“Survivor” (TV Show) Analysis: Who Earns the Jury Vote?

Brianna Beitling

```
## Warning: package 'SASmarkdown' was built under R version 3.4.3
```

I. Introduction

A. Study Design.

In 2000, “Survivor Fever” swept the nation as millions watched one of the first reality television shows unfold on their television screens. The commercials and advertisements preceding its inaugural episode led some to believe there might actually be some degree of brutality and danger to the situation- that perhaps even the contestants’ lives were at stake- a la “Battle Royale” or “The Hunger Games”. While that was, of course, inaccurate, the first season of Survivor did set itself up to become CBS’s pride and joy as the viewers learned the proper way to play the game along with the sixteen Americans dropped on the beach in Borneo. Strategy developed early as two tribes of eight faced off in challenges for reward and for immunity from the Tribal Council vote that would end one person’s shot at the million dollars and title of “sole survivor”.

Strategy, however, means little if a participant is able to reach the end of the show and not garner the favor of many of the contestants previously voted off, which is the criterion on which winning is hinged. This is best seen in the case of Russel Hantz, a contestant whose dirty tactics and lies propelled him to the end of the game twice in two consecutive seasons, only to face two bitter juries who refused to give him a majority of the votes (he earned two votes in season 19 and zero votes in season 20). This begs the question, then: what does it take to earn the jury’s votes at that final tribal council? Are the factors largely out of the contestant’s control, like their own age, career, etc., or does the way the contestant plays the game also have bearing?

This study will investigate this very question by examining the relationship between the percentage of the jury vote received (calculated as votes cast for a contestant, out of the total jury votes available per season) and various other variables. Variables expected to be related to jury votes received are: **age**, age of each contestant in years during time of filming; **votesagainst**, number of votes cast against the contestant over the course of the game (in attempt to vote the person out, not including final jury votes FOR the contestant to win); **indimmunities**, the number of individual immunities (safety from tribal council vote-offs) earned through excellent challenge performance; **tc_pct**, tribal council percent, calculated as the number of “correct” votes made by the contestant (votes for the person who ultimately gets booted at that tribal council) out of the number of tribal councils attended by that contestant; **tc_score**, tribal council score, calculated as the number of “correct” votes made by the contestant out of the number of tribal councils total in the season; **idols**, the number of hidden immunity idols (whose purpose is also to immunize a contestant from votes, if played) found by the contestant (which may represent idols found

alone or as a group within an alliance); **appstatus**, whether a contestant was an applicant to be on the show (1), a recruit (0), or recruited for applying to another show (0.5) (note: if the contestant had previously been on a season, they were treated as a recruit in each subsequent appearance since CBS recruited them to return); **collar**, whether the person has a white collar job (1), blue collar job (3), or a job that does not fit either category (referred to by the show itself as "no collar") (2); **season**, the season number in which the contestant participated– this is not expected to have an effect, but is included initially as a main effect for the following interaction term; season_TCPct, the interaction term between season and tribal council percent, as explained previously. This interaction term is included because of a general trend observed from the initial seasons, where the jury tends to be more bitter and thus not give their final jury votes to the contestants primarily responsible for orchestrating their vote-offs, toward rewarding correct gameplay and stragegy in later seasons, tending to give jury votes more frequently to players who did shoulder that responsibility.

Data was partially compiled from Jon Krause, "Survivor" analyst who has analyzed general trends in demographic data for multiple reality shows and who provided his data freely in a dropbox linked to from Rob Has a Podcast (http://robhasawebsite.com/rhappy-hour-survivor-big-brother-by-the-numbers/). The rest of the data was taken from various pages of the "Survivor Wiki" (http://survivor.wikia.com/wiki/Survivor:_Borneo through http://survivor.wikia.com/wiki/Survivor:_Millennials_vs._Gen_X) and appended onto the previous dataset. "Survivor" has aired 35 seasons to date (the 36th season is currently airing); this dataset includes contestant data from seasons 1-34.

B. Aims.

The purpose of the study is to **investigate the relationship** between percent of the jury vote received by finalists on the television show "Survivor" and various factors related to contestant demographics and gameplay.

C. Statistical Model.

A multiple linear regression model is considered. Let

$Y_i =$ the percent of the jury vote earned by the i^{th} finalist,

$X_{i1} =$ age in years of the i^{th} contestant,

$X_{i2} =$ season competed in by the i^{th} contestant,

$X_{i3} =$ total votes in the season cast against the i^{th} contestant,

$X_{i4} =$ individual immunities earned through immunity challenges by the i^{th} contestant,

$X_{i5} =$ tribal council percent, calculated as number of votes cast by the i^{th} contestant for the person who is subsequently booted out at that tribal council (not including votes at final 3) divided by the number of tribal councils participated in by the i^{th} contestant,

X_{i6} = tribal council score, calculated as number of votes cast by the i^{th} contestant for the person who is subsequently booted out at that tribal council (not including votes at final 3) divided by the number of tribal councils in the season (not including final 3 or final tribal councils),

X_{i7} = idols found by the i^{th} contestant (not necessarily equivalent to idols played by that contestant, and idols may be found alone or as part of a group/alliance),

X_{i8} = whether the i^{th} contestant was an applicant (1), or recruit (0) or hybrid (0) (applicant to another show, subsequently recruited to participate in survivor),

X_{i9} = whether the i^{th} contestant was a recruit (1), applicant (0), or hybrid (0) (applicant to another show, subsequently recruited to participate in survivor),

X_{i10} = "collar" job of the i^{th} contestant, where 1 = white collar, and 0 = "no" collar or blue collar,

X_{i11} = "collar" job of the i^{th} contestant, where 1 = "no" collar, and 0 = white collar or blue collar,

X_{i12} = interaction term between the tribal council percentage for the i^{th} contestant and the season on which the contestant participated.

