REALITY-COMPETITION PROGRAMS: ARE YOU IN OR ARE YOU OUT?

A THESIS

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE

MASTER OF SCIENCE

BY

REBECCA P. KURTZ

WITH DR. REBECCA PIERCE AS ADVISOR

BALL STATE UNIVERSITY

MUNCIE, INDIANA

MAY 2018

REALITY-COMPETITION PROGRAMS: ARE YOU IN OR ARE YOU OUT?

A THESIS

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE

MASTER OF SCIENCE

By

REBECCA PAULETTE KURTZ

Committee Approval:

....

Committee Chairman Date

....

Committee Member Date

....

Committee Member Date

Department Approval:

....

Department Chairperson Date

....

Dean of Graduate School Date

Ball State University
Muncie, Indiana
May 2018

Acknowledgment

I first want to thank my thesis advisor, Professor Rebecca Pierce. Prof. Pierce was an ideal advisor, offering unwavering support while also giving me room to explore and develop my researching skills.

I also want to thank those on my thesis committee, Prof. Rahmatullah Imon and Prof. Drew Lazar. Prof. Imon was always willing to step in with support and suggestions, and Prof. Lazar consistently had his door open and would make time to brainstorm ideas.

I am grateful to all of those in the Mathematical Science Department. The office staff, professors, and my peers were always just a door knock away. Whether it be a cup of coffee, a friendly smile, or fresh perspective, the collective presence of everyone provided an ideal working environment for a research project.

Lastly, I want to thank my family and fiancé Jacob Garcia. They have supported and encouraged me throughout my studies and the process of writing this thesis.

Abstract

THESIS: REALITY COMPETITION PROGRAMS: ARE YOU IN OR ARE YOU OUT?

STUDENT: Rebecca Paulette Kurtz

DEGREE: Master of Science

COLLEGE: Sciences and Humanities

DATE: MAY 2018

PAGES: 61

In the last few decades, reality competition shows have dominated the television industry. Have you ever wanted to know if you have what it takes to win *Project Runway*, *Face Off,* or *Top Chef*? These are just a few reality-competition shows out of hundreds that have helped average people get jump starts in their careers. Yet both reality television shows and game shows have had a controversial history of producer manipulation, cheating, and staging outcomes. This research explores the genre and two prediction methods to determine a contestant's probability of winning a reality-competition show.

These competitions can be classified as a categorical problem with two outcomes, winners and losers. A variety of different predictor variables are considered over different timepoints and methods. The results of a multinomial logistic regression model and a modified random forest model are compared using several assessment measures.

This analysis aims to answer the following questions: Is it possible to predict the winner of a reality-competition program? How early in the competition can a winner be determined? What are the primary factors used to determine the winner?

Contents

List of Figures

List of Tables

List of Equations

1. Introduction

The television industry has varied greatly over time. Genres have morphed, reformed, and produced new subgenres. The game show and reality show genres in particular have been ever changing.

Game shows first came into popularity in the late 1940's as quiz shows (Hoerschelmann, n.d.). They became popular due to their low budgets and tempting prizes for contestants. However, producers took advantage of the loose structure and began manipulating the outcomes by favoring popular contestants (Hoerschelmann, n.d.). This sparked the need for game show regulation. New laws were passed in 1958, which caused most of the 1950's game shows to be canceled, most notably *Twenty-One*, and *The $64,000 Question* (Hoerschelmann, n.d.).

To accommodate these regulations, alternative styles of game shows were developed between the 1960's and 1980's; however, many still had flaws, this time in favor of the contestants. In the 1970's the show *Let's Make a Deal* introduced one of the most famous statistical problems of this century, the Monty Hall Problem. A strategy to increase your odds of winning was soon developed (Selvin, 1975). For the Monty Hall problem, contestants can choose between three doors, one contains a valuable prize and the other two doors each contain a smaller mediocre prize. A contestant selects one door then the contents of one of the remaining doors is revealed, exposing a small prize. The contestant then has the option to switch to the remaining door, for which the contents are unknown. It can be proven that the contestant has a higher probability of winning the valuable prize if they switch doors (Selvin, 1975).

In addition, in 1984 one of the contestants discovered that *Press Your Luck*'s random generator was simply a repeated sequence of prize options (Crockett, 2016). The contestant was able to detect a pattern and determine when they would be successful in selecting the prize money. It was also suspected that shows with a studio audience could have audience members trying to signal answers to contestants; this was a problem in the show *Who Wants to be a Millionaire* (Reilly, 2016).

While the genre struggled with game show fairness for both contestants and the network, reality television began gaining popularity. The first reality television shows were *An American Family* in 1973 and *Cops* in 1989 (Hoerschelmann, n.d.). After *Cops*, the number of reality television shows skyrocketed. By 2006 reality shows made up approximately 41% of all production activity in Los Angeles, CA (Podlas, 2007). The United States is one of the leading producers of reality television, and in 2015 approximately 750 reality shows aired on primetime cable, 83% higher than the number of scripted shows (VanDerWerff, 2016).

Now nearly every type of television program has had an experimental reality-based counterpart, and game shows are no exception. This rise in reality television began a new sub-genre of reality television game shows, also known as reality-competition shows, which has been recognized as a separate Emmy category since 2003 (“Primetime Emmy”, 2017). The Television Academy defines a reality-competition program as a reality series with a minimum of 6 episodes that has a competition component (“Primetime Emmy”, 2017). The reality aspect offers a new twist on game shows that follow regulations. Reality-competition programs are among the most popular television shows of all time, with shows like *Survivor* and *American Idol* being some of the most watched shows in television history (Yahr, Moore, & Chow, 2015). These reality competition programs have been so successful, there are now sub-sub-genres for them.

Due to the rise and popularity of reality competition shows, the opportunity of producer manipulation, and the scandalous history of game shows in general, an investigation of the fairness and possible biases of these shows is warranted. Although there are communities and groups which aim to predict results for television programs, such as GoldDerby.com, they are typically based on a critic’s opinions rather than statistical methods.

This paper focuses on the reality-competition shows that follow an elimination format. The elimination format determines a single winner out of a select group of individuals by eliminating contestants in a series of competitions (“Tournament formats”, n.d.). The final winner typically receives cash, job advancing benefits, and/or a job. The elimination format is one of the most popular types of reality competition shows, with dozens of different variations. This analysis aims to answer the following questions: Is it possible to predict the winner of a reality-competition program? How early in the competition can a winner be determined? What are the primary factors used to determine the winner?

Although academic research on reality-competition programs is limited, there is research for similarly structured competitions, notably, the Academy Awards, or Oscars, and horse racing. Both the Oscars and horse racing have one winner from a set number of candidates. Thus, the response, or dependent variable, is binary, i.e. winner or loser. As with reality-competition programs, these situations have a select number of predictor variables, and the predictor variables can be continuous or categorical. Several methods have been proposed to model these various competition results, most notably, regression and machine learning techniques, each of which can be used for classification.

2. Literature Review

2A. Classification Using Regression Methods

Logistic regression is a regression method that limits the predicted value to a scale between zero and one, representing a probability. These probabilities can then be categorized into binary groups by using a cutoff value. The probabilities larger than the given criteria are categorized as a 1, and the rest are 0. Traditional regression methods do not generate results between zero and one, thus logistic regression effectively solves the problem of generating predicted values for a classification problem with two groups.

In 2004, Kaplan considered a logistic regression model to predict which film would win the Oscar for Best Picture. Kaplan considered a variety of dummy predictor variables related to statistics from previous Oscar results such as, whether a film had the most nominations in total, previous best director nominations, previous director wins, etc. Kaplan made three logistic regression models. The first model contained all the variables Kaplan considered. The second model contained only variables from the first model that were individually significant using a two-sided t-test. The third model contained variables he considered intuitively important.

Using Kaplan's method of dummy variables can help eliminate many potential problems which quantitative variables for competition statistics may present. The starting number of contestants for reality-competition shows may vary, thus each contestant may not have the same number of opportunities to increase their game show statistics. Creating a dummy variable helps make the contestants more comparable across different competitions and improve the accuracy of the model. Kaplan's model also used predictor variables such as age, height, and/or birthplace which may also be useful for predicting horse racing, or reality-competition game show results.

A drawback to Kaplan's models is that logistic regression cannot account for the within-race competition, or knowing which candidates compete with who. Using the independent variables, logistic regression aims to account for all available information which could possibly affect a candidate's chances of winning (Lessmann, Sung, & Johnson, 2010). However, in a race or competition, a candidate's chances are affected by the other competing candidates. If we have a particularly strong group of candidates in one competition and a weak group of contestants in another competition, the model could select multiple candidates from the strong group to win, and none from the weak group, which would not happen in practice. Thus, logistic regression has potentially large drawbacks.

