccessful. However, since colleges are widely variable in their statistics, this should not be a problem during this investigation. We removed one outlier observation because that college (Cazenovia) had a graduation rate of 118%. Our data had severe multicollinearity, so we transformed variables that were associated with each other (model 4). We then removed variables that were not significant to the regression (model 5)and added two interactions(model 6) to further attempt to reduce some moderate multicollinearity (model 6). The interactions added to the multicollinearity rather than than reducing it, so those were removed(model 7) and one variable was swapped for another, which did lower the multicollinearity to below moderate. We then took out the observations that were present in the reduced model but not in our original initial model(model 8) so we could compare the two.

Final Reduced Model F-test: $F = \frac{\frac{104775.55-104094.86}{670-665}}{156.53} = .8697 \sim F_{5,\,665}$ gives a p-value of .501

At the 5% significance with an F-statistic of 0.8697 and a p-value of 0.501 we do not have sufficient evidence to reject the null hypothesis and be convinced that the slopes of Accepted, Apps, FTUG, RmBrd, Phd, and SFRatio cannot effectively predict the graduation rate of students who graduated in 6 years from all American colleges in the mid-1990's.

Final Reduced Model Prediction Expression:

*Graduate=56.1041 + 4.2029*Funding[Private] + 0.0015*Enrolled + 0.1516*Top10+ 0.0008*Tuition + 0.1846*Alumni-0.0006 + 4.9091*StdScore + 0.3794*Region[Midwest]+ 5.2067*Region[Northeast] - 1.9904*Region[South] - 9.8914*Accepted/Apps*

Slope Interpretations:

Funding[Private]: At the 5% significance level, with a p-value of less than 0.0001, there is sufficient evidence to reject the null hypothesis and be convinced that when funding is private, it is associated with 4.2029 percentage points greater than public funding in average graduation rate of students who graduated in 6 years from all American colleges in the mid-1990's.

Enrolled: At the 5% significance level, with a p-value of 0.0291, there is sufficient evidence to reject the null hypothesis and be convinced that every increase in one student enrolled is associated with a 0.0015 percentage point increase in average graduation rate of students who graduated in 6 years from all American colleges in the mid-1990's. Thus there is a positive association between a college's number of enrolled students and its graduation rate.

Top10: At the 5% significance level, with a p-value of 0.0022, there is sufficient evidence to reject the null hypothesis and be convinced that every increase in one percentage of the college student body who was in the 10 % of their high school class is associated with a 0.1516 percentage point increase in average graduation rate of students who graduated in 6 years from all American colleges in the mid-1990's. Thus there is a positive association between a college's percentage of the college student body that was in the 10 % of their high school class and its graduation rate.

Tuition: At the 5% significance level, with a p-value of 0.0005, there is sufficient evidence to reject the null hypothesis and be convinced that every increase in one dollar in the cost of out-of-state student tuition is associated with a 0.0008 percentage point increase in average graduation rate of students who graduated in 6 years from all American colleges in the mid-1990's. Thus there is a positive association between a college's cost of out-of-state student tuition and its graduation rate.

Alumni: At the 5% significance level, with a p-value of 0.0003, there is sufficient evidence to reject the null hypothesis and be convinced that every increase in one percentage point of college alumni who donate is associated with a 0.1846 percentage point increase in average graduation rate of students who graduated in 6 years from all American colleges in the mid-1990's. Thus there is a positive association between a college's percentage of alumni who donate and its graduation rate.

Spending: At the 5% significance level, with a p-value of less than 0.0001, there is sufficient evidence to reject the null hypothesis and be convinced that every increase in one dollar of instructional expenditure per student is associated with a 0.0006 percentage point decrease in average graduation rate of students who graduated in 6 years from all American colleges in the mid-1990's. Thus there is a negative association between a college's amount of instructional expenditure per student and its graduation rate.

StdScore: At the 5% significance level, with a p-value of less than 0.0001, there is sufficient evidence to reject the null hypothesis and be convinced that every increase in one standard deviation in the average of standardized SAT and ACT scores for students at the school is associated with a 4.9091 percentage point increase in average graduation rate of students who

graduated in 6 years from all American colleges in the mid-1990's. Thus there is a positive association between the average of standardized SAT and ACT scores for students at the school and its graduation rate.