The **initial model** is given by

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \beta_6 X_{i6} + \beta_7 X_{i7} + \beta_8 X_{i8} + \beta_9 X_{i9} + \beta_{10} X_{i10} + \beta_{11} X_{i11} + \beta_{12} X_{i12} + \varepsilon_i$$

where $\varepsilon_i \sim iidN(0, \sigma^2)$, $i = 1,2,\ldots,83$, and $\beta_0, \beta_1, \ldots, \beta_{12}$, and σ^2 are the unknown model parameters.

Eighty-four contestants in seasons 1 through 34 have participated in the final tribal council as a candidate for the million dollars. Of those eighty-four, eighty-three are included in all analysis that includes the variable "appstatus", due to missing data (it is unknown whether Missy Payne in season 29 was an applicant or recruit).

Note that applicant status and collar are categorical variables each with 3 levels, so two parameters are added to the model for each. PROC GLM automatically treats one of the options (hybrid for applicant status and blue collar for collar) as the default, reflected in the model as a value of 0 for the two associated variable levels (applicant and recruit; and white collar and no collar, respectively). Thus, while in the data, appstatus of 0 = recruit, 0.5 = hybrid, and 1 = applicant, in the equation, recruits will have a 0 for X_{i8} and 1 for X_{i9}, applicants will have a 1 for X_{i8} and 0 for X_{i9}, and hybrids will have 0s for both values. Similarly, in the data collar of 1 = white collar, 2 = "no" collar, 3 = blue collar, but in the equation, white collar corresponds to a 1 for X_{i10} and 0 for X_{i11}, "no" collar corresponds to a 0 for X_{i10} and 1 for X_{i11}, and blue collar corresponds to a 0 for both values.

```
proc glm data = finaltribal;
    class Appstatus Collar;
```

```
    model Pct_Jury_Vote = Age Season VotesAgainst IndImmunities TC_Pct
TC_Score Idols AppStatus Collar Season*TC_Pct /solution;
run;
quit;
```

The GLM Procedure

Class Level Information

Class	Levels	Values
AppStatus	3	0 1 0.5
Collar	3	1 2 3

Number of Observations Read	84
Number of Observations Used	83

The GLM Procedure

Dependent Variable: Pct_Jury_Vote

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	12	32871.50632	2739.29219	3.45	0.0005
Error	70	55592.89834	794.18426		
Corrected Total	82	88464.40466			

R-Square	Coeff Var	Root MSE	Pct_Jury_Vote Mean
0.371579	71.14968	28.18128	39.60843

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Age	1	2726.14558	2726.14558	3.43	0.0681
Season	1	2156.44745	2156.44745	2.72	0.1039
VotesAgainst	1	1918.29111	1918.29111	2.42	0.1247
IndImmunities	1	3411.47413	3411.47413	4.30	0.0419
TC_Pct	1	1427.85152	1427.85152	1.80	0.1843
TC_Score	1	3.34802	3.34802	0.00	0.9484
Idols	1	12658.46791	12658.46791	15.94	0.0002
AppStatus	2	5401.21424	2700.60712	3.40	0.0390

```
Collar                     2    3168.23255    1584.11628     1.99  0.1437
Season*TC_Pct              1       0.03382       0.03382     0.00  0.9948


Source                    DF   Type III SS   Mean Square  F Value  Pr > F

Age                        1    5131.31739    5131.31739     6.46  0.0132
Season                     1      96.72606      96.72606     0.12  0.7281
VotesAgainst               1     991.09813     991.09813     1.25  0.2678
IndImmunities              1    1174.32708    1174.32708     1.48  0.2281
TC_Pct                     1     739.34646     739.34646     0.93  0.3379
TC_Score                   1     191.81175     191.81175     0.24  0.6246
Idols                      1   11687.51300   11687.51300    14.72  0.0003
AppStatus                  2    4062.75482    2031.37741     2.56  0.0847
Collar                     2    3164.06083    1582.03041     1.99  0.1441
Season*TC_Pct              1       0.03382       0.03382     0.00  0.9948

                                                 Standard
Parameter                  Estimate                 Error    t Value    Pr > |t|

Intercept              161.1781423 B     71.84070346       2.24      0.0280
Age                     -0.9167047       0.36064145      -2.54      0.0132
Season                  -0.9526839       2.72984314      -0.35      0.7281
VotesAgainst            -1.1220611       1.00442800      -1.12      0.2678
IndImmunities            2.9935403       2.46179114       1.22      0.2281
TC_Pct                  -0.7054044       0.73109670      -0.96      0.3379
TC_Score                -0.1718784       0.34973917      -0.49      0.6246
Idols                   19.0113373       4.95578266       3.84      0.0003
AppStatus     0         -5.5237532 B     18.61130917      -0.30      0.7675
AppStatus     1         10.1797503 B     18.69371927       0.54      0.5878
AppStatus     0.5        0.0000000 B       .                .          .
Collar        1        -13.7111791 B      8.93957656      -1.53      0.1296
Collar        2        -18.6937859 B      9.45163180      -1.98      0.0519
Collar        3          0.0000000 B       .                .          .
Season*TC_Pct           -0.0002065       0.03164422      -0.01      0.9948

NOTE: The X'X matrix has been found to be singular, and a generalized
      inverse was used to solve the normal equations.  Terms whose
      estimates are followed by the letter 'B' are not uniquely estimable.
```

II. Preliminary Analyses.

A. Bivariate Associations.

A scatterplot matrix indicates generally linear associations between the predictor variables and percent of jury vote received. Of course, the categorical variables do not look very linear in scatterplot form. Season and the interaction term utilizing season look less linear, but that is why an interaction term was created.

```
proc sgscatter data = finaltribal;
    matrix Pct_Jury_Vote Age Season VotesAgainst IndImmunities TC_Pct
TC_Score Idols AppStatus Collar season_TCPct;
run;
```