In contrast, Pardoe and Simonton (2007), and Lessmann et al. (2010) considered a multinomial logit model (MNL), or conditional logistic model, to predict the winners of the Oscars and horse races respectively. The multinomial model maintains connections between individuals in each group. We know which contestants are competing against each other and the multinomial model accounts for this information. The MNL model requires an additional assumption, known as the independence of irrelevant alternatives (IIA), which assumes that the Oscar nominees are not close substitutes for each other (Pardoe & Simonton, 2007). The property states that if we have five contestants competing, a sixth contestant's chance of winning would come proportionately from the shares of the existing contestants (McFadden, 1974). The conditions for the contestants on reality competition game shows are similar to the Oscars, and horse racing, thus the IIA assumption is suitable for this analysis.

Pardoe and Simonton's (2007) paper focused on four major Oscars: Best Picture, Best Director, Best Lead actor, and Best Lead actress. Pardoe and Simonton limited their variables primarily to award show statistics. The predictor variables for Best Lead actor and Best Lead actress included the number of previous Oscar wins and previous Oscar nominations, whether the film was also nominated for Best Picture, and whether the film was considered a Golden Globe drama. Other variables, such as age of nominees, running time, and release date, were also considered but ultimately excluded due to lack of significance. Pardoe and Simonton did not use a rigid cutoff for their variable selection, but instead a variety of measures. They considered, theory consistency, goodness of fit, predictive ability, robustness, theories from other studies, and parsimony. They estimated their variables using two approaches, classical maximum likelihood and Bayesian estimation. Both models produced similar accuracy rates: 67% for classical estimation and 69% for Bayesian estimation; however, the process for classical estimation was far simpler.

For their MNL model, Lessmann et al. (2010) used independent variables that remained constant, such as the horses' weight and height, as well as other variables that depend on the specific race, such as the critic's predicted odds for the day. They did not consider a variable selection method for the horse race predictions. Thus, all independent variables were included in their model.

Accounting for different groups competing with the MNL model provides an advantage over the traditional logit model. However, it unfortunately has a draw back. There is not a clear standard method for variable selection.

2B. Classification Using Machine Learning Methods

In contrast to regression techniques, machine learning algorithms model excel at detecting nonlinear relationships and have gained popularity, challenging these long-standing prediction methods. In particular, network systems and decision trees are tools used in machine learning algorithms that can be used for classification.

An artificial neural network (ANN) is a machine learning algorithm used to detect complex relationships. This algorithm gets its name because relationships detected between different pieces of information can often be graphed to resemble neurons (Haghighat, Rastegari, & Nourafza, 2013)

In 2017, Bonato used ANN to predict the results of the reality-competition show, *Survivor*. In *Survivor*, at each elimination round the current contestants vote off who they want to eliminate until the final round, in which the eliminated contestants then get to decide the overall winner. In particular, Bonato's model measured rivalry and alliances between contestants. This particular ANN only considers the current game and potential rivalries. Rivalries are measured by who votes for who to be eliminated, with more votes indicating a stronger rivalry. Most reality-competition shows are not structured in this format. Although contestants are often voted off by judges, the individual judges' votes are often not known. Thus, *Survivor* is an exception and this method would not be applicable to most reality-competition shows.

Although other ANN competition studies have been conducted, as analyzed by Haghighat, et al. (2013), another disadvantage is the method does not produce information regarding the amount of influence a variable had in the prediction process. ANN are frequently described as a black box prediction method because of this reason (Lessmann, et al., 2010). Which variables played a role in the prediction process are not measured.

In contrast, decision trees can start with the same variable setup considered for logistic regression and MNL regression while also detecting variable influence. In particular, the random forest method, RF, is a machine learning algorithm that uses decision trees for classification (James et al., 2017). In this learning algorithm, multiple copies of the original dataset are created using a bootstrap sample, which are then used to grow decision trees simultaneously. Each tree only considers a random subset of predictors for each split. The results of these decision trees are then combined into one predictive model. RFs have been praised for their intuitive process. The RF method can also detect more complex relationships then traditional logistic regression, and the variable influence is measured as the model is being created.

In addition to their MNL model, Lessmann, Sung, and Johnson (2010) used a RF model to predict horse race winners. The built in variable selection process simplified the analysis by simultaneously making predictions and evaluating variable importance.

Although machine learning algorithms rival traditional methods and have many benefits, the standard version of these algorithms are unable to account for within race competition. Consequently, Lessmann, Sung, Johnson (2010) introduced a two-stage modeling framework to accommodate for this problem in making horse race predictions. The first stage of the model uses random forests to predict the horse's normalized finishing positions in future races with fundamental variables, i.e. variables that remain constant. This is known as the horse's "strength". The second stage uses the MNL model with two variables, the horse's strength and the odds of winning, otherwise known as public opinion.

Thus, the two-stage model solves problems generated by the other predictive models. Namely, logistic regression and the RF model do not account for within race competition, and the MNL model does not detect nonlinear relationships.

In addition to the two-stage solution above, Karpievitch, Hill, Leclerc, Dabney, and Almeida (2009) proposed a subject-level bootstrap sampling method for their random forest model to control for their groups. Their study measured biological features with highly sensitive machines. Due to the sensitivity of the machines, it is typical to make multiple samples of the same object and average the results to make a single observation that will be added to their dataset, limiting the number of data values. In contrast, if they considered each sample individually and used a typical RF model, they would have correlated observations and risk over fitting the data because a test observation and a training observation may come from the same object. Thus, they purposed a cluster sampling method, grouping samples from the same object into clusters, and using a cluster bootstrap sample. The predicted values where then aggregated over the different trees and clusters.

When considering reality competition programs, it is not expected that the contestants are related to each other the way the biological samples are. However, if contestant A and contestant B are each competing in competition *j*, it does not make sense to use contestant A as a training observation and contestant B as a test observation, because if contestant A's results are known, then contestant B's results are also known. Although the biological experiment does not mimic the same set up as the competitive event, a cluster bootstrap sample achieves the same result, controlling for the known groups.

It is worth noting that boosting, another machine learning algorithm that relies on decision trees, is like the random forest method and is also used for predicting the outcome of competitive events. However, boosting creates decision trees sequentially instead of simultaneously, building subsequent trees based off the residuals of the previous trees (James, Witten, Hastie, & Tibshirani, 2017). Boosting and random forests are similar techniques, the key difference between them is that boosting considers more tuning parameters when building the model and boosting has a greater risk of overfitting (James, Witten, Hastie, & Tibshirani, 2017).

3. Methods

3A. Description of the Data

The two classification methods considered to answer the three questions posed in this research are the multinomial logit (MNL) model and the modified random forest (RF) model with a cluster sampling method. Each model is created at six different timepoints (t) in the series. Because each season can vary in the number of contestants and number of episodes, time is scaled to be a percent, where t_{25} would correspond to the episode closest to the point when the series is 25% complete, without exceeding 25%. For example, consider t_{25} for a competition with fifteen episodes, or elimination rounds. After the forth episode the series would be approximately 26.67% complete, and after the third episode the series would be approximately 20% complete. So, the information gathered at t_{25} was collected directly after the third episode, before the fourth airs. Only contestants that are still competing after the third episode are used to build the model. Note t_0 considers all contests before the start of the show, and t_{100} corresponds to the end of the last episode in the season. In addition, the timepoints t_{Ep1} and $t_{Ep2ndLast}$ correspond to the competition status after the first episode and after the second to last episode. A model for each method is created at the following six timepoints: t_0, t_{Ep1}, t_{25}, t_{50}, t_{75}, and $t_{Ep2ndLast}$.

This analysis uses over 300 contestants and 25 seasons from 5 different series. Data was gathered from programs that first aired between 2008 and 2017. The models are built from the five programs: *Project Runway, RuPaul's Drag Race, Top Chef, Skin Wars,* and *Face Off*.

The data were gathered using three different methods: recording the contestant information from the network directly either by watching the program or accessing the information from the program's official website, using published news and/or magazine articles, or using contestant information gathered by a third-party website such as fandom.com. Data gathered by any third-party website was cross checked with official network information using a stratified random sampling method, where a minimum of 25% of the information was verified. Table 3A.1 contains terms used to describe potential outcomes for contestants after each elimination round.

Table 3A.1: Key Terms for Reality-Competition Programs

Term	Definition
WIN	The contestant earns a WIN if they won an elimination challenge.
HIGH	A contestant was praised by the judges but did not win an elimination challenge.
LOW	A contestant received negative criticism but was not eliminated.
IN	A contestant did not receive negative criticism or praise by the judges and remains in the competition.
OUT	A contestant was eliminated from the competition.
TOP	A contestant fell into the WIN or HIGH category.
BOTTOM	A contestant fell into the OUT or the LOW category.
X	A contestant left the competition early for reasons unrelated to the competition. Typical reasons include medical issues or personal circumstances.
DQ	A contestant was disqualified.
SAVE	A contestant was eliminated but invited to compete again due to a special circumstance.

There are multiple variations of reality-competition shows. To be considered in this study, a show must meet the following criteria:

- Contestants compete for a minimum of 6 episodes before winning the grand prize.