Region[Midwest]: At the 5% significance level, with a p-value of 0.6715, there is not sufficient evidence to reject the null hypothesis and be convinced that the college being in a Midwestern state is any different than colleges in Western states in average graduation rate of students who graduated in 6 years from all American colleges in the mid-1990's. Thus there is not a an association between a college being in the American Midwest and its graduation rate.

Region[Northeast]: At the 5% significance level, with a p-value of less than 0.0001, there is sufficient evidence to reject the null hypothesis and be convinced that a college being in a Northeastern state is associated with 5.2067 percentage points greater than colleges in Western states in average graduation rate of students who graduated in 6 years from all American colleges in the mid-1990's.

Region[South]: At the 5% significance level, with a p-value of 0.0003, there is sufficient evidence to reject the null hypothesis and be convinced that a college being in a Southern state is associated with 1.9904 percentage points greater than colleges in Western states in average graduation rate of students who graduated in 6 years from all American colleges in the mid-1990's.

Accepted/Apps: Spending: At the 5% significance level, with a p-value of 0.0168, there is sufficient evidence to reject the null hypothesis and be convinced that every increase in one unit in the ratio of accepted applicants divided by total student applicants is associated with a 9.8914 percentage point decrease in average graduation rate of students who graduated in 6 years from all American colleges in the mid-1990's. Thus there is a negative association between a college's ratio of accepted applicants divided by total student applicants and its graduation rate.

Effect tests for categorical:

Funding: At the 5% significance level, with a p-value less than 0.0001, we have sufficient evidence to reject the null hypothesis and we are convinced that there is an association between funding and graduation rate of students who graduated in 6 years from all American colleges in the mid-1990's.

Region: At the 5% significance level, with a p-value less than 0.0001, we have sufficient evidence to reject the null hypothesis and we are convinced that there is an association between region and graduation rate of students who graduated in 6 years from all American colleges in the mid-1990's.

Appendix

- Initial Model 1:

Parameter Estimates

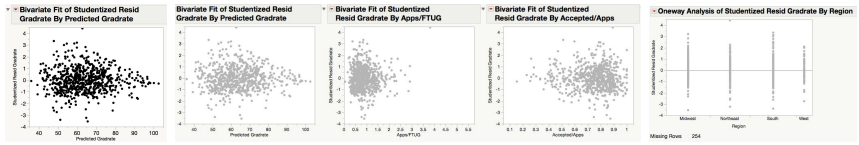
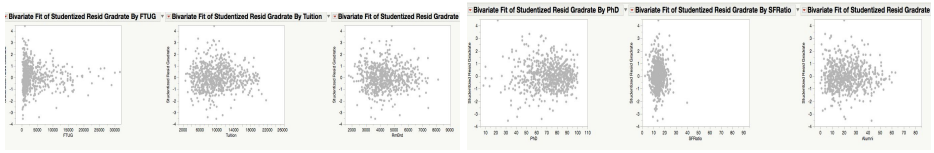
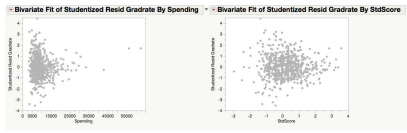
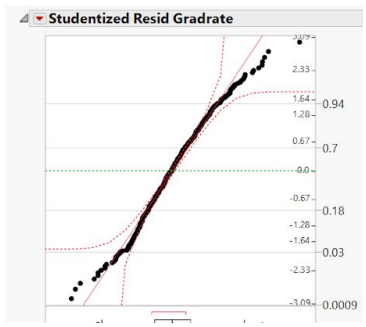
Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	45.367936	4.83406	9.39	<.0001*	.
Funding[Private]	3.6388619	0.929035	3.92	<.0001*	3.0140942
Apps	0.0004996	0.000487	1.03	0.3052	16.143788
Accepted	0.0002249	0.000899	0.25	0.8025	22.008626
Enrolled	0.0015771	0.002404	0.66	0.5119	20.952338
Top10	0.1496116	0.051687	2.89	0.0039*	3.7284107
FTUG	-0.000532	0.000421	-1.26	0.2066	17.28337
Tuition	0.0005936	0.000264	2.25	0.0247*	4.4355054
RmBrd	0.0012699	0.000673	1.89	0.0596	2.3217526
PhD	-0.006274	0.041342	-0.15	0.8794	1.9767053
SFRatio	0.0293308	0.1614	0.18	0.8559	1.7678126
Alumni	0.203735	0.052619	3.87	0.0001*	1.8014036
Spending	-0.000597	0.000156	-3.84	0.0001*	2.6492887
StdScore	5.0382727	1.007401	5.00	<.0001*	3.7902688
Region[Midwest]	0.557828	0.960635	0.58	0.5616	1.4502984
Region[Northeast]	4.7884909	1.015155	4.72	<.0001*	1.5419784
Region[South]	-1.725581	0.904054	-1.91	0.0567	1.3176714

