Date: June 21, 2021
To: Admissions Councilors
From: Charlie Evert
RE: **GRE & GPA in Predicting Admissions**

Introduction

People are admitted to Graduate School based upon their perceived ability to excel in Graduate School. These decisions are made using a wide variety of data, however the data set analyzed herein consists of GRE test scores, undergraduate GPA and the prestige of undergraduate programs along with whether applicants were admitted or not. These factors, in the real world, all hold weight in graduate school decision making, but the findings of this data set paint a different picture; since few are actually admitted to graduate school, just GPA or GRE scores are not very predictive by themselves of whether or not someone will be admitted into graduate school.

Decision makers can choose to base admissions decisions on more holistic data or on more objective performance related data; currently, there seems to be a predisposition on favoring more holistic attributes rather than GRE scores and GPA. These alternatives should be evaluated based upon the future success of admitted applicants; if admitted applicants graduate and become more successful in life or donate more to endowments based upon having higher GPA/GRE scores, then it logically follows that admissions counselors should admit more applicants with higher GPAs and GRE scores. The following is why admissions decision makers should objectively evaluate each applicant more based upon GRE and GPA scores than they have in the past.

Data Summary

The data indicates that some variables are slightly more relevant than others in pinpointing admissions decisions. GRE scores and GPA scores, when taken separately, are statistically significant in predicting admissions decisions, with a high degree of accuracy. Since most (273 of 400 total) were not admitted, any predictive model stemming from this data would have most applicants predicted to be not admitted. ADMIT (admissions decision) was dummy coded as 0 for not admitted and 1 for admitted and designated as a nominal data type in order to perform the following logistic regression.

Findings

Only GRE scores & GPA (separately) are statistically significant in determining admission to graduate school; while GRE scores & GPA separately are statistically significant in determining admissions, the best model according to JMP has serious flaws since, ideally, these variables should be able to factor in to the same model.

Additionally, the resultant models have very low R squared values, indicating very weak correlations. Thus, predicting admissions decisions based upon GPA or GRE scores is not reliable.

Limitations

Data that would improve a logistic regression model are as follows:

- Undergraduate school ranking data should be included, instead of a general indication of whether a program is prestigious or not.
    - This would allow for broader trends to be analyzed, since it is possible that the top few ranking schools could nearly guarantee Graduate Admissions despite low GRE scores and/or GPA.
        - For example, Harvard and Villanova hold different levels of prestige but would be combined in the same category for this data set.
- Data indicating when applicants applied (early decision, early action, etc).
- Data indicating whether an applicant was waitlisted or not, since this is also a relevant admission decision.
- Any other available data that admissions committees factor in for admissions decisions should be included.
- The success of admitted applicants should be tracked (in terms of their salaries and their contributions to their graduate school endowments); this data would indicate what data should be relied upon in admissions decisions moving forward.

Conclusions

This is a very weak model, and needs some sort of other data to become a more reliable model. With such a low R squared, while this model is statistically significant, this model should not be relied upon. Considering that every prediction made was that applicants would not be admitted, this model only really shows that graduate school applicants are far more likely to be rejected. While predictions were nearly 70% accurate, the fact that the only predictions made were that applicants would not be admitted is a shortfall of the model that, to me, discounts its usefulness.

In short, GRE & GPA are statistically significant in terms of predicting admissions to graduate school, but the prestige of one's undergraduate program is not. However, GRE & GPA, when combined for each applicant, seem to have little effect on admissions decisions.

Recommendations

In addition to gathering more information about each applicant, it would be beneficial to objectively admit more applicants. The data indicates that there are other factors in play for admissions decision, given that GPA and GRE together are not statistically significant in predicting admissions decisions. Thus, these factors should be known in order for decision makers to better know who to admit.

Graduate admissions decision makers should objectively evaluate each applicant more based upon GRE and GPA scores than they have in the past, since there are such weak correlations between GRE scores or GPA on admissions decisions; this may be seen as unfair to some that worked hard to gain these scores and averages.