- For each elimination, the judges' sort the contestants into five mutually exclusive categories: WIN, HIGH, IN, LOW, OUT.
- The series must average one elimination per episode/week.
- The competition is not based on a physical challenge.
- The season must be produced in the United States.
- All contestants must be 18 years old or older.

An example of the results of a competition are presented in Table 3A.2. This table corresponds to results of *Skin Wars* Season 2. The row names refer to a contestant, and the column names refer to an episode number. The cell values are the results of an elimination round for a contestant, after the end of the given episode. For convenience and ease of readability rows are organized by elimination order and columns are arranged by episode order. Only episodes in which contestants had the potential to be eliminated were considered. Special episodes that showed progress on a project without the potential for elimination were ignored, and episodes where there was an elimination challenge, but no one was eliminated due to a special circumstance were included. Notice, as more contestants are eliminated, less are categorized as "IN" after the elimination round. Only the highest and lowest scoring contestants are evaluated by the judges, typically limited to 3 contestants for either extreme. As the show progresses, there are less contestants to consider, thus less are categorized as "IN".

Table 3A.2: Summary of Results for Skin Wars Season 2

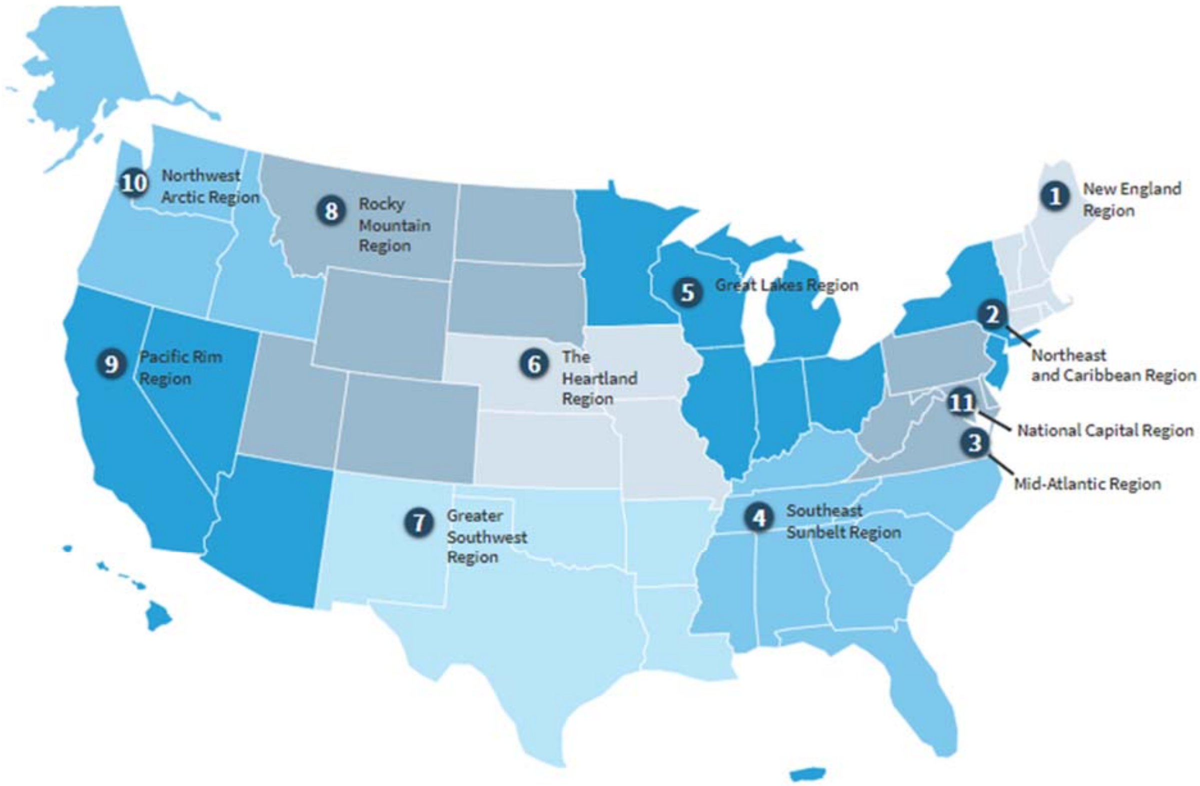
	EPISODES									
CONTESTANTS	1	2	3	4	5	6	7	8	9	10
Lana	WIN	IN	HIGH	LOW	LOW	HIGH	WIN	HIGH	LOW	WIN
Avi	IN	WIN	HIGH	HIGH	HIGH	WIN	LOW	WIN	WIN	OUT
Aryn	IN	HIGH	IN	WIN	HIGH	HIGH	LOW	LOW	HIGH	OUT
Cheryl Ann	HIGH	HIGH	LOW	LOW	LOW	IN	LOW	LOW	OUT	
Rio	IN	IN	WIN	LOW	HIGH	LOW	HIGH	OUT		
Kyle	HIGH	IN	HIGH	IN	WIN	LOW	OUT			
Dawn Marie	IN	LOW	LOW	HIGH	LOW	OUT				
Sammie	IN	IN	LOW	IN	OUT					
Fernello	LOW	IN	IN	OUT						
Rachel	LOW	LOW	OUT							
Rudy	IN	OUT								
Macio	OUT									

Only variables which vary across contestants are considered. Variables that are the same for all competitors within a competition are redundant because the MNL model and RF model control for within race competition variables. For example, the prize value or starting number of contestants would be the same for every contestant and hence, neither provide information that would indicate if one contestant is more likely than the other to win. Table 3A.3 describes the variables considered for each contestant in both models at all six timepoints. In addition, a figure summarizing the region variable is presented in Figure 3A.1 (GSA Regions, n.d.).

Table 3A.3: Description of Variables Considered at Every Timepoint

Variable Name	Description
Gender	A categorical variable that indicates gender of a contestant. (Male = “M”, Female=” F”)
Age	A continuous variable that indicates age of a contestant.
TopPop	A dummy variable that indicates if the location the contestant considers their place of residence is one of the top 100 most populated metropolitan cities of the world in 2016.
Region	A categorical variable that indicates the United States region (as defined by the US General Services Administration) the contestant considers to be their current place of residence. If the contestant does not consider themselves a resident of the United States, they are assigned a category of their own. This variable is nominal, with each region treated as a dummy variable for the MNL model. - (0) Other = Does not live in the United States - (1) NewEngland = Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island and Vermont. - (2) NortheastAndCaribbean = New Jersey, New York, the Commonwealth of Puerto Rico and the U.S. Virgin Islands. - (3) MidAtlantic = Delaware, Maryland, Pennsylvania, and Virginia (which included what is now West Virginia). For simplicity, this region will also include Washington DC. - (4) SouthEastSunbelt = Alabama, Florida, Georgia, Kentucky. Mississippi, North Carolina, South Carolina and Tennessee. - (5) GreatLakes = Illinois, Indiana, Michigan, Minnesota, Ohio, and Wisconsin, - (6) Heartland = Iowa, Kansas, Missouri and Nebraska - (7) GreaterSouthwest = Texas, Louisiana, Arkansas, Oklahoma and New Mexico - (8) RockyMountain = Colorado, Montana, North Dakota, South Dakota, Utah, Wyoming - (9) PacificRim = Arizona, California, Hawaii, Nevada, American Samoa, Guam, Northern Mariana Islands, Trust Territory of the Pacific Islands - (10) NorthwestArtic = Alaska, Idaho, Oregon, and Washington

Figure 3A.1: Visual Representation of the Region Variable



Variables related to the contestant's progress throughout the season change across the timepoints. After each elimination round, the show categorizes each contestant into ordered mutually exclusive groups. These categories are treated as scores. Table 3A.4 list the variables considered related to the scores the contestants have received up to a given timepoint in the season. The table describes each variable considered for the five timepoints t_{Ep1}, t_{25}, t_{50}, t_{75}, and $t_{Ep2ndLast}$. The variables in Table 3A.4 are not considered at t_0 because there is no competition information to consider at that point.

