Categorical Data Analysis: Project Step 3

Elizabeth Grimes
July 16, 2020

1 Introduction

In this paper, we will answer the following questions in preparation for completing our term project.

1. Use multinomial logistic regression if your response variable has more than two categories.

2. Use ordinal logistic regression if your response variable has more than two categories.

3. Use log linear Poisson model if you have at least two categorical variables.

4. Use the Poisson and Negative Binomial regression models with variable selection, and compare the best Poisson regression model, using LRT and backward variable selection, with the best Negative Binomial regression model, using LRT and backward varaiable selection to find it, with AIC.

5. Interpret your best model in problem 4 and discuss whether the fitted model is reasonable and how you can improve it and use it.

2 Variable Selection Methodology

Since R is unable to perform regressions using all 150 variables in our data set, we select a subset of variables that is small enough to perform regressions. We select 22 measures that are all per-game or per-attempt averages. They are listed in Table 1, along with a description of each. However, this is still too many variables, and the algorithm does not converge, so we seek to eliminate extra variables. We find the correlations between each pair of variables in our subset, and for pairs with correlations that are significant at the 0.05 level, we remove one of them. There were many pairs of correlations that were significant at this level. We began by first examining the correlations that were significant at the 0.001 level. We chose the variable that correlated with the most other variables for removal, then performed the correlations again. Repeating this process, we were able to reduce the variables down to 5, which are listed in Table 2.

3 Question 1: Multinomial Regression

Our response variable is the number of wins each team achieved during the 2019 season. For previous types of regression, we have used the win record for each team as a binary variable. However, we can divide the number of wins into categories in order to perform multinomial regression. Since 10 wins or more is considered a highly successful season, and 6 wins or more makes a team bowl-eligible, we will divide our win records into three categories: 10 wins or more, 6-9 wins, and 0-5 wins. We perform multinomial regression considering all possible interactions, and use backwards variable

Table 1: Initial Variable Choices and Descriptions

Variable Name	Description
Off.Yards.per.Game	Offensive Yards / Number of Games
Yards.Per.Game.Allowed	Yards Allowed / Number of Games
X4th.Percent	4th Down Conversions / 4th Down Attempts
Opponent.4th.Percent	Opponent 4th Down Conversions / Opponent 4th Down Attempts
Avg.Yards.per.Kickoff.Return.Allowed	Net Kickoff Return Yards Allowed / Kickoff Returns Allowed
Avg.Yard.per.Kickoff.Return	Net Kickoff Return Yards / Number of Kickoffs Returned
Pass.Yards.Per.Game	Pass Yards / Number of Games
Pass.Yards.Per.Game.Allowed	Opponent Pass Yards Allowed / Number of Games
Penalty.Yards.Per.Game	Penalty Yards / Number of Games
Avg.Yards.Per.Punt.Return	Net Punt Return Yards / Number of Punts Returned
Avg.Yards.Allowed.per.Punt.Return	Net Punt Return Yards Allowed / Number of Punt Returns Allowed
Redzone.Points	Redzone Scores / Redzone Attempts
Redzone.Points.Allowed	Opponent Redzone Scores / Opponent Redzone Attempts
Rush.Yards.Per.Game.Allowed	Opponent Rush Yards Allowed / Number of Games
Rushing.Yards.per.Game	Rush Yards / Number of Games
Average.Sacks.per.Game	Sacks / Number of Games
Avg.Points.per.Game.Allowed	Points Allowed / Number of Games
Points.Per.Game	Total Points / Number of Games
Tackle.For.Loss.Per.Game	Total Tackles for Loss / Number of Games
X3rd.Percent	3rd Down Conversions / 3rd Down Attempts
Avg.Turnover.Margin.per.Game	Turnover Margin / Number of Games

Table 2: Final variables selected after correlations removed

Variable	Alternate Name
Pass.Yards.Per.Game.Allowed	x1
Penalty.Yards.Per.Game	x2
Avg.Yards.Per.Punt.Return	x3
Avg.Yards.Allowed.per.Punt.Return	x4
Rushing.Yards.per.Game	x5

selection to fit our model. The coefficients for the initial regression equation are displayed in Table 3. The list of variables removed in each step of variable selection is displayed in Table 4. The final regression model was found after 13 steps, and the coefficients for each variable are displayed in Table 5. The model has a residual deviance of 182.6571 on 220 degrees of freedom. There is a Hauck-Donner effect detected in the following estimates: '(Intercept):1', '(Intercept):2', 'x3:2', 'x2:x4:1', 'x2:x4:2', 'x3:x4:1', 'x2:x5:1', 'x2:x5:2', 'x3:x5:1', 'x4:x5:1', 'x4:x5:2', 'x4:x1:1', 'x4:x1:2', 'x5:x1:1', 'x5:x1:2', 'x2:x1:1', 'x2:x1:2', 'x3:x4:x5:2', 'x2:x4:x5:x1:1', 'x2:x4:x5:x1:2'.