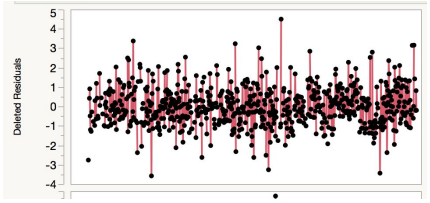
Summary of Fit

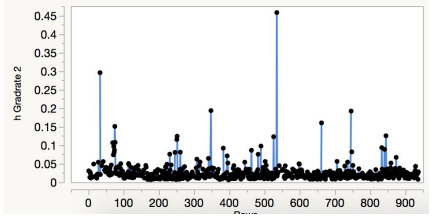
RSquare	0.463308
RSquare Adj	0.450414
Root Mean Square Error	12.69044
Mean of Response	64.00732
Observations (or Sum Wgts)	683

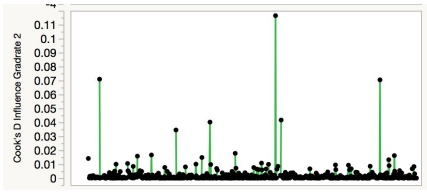
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	16	92591.53	5786.97	35.9334
Error	666	107257.43	161.05	Prob > F
C. Total	682	199848.96		<.0001*

-
    - Graphs for Assumptions
        - 
        - 
        - 
        Residual Plots (1)
        - 
        Normal Probability Plot (2)
        - Linearity is not violated since there are not curves in the residual plots (1)
        - Normality is not violated since the points stay fairly close to the diagonal (2)
        - Indepence may or may not be violated as it is unclear whether or not this was a random sample of American colleges, and some of the colleges come from the same university system (such as Cal State or SUNY).
        - Equal variance looked somewhat concerning as there was a slight fan shape to the residual plots (1)
    - Multicollinearity

        - There was high multicollinearity between the number of applications a college received (VIF = 16.144), the amount of student they accepted (VIF = 22.009), the number of students that actually enrolled (VIF = 20.952), and the amount of full time undergraduates at the school (VIF = 17.283).
        - This makes sense generally, the greater the number of undergraduates at a college, the more spots they have to take new students, so the number of apps, accepted, and enrolled students will also be higher.
    - Influential Observations
        - Outliers: Critical value = (1- .05/2(683)) = .999963, t_{665} = 3.9881
            - 

(3)
            - Row 551 (Cazenovia College) was an outlier with a Grad Rate of 118%. We determined that this observation needed to be taken out of the graph because that was clear a typo as a graduation rate of 118% is not possible
        - High Leverage: Cutoff = 3p/n = 3(17)/683 = 0.0747
            - 

(4)
            - There are many high leverage points, but this is just an influence of possible high leverage.
        - Influence: 50th percentile of $F_{17, 666}$ = .9620
            - 

(5)
            - Row 535 (Rutgers at New Brunswick) is the only influential point. This is because it has extremely high number of applications and accepted students, however, we decided to keep this point in the regression as out fix for multicollinearity includes dividing the number of apps by the number of accepted students, thus minimizing the effect of this influential point
- New Initial Model B without the outlier