Table 3A.4: Description of Dummy Variables that Vary Over Time

Variable Name	Description
TOP_Ep1	Indicates if a contestant was considered TOP for the first challenge (1 = TOP, 0 = Not in the TOP)
BOTTOM_Ep1	Indicates if a contestant was considered BOTTOM for the first challenge (1 = BOTTOM, 0 = Not in the BOTTOM)
LeastBOTTOM	Indicates if a contestant was in the BOTTOM the least amount of times up to a given timepoint (1 = Yes, 0 = No)
MostBOTTOM	Indicates if a contestant was in the BOTTOM the most amount of times up to a given timepoint (1 = Yes, 0 = No)
AnyBOTTOM	Indicates if a contestant was in the BOTTOM at any time up to a given timepoint (1 = Yes, 0 = No)
LeastHIGH	Indicates if a contestant was in the HIGH the least amount of times up to a given timepoint (1 = Yes, 0 = No)
MostHIGH	Indicates if a contestant was in the HIGH the most amount of times up to a given timepoint (1 = Yes, 0 = No)
AnyHIGH	Indicates if a contestant was in the HIGH at any time up to a given timepoint (1 = Yes, 0 = No)
LeastWIN	Indicates if a contestant was in the WIN the least amount of times up to a given timepoint (1 = Yes, 0 = No)
MostWIN	Indicates if a contestant was in the WIN the most amount of times up to a given timepoint (1 = Yes, 0 = No)
AnyWIN	Indicates if a contestant was in the WIN at any time up to a given timepoint (1 = Yes, 0 = No)

Several other variables were also considered but ultimately excluded for various reasons. Some variables corresponding to the competition were simply excluded due to the lack of time and resources to collect and record the data. These variables include the number of judges per round, whether a guest judge was present, results of non-elimination challenges, whether the elimination challenge was a team challenge, whether a contestant had immunity or an advantage, etc. Other variables related to the individual contestants were excluded due to lack of information, or inconsistent information. For example, some programs provide the number of years of experience a contestant has, but other programs might mention current occupation. Personal contestant information varied greatly between programs; thus, variables related to the following were not considered: education level, years' experience, current occupation, hometown, sexual orientation, marital status, religious affiliation, race, etc.

Two dummy variables were considered but ultimately excluded due to lack of observations, Returner and Saved. The variable Returner indicated if a contestant had competed on a reality-competition program before, and the variable Saved indicated if a contestant was eliminated in a reality-competition but was invited back to compete in the same competition due to a special circumstance. Only one contestant in the dataset was a returner, and of the saved contestants, only one remained in the competition long enough to be considered at the next timepoint before being eliminated again, thus these variables were excluded.

3B. Multinomial Logistic Regression Model

Logistic regression is often used to model regression problems with a binary response variable. Consider the standard multiple logistic regression model in Equation 3B.1 based on Agresti (2013),

$$\hat{\pi}_j(\mathbf{x}_i) = \exp(\ \beta_0 + \boldsymbol{\beta}^T\mathbf{x}_i\ ) \qquad \textit{Equation 3B.1}$$

where $\pi(\mathbf{x}_i)$ is the probability that contestant *i* wins a competition. Let the vector $\mathbf{x}_i = (x_{i,1}, x_{i,2}, \ldots x_{i,P})$ contain contestant i's values for each of the variables, with P being the total number of variables being considered. Let $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots \beta_P)$ be the vector of coefficients for each variable, where β_0 represents the y-intercept, β_1 is the coefficient for the first variable, β_2 is the coefficient for the second variable, and so on for all variables. This model produces the probability that a given contestant wins any reality-competition program (Lessmann, et al., 2010).

The logistic regression model in Equation 3B.1 is a generalized linear model (GLM) as defined by Nelder and Wedderburn (1972). The parameters for this GLM model are estimated using the maximum likelihood estimation method (James et al., 2017). GLMs require the following properties: a random component, a systematic component, and a link function (Agresti, 2013). The model in equation (3B.1) satisfies these three properties. The random component is the response variable, which must come from independent observations of a distribution that is part of the exponential family. In this case, we are considering the binomial distribution (Agresti, 2013). The systematic component refers to the explanatory variables discussed in section 3A. The link function is a monotonic, differentiable function that combines the first two components (Agresti, 2013). For a binomial response variable, the link function is the log odds of the response, given in Equation 3B.2 (Agresti, 2013).

$$log\left(\frac{\pi(\mathbf{x}_i)}{1-\pi(\mathbf{x}_i)}\right) \qquad \textit{Equation 3B.2}$$

As stated earlier, the traditional logistic regression model in Equation 3B.1 does not account for the within race competition. The model returns the probability of a contestant winning any reality-competition program. The model does not account for only one winner being selected from a discrete set of known contestants. McFadden (1974) first purposed a solution to this problem, known then as a conditional logit model. It was later shown this model was a special case of the MNL model (Agresti, 2017). This special case of the MNL model requires an assumption known as the independence of irrelevant alternatives (IIA) (McFadden, 1974). This assumption indicates the odds of choosing contestant A over contestant B does not depend on the other contestants or their explanatory variables (Agresti, 2017). All options are assumed to be distinct, weighed independently in the competition, and are not close substitutes for each other (McFadden, 1974). For example, consider a competition with 3 male and 2 female competitors, if we suspect that adding another female to the competition would only decrease the probability of the original 2 female competitors of winning, the IIA assumption is violated. The IIA assumption assumes that each contestant’s probability of winning would be proportionally affected by the presence of an additional contestant (McFadden, 1974). Thus, all 5 of the original contestants' probabilities should decrease. All reality-competition programs in this analysis aim to judge each contestant independently, regardless of background. Several shows have set up mechanisms to further accomplish independence by limiting the judges' interactions with the competitors, so the product is judged without knowing which competitor created it. Therefore, it is reasonable to assume the IIA assumption is met. Consider the MNL model formula in Equation 3B.3.

$$\hat{\pi}_j(\mathbf{x}_{i,j}) = \frac{\exp(\ \beta_0 + \boldsymbol{\beta}^T\mathbf{x}_{i,j}\ )}{\sum_{i=1}^{n_j}\exp(\ \beta_0 + \boldsymbol{\beta}^T\mathbf{x}_{i,j}\ )} \qquad \text{Equation 3B.3}$$

Let $\pi_j(\mathbf{x}_{i,j})$ be the probability of contestant *i* winning competition *j*. Let the vector $\mathbf{x}_{i,j}$ = $(x_{i,j,1}, x_{i,j,2}, \dots x_{i,j,P})$ correspond to the variable values for contestant *i* in competition *j*, and n_j represent the total amount of contestants in competition *j*.

The MNL model effectively scales the probabilities. Let $\hat{Y}_{i,j}$ be a logical value that is equal to 1 when $\hat{\pi}_j(\mathbf{x}_{i,j}) > \hat{\pi}_j(\mathbf{x}_{k,j})$ for all $k \neq i$, otherwise $\hat{Y}_{i,j}$ is equal to 0. Thus, the contestant with the highest probability $\hat{\pi}_j(\mathbf{x}_{i,j})$ is categorized as the winner for competition *j*, because we know there can only be one winner per competition.

Each MNL model is generated using the forward stepwise subset selection procedure. The forward stepwise procedure is an algorithm designed to be computationally efficient, and resistant to over fitting the model. Below is the outline for the procedure from James et al.

Forward Stepwise Selection

1. Let M_o denote the null model, which contains no predictors.
2. For $k = 0, \dots, P - 1$:
    a. Consider all $P - k$ models that augment the predictors in M_k with one additional predictor.
    b. Choose the best among these $P - k$ models and call it M_{k+1}. Here best is defined as having smallest RSS or highest R^2.

3. Using cross validated prediction error, select a single best model from among $M_0, \ldots, M_P$. (James et al., 2017)

The predicted values are generated using a leave-one-out (LOO) cross validation approach, separating observations by competition rather than by contestant (James, et al., 2017). For each competition j, all observations not included in competition j will be used to build a model using the forward stepwise selection method. This set is known as the training set. The remaining n_j contestants will be used to test the data set, known as the test set. For each competition, this process is repeated for each timepoint t.

3C. Random Forest Model

The random forest model is a tree-based method of classification. A classification tree is built using an approach known as recursive binary splitting (James et al., 2017). In recursive binary splitting, all contestants are split into two groups based on a variable splitting mechanism. Several splitting criteria have been purposed, however a variable's Gini index, detailed in Equation 3C.1, is one of the most common methods (James et al. 2017).

$$G_p = \hat{m}_{p,0} - \hat{m}_{p,0}^2 + \hat{m}_{p,1} - \hat{m}_{p,1}^2 \qquad \textit{Equation 3C.1}$$

Let G_p be variable p's Gini index at a given stage in the algorithm. Let $\hat{m}_{p,k}$ represent the proportion of training observations with variable p that are in the response category k. The response categories are $k = 0$ the contestant loses the competition, and $k = 1$ the contestant wins the competition. The closer the $\hat{m}_{p,k}$'s are to 0 or 1, the smaller G_p will be, which indicates the importance of the variable (James et al., 2017).

The data set continues to be divided into nodes, or groups, based off the Gini index until a node satisfies a stopping criteria. A stopping criteria could be, but is not limited to, a minimum node size, a maximum number of nodes, or a minimum Gini index decrease (James et al., 2017). When using a classification tree to make predictions, the data must be split into a test and training set as before. The data values in the training set are used to create the tree. Contestants in the test set are then sorted into groups determined by the decision tree. The predicted class of the test contestants corresponds to the final node they were categorized in. If the final node had more losers, class 0, then contestant is predicted to lose. Consider the decision trees in Figure 3C.1 and Figure 3C.2.