Table 3: Coefficients for Multinomial Regression, Step 1

	Estimate	Std. Error	z value	Pr($>|z|$)
(Intercept):1	-1.92E+02	1.48E+03	NA	NA
(Intercept):2	-1.57E+03	1.09E+03	NA	NA
x1:1	1.04E+00	6.70E+00	0.155	0.877
x1:2	6.50E+00	4.68E+00	1.39	0.165
x2:1	5.52E+00	2.78E+01	0.199	0.842
x2:2	3.15E+01	2.03E+01	1.549	0.121
x3:1	-1.59E+01	1.76E+02	NA	NA
x3:2	8.42E+01	1.34E+02	0.627	0.531
x4:1	-3.53E+01	1.79E+02	NA	NA
x4:2	1.53E+02	1.38E+02	1.111	0.267
x5:1	2.14E+00	8.89E+00	0.241	0.81
x5:2	9.83E+00	7.06E+00	NA	NA
x1:x2:1	-3.01E-02	1.26E-01	NA	NA
x1:x2:2	-1.27E-01	8.64E-02	NA	NA
x1:x3:1	9.21E-02	7.97E-01	0.116	0.908
x1:x3:2	-3.09E-01	5.68E-01	NA	NA
x2:x3:1	-6.67E-02	3.22E+00	NA	NA
x2:x3:2	-2.13E+00	2.47E+00	NA	NA
x1:x4:1	1.45E-01	8.16E-01	0.178	0.859
x1:x4:2	-6.03E-01	5.91E-01	NA	NA
x2:x4:1	2.51E-01	3.35E+00	0.075	0.94
x2:x4:2	-3.13E+00	2.60E+00	NA	NA
x3:x4:1	1.04E+01	2.16E+01	0.479	0.632
x3:x4:2	-6.96E+00	1.70E+01	NA	NA
x1:x5:1	-1.01E-02	4.01E-02	NA	NA
x1:x5:2	-4.06E-02	3.01E-02	NA	NA
x2:x5:1	-5.08E-02	1.68E-01	NA	NA
x2:x5:2	-1.97E-01	1.32E-01	NA	NA
x3:x5:1	1.41E-03	1.05E+00	0.001	0.999
x3:x5:2	-5.09E-01	8.49E-01	NA	NA
x4:x5:1	1.10E-01	1.07E+00	0.102	0.918
x4:x5:2	-9.65E-01	8.99E-01	NA	NA
x1:x2:x3:1	8.26E-05	1.46E-02	0.006	0.995
x1:x2:x3:2	7.82E-03	1.04E-02	0.752	0.452
x1:x2:x4:1	-6.55E-04	1.52E-02	NA	NA
x1:x2:x4:2	1.21E-02	1.10E-02	1.093	0.274
x1:x3:x4:1	-4.98E-02	9.89E-02	NA	NA

x1:x3:x4:2	2.13E-02	7.22E-02	0.295	0.768
x2:x3:x4:1	-1.26E-01	3.91E-01	-0.323	0.746
x2:x3:x4:2	2.03E-01	3.16E-01	0.642	0.521
x1:x2:x5:1	2.50E-04	7.57E-04	0.331	0.741
x1:x2:x5:2	7.96E-04	5.61E-04	1.419	0.156
x1:x3:x5:1	-1.53E-04	4.73E-03	NA	NA
x1:x3:x5:2	1.91E-03	3.58E-03	0.534	0.593
x2:x3:x5:1	2.04E-03	1.92E-02	0.106	0.915
x2:x3:x5:2	1.30E-02	1.56E-02	0.828	0.408
x1:x4:x5:1	-4.61E-04	4.87E-03	NA	NA
x1:x4:x5:2	3.82E-03	3.84E-03	0.995	0.32
x2:x4:x5:1	2.79E-04	2.02E-02	0.014	0.989
x2:x4:x5:2	1.98E-02	1.70E-02	1.163	0.245
x3:x4:x5:1	-5.15E-02	1.28E-01	-0.401	0.688
x3:x4:x5:2	4.16E-02	1.09E-01	0.381	0.703
x1:x2:x3:x4:1	6.04E-04	1.78E-03	0.34	0.734
x1:x2:x3:x4:2	-6.81E-04	1.33E-03	NA	NA
x1:x2:x3:x5:1	-7.64E-06	8.67E-05	NA	NA
x1:x2:x3:x5:2	-4.85E-05	6.59E-05	NA	NA
x1:x2:x4:x5:1	-3.42E-06	9.17E-05	NA	NA
x1:x2:x4:x5:2	-7.66E-05	7.22E-05	NA	NA
x1:x3:x4:x5:1	2.49E-04	5.84E-04	0.427	0.669
x1:x3:x4:x5:2	-1.34E-04	4.63E-04	NA	NA
x2:x3:x4:x5:1	5.76E-04	2.33E-03	NA	NA
x2:x3:x4:x5:2	-1.24E-03	2.03E-03	NA	NA
x1:x2:x3:x4:x5:1	-2.77E-06	1.06E-05	-0.262	0.793
x1:x2:x3:x4:x5:2	4.31E-06	8.57E-06	0.503	0.615 [b]

Table 4: Backwards variable selection for multinomial regression

Step	Variable	Df	Deviance	AIC	LRT	Pr(>Chi)
2	x1:x2:x3:x4:x5	2	167.52	291.52	0.61158	0.7365 [t]
3	x2:x3:x4:x1	2	167.71	287.71	0.1825	0.91277
4	x2:x3:x5:x1	2	169.85	285.85	2.1426	0.34257
5	x2:x3:x1	2	169.97	281.96	0.1157	0.94377
6	x2:x3:x4:x5	2	172.97	280.97	3.0011	0.22301
7	x2:x3:x5	2	173.08	277.08	0.1128	0.94515
8	x2:x3:x4	2	174.68	274.68	1.6001	0.4493
9	x3:x4:x5:x1	2	176.54	272.54	1.8567	0.3952
10	x3:x5:x1	2	177.26	269.26	0.7251	0.6959
11	x3:x4:x1	2	178.35	266.36	1.0944	0.57858
12	x3:x1	2	178.73	262.73	0.3767	0.82832
13	x2:x3	2	182.66	262.66	3.9252	0.14049 [b]

