Team 47: Arushi Dheer, Roland Graham, Manya Shah, Esther Sun, Leo Zhou

Part A Model Output/Summary

**Model Output**

Call:

glm(formula = completedadm ~ gpa + schoolsel\_chr + major1 + minor +

essay3length + essayuniquewords + appdeadline + submitteddate +

attendedevent, family = binomial(), data = train)

Deviance Residuals:

Min 1Q Median 3Q Max

-2.5569 0.3837 0.5489 0.7076 1.3257

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) 6.8520608 0.3443831 19.897 < 2e-16 \*\*\*

gpa -1.3397366 0.0472965 -28.326 < 2e-16 \*\*\*

schoolsel\_chrless\_selective -0.3181959 0.1013318 -3.140 0.001689 \*\*

schoolsel\_chrmore\_selective -0.4428198 0.0916795 -4.830 1.36e-06 \*\*\*

schoolsel\_chrmost\_selective -0.6077891 0.1007557 -6.032 1.62e-09 \*\*\*

schoolsel\_chrselective -0.3541583 0.0923165 -3.836 0.000125 \*\*\*

schoolsel\_chrunknown -0.2700107 0.0962949 -2.804 0.005047 \*\*

major1art 0.2865888 0.2563294 1.118 0.263546

major1biology -0.2393804 0.1877448 -1.275 0.202298

major1business -0.5482733 0.1876762 -2.921 0.003485 \*\*

major1chemistry -0.0372809 0.2610779 -0.143 0.886451

major1communications -0.3306781 0.1899165 -1.741 0.081653 .

major1criminal justice -0.3619478 0.2102828 -1.721 0.085207 .

major1economics -0.5359064 0.2035331 -2.633 0.008463 \*\*

major1education (elementary) -0.1240171 0.1935354 -0.641 0.521654

major1education (general) -0.2402841 0.2482642 -0.968 0.333116

major1education (secondary) -0.3411445 0.2110155 -1.617 0.105947

major1education early childhood -0.1450263 0.2337900 -0.620 0.535042

major1english -0.1052508 0.1832863 -0.574 0.565804

major1english literature 0.0059308 0.2194361 0.027 0.978438

major1history -0.1439694 0.1844795 -0.780 0.435150

major1international studies/relations -0.3930733 0.2028902 -1.937 0.052700 .

major1journalism -0.1000069 0.2361083 -0.424 0.671884

major1management -0.3627983 0.2319672 -1.564 0.117816

major1marketing -0.7249892 0.2260293 -3.208 0.001339 \*\*

major1mathematics -0.2360314 0.2109972 -1.119 0.263291

major1other -0.3783994 0.1682817 -2.249 0.024537 \*

major1philosophy -0.4392366 0.2477997 -1.773 0.076304 .

major1political science -0.4403925 0.1834101 -2.401 0.016344 \*

major1psychology -0.3530964 0.1753749 -2.013 0.044075 \*

major1sociology -0.4446337 0.1910241 -2.328 0.019932 \*

major1spanish -0.0484829 0.2259609 -0.215 0.830108

minorart 0.3906836 0.3395911 1.150 0.249957

minorbiology 0.2854164 0.3055944 0.934 0.350319

minorbusiness -0.3052087 0.2592841 -1.177 0.239147

minorchemistry 0.1009044 0.2953923 0.342 0.732656

minorcommunications 0.0706637 0.2996602 0.236 0.813578

minorcriminal justice 0.0874065 0.3256214 0.268 0.788369

minoreconomics -0.0849433 0.3015938 -0.282 0.778213

minoreducation (elementary) 0.1969001 0.3990559 0.493 0.621719

minoreducation (general) 0.5220671 0.2775271 1.881 0.059953 .

minoreducation (secondary) 0.0881170 0.3399141 0.259 0.795455

minoreducation early childhood 0.3284266 0.4603921 0.713 0.475621

minorenglish 0.3567899 0.2772729 1.287 0.198170

minorenglish literature 0.3275086 0.4876099 0.672 0.501799

minorhistory 0.2403876 0.2695711 0.892 0.372532

minorinternational studies/relations 0.2368086 0.3494287 0.678 0.497961

minorjournalism 0.2945857 0.4298130 0.685 0.493104

minormanagement 0.1200625 0.3720038 0.323 0.746888

minormarketing -0.4892484 0.3327574 -1.470 0.141484

minormathematics 0.3019532 0.2932116 1.030 0.303098

minornone 0.1530118 0.2308437 0.663 0.507435

minorother 0.1383391 0.2335601 0.592 0.553645

minorphilosophy 0.8615090 0.3448027 2.499 0.012470 \*

minorpolitical science 0.0023842 0.2764329 0.009 0.993118

minorpsychology -0.0487653 0.2525404 -0.193 0.846881

minorsociology 0.1753016 0.2682229 0.654 0.513391

minorspanish 0.0875617 0.2528649 0.346 0.729133

essay3length 0.0017411 0.0004244 4.102 4.09e-05 \*\*\*

essayuniquewords -0.0026313 0.0003804 -6.918 4.59e-12 \*\*\*

appdeadlineJanuary -0.1753950 0.0676175 -2.594 0.009489 \*\*

appdeadlineMarch -0.2209285 0.0665606 -3.319 0.000903 \*\*\*

appdeadlineOctober 0.1700687 0.0680240 2.500 0.012415 \*

appdeadlineSeptember 0.1603240 0.0807668 1.985 0.047142 \*

submitteddate -0.0075384 0.0009980 -7.554 4.23e-14 \*\*\*

attendedevent 0.3084475 0.0626983 4.920 8.67e-07 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 21738 on 21872 degrees of freedom

Residual deviance: 20431 on 21807 degrees of freedom

AIC: 20563

**Model Summary**

This model uses binary logistic regression in order to determine which explanatory variables are significant in predicting the outcome variable: whether the applicant completed all steps of the admissions process. The variables that are included in the model are GPA, school selectivity, major 1, minor, essay 3 length, essay unique words, application deadline, submitted date, and whether the applicant attended a recruiting event.

If the p-value of an explanatory variable is less than 0.05, then it is significant in predicting the outcome variable. Most variables in the model are significant, however the categorical variables “major1” and “minor” have some levels that are not significant. By performing a likelihood ratio test, however, we can see that eliminating “major1” or “minor” would present a significant change in the model and that they should be kept in.

If a variable’s coefficient is positive, then a one unit increase for that variable leads to an increased probability by a factor of the coefficient that an applicant completed the application and vice versa for a negative coefficient.

The Deviance Residuals have a positive first quartile, median, and third quartile values which show that the residuals of this model are mostly positive. This means that the predicted outcome values tend to be lower than the actual outcome values.

The Null Deviance and Residual Deviance can be used to calculated a Chi-Square value that can the usefulness of the model. This value is . Using an online calculator, the p-value of this chi-squared value is zero, which means that this model is useful in predicting the outcome variable given that it is below 0.05.