Figure 3C.1: Decision Tree Example with Minimum Node Size of 20 at t_{75}

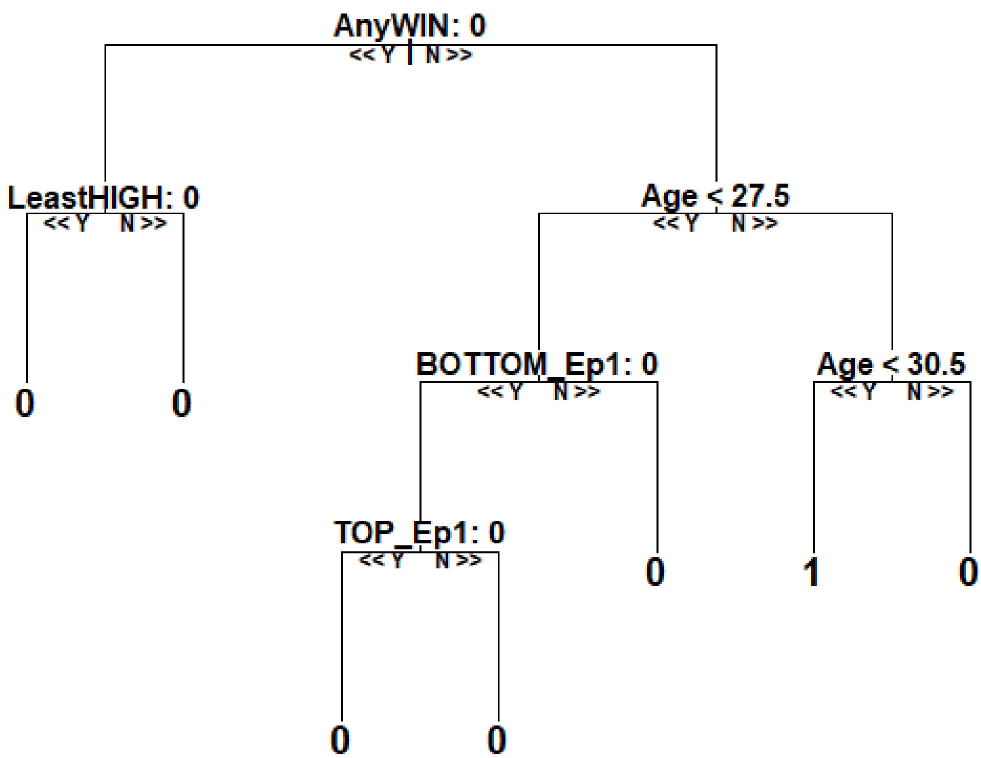


Figure 3C.2: Decision Tree Example with No Minimum Node Size at t_{75}

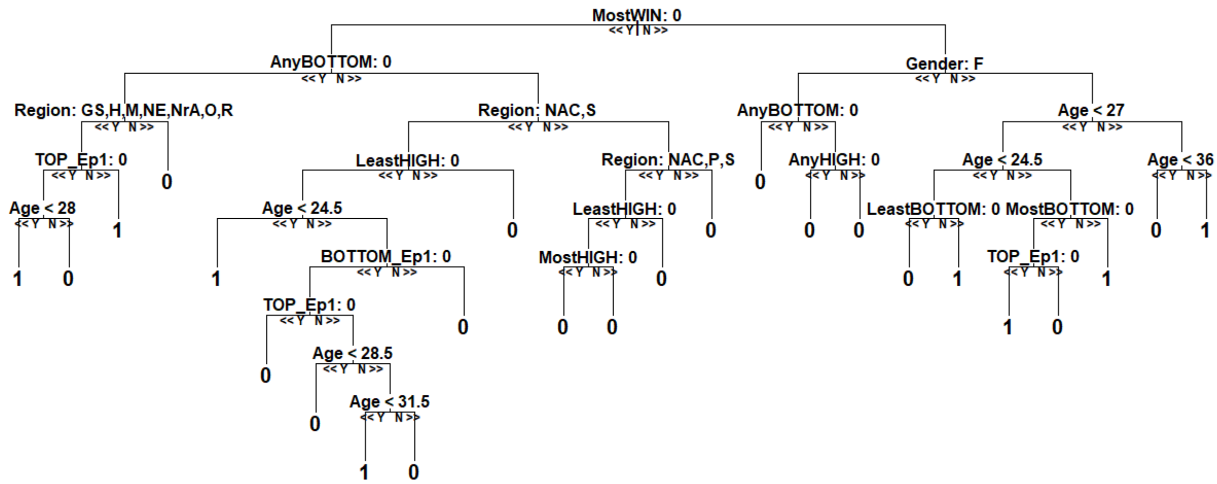


Figure 3C.1 contains a decision tree with a minimum node size of twenty. Hence when a node has less than twenty observations, it stops splitting. At the top of each node is text indicating the variable name, followed by a colon and the option it checked. Where variable name corresponds to the variable with the lowest Gini index at that point, and option is a value for that variable that the algorithm checked. For example, the top most node in Figure 3C.1 is the dummy variable AnyWIN, which had the lowest Gini index out of all the variables considered at that point. Thus, if a contestant had a win at any point before t_{75}, AnyWin would equal one, the algorithm would next check to see if a contestant's age is under 27.5. Notice that when checking for LeastHIGH: 0 the contestant is categorized as losing the competition regardless of the response. This is because the Gini index is a measure of node purity, that is, a measure that indicates if a node contains predominantly observations from a single class (James et al., 2017). Improving node purity does not improve the classification rate, but it does improve the Gini index.

Figure 3C.2 contains a decision tree with no minimum node size, thus a node stops splitting only when the Gini index cannot be improved, or when all observations in a node are of the same class. Notice also that the Age and Region variables appear in multiple places with different value checks. This is because Age is a continuous variable and Region is nominal variable with eleven categories. The algorithm selects the best binary split when multiple options are available.

Decision trees can vary greatly by changing the input slightly, thus resulting in a high variance (James et al., 2017). There are methods, known as pruning, to control for this variance. For instance, setting a large minimum node size, or limiting the number of nodes created. Alternatively, boosting and random forests are also methods used to control for the high variance (James et al., 2017).

Random forests are based on the idea of bootstrapping and decision trees. Taking B repeated samples from the full data set and making a tree with each sample, thus creating a total of B trees (James et al., 2017). Only a random subset of $\sqrt{P}$ variables is considered at each node. Using several bootstrapped samples reduces variance and using a subset of the predictors reduces the test error when there is correlation among the predictor variables (James et al., 2017). There are no set guidelines on determining the number of B trees to create, however, in practice, the model's performance stabilizes when there are one hundred or more (James et al., 2017).

To control for competition groups the contestants considered in the B repeated samples will be gathered based on a cluster sampling method. Specifically, a repeated sample of the J competitions will first be sampled, and all contestants within the selected competitions will be used to create the respective classification tree, the training set. This will be repeated until there are B trees. For example, if there are 25 competitions considered, then a simple random sample with replacement of the 25 competitions will be conducted. On average approximately $\frac{1}{3}$ of the competitions will not be selected, the competitors in these competitions will make up the test set (James et al., 2017). The contestants in the sampled competitions make up the training set.

The predicted value is calculated using the out-of-bag (OOB) method introduced by Breiman (2001). As stated previously, on average each bootstrapped sample omits $\frac{1}{3}$ of the observations, these are referred to as the OOB observations (Liaw & Wiener 2002). Predictions can be made for the OOB observations and these can be thought of as votes. These OOB values are aggregated for each observation to determine their probability, and hence, their predicted value (Liaw & Wiener 2002). We record the OOB values and the contestant that was classified as the winner the most number of times is then categorized as the winner, and the remaining contestants in a competition are categorized as the losers. A formula to calculate this probability is illustrated in Equation 3C.2

$$\hat{\pi}_j(\mathbf{x}_{i,j}) = \frac{1}{|B_j|}\sum_{b \in B_j} \hat{Y}_{i,j,b} \qquad \text{Equation 3C.2}$$

Let $\hat{\pi}_j(\mathbf{x}_{i,j})$ be the probability of contestant i winning competition j. Let $\hat{Y}_{i,j,b}$ be the predicted value of contestant i in competition j for tree b, where b is an element of the set of all trees that do not include competition j in the training set, B_j.

As before, let $\hat{Y}_{i,j}$ be a logical value that is equal to 1 when $\hat{\pi}_j(\mathbf{x}_{i,j}) > \hat{\pi}_j(\mathbf{x}_{k,j})$ for all $k \neq i$, otherwise $\hat{Y}_{i,j}$ is equal to 0. Thus, the contestant with the highest probability is categorized as the winner for competition *j*, because we know there can only be one winner per competition. Note that the LOO cross-validation method introduced for the MNL model is essentially equivalent to the OOB method (James et al., 2017).