Table 5: Coefficients for multinomial regression after backwards variable selection

	Estimate	Std. Error	z value	$\Pr(>\|z\|)$
(Intercept):1	-1.40E+02	2.77E+02	NA	NA
(Intercept):2	-8.43E+02	2.80E+02	NA	NA
x2:1	1.70E+00	5.09E+00	0.334	0.73844
x2:2	1.39E+01	4.88E+00	2.846	0.00443
x3:1	1.54E+00	1.35E+00	1.14	0.25415
x3:2	-2.30E+00	1.21E+00	NA	NA
x4:1	2.07E+01	3.48E+01	0.595	0.55173
x4:2	9.17E+01	3.42E+01	2.679	0.00738
x5:1	1.18E+00	1.52E+00	0.777	0.4373
x5:2	5.07E+00	1.67E+00	3.036	0.0024
x1:1	8.34E-01	1.25E+00	0.669	0.50367
x1:2	3.84E+00	1.25E+00	3.079	0.00208
x2:x4:1	-2.81E-01	6.16E-01	NA	NA
x2:x4:2	-1.45E+00	5.83E-01	NA	NA
x3:x4:1	-2.01E-01	1.55E-01	NA	NA
x3:x4:2	1.69E-01	1.43E-01	1.182	0.23706
x2:x5:1	-1.57E-02	2.79E-02	NA	NA
x2:x5:2	-8.35E-02	2.92E-02	NA	NA
x3:x5:1	-1.09E-02	7.86E-03	NA	NA
x3:x5:2	1.42E-02	7.55E-03	1.886	0.05924
x4:x5:1	-1.56E-01	1.85E-01	NA	NA
x4:x5:2	-5.49E-01	2.03E-01	NA	NA
x4:x1:1	-1.19E-01	1.58E-01	NA	NA
x4:x1:2	-4.10E-01	1.51E-01	NA	NA
x5:x1:1	-6.08E-03	6.83E-03	NA	NA
x5:x1:2	-2.29E-02	7.41E-03	NA	NA
x2:x1:1	-1.32E-02	2.26E-02	NA	NA
x2:x1:2	-6.14E-02	2.13E-02	NA	NA
x2:x4:x5:1	2.20E-03	3.29E-03	0.667	0.50488
x2:x4:x5:2	8.58E-03	3.45E-03	2.491	0.01275
x3:x4:x5:1	1.42E-03	9.20E-04	1.548	0.1217
x3:x4:x5:2	-1.09E-03	9.11E-04	NA	NA
x4:x5:x1:1	7.92E-04	8.39E-04	0.944	0.34539
x4:x5:x1:2	2.43E-03	8.91E-04	2.732	0.00629
x2:x4:x1:1	1.91E-03	2.76E-03	0.692	0.48911
x2:x4:x1:2	6.37E-03	2.55E-03	2.504	0.01229
x2:x5:x1:1	9.68E-05	1.25E-04	0.777	0.43731
x2:x5:x1:2	3.65E-04	1.27E-04	2.88	0.00398
x2:x4:x5:x1:1	-1.28E-05	1.49E-05	NA	NA
x2:x4:x5:x1:2	-3.72E-05	1.50E-05	NA	NA

4 Question 2: Ordinal Logistic Regression

Because the three categories for numbers of wins are ordinal, we can also perform ordinal logistic regression. We use the same five variables that we chose in section 2. The initial regression model is displayed in Table 6. The threshold coefficients are 559.7 for the categories 10 or More—5 or Less and 561.8 for the categories 5 or Less—6 to 9.

During variable selection, the variables we are ultimately left with are not the same as the ones we are left with for the multinomial regression. The selection is displayed in Table 7. The coefficients for the final model are displayed in Table 8. The model has threshold coefficients of 94.71 for the categories 10 or More—5 or Less and 96.76 for the categories 5 or Less—6 to 9.

Table 6: Coefficients for ordinal logistic regression, step 1

	Estimate	Std. Error	z value	Pr(> $|z|$)
(Intercept):1	-1.92E+02	1.48E+03	NA	NA
(Intercept):2	-1.57E+03	1.09E+03	NA	NA
x1:1	1.04E+00	6.70E+00	0.155	0.877
x1:2	6.50E+00	4.68E+00	1.39	0.165
x2:1	5.52E+00	2.78E+01	0.199	0.842
x2:2	3.15E+01	2.03E+01	1.549	0.121
x3:1	-1.59E+01	1.76E+02	NA	NA
x3:2	8.42E+01	1.34E+02	0.627	0.531
x4:1	-3.53E+01	1.79E+02	NA	NA
x4:2	1.53E+02	1.38E+02	1.111	0.267
x5:1	2.14E+00	8.89E+00	0.241	0.81
x5:2	9.83E+00	7.06E+00	NA	NA
x1:x2:1	-3.01E-02	1.26E-01	NA	NA
x1:x2:2	-1.27E-01	8.64E-02	NA	NA
x1:x3:1	9.21E-02	7.97E-01	0.116	0.908
x1:x3:2	-3.09E-01	5.68E-01	NA	NA
x2:x3:1	-6.67E-02	3.22E+00	NA	NA
x2:x3:2	-2.13E+00	2.47E+00	NA	NA
x1:x4:1	1.45E-01	8.16E-01	0.178	0.859
x1:x4:2	-6.03E-01	5.91E-01	NA	NA
x2:x4:1	2.51E-01	3.35E+00	0.075	0.94
x2:x4:2	-3.13E+00	2.60E+00	NA	NA
x3:x4:1	1.04E+01	2.16E+01	0.479	0.632
x3:x4:2	-6.96E+00	1.70E+01	NA	NA
x1:x5:1	-1.01E-02	4.01E-02	NA	NA
x1:x5:2	-4.06E-02	3.01E-02	NA	NA
x2:x5:1	-5.08E-02	1.68E-01	NA	NA
x2:x5:2	-1.97E-01	1.32E-01	NA	NA
x3:x5:1	1.41E-03	1.05E+00	0.001	0.999
x3:x5:2	-5.09E-01	8.49E-01	NA	NA
x4:x5:1	1.10E-01	1.07E+00	0.102	0.918
x4:x5:2	-9.65E-01	8.99E-01	NA	NA
x1:x2:x3:1	8.26E-05	1.46E-02	0.006	0.995
x1:x2:x3:2	7.82E-03	1.04E-02	0.752	0.452
x1:x2:x4:1	-6.55E-04	1.52E-02	NA	NA