The Pearson correlation coeffecients for all pairwise associations are shown below. Most of the variables have a weak to moderate correlation with percent of jury vote received. It is evident already that some variables will likely not be utilized in the final model, but they will be used in variable selection regardless.

```
proc corr data = finaltribal;
    var Pct_Jury_Vote Age Season VotesAgainst IndImmunities TC_Pct TC_Score
Idols AppStatus Collar season_TCPct;
run;
```

The CORR Procedure

11 Variables: Pct_Jury_Vote Age Season VotesAgainst IndImmunities TC_Pct TC_Score Idols AppStatus Collar Season_TCPct

Simple Statistics

Variable	N	Mean	Std Dev	Sum
Pct_Jury_Vote	84	39.28571	32.78086	3300
Age	84	32.96429	9.40577	2769
Season	84	18.36905	9.33291	1543
VotesAgainst	84	3.83333	3.36113	322.00000
IndImmunities	84	1.54762	1.39193	130.00000
TC_Pct	84	85.51275	12.56329	7183
TC_Score	84	60.97707	13.01249	5122
Idols	84	0.42857	0.71618	36.00000
AppStatus	83	0.45181	0.49148	37.50000
Collar	84	1.69048	0.76009	142.00000
Season_TCPct	84	1519	731.47431	127633

Simple Statistics

Variable	Minimum	Maximum
Pct_Jury_Vote	0	100.00000
Age	19.00000	57.00000
Season	1.00000	33.00000

VotesAgainst	0	17.00000
IndImmunities	0	5.00000
TC_Pct	50.00000	100.00000
TC_Score	33.33333	93.33333
Idols	0	3.00000
AppStatus	0	1.00000
Collar	1.00000	3.00000
Season_TCPct	90.00000	2844

Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

	Pct_ Jury_ Vote	Age	Season	Votes Against	Ind Immunities	TC_Pct
Pct_Jury_Vote	1.00000	-0.18731	-0.17792	-0.20102	0.24650	-0.04065
		0.0880	0.1054	0.0667	0.0238	0.7136
	84	84	84	84	84	84
Age	-0.18731	1.00000	0.09252	0.16635	-0.12088	-0.21247
	0.0880		0.4026	0.1304	0.2734	0.0523
	84	84	84	84	84	84
Season	-0.17792	0.09252	1.00000	0.21284	-0.12147	-0.44318
	0.1054	0.4026		0.0519	0.2710	<.0001
	84	84	84	84	84	84
VotesAgainst	-0.20102	0.16635	0.21284	1.00000	-0.10387	-0.01539
	0.0667	0.1304	0.0519		0.3471	0.8895
	84	84	84	84	84	84
IndImmunities	0.24650	-0.12088	-0.12147	-0.10387	1.00000	-0.08154
	0.0238	0.2734	0.2710	0.3471		0.4609
	84	84	84	84	84	84
TC_Pct	-0.04065	-0.21247	-0.44318	-0.01539	-0.08154	1.00000
	0.7136	0.0523	<.0001	0.8895	0.4609	
	84	84	84	84	84	84
TC_Score	-0.04952	-0.20950	-0.23207	0.15040	-0.03488	0.65862
	0.6546	0.0558	0.0337	0.1721	0.7528	<.0001
	84	84	84	84	84	84
Idols	0.24031	0.10246	0.28068	0.06006	-0.06906	0.12704
	0.0277	0.3537	0.0097	0.5873	0.5325	0.2495
	84	84	84	84	84	84

AppStatus	0.29109	0.07891	-0.29311	-0.01194	0.14756	0.04695
	0.0076	0.4783	0.0072	0.9147	0.1831	0.6734
	83	83	83	83	83	83
Collar	0.18622	-0.04875	-0.08051	-0.07703	0.25324	0.09200
	0.0899	0.6597	0.4666	0.4862	0.0201	0.4052
	84	84	84	84	84	84
Season_TCPct	-0.19632	0.01965	0.93943	0.23197	-0.13800	-0.13609
	0.0735	0.8592	<.0001	0.0337	0.2106	0.2171
	84	84	84	84	84	84

Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

	TC_Score	Idols	App Status	Collar	Season_ TCPct
Pct_Jury_Vote	-0.04952	0.24031	0.29109	0.18622	-0.19632
	0.6546	0.0277	0.0076	0.0899	0.0735
	84	84	83	84	84
Age	-0.20950	0.10246	0.07891	-0.04875	0.01965
	0.0558	0.3537	0.4783	0.6597	0.8592
	84	84	83	84	84
Season	-0.23207	0.28068	-0.29311	-0.08051	0.93943
	0.0337	0.0097	0.0072	0.4666	<.0001
	84	84	83	84	84
VotesAgainst	0.15040	0.06006	-0.01194	-0.07703	0.23197
	0.1721	0.5873	0.9147	0.4862	0.0337
	84	84	83	84	84
IndImmunities	-0.03488	-0.06906	0.14756	0.25324	-0.13800
	0.7528	0.5325	0.1831	0.0201	0.2106
	84	84	83	84	84
TC_Pct	0.65862	0.12704	0.04695	0.09200	-0.13609
	<.0001	0.2495	0.6734	0.4052	0.2171
	84	84	83	84	84
TC_Score	1.00000	0.10561	0.14423	-0.04321	-0.02152
		0.3390	0.1933	0.6963	0.8459
	84	84	83	84	84
Idols	0.10561	1.00000	-0.07817	0.00316	0.36819
	0.3390		0.4824	0.9772	0.0006