3D. Evaluation Methods

A primary goal of this study is to predict the winners of a reality-competition program. Thus, the variables, method, and timepoint at which a winner can be predicted are evaluated. Variable selection and evaluation methods vary greatly between the MNL and RF models. Using the LOO approach for the MNL model for twenty-five competitions generates twenty-five different models for each timepoint. Thus, analyzing the individual variable significance is not efficient nor useful. Alternatively, the percentage of models that selected a variable at a specific timepoint will be used to evaluate the variable's relevance for the MNL model. Variable importance is defined as the average of the summed total amount that the Gini index decreases by splitting over a predictor (James, et al., 2017). For the RF method, the variable importance measurement will be recorded. The variable importance measure can also be described as the mean decrease in Gini index for each variable (James, et al., 2017). The larger the variable importance measure, the more important the variable.

Clearly, the percentage of MNL models that include a given variable at a specific timepoint and the variable importance measurement for the RF method are not equivalent measurements. Furthermore, neither measurement has a clear cutoff or threshold to determine variable significance. However, both measurements rank the variables in order of influence.

Although one of the main objectives of this analysis is to evaluate if the winner of a reality-competition program can be determined, prediction accuracy is limited in its ability to evaluate a model performance and should not be the only method used. A single method for determining the best model for classification is an open problem in statistics, and many methods are often needed to evaluate classification predictions (Gorunescu, 2011). Thus, the accuracy rate, specificity, sensitivity, and Cohen's Kappa value are evaluated to assess a model's performance. When generating these values, it is helpful to visualize the formulas using a confusion matrix.

Table 3D.1: Confusion Matrix

		Observed Values	
		$Win\ (Y = 1)$	$Lose\ (Y = 0)$
Predicted Values	$Win\ (\hat{Y} = 1)$	a	b
	$Lose\ (\hat{Y} = 0)$	c	d

Table 3D.1 displays a confusion matrix, a table used to compare predicted and observed values for classification problems (James, et al., 2017). Let a represent the number of contestants correctly classified as winners, b be the number correctly classified as losers, c the number incorrectly classified as losers, and d the number incorrectly classified as winners. The accuracy rate for a given model can be calculated with Equation 3D.1 for both the MNL and RF model.

$$Prediction\ Accuracy = \frac{(a + d)}{(a + b + c + d)} \quad \text{Equation 3D.1}$$

Specificity is the percentage of contestants that lost who were correctly classified (Agresti, 2013). Again, using the variables in Table 3D.1, the equation for specificity is defined in Equation 3D.2.

$$Specificity = \frac{(d)}{(b+d)} \quad \text{Equation 3D.2}$$

In contrast, sensitivity is the percentage of contestants who won that were correctly classified (Agresti, 2013). Using the variables in Table 3D.1, the equation for sensitivity is defined in Equation3D.3.

$$Senesitivity = \frac{(a)}{(a+c)} \quad \text{Equation 3D.3}$$

Both the sensitivity and specificity measurements are on a scale between 0 and 1. Which measure is best will depend on the problem at hand, and which errors would be less detrimental. In general, it is ideal to maximize both of these values (Gorunescu, 2011).

In addition, Cohen's Kappa is also a useful measure when comparing classification techniques. Cohen's Kappa compares matched data sets and returns a value that corresponds to how well the data sets agree, or are similar (Agresti, 2013). Cohen's Kappa, Equation 3D.6, is calculated using the expected and observed values for a classification problem, defined in Equations 3D.4 and 3D.5, respectively.

$$Expected\ Value\ (Exp) = \frac{(a+c)(a+b)+(b+d)(c+d)}{(a+b+c+d)} \quad \text{Equation 3D.4}$$

$$Observed\ Value\ (Obs) = \frac{(a+d)}{(a+b+c+d)} \quad \text{Equation 3D.5}$$

$$Cohen's\ Kappa = \frac{(Obs-Exp)}{(1-Exp)} \quad \text{Equation 3D.6}$$

When comparing actual versus predicted classes, a value of 1 indicates the data sets match for 100% of the observations, a value of 0 indicates that the data sets are random, and a negative value means that the predicted values are worse than random guessing (Agresti, 2013). Furthermore, Cohen's Kappa can also be used to compare the predicted values of the two different models. When comparing MNL predictions to RF predictions a value of 1 would mean that all the predicted values match, a value of 0 would mean the datasets would be random, and a negative value would mean the MNL and RF predictions were more different then similar.

Sensitivity, specificity, and Cohen's Kappa each produce a relative value in a set range which can be compared across time and methods. In contrast, when considering variable influence for the MNL model, we can only compare a variable across the MNL timepoints because the percentage of models that selected a certain variable would not have the same meaning for a RF model. In addition, the variable importance for the RF model can only be used to compare models across the RF timepoints, because the Gini index measurement is not applicable to the MNL model. Thus, neither measure is applicable to assess variables across models. Accuracy rate has the opposite problem in that it can be compared across models, but it cannot be well compared across time for this problem. Consider a dataset with a thousand observations generating only one winner, i.e. a success rate of 0.001. Consider a set of predicted values which misclassifies this one winner, the accuracy rate for the model would be 0.998. In contrast, consider a dataset with ten observations with a single success and a generated set of predicted values which misclassifies the winner. The prediction accuracy rate would be 0.80. Both sets of predicted values are essentially random, but the accuracy rate implies that one is more accurate than the other. The proportion of contestants who win in the data set changes as the show progresses further in the season as contestants get eliminated. The accuracy rate is useful when comparing two models created with the same datasets, but not as the characteristics of the datasets change. Accuracy and the variable assessment methods are useful, but sensitivity, specificity, and Cohen's Kappa are more transferable.

4. Results

4A. Multinomial Logistic Results

For the MNL model, to find the accuracy rate using the LOO approach, begin with separating the data into a test set and a training set. Recall that the test set consists of all contestants from a given competition *j*, and the training set consists of the remaining contestants. This is repeated for each competition creating a total of *J* models. For example, consider Project Runway Season 8, as the test set.

Table 4A.1: Coefficient Estimates at Timepoint t_{Ep1} using Project Runway Season 8 as the Test Set

Coeffient Name	Coeffient Estimate
(Intercept)	-0.14322
BOTTOM_Ep1	-1.36997
TOP_Ep1	1.79397
RegionNorthwestArctic	1.12844
MostHIGH	-1.71312
Age	-0.08818

The probability a contestant wins Project Runway Season 8 at timepoint t_{Ep1} is calculated using the coefficients in Table 4A.1 and the MNL regression formula in Equation 3B.2. Thus, the probabilities for the sixteen contestants in the test set are (0.604, 0.0282, 0.062, 0.044, 0.0189, 0.0334, 0.0366, 0.00859, 0.03074, 0.0066, 0.0058, 0.0475, 0.0166, 0.0435, 0.00787, 0.0066). The contestant with highest probability of winning is catorgized as the winner because we know there is only one winner per competition. For Project Runway Season 8, contestant 1 is categorized as the winner, $\hat{Y}_{1,24} = 1$, and the rest are categorized as a loser, $\hat{Y}_{i,24} = 0$ for i = [2, 16]. This process is repeated for each competition. For competitions with ties, one of the tied contestants will be randomly selected as the winner. For this particular case, contestant $\hat{Y}_{1,24}$ was correctly classified as the winner.

Notice, because there are *J* competitions there will be *J* models, all of which could select different variables. Table 4A.2 contains a summary of all the models created. For each timepoint the table contains the total number of contestants, the percentage of winners, LOO accuracy rate, the total number of variables considered across all models, and the largest number of variables considered for a single model. In addition, the sensitivity, specificity, and Cohen's Kappa value for the the predicted values versus the actual values are also reported. Recall, the ideal model will maximize sensitivity, specificity, and Cohen's Kappa across methods and timepoints, but accuracy rate should be maximized across methods but not necessarily time. The Cohen's Kappa value in Table 4A.2 compares the MNL predicted results versus the actual results, meaning the value indicates the percentage of predicted values that were correct after controlling for random guessing.