x1:x2:x4:2	1.21E-02	1.10E-02	1.093	0.274
x1:x3:x4:1	-4.98E-02	9.89E-02	NA	NA
x1:x3:x4:2	2.13E-02	7.22E-02	0.295	0.768
x2:x3:x4:1	-1.26E-01	3.91E-01	-0.323	0.746
x2:x3:x4:2	2.03E-01	3.16E-01	0.642	0.521
x1:x2:x5:1	2.50E-04	7.57E-04	0.331	0.741
x1:x2:x5:2	7.96E-04	5.61E-04	1.419	0.156
x1:x3:x5:1	-1.53E-04	4.73E-03	NA	NA
x1:x3:x5:2	1.91E-03	3.58E-03	0.534	0.593
x2:x3:x5:1	2.04E-03	1.92E-02	0.106	0.915
x2:x3:x5:2	1.30E-02	1.56E-02	0.828	0.408
x1:x4:x5:1	-4.61E-04	4.87E-03	NA	NA
x1:x4:x5:2	3.82E-03	3.84E-03	0.995	0.32
x2:x4:x5:1	2.79E-04	2.02E-02	0.014	0.989
x2:x4:x5:2	1.98E-02	1.70E-02	1.163	0.245
x3:x4:x5:1	-5.15E-02	1.28E-01	-0.401	0.688
x3:x4:x5:2	4.16E-02	1.09E-01	0.381	0.703
x1:x2:x3:x4:1	6.04E-04	1.78E-03	0.34	0.734
x1:x2:x3:x4:2	-6.81E-04	1.33E-03	NA	NA
x1:x2:x3:x5:1	-7.64E-06	8.67E-05	NA	NA
x1:x2:x3:x5:2	-4.85E-05	6.59E-05	NA	NA
x1:x2:x4:x5:1	-3.42E-06	9.17E-05	NA	NA
x1:x2:x4:x5:2	-7.66E-05	7.22E-05	NA	NA
x1:x3:x4:x5:1	2.49E-04	5.84E-04	0.427	0.669
x1:x3:x4:x5:2	-1.34E-04	4.63E-04	NA	NA
x2:x3:x4:x5:1	5.76E-04	2.33E-03	NA	NA
x2:x3:x4:x5:2	-1.24E-03	2.03E-03	NA	NA
x1:x2:x3:x4:x5:1	-2.77E-06	1.06E-05	-0.262	0.793
x1:x2:x3:x4:x5:2	4.31E-06	8.57E-06	0.503	0.615 [b]

Table 7: Backwards variable selection for ordinal logistic regression

Step	Variable	Df	AIC	LRT	Pr(>Chi)
2	x1:x2:x3:x4:x5	1	300.8	3.7319	0.05338
3	x2:x4:x5:x1	1	298.8	0.00004	0.9949
4	x2:x3:x5:x1	1	296.86	0.0542	0.81587
5	x2:x5:x1	1	294.86	0.0057	0.93978
6	x2:x3:x4:x5	1	293.1	0.2373	0.62619
7	x2:x3:x4:x1	1	291.63	0.5332	0.46526
8	x2:x3:x1	1	290.05	0.4134	0.5202
9	x2:x1	1	286.75	0.3393	0.56023
10	x2:x3:x5	1	285.63	0.8867	0.3464
11	x2:x4:x5	1	284.17	0.5399	0.46249
12	x2:x3:x4	1	282.9	0.7314	0.39244
13	x2:x5	1	282.37	1.4659	0.226

Table 8: Coefficients for ordinal logistic regression after backwards variable selection

	Estimate	Std. Error	z value	Pr(> $\|z\|$)
x2	2.18E-01	NA	NA	NA
x3	1.56E+01	NA	NA	NA
x4	1.18E+01	NA	NA	NA
x5	4.58E-01	NA	NA	NA
x1	3.81E-01	NA	NA	NA
x2:x3	-1.48E-02	NA	NA	NA
x2:x4	-1.51E-02	NA	NA	NA
x3:x4	-1.79E+00	NA	NA	NA
x3:x5	-8.58E-02	NA	NA	NA
x4:x5	-6.19E-02	NA	NA	NA
x3:x1	-6.60E-02	NA	NA	NA
x4:x1	-5.04E-02	NA	NA	NA
x5:x1	-2.11E-03	NA	NA	NA
x3:x4:x5	1.07E-02	NA	NA	NA
x3:x4:x1	8.21E-03	NA	NA	NA
x3:x5:x1	3.87E-04	NA	NA	NA
x4:x5:x1	2.90E-04	NA	NA	NA
x3:x4:x5:x1	-4.95E-05	NA	NA	NA

5 Question 3: Log Linear Poisson Model

In order to utilize the log linear Poisson modeling technique, we must convert our explanatory variables to categorical variables. For each variable, we obtained descriptive statistics to find the median. We then divided the observations into "high" and "low" categories for each variable. In addition, we used the "winning record" response variable from our logistic and probit regression models. A frequency table can be seen in Table 9.