	84	84	83	84	84
AppStatus	0.14423	-0.07817	1.00000	0.07485	-0.30570
	0.1933	0.4824		0.5012	0.0049
	83	83	83	83	83
Collar	-0.04321	0.00316	0.07485	1.00000	-0.05760
	0.6963	0.9772	0.5012		0.6028
	84	84	83	84	84
Season_TCPct	-0.02152	0.36819	-0.30570	-0.05760	1.00000
	0.8459	0.0006	0.0049	0.6028	
	84	84	83	84	84

B. Screening of Covariates and Verification of Assumptions

Based on automatic variable selection methods in combination with criterion-based statistics, the following variables were removed from the model: season, votesagainst, indimmunities, and tc_score. Residual plots, statistics such as Cook's D, and values of the residuals themselves confirmed no cause for concern with regard to outliers and points of high influence. Additionally, the remainder of the assumptions of the linear model (normality of resdiuals, homoscedasticity, linearity, and independence) were confirmed to apparently hold. Multicollinearity might otherwise pose minor issues in this model, but variables amongst variables most highly collinear with one another (such as tc_pct and tc_score), at least one was removed. The remaining variables only appear to have weak relationships at best.

C. Final Model

The **final model** is given by

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \beta_6 X_{i6} + \beta_7 X_{i7} + \beta_8 X_{i8} + \varepsilon_i$$

where $\varepsilon_i \sim iidN(0, \sigma^2)$, $i = 1,2,\ldots,83$, and $\beta_0, \beta_1, \ldots, \beta_8$, and σ^2 are the unknown model parameters.

III. Statistical Analysis.

Relevant output for the fitted model, including ANOVA table, parameter estimates, and 95% confidence intervals for the parameters, is displayed below. The F value and associate p-value (4.77 and <.0001 respectively) confirm that at least one of the predictors is significantly related to percent of jury vote received. The $R^2 = .340304$, meaning that approximately 34% of the variation in percent of jury vote received by finalists can be explained by the collective effect of the predictors in the model. While this is not a relatively high R^2, when taking the amount of luck and variety associated with "Survivor" into account (including the fact that a person can win one season and do very poorly in another), it is easy to see that it would be impossible to account for all, or even a majority,

of variation when creating a model for a reality show based upon people, relationship, and psychology.

```
proc glm data = finaltribal;
    class Appstatus Collar;
    model Pct_Jury_Vote = Age TC_Pct Idols AppStatus Collar season*TC_Pct
/solution clparm alpha = .05;
run;
quit;
```

The GLM Procedure

Class Level Information

Class	Levels	Values
AppStatus	3	0 1 0.5
Collar	3	1 2 3

Number of Observations Read	84
Number of Observations Used	83

The GLM Procedure

Dependent Variable: Pct_Jury_Vote

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	30104.76747	3763.09593	4.77	<.0001
Error	74	58359.63719	788.64375		
Corrected Total	82	88464.40466			

R-Square	Coeff Var	Root MSE	Pct_Jury_Vote Mean
0.340304	70.90107	28.08280	39.60843

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Age	1	2726.14558	2726.14558	3.46	0.0670
TC_Pct	1	777.97861	777.97861	0.99	0.3238
Idols	1	6631.89645	6631.89645	8.41	0.0049

AppStatus	2	11022.66238	5511.33119	6.99	0.0017
Collar	2	3052.33292	1526.16646	1.94	0.1516
TC_Pct*Season	1	5893.75152	5893.75152	7.47	0.0078

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Age	1	6688.54031	6688.54031	8.48	0.0047
TC_Pct	1	4708.13109	4708.13109	5.97	0.0169
Idols	1	13193.78857	13193.78857	16.73	0.0001
AppStatus	2	4644.78463	2322.39231	2.94	0.0588
Collar	2	3342.45026	1671.22513	2.12	0.1274
TC_Pct*Season	1	5893.75152	5893.75152	7.47	0.0078

Parameter		Estimate		Standard Error	t Value	Pr > |t|
Intercept		144.7476841	B	37.14161864	3.90	0.0002
Age		-1.0105315		0.34699610	-2.91	0.0047
TC_Pct		-0.6562001		0.26856686	-2.44	0.0169
Idols		19.9941753		4.88831482	4.09	0.0001
AppStatus	0	2.8460366	B	17.35227484	0.16	0.8702
AppStatus	1	18.8746603	B	17.72147454	1.07	0.2903
AppStatus	0.5	0.0000000	B	.	.	.
Collar	1	-16.3696105	B	8.67333234	-1.89	0.0630
Collar	2	-17.8531981	B	9.36326466	-1.91	0.0604
Collar	3	0.0000000	B	.	.	.
TC_Pct*Season		-0.0133077		0.00486797	-2.73	0.0078

Parameter		95% Confidence Limits	
Intercept		70.7413935	218.7539747
Age		-1.7019363	-0.3191266
TC_Pct		-1.1913312	-0.1210689
Idols		10.2539954	29.7343552
AppStatus	0	-31.7291257	37.4211990
AppStatus	1	-16.4361485	54.1854691
AppStatus	0.5	.	.
Collar	1	-33.6516024	0.9123815
Collar	2	-36.5099101	0.8035140
Collar	3	.	.
TC_Pct*Season		-0.0230074	-0.0036081

NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.