Appendix Organization

Appendix	**Purpose**
A	Data
B	Logistic Plot, N & R Squared
C	Parameter Estimates
D	Confusion Matrix
E	ROC Curves
F	Lack of Fit Tables
G	Effect Summaries
H	Overall Descriptive Statistics

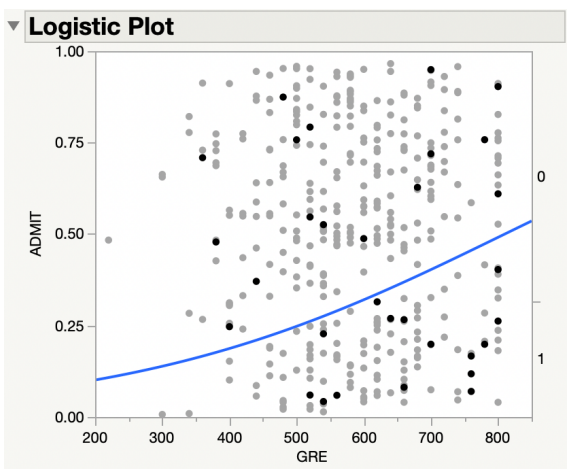
For Appendices B, C, E & F, figures related to GRE are on the left, while figures related to GPA are on the right.

Appendix A (Data)

ADMIT	Above Average GRE?	GRE	TOPNOTCH	GPA	Lin[1]	Prob[1]	Prob[0]	Most Likely ADMIT
0	0	380	0	3.61	-1.540103831	0.1765201813	0.8234798187	0
1	1	660	1	3.67	-0.537084561	0.3688660479	0.6311339521	0
1	1	800	1	4	-0.035574926	0.4911072064	0.5088927936	0
1	1	640	0	3.19	-0.608728795	0.3523492316	0.6476507684	0
0	1	520	0	2.93	-1.038594196	0.2614213358	0.7385786642	0
1	1	760	0	3	-0.178863393	0.4554029845	0.5445970155	0
1	1	560	0	2.98	-0.895305729	0.290016126	0.709983874	0
0	0	400	0	3.08	-1.468459598	0.1871768604	0.8128231396	0
1	1	540	0	3.39	-0.966949963	0.2754888566	0.7245111434	0
0	1	700	1	3.92	-0.393796094	0.4028038034	0.5971961966	0
0	1	800	1	4	-0.035574926	0.4911072064	0.5088927936	0
0	0	440	0	3.22	-1.325171131	0.2099592387	0.7900407613	0
1	1	760	0	4	-0.178863393	0.4554029845	0.5445970155	0
0	1	700	0	3.08	-0.393796094	0.4028038034	0.5971961966	0
1	1	700	1	4	-0.393796094	0.4028038034	0.5971961966	0
0	0	480	0	3.44	-1.181882663	0.2347138572	0.7652861428	0
0	1	780	0	3.87	-0.107219159	0.4732208595	0.5267791405	0
0	0	360	0	2.56	-1.611748065	0.1663460592	0.8336539408	0
0	1	800	0	3.75	-0.035574926	0.4911072064	0.5088927936	0
1	1	540	0	3.81	-0.966949963	0.2754888566	0.7245111434	0
0	0	500	0	3.17	-1.11023843	0.2478264406	0.7521735594	0
1	1	660	1	3.63	-0.537084561	0.3688660479	0.6311339521	0
0	1	600	0	2.82	-0.752017262	0.3203819085	0.6796180915	0
0	1	680	0	3.19	-0.465440327	0.3856960232	0.6143039768	0
1	1	760	0	3.35	-0.178863393	0.4554029845	0.5445970155	0
1	1	800	0	3.66	-0.035574926	0.4911072064	0.5088927936	0
1	1	620	1	3.61	-0.680373028	0.3361780517	0.6638219483	0
1	1	520	0	3.74	-1.038594196	0.2614213358	0.7385786642	0
1	1	780	0	3.22	-0.107219159	0.4732208595	0.5267791405	0
0	1	520	0	3.29	-1.038594196	0.2614213358	0.7385786642	0

First, I transformed the data to include an additional column titled “Above Average GRE?”. I created a calculated field, in which a score of 1 indicates a score greater than 500, and a score of 0 indicates a score less than 500. This correlation had a lesser R squared value than GRE scores as a whole, so I did not move forward with using this variable. An interesting thing to note is that every prediction that the resultant model made, for both GRE & GPA as separate variables, is a prediction for no admission; this model is largely accurate despite it only making predictions of no admission, as can be seen in the prediction matrix later.