Table 9: Frequency table for variables x1, x2, x3, x4, x5, WR

x1	x2	x3	x4	x5	WR	Freq
low	low	low	low	low	No	2
					Yes	3
				high	No	2
					Yes	2
			high	low	No	2
					Yes	1
				high	No	1
					Yes	5
		high	low	low	No	1
					Yes	2
				high	No	2
					Yes	2
			high	low	No	2
					Yes	2
				high	No	0

					Yes	3
	high	low	low	low	No	2
					Yes	2
				high	No	1
					Yes	2
			high	low	No	2
					Yes	4
				high	No	0
					Yes	2
		high	low	low	No	1
					Yes	1
				high	No	2
					Yes	4
			high	low	No	3
					Yes	1
				high	No	0
					Yes	6
high	low	low	low	low	No	3
					Yes	0
				high	No	1
					Yes	2
			high	low	No	3
					Yes	1
				high	No	4
					Yes	2
		high	low	low	No	4
					Yes	1
				high	No	3
					Yes	5
			high	low	No	1
					Yes	2
				high	No	0
					Yes	1
	high	low	low	low	No	4
					Yes	1
				high	No	1
					Yes	0
			high	low	No	2
					Yes	1
				high	No	4
					Yes	4
		high	low	low	No	3
					Yes	2
				high	No	0
					Yes	4

			high	low	No	4
					Yes	2
				high	No	0
					Yes	0

We perform a log linear regression using a three-way interaction model. The coefficients of the initial regression model are displayed in Table 10. This model has a residual deviance of 26.867 on 22 degrees of freedom, and an AIC of 256.71. Next, we perform backwards variable selection on the model, the results of which are displayed in Table 11. The coefficients for the final Poisson model are displayed in Table 12. This model has a residual deviance of 35.546 on 45 degrees of freedom, and an AIC of 219.39. Notably, the variable x2 has been completely removed from this model.

Table 10: Log linear Poisson regression, three-way interaction

	Estimate	Std. Error	z value	$Pr(>\|z\|)$
(Intercept)	0.97685	0.53499	1.826	0.0679 [t]
x1high	-0.14604	0.69028	-0.212	0.8324
x2high	-0.18439	0.7134	-0.258	0.7961
x3high	-0.78398	0.76739	-1.022	0.307
x4high	-0.47181	0.72635	-0.65	0.516
x5high	-0.51582	0.75344	-0.685	0.4936
WRYes	-0.01899	0.71817	-0.026	0.9789
x1high:x2high	0.44148	0.82668	0.534	0.5933
x1high:x3high	1.36027	0.82543	1.648	0.0994
x1high:x4high	0.53395	0.80979	0.659	0.5097
x1high:x5high	0.59473	0.84846	0.701	0.4833
x1high:WRYes	-1.47856	0.95517	-1.548	0.1216
x2high:x3high	0.20456	0.83702	0.244	0.8069
x2high:x4high	0.09072	0.81117	0.112	0.911
x2high:x5high	-0.60583	0.89353	-0.678	0.4978
x2high:WRYes	0.25335	0.87962	0.288	0.7733
x3high:x4high	0.8674	0.89038	0.974	0.33
x3high:x5high	0.89961	0.91274	0.986	0.3243
x3high:WRYes	-0.62833	0.92422	-0.68	0.4966
x4high:x5high	0.06165	0.89182	0.069	0.9449
x4high:WRYes	0.35736	0.8718	0.41	0.6819
x5high:WRYes	0.21805	0.8997	0.242	0.8085
x1high:x2high:x3high	-0.66354	0.7769	-0.854	0.3931
x1high:x2high:x4high	0.20933	0.76086	0.275	0.7832
x1high:x2high:x5high	-0.48129	0.78195	-0.616	0.5382
x1high:x2high:WRYes	0.11563	0.78625	0.147	0.8831
x1high:x3high:x4high	-1.72077	0.81593	-2.109	0.0349
x1high:x3high:x5high	-1.7425	0.86431	-2.016	0.0438
x1high:x3high:WRYes	1.41072	0.82856	1.703	0.0886
x1high:x4high:x5high	0.30877	0.84142	0.367	0.7136
x1high:x4high:WRYes	-0.43476	0.8099	-0.537	0.5914
x1high:x5high:WRYes	0.29172	0.81993	0.356	0.722
x2high:x3high:x4high	0.30043	0.76835	0.391	0.6958

x2high:x3high:x5high	0.47872	0.81005	0.591	0.5545
x2high:x3high:WRYes	-0.24712	0.8019	-0.308	0.758
x2high:x4high:x5high	0.15711	0.78197	0.201	0.8408
x2high:x4high:WRYes	-0.19369	0.78658	-0.246	0.8055
x2high:x5high:WRYes	0.27994	0.78509	0.357	0.7214
x3high:x4high:x5high	-2.40079	0.88703	-2.707	0.0068
x3high:x4high:WRYes	0.44711	0.87903	0.509	0.611
x3high:x5high:WRYes	1.13061	0.89474	1.264	0.2064
x4high:x5high:WRYes	0.82388	0.86682	0.95	0.3419 [b]