A. Utilizing the Model for Prediction

To test the utility of the model in predicting voting outcomes for "Survivor" finalists, the three finalists from the previous season (not included in the full dataset) were tested. Outcomes were compared to actual values. As evidenced, below, the model for this season underestimated the percent of jury vote received for each of the three data points. It was predicted that Troyzan would receive -5.46% of the vote (technically impossible); he actually received 0%. Brad was predicted to receive 3.10% of the jury vote; he actually received 30%. Last, Sarah was predicted to receive 14.47% of the jury vote, but she actually received 70%. Other measures could possibly be taken in "better" fitting models to account for the fact that percentages within the same season should have summed to 100.00% and to eliminate any bias retained in the final model presented. However, while the values predicted were less than the actual values, the predicted values would have correctly pointed to the winner, as predicted values ascended accordingly with the actual values.

```
proc import out= survpred
            datafile= "C:\Users\Brianna\Documents\Stat 840\Prediction
Dataset.csv"
            dbms=CSV REPLACE;
     getnames=yes;
     datarow=2;
run;

data survpredvals;
    set survpred;
    predictval = 144.7476841 - 1.0105315*Age -0.6562001*TC_Pct +
19.9941753*Idols + 2.8460366*Recruit + 18.8746603*Applicant -
16.3696105*WhiteCollar -17.8531981*NoCollar-0.0133077*(TC_Pct*Season);
    predvsactual = predictval - Pct_Jury_Vote;
run;

proc print data = survpredvals;
    var Name Age Season TC_Pct Idols Pct_Jury_Vote AppStatus Collar
predictval predvsactual;
run;
```

Obs	Name	Age	Season	TC_Pct	Idols
1	Troyzan Roberston	54	34	72.72727273	0
2	Brad Culpepper	47	34	72.72727273	0
3	Sarah Lacina	32	34	90.90909091	0

Obs	Pct_Jury_Vote	AppStatus	Collar	predictval	predvsactual
1	0	0	2	-5.4581	-5.4581
2	30	0	1	3.0992	-26.9008
3	70	0	3	14.4693	-55.5307

IV. Summary of Findings.

1. There appears to be a negative relationship between age and percent of the jury vote received. Older contestants receive fewer of the votes available overall.

2. There is a negative relationship between percent of "correct" votes at tribal councils (that is, votes for the contestant that ultimately gets booted) and percent of jury votes received. This is consistent with the idea of a "bitter jury" that permeated most of the early-middle seasons of "Survivor", where the jury did not reward finalists responsible for their own removal from the game, as a whole. This has changed some in recent years, which is why the interaction term (described in (6)) was included.

3. There is a positive relationship between idols found and percent of jury vote received. Hidden immunity idols were not in the early seasons of "Survivor"; rather, they were first introduced in the middle seasons of the game. Contestants who found idols rarely had them fall into their laps; they were generally playing very aggressively, seeking out idols, and in many cases finding them even before a clue to their location was given. There are likely many reasons, then, why finding an idol is positively correlated with percent of the jury vote received, not all of which are directly described by the actual variable used here.

4. With regard to whether a contestant is an applicant or a recruit, it appears that applicants are more likely to receive jury votes than recruits. Applicants are more likely familiar with the game, whereas recruits are frequently cast by "Survivor" producers because they will make good characters. Perhaps, then, this is a reflection of friendships made with the jury by applicants, their general capability for strategy in the game, or their perception as a more genuine player.

5. Contestants with blue collar jobs are more likely to win than finalists in white collar or "no" collar jobs. Perhaps these contestants are seen as more relatable by the majority of the jury, or perhaps another reason entirely accounts for this relationship.

6. There was generally a negative association between percent of jury votes received and the interaction term between tribal council percent and season. Looking at values in the dataset, it is apparent that the interaction is indeed significant in the hypothesized way– that is, that finalists in earlier seasons receive less of the jury votes when their tribal council percentage is higher, whereas contestants in later seasons receive more of the jury votes (relative to previous seasons, at least) when their tribal council percentage is higher. This corresponds to a shifting attitude toward rewarding strategy and dominant play.

Clearly, a lot of variation still exists in this model. It is difficult to account for every variable that is responsible for the percentage of the jury vote received by jury members, especially since jury members themselves have a large degree of variation, and frequently reward players for different reasons in different seasons. Seasons 19 and 20 saw juries so bitter at their treatment in the game by Russell Hantz that he received very few votes despite dominating strategically and finding many hidden immunity idols. On the other hand, Kim Spradlin dominated season 24 and was responsible for most of the jury members' oustings,

but still easily won in the final tribal council 7-2. In season 15, Amanda Kimmel played a game easily on par with winner, Todd Herzog, but her botched arguments and inability to take responsibility for her actions at the final tribal annoyed jury members and potentially cost her the game. And finally, there are seasons like season 33 where revelations at the final tribal council reward sympathy points to the ultimate winner (Adam Klein revealed that his mother was dying) resulting in a 10-0-0 vote which would likely not have been quite so unanimous otherwise (runner-up Ken McNickle played a game which in any other season likely would have rewarded him with at least a portion of the jury vote).

Overall, it is difficult to predict something as fickle as a "Survivor" jury's whims. However, this model does explain approximately 34% of the variation in jury vote, and given the nature of the game, that does offer some degree of value (such as the ability of the model to correctly predict the order of placement via predicted percentages received of the finalists in season 34).

V. Appendix

A. Diagnostics for Predictors

The purpose of this section is to check predictors for outlier values. Boxplots created show some skewness, but overall no major problems with outliers.

```
data finaltribal;
set finaltribal;
group = 1;
run;

proc boxplot data = finaltribal;
plot Pct_Jury_Vote*group;
plot Age*group;
plot TC_Pct*group;
plot Idols*group;
plot AppStatus*group;
plot Collar*group;
plot season_TCPct*group;
run;
```

B. Screening of Predictors

1. **Added variable plots** for each predictor are included below. The categorical variables (applicant and collar) do not look impressively linear, but that is unsurprising. The others, while not strongly linear, do show weak to moderate linearity. Overall, the plots indicate no need for transformations, and that the predictors do have value in the final model.

```
proc reg data = finaltribal;
model Pct_Jury_Vote = Age TC_Pct Idols Applicant Recruit BlueCollar
WhiteCollar season_TCPct /partial;
run;
quit;
```