Appendix B (Logistic Plot, N & R Squared)



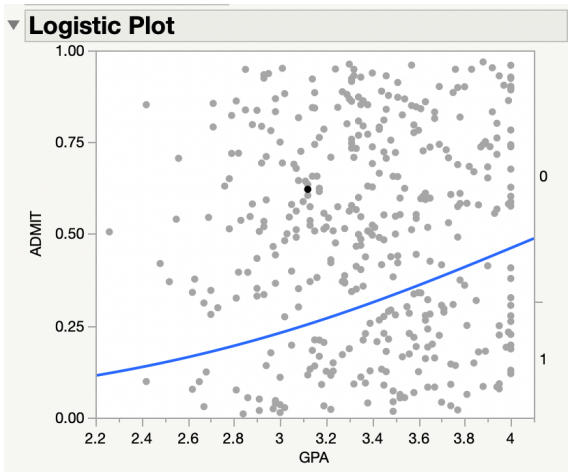
Converged in Gradient, 4 iterations

Iterations

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	6.96019	1	13.92038	0.0002*
Full	243.02807			
Reduced	249.98826			

RSquare (U)	0.0278
AICc	490.086
BIC	498.039
Observations (or Sum Wgts)	400



Converged in Gradient, 4 iterations

Iterations

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	6.50445	1	13.0089	0.0003*
Full	243.48381			
Reduced	249.98826			

RSquare (U)	0.0260
AICc	490.998
BIC	498.951
Observations (or Sum Wgts)	400

There are 400 observations (with none missing) for each model. Both models have very low R squared values of less than 3%, indicating that both GRE & GPA, when taken separately, have very little influence on the variation of Admissions Decisions.

Appendix C (Parameter Estimates)

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	-2.9013443	0.6060379	22.92	<.0001*
GRE	0.00358221	0.000986	13.20	0.0003*

For log odds of 1/0

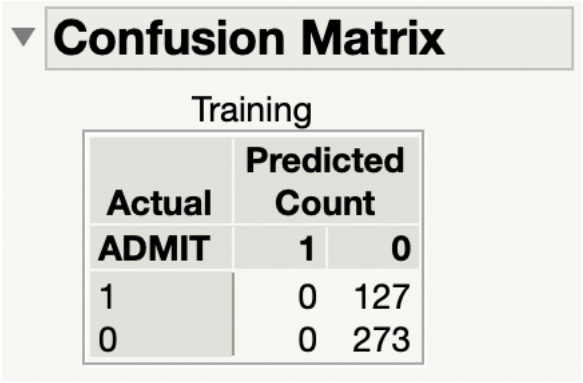
Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	-4.3575873	1.0353175	17.72	<.0001*
GPA	1.05110873	0.2988695	12.37	0.0004*

For log odds of 1/0

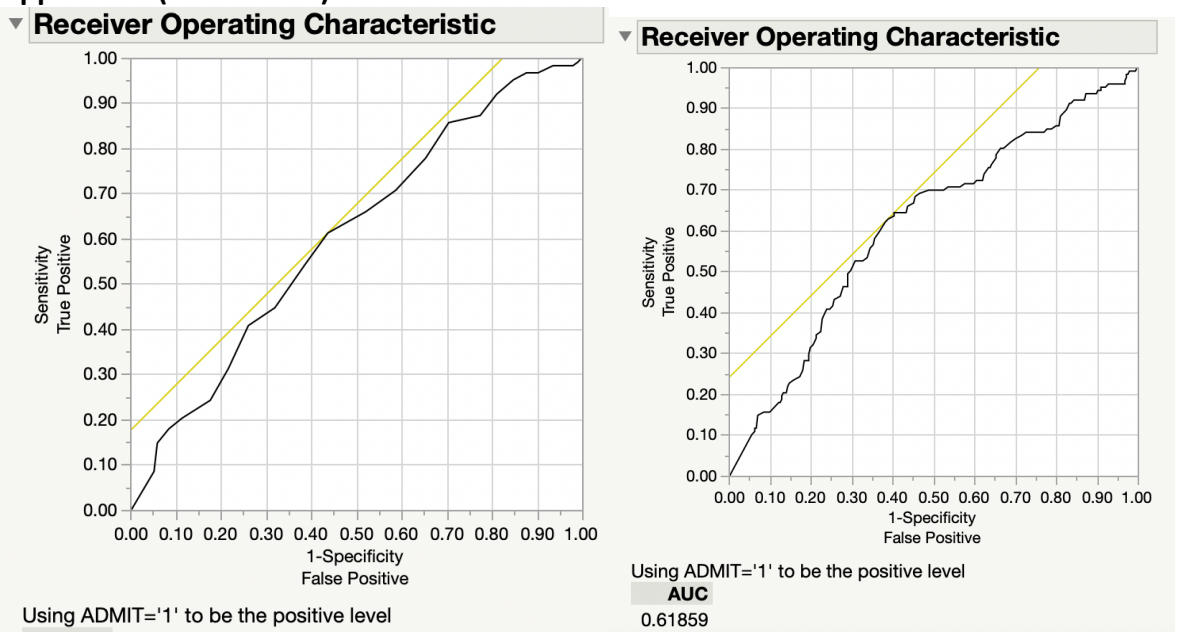
For these parameter estimates, both models are statistically significant due to them having P values of less than .005. This means that these predictive models are not explainable by chance alone, and there is some correlation between the dependent and independent variables.