Table 11: Backwards variable selection for log linear Poisson model

Step	Variable	Df	Deviance	AIC	LRT	Pr(>Chi)
2	x1:x2:WR	1	26.889	254.73	0.0216	0.883054 [t]
3	x2:x4:x5	1	26.927	252.77	0.0382	0.845034
4	x2:x4:WR	1	26.964	250.81	0.0373	0.84679
5	x2:x3:WR	1	27.036	248.88	0.0716	0.78906
6	x1:x5:WR	1	27.151	246.99	0.1154	0.734023
7	x2:x3:x4	1	27.259	245.1	0.1078	0.742607
8	x1:x2:x4	1	27.381	243.22	0.1224	0.72641
9	x1:x4:x5	1	27.546	241.39	0.1642	0.685345
10	x2:x5:WR	1	27.723	239.56	0.1779	0.673157
11	x1:x4:WR	1	27.948	237.79	0.2242	0.635836
12	x3:x4:WR	1	28.168	236.01	0.2201	0.639002
13	x2:x3:x5	1	28.48	234.32	0.3126	0.576083
14	x2:WR	1	28.874	232.72	0.3934	0.53054
15	x1:x2:x5	1	29.304	231.15	0.4299	0.512035
16	x2:x4	1	29.88	229.72	0.576	0.447901
17	x4:x5:WR	1	30.73	228.57	0.8503	0.356464
18	x1:x2:x3	1	31.767	227.61	1.037	0.308527
19	x1:x2	1	31.814	225.66	0.047	0.828302
20	x2:x3	1	31.938	223.78	0.1239	0.724886
21	x2:x5	1	32.708	222.55	0.77	0.38022
22	x2	1	32.708	220.55	0	1
23	x3:x5:WR	1	33.975	219.82	1.2669	0.260357
24	x4:WR	1	35.546	219.39	1.579	0.2100745 [b]

6 Question 4: Poisson and Negative Binomial Regression

Next, we consider Poisson and negative binomial regression for the data. Since both forms of regression consider counts, we will use the number of wins as our response variable (ranging from 0 to 15). It is not likely that this variable will have a Poisson distribution, as it is very rare that a team wins NO football games during a season. Similarly, it is also very rare that a team has no losses during a season.

Table 12: Log linear Poisson model after backwards variable selection

	Estimate	Std. Error	z value	Pr(> $|z|$)
(Intercept)	0.91707	0.36207	2.533	0.01131 [t]
x1high	0.10161	0.47993	0.212	0.83233
x3high	-0.93559	0.61151	-1.53	0.12603
x4high	-0.2648	0.42105	-0.629	0.52942
x5high	-1.40837	0.53433	-2.636	0.00839
WRYes	0.03474	0.40345	0.086	0.93138
x1high:x3high	1.11632	0.74102	1.506	0.13195
x1high:x4high	0.44789	0.51081	0.877	0.38058
x1high:x5high	0.73517	0.55026	1.336	0.18154
x1high:WRYes	-1.50264	0.55292	-2.718	0.00657
x3high:x4high	1.14442	0.66844	1.712	0.08688
x3high:x5high	1.66393	0.69671	2.388	0.01693
x3high:WRYes	-0.09471	0.54642	-0.173	0.8624
x4high:x5high	0.72184	0.51131	1.412	0.15803
x5high:WRYes	1.28765	0.39401	3.268	0.00108
x1high:x3high:x4high	-1.68663	0.77067	-2.189	0.02863
x1high:x3high:x5high	-1.75125	0.80536	-2.174	0.02967
x1high:x3high:WRYes	1.18548	0.7823	1.515	0.12968
x3high:x4high:x5high	-1.96058	0.771	-2.543	0.01099 [b]

6.1 Poisson Regression

The coefficients for the initial Poisson regression are displayed in Table 13. This model has a residual deviance of 122.94 on 98 degrees of freedom, and an AIC of 660.46. We perform backwards variable selection on the regression, the details of which are displayed in Table 14. The coefficients for the final Poisson regression model are displayed in Table 15. This model has a null deviance of 200.86 on 129 degrees of freedom, a residual deviance of 131.03 on 109 degrees of freedom, and an AIC of 646.56. This is extremely high and reflects the initial hypothesis that a Poisson regression is not suitable for this data.

Table 13: Poisson regression model, initial fit

	Estimate	Std. Error	z value	Pr(> $|z|$)
(Intercept)	6.98E+01	4.16E+01	1.68	0.0929 [t]
x1	-2.68E-01	1.89E-01	-1.42	0.1555
x2	-1.42E+00	8.03E-01	-1.766	0.0774
x3	-1.47E+00	6.16E+00	-0.239	0.8111
x4	-9.30E+00	6.49E+00	-1.431	0.1523
x5	-4.48E-01	2.27E-01	-1.98	0.0477
x1:x2	5.48E-03	3.62E-03	1.513	0.1303
x1:x3	4.28E-03	2.71E-02	0.158	0.8745
x2:x3	5.64E-02	1.11E-01	0.51	0.6103
x1:x4	3.70E-02	2.93E-02	1.263	0.2064
x2:x4	1.82E-01	1.20E-01	1.518	0.129
x3:x4	3.81E-01	8.56E-01	0.444	0.6567
x1:x5	1.88E-03	1.03E-03	1.817	0.0693