Table 4A.2: Multinomial Logistic Regression Results

	t_0	t_{Ep1}	t_{25}	t_{50}	t_{75}	$t_{Ep2ndLast}$
Number of Contestants	362	338	294	208	141	80
Percentage of Winners	0.0691	0.0740	0.0850	0.1154	0.1773	0.3125
LOO Accuracy Rate	0.9061	0.8876	0.8707	0.8413	0.7589	0.6000
Number of Variables Considered Across All Models	2	8	11	18	15	15
Largest Number of Variables in a Single Model	2	6	6	16	13	9
Sensitivity	0.3200	0.2400	0.2400	0.3333	0.3200	0.3600
Specificity	0.9496	0.9393	0.9294	0.9076	0.8535	0.7091
Cohen's Kappa: Actual vs MNL predicted	0.2696	0.1793	0.1694	0.2367	0.1734	0.0691

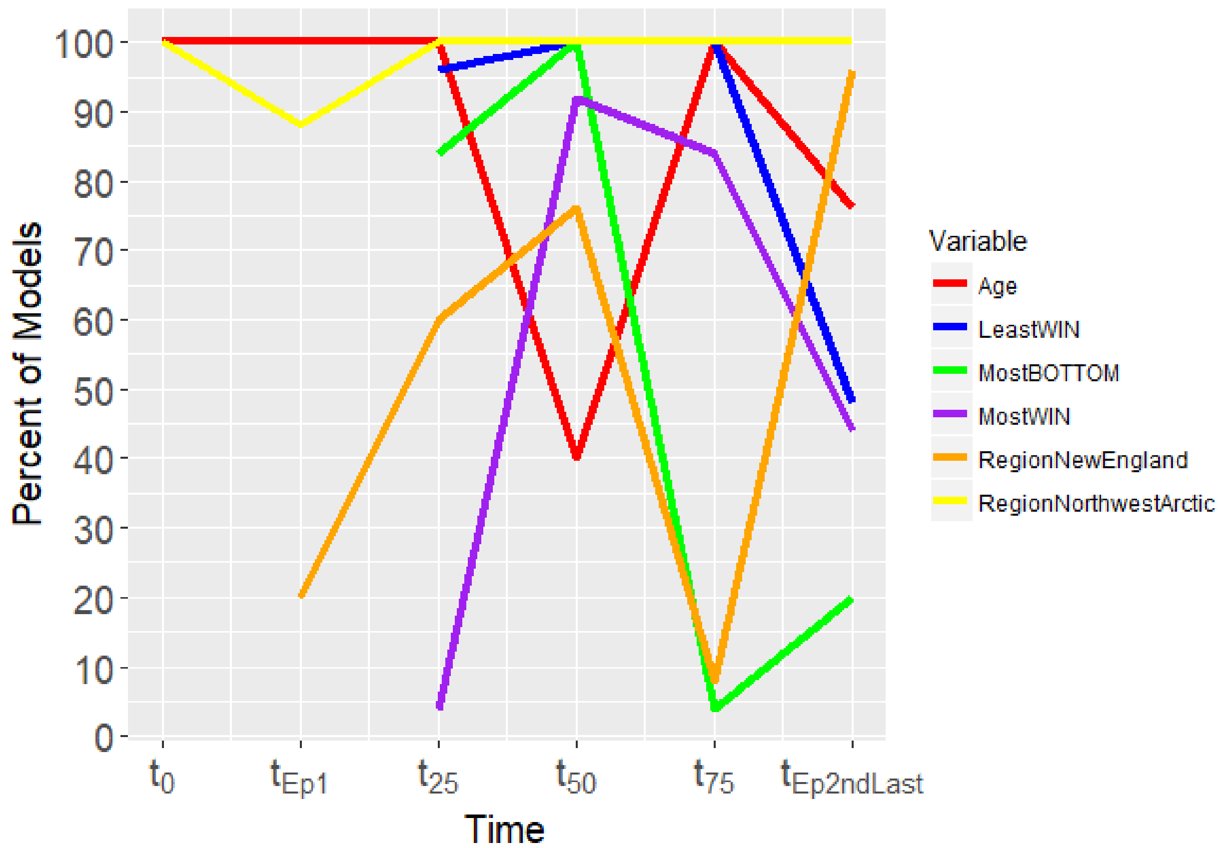
A summary of the average number of times a variable was selected is presented in Table 4A.2. Recall, for the MNL method, the nominal variable Region was transformed into a series of dummy variables. Thus, a variable with the form "RegionName", corresponds to a dummy region variable checking for the region name. For example, the variable "RegionGreatLakes" is checking to see if the contestant is from the Great Lakes region.

Table 4A.3: Average Variable Selection Percentage

Variable	Average Variable Selection Percentage	Variable	Average Variable Selection Percentage
RegionNorthwestArctic	98	RegionSoutheastSunbelt	5
Age	86	RegionRockyMountain	3
LeastWIN	69	RegionMidAtlantic	3
RegionNewEngland	52	LeastBOTTOM	2
MostWIN	44	RegionGreatLakes	2
MostBOTTOM	42	RegionNortheastAndCaribbean	2
TOP_Ep1	36	RegionPacificRim	2
LeastHIGH	28	GenderM	2
MostHIGH	19	AnyBOTTOM	2
BOTTOM_Ep1	19	LargeMetroArea	0
AnyHIGH	7	RegionHeartland	0
RegionOther	6	AnyWIN	0

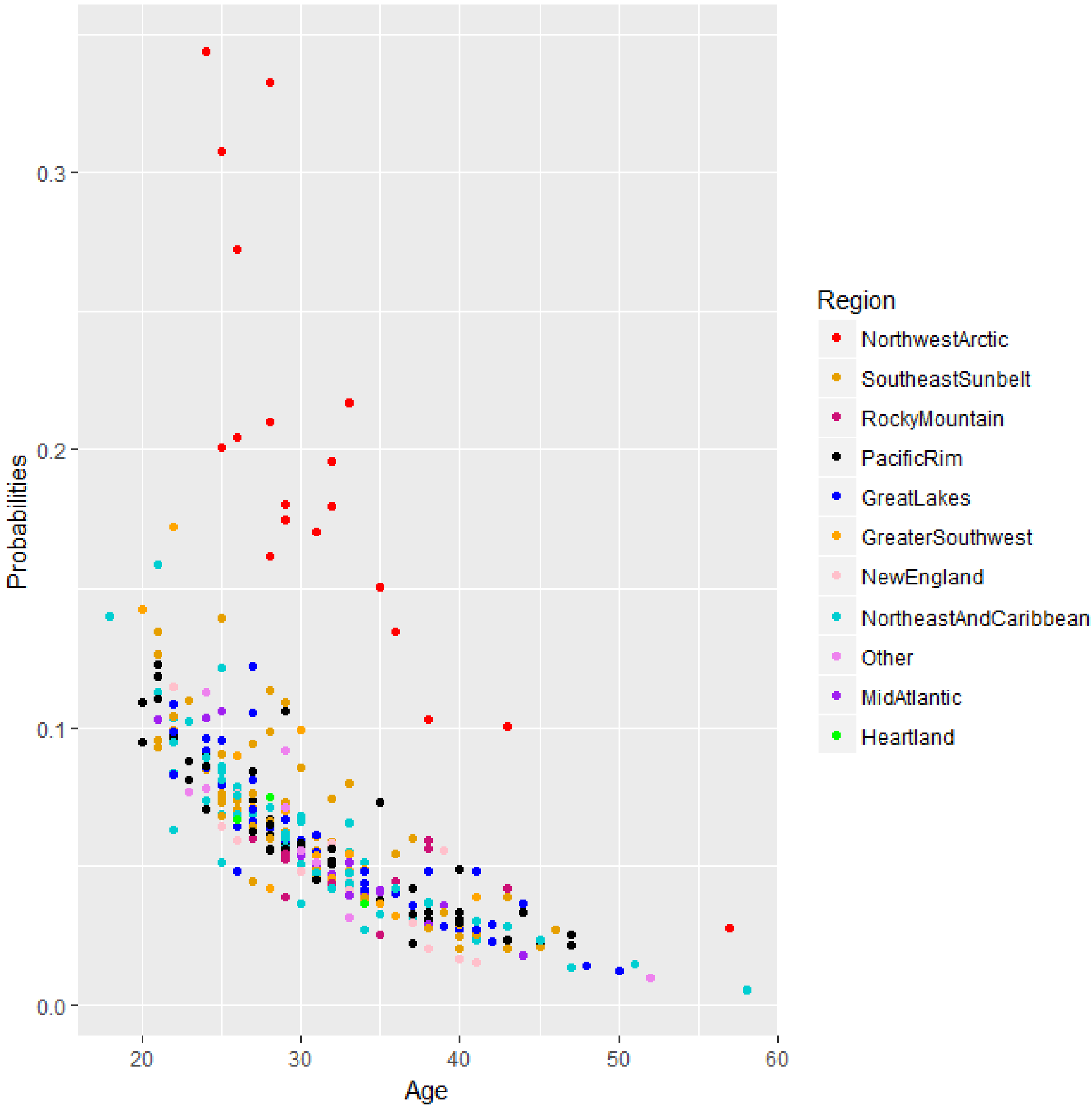
Figure 4A.1 graphs the percent of models that selected the top six overall occuring variables overtime. Notice in this graph the lines disappear at certain timepoints, this indicates that the variable was either not applicable for a specific timepoint, or not selected by the forward stepwise procedure.

Figure 4A.1: Percent of Models that Selected a Top Overall Occuring Variable Over Time



A scatter plot of the most commonly occuring variables at t_0, Age and Region, is presented in Figure 4A.3. Each point represents a contestant, the y-axis corresponds to the probabilities, the x-axis corresponds to age, and the color represents the contestant's region.