Appendix D (Confusion Matrix – Same for GRE & GPA)



The confusion matrices for both GRE and GPA are identical; both models predicted 100% of the time that applicants would not be admitted. Since 273 of the 400 total applicants (68.25%) were not admitted, this blanket prediction that nobody would be admitted was 68.25% correct; any prediction made with around 70% accuracy should be trusted to an extent.

Appendix E (ROC Curves)



Both ROC curves are below the accuracy of a random guess, but with approximately a .4 level of specificity both curves encroach the accuracy of a random guess. Should the ROC curves be upward and leftward of the random guess line, then each model could be seen as reliable. Since the ROC curves are below and to the right of the random guess line, this proves that the model is not so reliable in its predictions.

Appendix F (Lack of Fit Tables)

Lack Of Fit

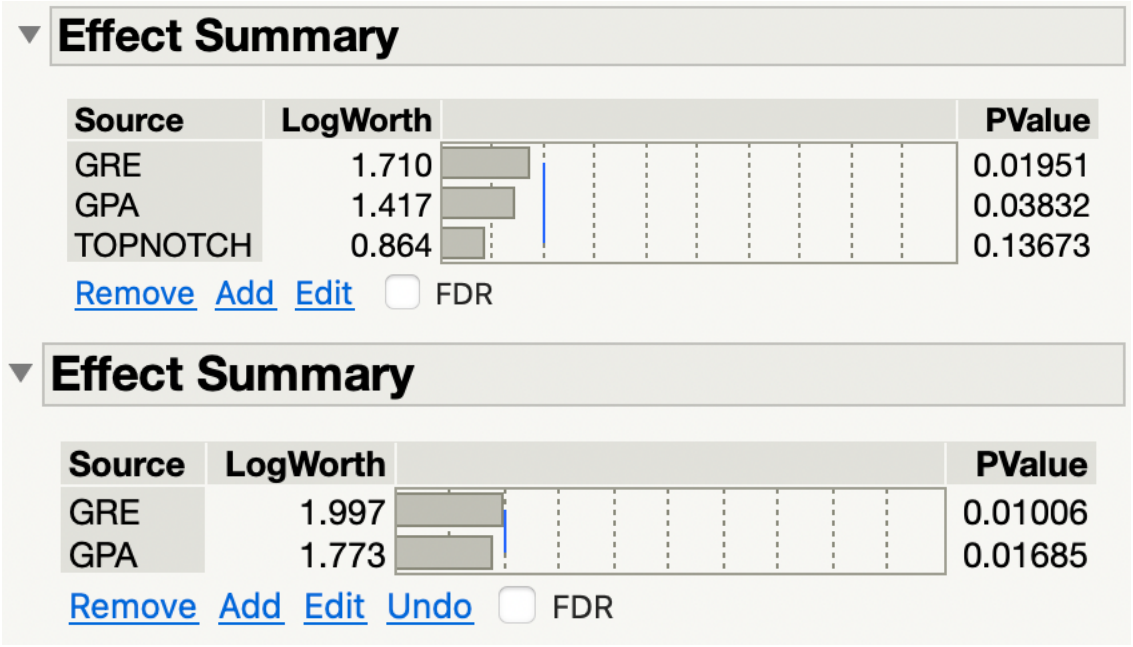
Source	DF	-LogLikelihood	ChiSquare
Lack Of Fit	24	16.20856	32.41712
Saturated	25	226.81951	Prob>ChiSq
Fitted	1	243.02807	0.1170

Lack Of Fit

Source	DF	-LogLikelihood	ChiSquare
Lack Of Fit	130	75.71356	151.4271
Saturated	131	167.77025	Prob>ChiSq
Fitted	1	243.48381	0.0963