x2:x5	9.55E-03	4.56E-03	2.096	0.0361
x3:x5	2.01E-02	3.23E-02	0.622	0.5342
x4:x5	5.83E-02	3.70E-02	1.578	0.1146
x1:x2:x3	-1.98E-04	4.88E-04	-0.406	0.6846
x1:x2:x4	-7.11E-04	5.35E-04	-1.329	0.1839
x1:x3:x4	-1.36E-03	3.79E-03	-0.359	0.7195
x2:x3:x4	-9.25E-03	1.48E-02	-0.627	0.5307
x1:x2:x5	-3.90E-05	2.07E-05	-1.888	0.059
x1:x3:x5	-8.60E-05	1.43E-04	-0.599	0.5489
x2:x3:x5	-5.66E-04	5.95E-04	-0.951	0.3414
x1:x4:x5	-2.42E-04	1.67E-04	-1.451	0.1468
x2:x4:x5	-1.16E-03	6.96E-04	-1.667	0.0956
x3:x4:x5	-3.18E-03	4.74E-03	-0.67	0.5027
x1:x2:x3:x4	3.40E-05	6.48E-05	0.524	0.6001
x1:x2:x3:x5	2.38E-06	2.65E-06	0.899	0.3685
x1:x2:x4:x5	4.72E-06	3.12E-06	1.513	0.1302
x1:x3:x4:x5	1.31E-05	2.11E-05	0.621	0.5347
x2:x3:x4:x5	7.43E-05	8.21E-05	0.905	0.3653
x1:x2:x3:x4:x5	-3.04E-07	3.64E-07	-0.835	0.404 [b]

Table 14: Backwards variable selection for Poisson regression

Step	Variable	Df	Deviance	AIC	LRT	Pr(>Chi)
2	x1:x2:x3:x4:x5	1	123.63	659.16	0.69806	0.4034 [t]
3	x1:x2:x3:x5	1	123.76	657.28	0.12391	0.7248
4	x1:x3:x4:x5	1	124.55	656.07	0.79036	0.37399
5	x2:x3:x4:x5	1	124.84	654.36	0.29064	0.5898
6	x2:x3:x5	1	124.86	652.39	0.02177	0.88269
7	x3:x4:x5	1	125.27	650.79	0.40743	0.52328
8	x1:x3:x5	1	126.92	650.45	1.6578	0.19791
9	x3:x5	1	127.66	649.19	0.73691	0.39065
10	x1:x2:x3:x4	1	129.46	648.99	1.8	0.1797
11	x1:x3:x4	1	129.78	647.3	0.3164	0.57376
12	x2:x3:x4	1	131.03	646.56	1.2529	0.26299 [b]

6.2 Negative Binomial Regression

The coefficients for the initial model using negative binomial regression are listed in Table 16. This model has a null deviance of 200.85 on 129 degrees of freedom, a residual deviance of 122.93 on 98 degrees of freedom, and an AIC of 662.46. It also has a theta of 91041 and a standard error of 1233911, and we are warned that the iteration limit was reached while fitting theta. We continue fitting the model using backwards variable selection, and the results of this process are displayed in Table 17. The coefficients for the final negative binomial regression model are displayed in Table 18. This model has a residual deviance of 131.02 on 109 degrees of freedom, and an AIC of 648.56. It has a theta of 77291 and a standard error of 1208338, and once again we are warned that the iteration limit was reached.

Table 15: Coefficients for Poisson regression after backwards variable selection

	Estimate	Std. Error	z value	Pr(> $|z|$)
(Intercept)	4.49E+01	1.63E+01	2.763	0.00572 [t]
x1	-1.82E-01	7.18E-02	-2.532	0.01133
x2	-7.75E-01	2.99E-01	-2.595	0.00946
x3	6.89E-01	5.24E-01	1.315	0.18845
x4	-5.83E+00	2.32E+00	-2.507	0.01216
x5	-2.35E-01	8.67E-02	-2.708	0.00677
x1:x2	3.18E-03	1.31E-03	2.423	0.01539
x1:x3	-3.58E-03	2.37E-03	-1.509	0.13139
x2:x3	-1.26E-02	9.18E-03	-1.377	0.16836
x1:x4	2.51E-02	1.04E-02	2.413	0.01581
x2:x4	1.05E-01	4.22E-02	2.488	0.01284
x3:x4	4.36E-03	2.74E-03	1.592	0.11133
x1:x5	1.02E-03	3.90E-04	2.609	0.00907
x2:x5	4.31E-03	1.61E-03	2.674	0.00751
x4:x5	2.92E-02	1.25E-02	2.337	0.01946
x1:x2:x3	6.36E-05	4.22E-05	1.509	0.13125
x1:x2:x4	-4.47E-04	1.87E-04	-2.386	0.01704
x1:x2:x5	-1.80E-05	7.25E-06	-2.487	0.01288
x1:x4:x5	-1.28E-04	5.69E-05	-2.244	0.02482
x2:x4:x5	-5.25E-04	2.29E-04	-2.289	0.02209
x1:x2:x4:x5	2.25E-06	1.04E-06	2.162	0.03061 [b]