- Model 2:
  - Take the Square Root of Grad Rate to try and fix equal variance
  - **Parameter Estimates**

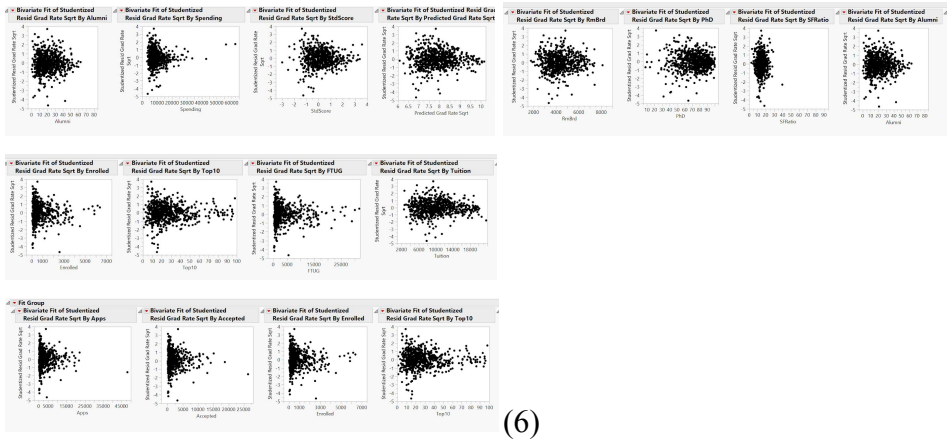
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	6.6858132	0.315456	21.19	<.0001*
Funding[Private]	0.2450802	0.060551	4.05	<.0001*
Apps	2.2326e-5	3.178e-5	0.70	0.4826
Accepted	3.5857e-5	5.882e-5	0.61	0.5423
Enrolled	-4.038e-5	0.000157	-0.26	0.7967
Top10	0.0086249	0.003371	2.56	0.0107*
FTUG	-1.025e-5	2.744e-5	-0.37	0.7088
Tuition	3.7652e-5	1.719e-5	2.19	0.0288*
RmBrd	7.7972e-5	4.387e-5	1.78	0.0759
PhD	0.0012518	0.002714	0.46	0.6448
SFRatio	0.0017388	0.010519	0.17	0.8688
Alumni	0.0124624	0.003431	3.63	0.0003*
Spending	-0.000041	1.014e-5	-4.04	<.0001*
StdScore	0.349625	0.065681	5.32	<.0001*
Region[Midwest]	0.0650697	0.062661	1.04	0.2994
Region[Northeast]	0.2911399	0.066264	4.39	<.0001*
Region[South]	-0.119116	0.058926	-2.02	0.0436*

  - **Summary of Fit**

RSquare	0.47141
RSquare Adj	0.458693
Root Mean Square Error	12.51134
Mean of Response	63.92815
Observations (or Sum Wgts)	682

  - **Analysis of Variance**

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	16	92834.62	5802.16	37.0666
Error	665	104094.86	156.53	Prob > F
C. Total	681	196929.48		<.0001*

  - Graphs for Assumptions:
    - 
    - (6)
    - This clearly did not make our equal variance requirement better.
- Model 3:
  - Squaring Grad Rate to try and fix equal variance