Model: MODEL1
Dependent Variable: Pct_Jury_Vote
Number of Observations Read
84
Number of Observations Used

Analysis of Variance

Source	DF	Sum ofSquares	MeanSquare	F Value	Pr > F
Model	8	30828	3853.48909	4.95	<.0001
Error	75	58363	778.16814		
Corrected Total	83	89191			

Root MSE	R-Square	Dependent Mean	Adj R-Sq	Coeff Var
27.89567	0.3456	39.28571	0.2758	71.00715

Parameter Estimates

Variable	DF	ParameterEstimate	StandardError	t Value

Pr > |t|
Intercept
1
126.23709
32.82112
3.85
0.0002
Age
1
-1.01205
0.34381
-2.94
0.0043
TC_Pct
1
-0.65385
0.26405
-2.48
0.0155
Idols
1
20.04346
4.78983
4.18
<.0001
Applicant
1
19.39790
15.43470
1.26
0.2127
Recruit
1
3.36282
15.07355
0.22
0.8241
BlueCollar
1
17.81499

9.28031
1.92
0.0587
WhiteCollar
1
1.44806
6.96140
0.21
0.8358
Season_TCPct
1
-0.01331
0.00483
-2.75
0.0074

Model: MODEL1
Dependent Variable: Pct_Jury_Vote

Model: MODEL1
Partial Regression Residual Plot

2. A scatterplot matrix and correlation matrix were created for the final predictors to assess for multicollinearity. Multicollinearity, or a high degree of association amongst

predictor variables, can create a model that is not stable. These matrices, displayed below, show a moderate degree of association between the tribal council percent x season interaction term, and both idols and applicant status. All other variables have weak associations at most. The degree of potentiall muticollinearity presented by these findings does not pose great cause for concern.

```
proc sgscatter data = finaltribal;
    matrix Pct_Jury_Vote Age TC_Pct Idols AppStatus Collar season_TCPct;
run;
```

```
proc corr data = finaltribal;
    var Pct_Jury_Vote Age TC_Pct Idols AppStatus Collar season_TCPct;
run;
```

The CORR Procedure

7 Variables: Pct_Jury_Vote Age TC_Pct Idols AppStatus Collar Season_TCPct

Simple Statistics

Variable	N	Mean	Std Dev	Sum
Pct_Jury_Vote	84	39.28571	32.78086	3300
Age	84	32.96429	9.40577	2769
TC_Pct	84	85.51275	12.56329	7183
Idols	84	0.42857	0.71618	36.00000
AppStatus	83	0.45181	0.49148	37.50000
Collar	84	1.69048	0.76009	142.00000
Season_TCPct	84	1519	731.47431	127633

Simple Statistics

Variable	Minimum	Maximum
Pct_Jury_Vote	0	100.00000
Age	19.00000	57.00000
TC_Pct	50.00000	100.00000
Idols	0	3.00000
AppStatus	0	1.00000
Collar	1.00000	3.00000
Season_TCPct	90.00000	2844

Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0

Number of Observations

	Pct_ Jury_ Vote	Age	TC_Pct	Idols
Pct_Jury_Vote	1.00000	-0.18731	-0.04065	0.24031
		0.0880	0.7136	0.0277
	84	84	84	84
Age	-0.18731	1.00000	-0.21247	0.10246
	0.0880		0.0523	0.3537
	84	84	84	84
TC_Pct	-0.04065	-0.21247	1.00000	0.12704
	0.7136	0.0523		0.2495
	84	84	84	84
Idols	0.24031	0.10246	0.12704	1.00000
	0.0277	0.3537	0.2495	
	84	84	84	84
AppStatus	0.29109	0.07891	0.04695	-0.07817
	0.0076	0.4783	0.6734	0.4824
	83	83	83	83
Collar	0.18622	-0.04875	0.09200	0.00316
	0.0899	0.6597	0.4052	0.9772
	84	84	84	84
Season_TCPct	-0.19632	0.01965	-0.13609	0.36819
	0.0735	0.8592	0.2171	0.0006
	84	84	84	84

Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

	App Status	Collar	Season_ TCPct
Pct_Jury_Vote	0.29109	0.18622	-0.19632
	0.0076	0.0899	0.0735
	83	84	84
Age	0.07891	-0.04875	0.01965
	0.4783	0.6597	0.8592
	83	84	84

TC_Pct	0.04695	0.09200	-0.13609
	0.6734	0.4052	0.2171
	83	84	84
Idols	-0.07817	0.00316	0.36819
	0.4824	0.9772	0.0006
	83	84	84
AppStatus	1.00000	0.07485	-0.30570
		0.5012	0.0049
	83	83	83
Collar	0.07485	1.00000	-0.05760
	0.5012		0.6028
	83	84	84
Season_TCPct	-0.30570	-0.05760	1.00000
	0.0049	0.6028	
	83	84	84

3. Automatic variable selection was used to determine which model was best utilized as the final model. Many criteria were examined, and the model which seemed best according to the most criteria was used. The model which minimized AIC, minimized Mallow's Cp, and maximized R^2adj corresponds to the final model which was selected.

```
ods graphics on;
proc glmselect data = finaltribal plot = CriterionPanel;
    class AppStatus Collar;
    model Pct_Jury_Vote = Age Season VotesAgainst IndImmunities TC_Pct
TC_Score Idols AppStatus Collar season*TC_Pct
    /selection = stepwise(select = cp) stats = all;
run;
quit;
ods graphics off;
```

Data Set
WORK.FINALTRIBAL
Dependent Variable
Pct_Jury_Vote
Selection Method
Stepwise
Select Criterion
C(p)
Stop Criterion
C(p)
Effect Hierarchy Enforced