Table 16: Initial negative binomial regression model

	Estimate	Std. Error	z value	Pr(> $|z|$)
(Intercept)	6.98E+01	4.16E+01	1.68	0.0929
x1	-2.68E-01	1.89E-01	-1.42	0.1556
x2	-1.42E+00	8.03E-01	-1.766	0.0774
x3	-1.47E+00	6.16E+00	-0.239	0.8111
x4	-9.30E+00	6.49E+00	-1.431	0.1523
x5	-4.48E-01	2.27E-01	-1.98	0.0477
x1:x2	5.48E-03	3.62E-03	1.513	0.1303
x1:x3	4.28E-03	2.71E-02	0.158	0.8745
x2:x3	5.64E-02	1.11E-01	0.51	0.6103
x1:x4	3.70E-02	2.93E-02	1.263	0.2065
x2:x4	1.82E-01	1.20E-01	1.518	0.129
x3:x4	3.81E-01	8.56E-01	0.444	0.6567
x1:x5	1.88E-03	1.03E-03	1.817	0.0693
x2:x5	9.55E-03	4.56E-03	2.096	0.0361
x3:x5	2.01E-02	3.23E-02	0.622	0.5342
x4:x5	5.83E-02	3.70E-02	1.578	0.1147
x1:x2:x3	-1.98E-04	4.88E-04	-0.406	0.6846
x1:x2:x4	-7.11E-04	5.35E-04	-1.329	0.1839
x1:x3:x4	-1.36E-03	3.79E-03	-0.359	0.7195
x2:x3:x4	-9.25E-03	1.48E-02	-0.627	0.5307

x1:x2:x5	-3.90E-05	2.07E-05	-1.888	0.059
x1:x3:x5	-8.60E-05	1.43E-04	-0.599	0.5489
x2:x3:x5	-5.66E-04	5.95E-04	-0.951	0.3414
x1:x4:x5	-2.42E-04	1.67E-04	-1.451	0.1468
x2:x4:x5	-1.16E-03	6.96E-04	-1.666	0.0956
x3:x4:x5	-3.18E-03	4.74E-03	-0.67	0.5027
x1:x2:x3:x4	3.40E-05	6.48E-05	0.524	0.6001
x1:x2:x3:x5	2.38E-06	2.65E-06	0.899	0.3685
x1:x2:x4:x5	4.72E-06	3.12E-06	1.513	0.1302
x1:x3:x4:x5	1.31E-05	2.11E-05	0.621	0.5347
x2:x3:x4:x5	7.43E-05	8.21E-05	0.905	0.3653
x1:x2:x3:x4:x5	-3.04E-07	3.64E-07	-0.835	0.404

Table 17: Backwards variable selection, negative binomial model

Step	Variable	Df	Deviance	AIC	LRT	Pr(>Chi)
2	x1:x2:x3:x4:x5	1	123.62	659.16	0.69797	0.4035 [t]
3	x2:x3:x5:x1	1	123.75	657.28	0.12393	0.7248
4	x2:x3:x4:x5	1	124.24	655.77	0.48715	0.4852
5	x2:x3:x5	1	124.34	653.88	0.10505	0.74586
6	x3:x4:x5:x1	1	124.85	652.39	0.51051	0.47492
7	x3:x4:x5	1	125.26	650.79	0.40739	0.5233
8	x3:x5:x1	1	126.92	650.45	1.6576	0.19793
9	x3:x5	1	127.65	649.19	0.73683	0.39068
10	x2:x3:x4:x1	1	129.45	648.99	1.7999	0.1797
11	x3:x4:x1	1	129.77	647.31	0.3164	0.57377
12	x2:x3:x4	1	131.02	646.56	1.2529	0.26299 [b]

7 Question 5: Model Comparison

Both the Poisson regression and the negative binomial regression were extremely ill-suited to the college football win data, as we hypothesized. The Poisson regression had an AIC of 646.56, and the negative binomial regression had an AIC of 648.56. Both of these are extremely high, though the Poisson regression is slightly better. The other three types of regression performed in this project used slightly different data. The multinomial and ordinal logistic regressions grouped the number of wins into categories, and the AIC for those regressions were 262.66 and 282.37, respectively. The log linear regression considered all six variables as binary variables. The AIC is lowest in this regression (219.39) but we cannot compare that to any of the other regressions, because the models are not comparable and are not using the same data.

Table 18: Coefficients for negative binomial regression after backwards variable selection

	Estimate	Std. Error	z value	$\Pr(>\lvert z\rvert)$
(Intercept)	4.49E+01	1.63E+01	2.763	0.00573
x2	-7.75E-01	2.99E-01	-2.595	0.00946
x3	6.89E-01	5.24E-01	1.315	0.18847
x4	-5.83E+00	2.32E+00	-2.507	0.01216
x5	-2.35E-01	8.67E-02	-2.708	0.00677
x1	-1.82E-01	7.18E-02	-2.532	0.01134
x2:x3	-1.26E-02	9.18E-03	-1.377	0.16838
x2:x4	1.05E-01	4.22E-02	2.488	0.01284
x3:x4	4.36E-03	2.74E-03	1.592	0.11135
x2:x5	4.31E-03	1.61E-03	2.673	0.00751
x4:x5	2.92E-02	1.25E-02	2.337	0.01946
x3:x1	-3.58E-03	2.37E-03	-1.509	0.1314
x4:x1	2.51E-02	1.04E-02	2.413	0.01582
x5:x1	1.02E-03	3.90E-04	2.609	0.00907
x2:x1	3.18E-03	1.31E-03	2.423	0.01539
x2:x4:x5	-5.25E-04	2.29E-04	-2.289	0.02209
x4:x5:x1	-1.28E-04	5.69E-05	-2.244	0.02482
x2:x4:x1	-4.47E-04	1.87E-04	-2.386	0.01705
x2:x5:x1	-1.80E-05	7.25E-06	-2.487	0.01289
x2:x3:x1	6.36E-05	4.22E-05	1.509	0.13127
x2:x4:x5:x1	2.25E-06	1.04E-06	2.162	0.03061