- ○ **Parameter Estimates**

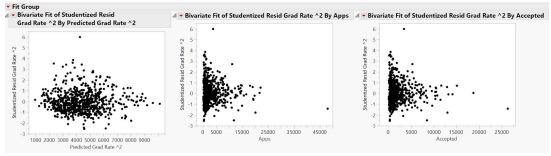
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	1752.4188	610.7088	2.87	0.0042*
Funding[Private]	439.60784	117.2247	3.75	0.0002*
Apps	0.1433011	0.061524	2.33	0.0201*
Accepted	-0.170162	0.113866	-1.49	0.1355
Enrolled	0.6098413	0.303258	2.01	0.0447*
Top10	19.989701	6.525232	3.06	0.0023*
FTUG	-0.114303	0.053117	-2.15	0.0318*
Tuition	0.0800229	0.03327	2.41	0.0164*
RmBrd	0.1715363	0.084921	2.02	0.0438*
PhD	1.1669703	5.2545	0.22	0.8243
SFRatio	4.5611454	20.36422	0.22	0.8228
Alumni	25.266049	6.643026	3.80	0.0002*
Spending	-0.072888	0.019635	-3.71	0.0002*
StdScore	611.3806	127.156	4.81	<.0001*
Region[Midwest]	49.07319	121.3093	0.40	0.6860
Region[Northeast]	585.68144	128.2845	4.57	<.0001*
Region[South]	-169.9147	114.0786	-1.49	0.1368

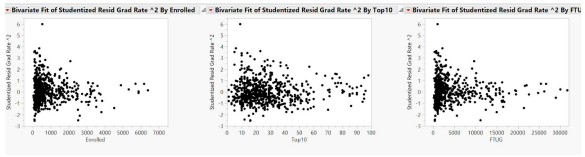
Summary of Fit

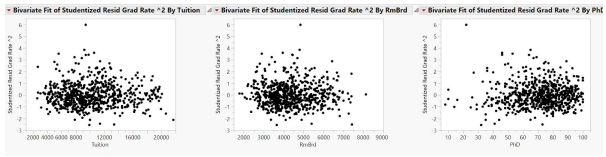
RSquare	0.478601
RSquare Adj	0.466056
Root Mean Square Error	1601.161
Mean of Response	4375.562
Observations (or Sum Wgts)	682

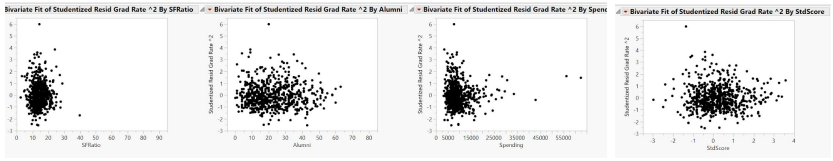
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	16	1564931886	97808243	38.1510
Error	665	1704870728	2563715.4	Prob > F
C. Total	681	3269802614		<.0001*

- ○ Graphs for Assumptions
  - ■ 







(7)

    - ■ As one can see, this transformation also didn't help the unequal variance. Thus we decided to use the original variable of Grad Rate for the rest of or exploration as we do not have any other methods that we could attempt to try and fix the violation of the equal variance assumptions.
- Model 4:
  - ○ Changes to try and fix Multicollinearity:
    - ■ Replace Enrolled with Enrolled/Accepted
    - ■ Replace Accepted with Accepted/Apps
    - ■ Replace Apps with Apps/FTUG

- ○ **Parameter Estimates**

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	56.905987	7.852269	7.25	<.0001*	.
Funding[Private]	4.0039095	0.924474	4.33	<.0001*	3.0664764
Top10	0.158484	0.050366	3.15	0.0017*	3.6344553
FTUG	0.0001795	0.000144	1.24	0.2137	2.0857253
Tuition	0.0006355	0.000264	2.41	0.0162*	4.5615784
RmBrd	0.0011027	0.000674	1.64	0.1024	2.3935388
PhD	0.0129507	0.04099	0.32	0.7521	1.9710436
SFRatio	0.0161202	0.160268	0.10	0.9199	1.792005
Alumni	0.1959956	0.052209	3.75	0.0002*	1.8231859
Spending	-0.000615	0.000155	-3.97	<.0001*	2.7025993
StdScore	4.8206641	1.003886	4.80	<.0001*	3.8539183
Region[Midwest]	1.213357	0.954393	1.27	0.2041	1.4714549
Region[Northeast]	4.5100161	1.013334	4.45	<.0001*	1.5750043
Region[South]	-1.691551	0.885589	-1.91	0.0566	1.2996468
Apps/FTUG	-1.417945	1.978957	-0.72	0.4739	3.2259001
Accepted/Apps	-10.88102	4.775682	-2.28	0.0230*	2.0995183
Enrolled/Accepted	-6.449426	5.113316	-1.26	0.2076	2.0787358