None

Number of Observations Read
84
Number of Observations Used
83

Class Level Information
Class
Levels
Values
AppStatus
3
0 1 0.5
Collar
3
1 2 3

Dimensions
Number of Effects
11
Number of Parameters
15

Stepwise Selection Summary
Step
EffectEntered
EffectRemoved
NumberEffects In
NumberParms In
ModelR-Square
Adjusted R-Square
AIC
AICC
BIC
CP
SBC
PRESS

ASE
F Value
Pr > F
0
Intercept
1
1
0.0000
0.0000
665.6357
665.7857
581.9961
30.3903
583.0546
90635.2296
1065.8362
0.00
1.0000
1
AppStatus
2
3
0.0866
0.0637
662.1208
662.6337
577.9474
24.7480
584.3774
88804.8116
973.5742
3.79
0.0267
2
Idols
3
4
0.1641
0.1324
656.7581

657.5373
572.8658
18.1109
581.4335
88682.3586
890.9302
7.33
0.0083
3
Age
4
5
0.2148
0.1745
653.5684
654.6737
570.0525
14.4673
580.6626
84063.7427
836.9296
5.03
0.0277
4
Season*TC_Pct
5
6
0.2625
0.2146
650.3673
651.8606
567.4906
11.1544
579.8803
82894.5767
786.0933
4.98
0.0286
5
TC_Pct

6
7
0.3025
0.2475
647.7323
649.6782*
565.6794*
8.6924
579.6642*
82825.1242
743.3986
4.36
0.0400
6
Collar
7
9
0.3403
0.2690*
647.1098*
650.1653
566.4072
8.4837*
583.8793
82610.2980*
703.1282
2.12
0.1274

- Optimal Value of Criterion

  Selection stopped at a local minimum of the C(p) criterion.

  Stop Details
  CandidateFor
  Effect
  CandidateC(p)
  CompareC(p)
  Entry
  IndImmunities
  8.8294

>
8.4837
Removal
Collar
8.6924
>
8.4837

Selected Model

The selected model is the model at the last step (Step 6).
Effects:
Intercept Age TC_Pct Idols AppStatus Collar Season*TC_Pct

Analysis of Variance
Source
DF
Sum ofSquares
MeanSquare
F Value
Model
8
30105
3763.09593
4.77
Error
74
58360
788.64375
Corrected Total
82
88464

Root MSE
28.08280
Dependent Mean

39.60843
R-Square
0.3403
Adj R-Sq
0.2690
AIC
647.10975
AICC
650.16531
BIC
566.40724
C(p)
8.48375
PRESS
82610
SBC
583.87932
ASE
703.12816

Parameter Estimates
Parameter
DF
Estimate
StandardError
t Value
Intercept
1
144.747684
37.141619
3.90
Age
1
-1.010531
0.346996
-2.91
TC_Pct
1
-0.656200
0.268567

-2.44
Idols
1
19.994175
4.888315
4.09
AppStatus 0
1
2.846037
17.352275
0.16
AppStatus 1
1
18.874660
17.721475
1.07
AppStatus 0.5
0
0
.
.
Collar 1
1
-16.369610
8.673332
-1.89
Collar 2
1
-17.853198
9.363265
-1.91
Collar 3
0
0
.
.
Season*TC_Pct
1
-0.013308
0.004868

-2.73

4. No variables appeared to have concerns with multicollinearity above; this is confirmed by finding VIF values, none of which exceed the rule of thumb value of 10.

5. The final model fitted model is displayed below, assuming no violations of model assumptions are found.

```
proc glm data = finaltribal;
    class Appstatus Collar;
    model Pct_Jury_Vote = Age TC_Pct Idols AppStatus Collar season*TC_Pct
/solution;
run;
quit;
```

The GLM Procedure

Class Level Information

Class	Levels	Values
AppStatus	3	0 1 0.5
Collar	3	1 2 3

Number of Observations Read	84
Number of Observations Used	83

The GLM Procedure

Dependent Variable: Pct_Jury_Vote

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	30104.76747	3763.09593	4.77	<.0001
Error	74	58359.63719	788.64375		
Corrected Total	82	88464.40466			

R-Square	Coeff Var	Root MSE	Pct_Jury_Vote Mean
0.340304	70.90107	28.08280	39.60843

```
Source                     DF     Type I SS   Mean Square  F Value  Pr > F

Age                         1    2726.14558    2726.14558     3.46  0.0670
TC_Pct                      1     777.97861     777.97861     0.99  0.3238
Idols                       1    6631.89645    6631.89645     8.41  0.0049
AppStatus                   2   11022.66238    5511.33119     6.99  0.0017
Collar                      2    3052.33292    1526.16646     1.94  0.1516
TC_Pct*Season               1    5893.75152    5893.75152     7.47  0.0078


Source                     DF   Type III SS   Mean Square  F Value  Pr > F

Age                         1    6688.54031    6688.54031     8.48  0.0047
TC_Pct                      1    4708.13109    4708.13109     5.97  0.0169
Idols                       1   13193.78857   13193.78857    16.73  0.0001
AppStatus                   2    4644.78463    2322.39231     2.94  0.0588
Collar                      2    3342.45026    1671.22513     2.12  0.1274
TC_Pct*Season               1    5893.75152    5893.75152     7.47  0.0078

                                                Standard
Parameter                       Estimate           Error    t Value    Pr > |t|

Intercept                144.7476841 B     37.14161864       3.90      0.0002
Age                       -1.0105315       0.34699610      -2.91      0.0047
TC_Pct                    -0.6562001       0.26856686      -2.44      0.0169
Idols                     19.9941753       4.88831482       4.09      0.0001
AppStatus     0            2.8460366 B     17.35227484       0.16      0.8702
AppStatus     1           18.8746603 B     17.72147454       1.07      0.2903
AppStatus     0.5          0.0000000 B       .                .          .
Collar        1          -16.3696105 B      8.67333234      -1.89      0.0630
Collar        2          -17.8531981 B      9.36326466      -1.91      0.0604
Collar        3            0.0000000 B       .                .          .
TC_Pct*Season             -0.0133077       0.00486797      -2.73      0.0078

NOTE: The X'X matrix has been found to be singular, and a generalized
      inverse was used to solve the normal equations.  Terms whose
      estimates are followed by the letter 'B' are not uniquely estimable.
```

C. Residual Diagnostics

Residuals were checked to ensure that no model assumptions were violated.