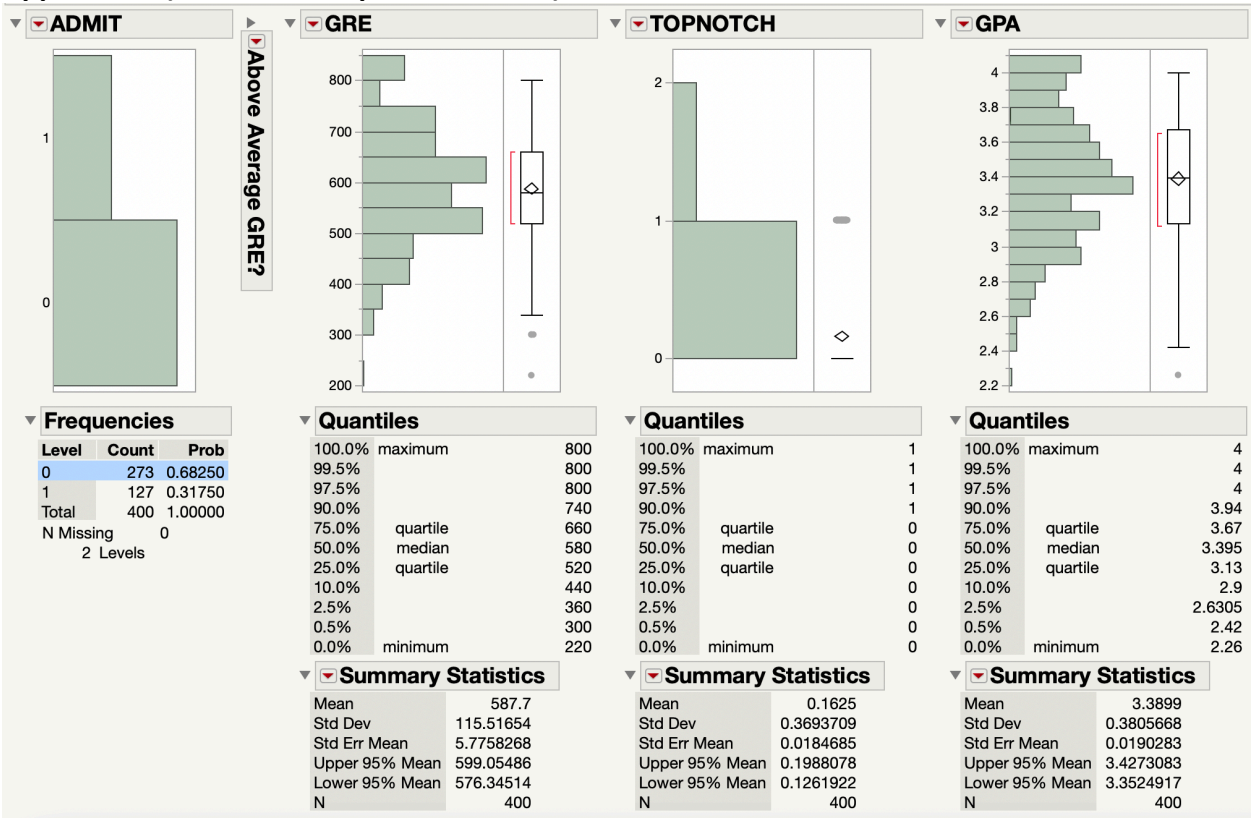
The lack of fit tables show that the model is a poor fit for the data. The model with GPA as the independent variable (on the right) has a slightly lower & therefore better P value. A P value of less than .005 would indicate a model that first the data well, so both models do not fit the data very well.

Appendix G (Effect Summary Prior to Eliminating Insignificant Variables)



Above demonstrates that GRE and GPA cannot be used for the same predictive model, given that they have P values of over .005. TOPNOTCH (whether an applicants' undergraduate program is prestigious or not) was eliminated prior to analyzing GRE and GPA together, since its P value is much greater than .005.

Appendix H (Overall Descriptive Statistics)



Overall, admissions decisions are greatly skewed towards not admitting applicants.

The means are slightly greater than the medians for GRE and whether an applicant comes from a top-notch undergraduate program; this indicates a positive skew. The converse is true for GPA, indicating a negative skew.

Interestingly, TOPNOTCH and ADMIT have similar distributions, indicating possible similarity between both variables.

Both GRE and GPA histograms indicate normal distributions, which indicates that the data fits natural phenomena instead of being steered or inaccurate in its collection.