Summary of Fit

RSquare	0.478601
RSquare Adj	0.466056
Root Mean Square Error	1601.161
Mean of Response	4375.562
Observations (or Sum Wgts)	682

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	16	1564931886	97808243	38.1510
Error	665	1704870728	2563715.4	Prob > F
C. Total	681	3269802614		<.0001*

  - This fixed our severe multicollinearity issues, and left us with only a few variables showing moderate VIFs.
  - However, now we start the process of variable reduction and notice that several variables are not significant no matter what combination of variables we have
  -
- Model 5:
  - After deliberation and looking at the stepwise regression outputs, we decided to remove the following insignificant variables: enrolled/accepted, apps/FTUG, FTUG, SFRatio, and PhD.
  - **Parameter Estimates**

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	51.857416	4.471735	11.60	<.0001*	.
Funding[Private]	3.1059386	0.722091	4.30	<.0001*	1.9032199
Top10	0.1634316	0.047612	3.43	0.0006*	3.3908485
Tuition	0.0006627	0.000252	2.63	0.0088*	4.264676
RmBrd	0.0014138	0.000646	2.19	0.0291*	2.2616091
Alumni	0.1825453	0.049705	3.67	0.0003*	1.6777462
Spending	-0.000606	0.000141	-4.28	<.0001*	2.3518208
StdScore	5.0590492	0.957891	5.28	<.0001*	3.6125039
Region[Midwest]	1.397837	0.92738	1.51	0.1322	1.4229785
Region[Northeast]	4.3704387	0.977414	4.47	<.0001*	1.5097207
Region[South]	-1.871411	0.864373	-2.17	0.0307*	1.2707002
Accepted/Apps	-8.604371	4.093558	-2.10	0.0359*	1.5837406

Summary of Fit

RSquare	0.468779
RSquare Adj	0.460211
Root Mean Square Error	12.5089
Mean of Response	63.94669
Observations (or Sum Wgts)	694

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	11	94170.68	8560.97	54.7123
Error	682	106714.35	156.47	Prob > F
C. Total	693	200885.03		<.0001*

  - This helped with our variable significance issues, but still left us with moderate VIFs that we wanted to fix
- Model 6:
  - We tried to make interactions between StdScore and Top10 (as good students with high scores on the SAT or ACT are probably more likely to be in the top 10% of their high school class)as well as Funding and Tuition (and schools with

private funding are more likely to have higher tuition)to decrease all of the VIF's, but this did not help, and the interactions were not significant

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	63.365308	4.81306	13.17	<.0001*	.
RmBrd	0.0014324	0.000653	2.19	0.0286*	2.2536607
Alumni	0.1912996	0.050923	3.76	0.0002*	1.7415622
Region[Midwest]	1.1652128	0.93285	1.25	0.2121	1.4115422
Region[NorthEast]	4.5323125	0.995704	4.55	<.0001*	1.5269123
Region[South]	-1.707968	0.874736	-1.95	0.0513	1.273187
Accepted/Apps	-8.352113	4.266845	-1.96	0.0507	1.6828282
Centered Top10 2	0.142924	0.05355	2.67	0.0078*	4.1253186
Centered StdScore	5.2794556	0.990529	5.33	<.0001*	3.7674336
(Centered Top10 2-1.0405)*(Centered StdScore-0.16829)	0.0127035	0.02616	0.49	0.6274	2.372231
Funding[Private]	1.9427962	0.947587	2.05	0.0407*	3.2349369
Centered Tuition 2	0.0009625	0.00029	3.32	0.0009*	5.5277567
Funding[Private]*(Centered Tuition 2-647.697)	-0.000429	0.000228	-1.88	0.0603	2.5483156
Spending	-0.000645	0.000157	-4.11	<.0001*	2.7813079

- The interactions made multicollinearity worse so they were taken out.