First, residual values were checked to ensure that no outliers were present. In the case of this model, only one outlier was found. Cook's D statistics were calculated for each residual to check for high levels of influence. This would only be a problem for any outliers found. No Cook's D values were concerning or merited further investigation.

```
ods graphics on;
proc glm data = finaltribal plots = (diagnostics residuals);
    class Appstatus Collar;
```

```
    model Pct_Jury_Vote = Age TC_Pct Idols AppStatus Collar season*TC_Pct;
    output out=pred r=resid cookd=cooksd rstudent = rstud;
run;
quit;
ods graphics off;

proc print data = pred;
    var Name rstud;
    where abs(rstud) >= 3;
run;

proc print data = pred;
    var Name cooksd;
    where abs(cooksd) >= 1;
run;
```

Class Level Information
Class
Levels
Values
AppStatus
3
0 1 0.5
Collar
3
1 2 3

Number of Observations Read
84
Number of Observations Used
83

Dependent Variable: Pct_Jury_Vote
Source
DF
Sum of Squares
Mean Square
F Value
Pr > F
Model

8
30104.76747
3763.09593
4.77
<.0001
Error
74
58359.63719
788.64375
Corrected Total
82
88464.40466

R-Square
Coeff Var
Root MSE
Pct_Jury_Vote Mean
0.340304
70.90107
28.08280
39.60843

Source
DF
Type I SS
Mean Square
F Value
Pr > F
Age
1
2726.14558
2726.14558
3.46
0.0670
TC_Pct
1
777.97861
777.97861
0.99
0.3238

Idols
1
6631.89645
6631.89645
8.41
0.0049
AppStatus
2
11022.66238
5511.33119
6.99
0.0017
Collar
2
3052.33292
1526.16646
1.94
0.1516
TC_Pct*Season
1
5893.75152
5893.75152
7.47
0.0078

Source
DF
Type III SS
Mean Square
F Value
Pr > F
Age
1
6688.54031
6688.54031
8.48
0.0047
TC_Pct
1
4708.13109

4708.13109
5.97
0.0169
Idols
1
13193.78857
13193.78857
16.73
0.0001
AppStatus
2
4644.78463
2322.39231
2.94
0.0588
Collar
2
3342.45026
1671.22513
2.12
0.1274
TC_Pct*Season
1
5893.75152
5893.75152
7.47
0.0078

Obs
Name
rstud
43
Natalie White
3.31973

In addition, the predicted versus residual plot displayed above confirms many assumptions: it looks randomly distributed, with a mean of zero, no prominent variance problems as confirmed by no clear funneling of residual values, and no apparent issues with independence of error terms.

Further, the Q-Q plot above supports normality. Normality is confirmed by the Shapiro-Wilk test below, with p-value of .1041, which is less than .05, and thus not sufficient to reject the assumption that the error terms are normally distributed.

```
proc glm data = finaltribal noprint;
    class Appstatus Collar;
    model Pct_Jury_Vote = Age TC_Pct Idols AppStatus Collar season*TC_Pct
/solution;
    output out=pred r=resid cookd=cooksd rstudent = rstud;
run;
quit;

proc univariate data = pred normal;
    var resid;
run;
```

```
                        The UNIVARIATE Procedure
                            Variable:  resid

                                 Moments

      N                          83    Sum Weights                83
      Mean                        0    Sum Observations            0
      Std Deviation      26.6777603    Variance           711.702893
      Skewness           0.27190809    Kurtosis           -0.2728154
      Uncorrected SS     58359.6372    Corrected SS       58359.6372
      Coeff Variation             .    Std Error Mean     2.92826461

                      Basic Statistical Measures

              Location                     Variability

          Mean     0.000000     Std Deviation           26.67776
          Median   2.223321     Variance               711.70289
          Mode      .           Range                  119.97614
                                 Interquartile Range     40.73790

                      Tests for Location: Mu0=0

             Test           -Statistic-    -----p Value------

             Student's t    t         0    Pr > |t|    1.0000
             Sign           M       0.5    Pr >= |M|   1.0000
             Signed Rank    S       -34    Pr >= |S|   0.8784
```

Tests for Normality

Test		--Statistic---		-----p Value------
Shapiro-Wilk	W	0.974926	Pr < W	0.1041
Kolmogorov-Smirnov	D	0.093797	Pr > D	0.0715
Cramer-von Mises	W-Sq	0.107245	Pr > W-Sq	0.0912
Anderson-Darling	A-Sq	0.667333	Pr > A-Sq	0.0824

Quantiles (Definition 5)

Level	Quantile
100% Max	66.57819
99%	66.57819
95%	35.21625
90%	28.28630
75% Q3	20.17381
50% Median	2.22332
25% Q1	-20.56410
10%	-32.02746
5%	-38.85457
1%	-53.39795
0% Min	-53.39795

Extreme Observations

------Lowest-----		-----Highest-----	
Value	Obs	Value	Obs
-53.3980	29	35.2163	67
-52.1803	36	59.0399	30
-50.2842	44	60.8526	64
-43.6643	31	61.1095	52
-38.8546	42	66.5782	43

Missing Values

		-----Percent Of-----	
Missing Value	Count	All Obs	Missing Obs
.	1	1.19	100.00