- Model 7

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	55.737857	3.919167	14.22	<.0001*	.
Funding[Private]	4.8458625	0.82991	5.84	<.0001*	2.61916
Enrolled	0.0015715	0.000687	2.29	0.0224*	1.7999125
Top10	0.1570598	0.048168	3.26	0.0012*	3.5234792
Tuition	0.0007522	0.000229	3.28	0.0011*	3.7242323
Alumni	0.1698084	0.049296	3.44	0.0006*	1.6895857
Spending	-0.000623	0.00014	-4.46	<.0001*	2.3080435
StdScore	5.4476394	0.948293	5.74	<.0001*	3.6622864
Region[Midwest]	0.4351037	0.880608	0.49	0.6214	1.3066073
Region[Northeast]	5.2895043	0.90155	5.87	<.0001*	1.329984
Region[South]	-1.969889	0.849656	-2.32	0.0207*	1.2387376
Accepted/Apps	-8.889392	4.039044	-2.20	0.0281*	1.581469

Summary of Fit

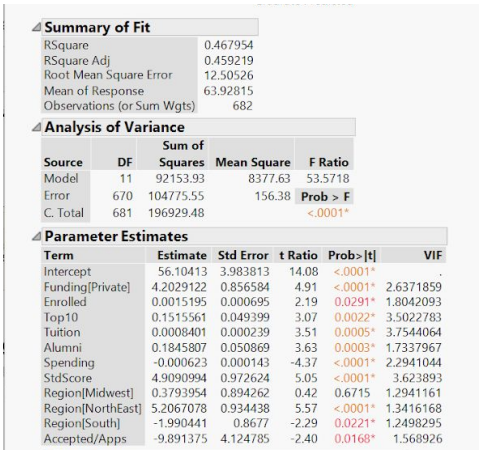
RSquare	0.484854
RSquare Adj	0.47677
Root Mean Square Error	12.5514
Mean of Response	63.51893
Observations (or Sum Wgts)	713

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	11	103940.02	9449.09	59.9799
Error	701	110433.97	157.54	Prob > F
C. Total	712	214373.99		<.0001*

- We went back to model 5, and after looking at other variables, we determined that if Room and Board and Enrolled were both in the regression, they were both ever so slightly insignificant. However, if either one was taken out, the other became significant. Since removing Room and Board fixed our moderate multicollinearity and removing enrolled did not, we determined that it was best to remove Room and Board.

- Model 8 - Final Model
  - Our data set had a lot of missing data, thus when we took variables out of the regression, some observations were added back into the regression. This happened because we had taken out the variables in which they were missing points, so now they had "complete" data and could be run in the regression. However, in order to compare to our initial model, we needed to exclude all the observations that were missing data originally so that we had the same number of observations in both the reduced and full model.

- 

Summary of Fit

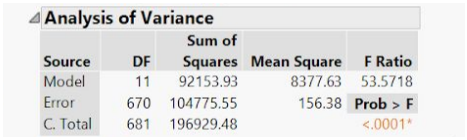
RSquare	0.467954
RSquare Adj	0.459219
Root Mean Square Error	12.50526
Mean of Response	63.92815
Observations (or Sum Wgts)	682

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	11	92153.93	8377.63	53.5718
Error	670	104775.55	156.38	Prob > F
C. Total	681	196929.48		<.0001*

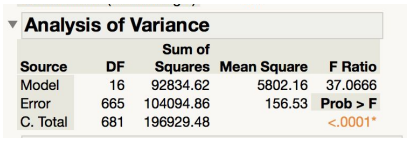
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	56.10413	3.983813	14.08	<.0001*	.
Funding[Private]	4.2029122	0.856584	4.91	<.0001*	2.6371859
Enrolled	0.0015195	0.000695	2.19	0.0291*	1.8042093
Top10	0.1515561	0.049399	3.07	0.0022*	3.5022783
Tuition	0.0008401	0.000239	3.51	0.0005*	3.7544064
Alumni	0.1845807	0.050869	3.63	0.0003*	1.7337967
Spending	-0.000623	0.000143	-4.37	<.0001*	2.2941044
StdScore	4.9090994	0.972624	5.05	<.0001*	3.623893
Region[Midwest]	0.3793954	0.894262	0.42	0.6715	1.2941161
Region[NorthEast]	5.2067078	0.934438	5.57	<.0001*	1.3416168
Region[South]	-1.990441	0.8677	-2.29	0.0221*	1.2498295
Accepted/Apps	-9.891375	4.124785	-2.40	0.0168*	1.568926

- Reduced Model (8) Full Model (9)
- 

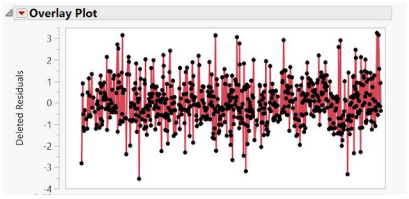
Analysis of Variance

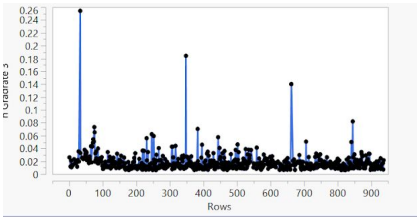
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	11	92153.93	8377.63	53.5718
Error	670	104775.55	156.38	Prob > F
C. Total	681	196929.48		<.0001*

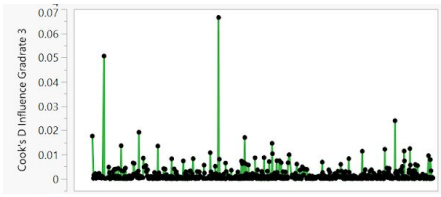


Analysis of Variance

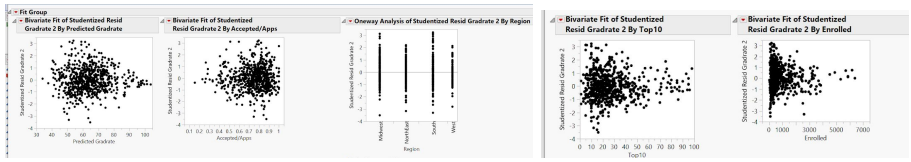
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	16	92834.62	5802.16	37.0666
Error	665	104094.86	156.53	Prob > F
C. Total	681	196929.48		<.0001*

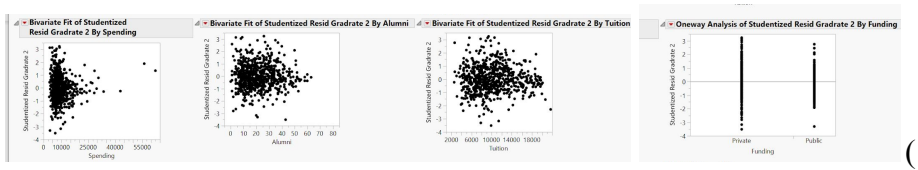
- Influential Observations
  - Outlier test: (1 - a/2n) = (1 - 0.05/2(682)) = .999963 ; t_{665} = 3.9881
    - 

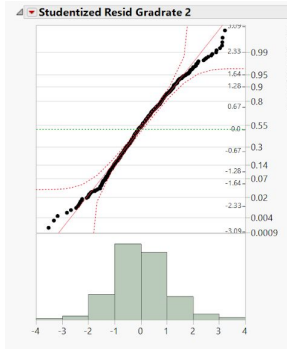
(10)
      - There are no outliers in the data
  - High leverage test 3p/n = 3(12)/682 = 0.528
    - 

(11)
      - There are no high leverage points in the data
  - High influence points: n - p = 682 - 12 = 670
    - 

(12)
    - compare to F(11,670), has to be above .94
    - There are no highly influential points

- Check assumptions
  - 

  - 

(13)
  - 

(14)
  - Normality looks ok, as there is no major deviation away from the line of fit.
  - Linearity is also passing, there is no large curve in the residual plots.
  - Equal variance is okay, there is a slight fan shape, but nothing too major
  - Indepence may or may not be violated as it is unclear whether or not this was a random sample of American colleges, and some of the colleges come from the same university system (such as Cal State or SUNY).