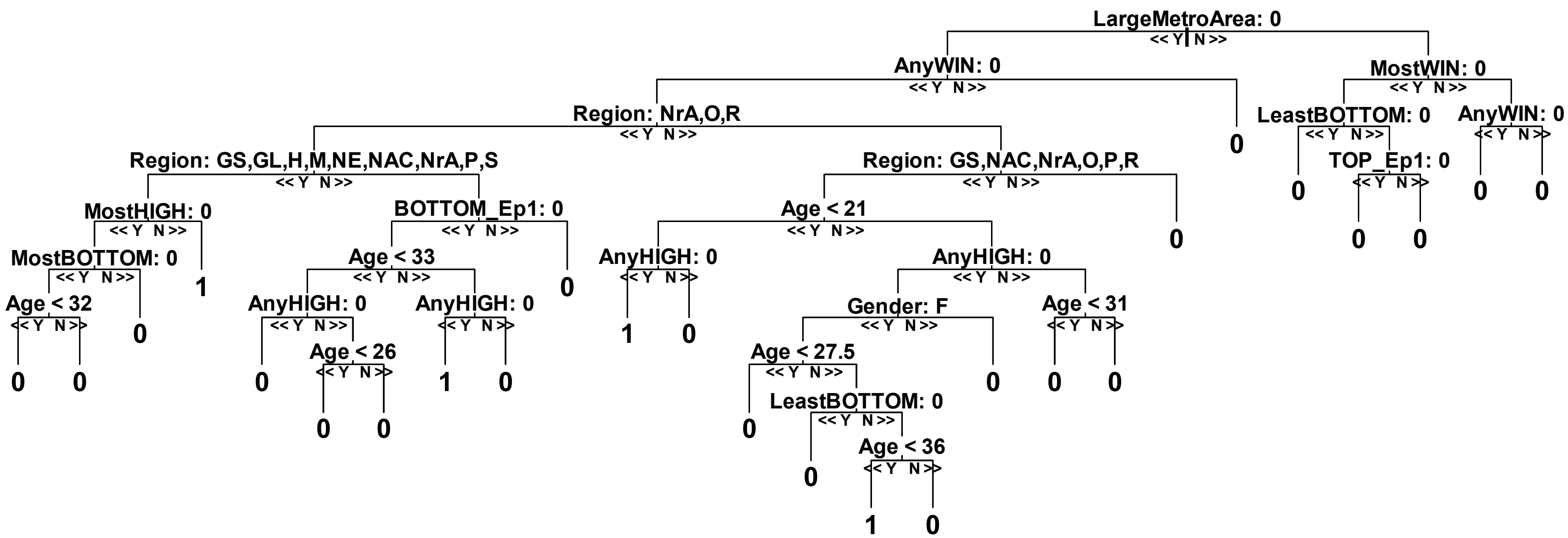
Figure 4A.2: Probability vs Age for the MNL Method at t_0



4B. Random Forest Results

For a random forest model, to find the accuracy rate using the OOB method start by dividing the data into a test and training set. Recall, to control for the different compeition groups, the training set is created using a cluster sampling method with replacement. This cluster sample starts with a simple random sample with replacement of the different competitions. All the contestants in the sampled competitions will then be placed in the training set, and the contestants in the groups that were not sampled make up the test set. Consider the example decision tree in Figure 4B.1.

Figure 4B.1: Decision Tree at t_{Ep1}



The decision tree in Figure 4B.1 was built using 16 different competitions composed of 213 unique contestants from the training set. The test set consists of 9 different competitions composed of 125 total contestants. Each contestant in the test set is then categorized as winning (1) or losing (0). This process is repeated 500 times, $B = 500$. The outcome a contestant falls into for each tree is considered a vote, which then turns into a probability. The contestant with the highest probability of winning within their competition is then categorized as the sole winner for their competition. In the event of a tie, one of the tied contestants is randomly selected.

Table 4B.1 contains a summary of the RF model results. For each timepoint, the table reports the total number of contestants, the proportion of winners, and the OOB accuracy rate. In addition, the sensitivity, specificity, and Cohen's Kappa value are also reported. The Cohen's Kappa value in Table 4B.1 compares the actual results versus the RF predicted results, meaning the value indicates the percentage of predicted values that were correct after controlling for random guessing. A negative Cohen's Kappa value indicates the predicted values were worse then random guessing.

Table 4B.1: Random Forest Results

	t_0	t_{Ep1}	t_{25}	t_{50}	t_{75}	$t_{Ep2ndLast}$
Number of Contestants	362	338	294	208	141	80
Percentage of Winners	0.0691	0.0740	0.0850	0.1154	0.1773	0.3125
OOB Accuracy Rate	0.8729	0.8698	0.8299	0.8029	0.6879	0.6500
Sensitivity	0.0800	0.1200	0.0000	0.1667	0.1200	0.4400
Specificity	0.9318	0.9297	0.9071	0.88587	0.8103	0.7455
Cohen's kappa: Actual vs RF predicted	0.0118	0.0497	-0.0929	0.0516	-0.0697	0.1855

Figure 4B.2 contains a graph of the variable importance measurement, the average Gini index decrease, over time. Notice that the variables considered are slightly different then the MNL approach because the RF model can accommodate nominal variables with more than two categories.

Figure 4B.2: Variable Importance Over Time

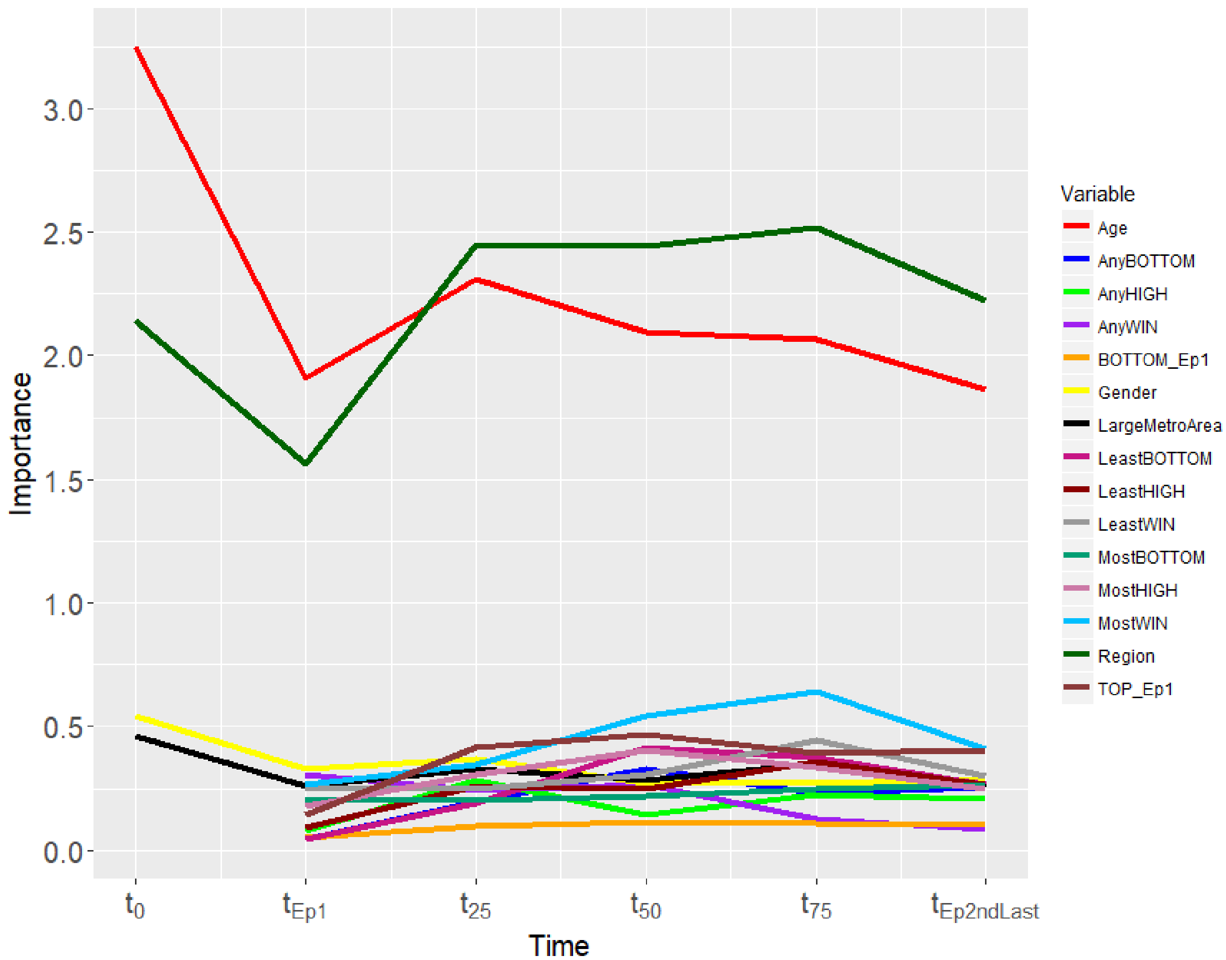
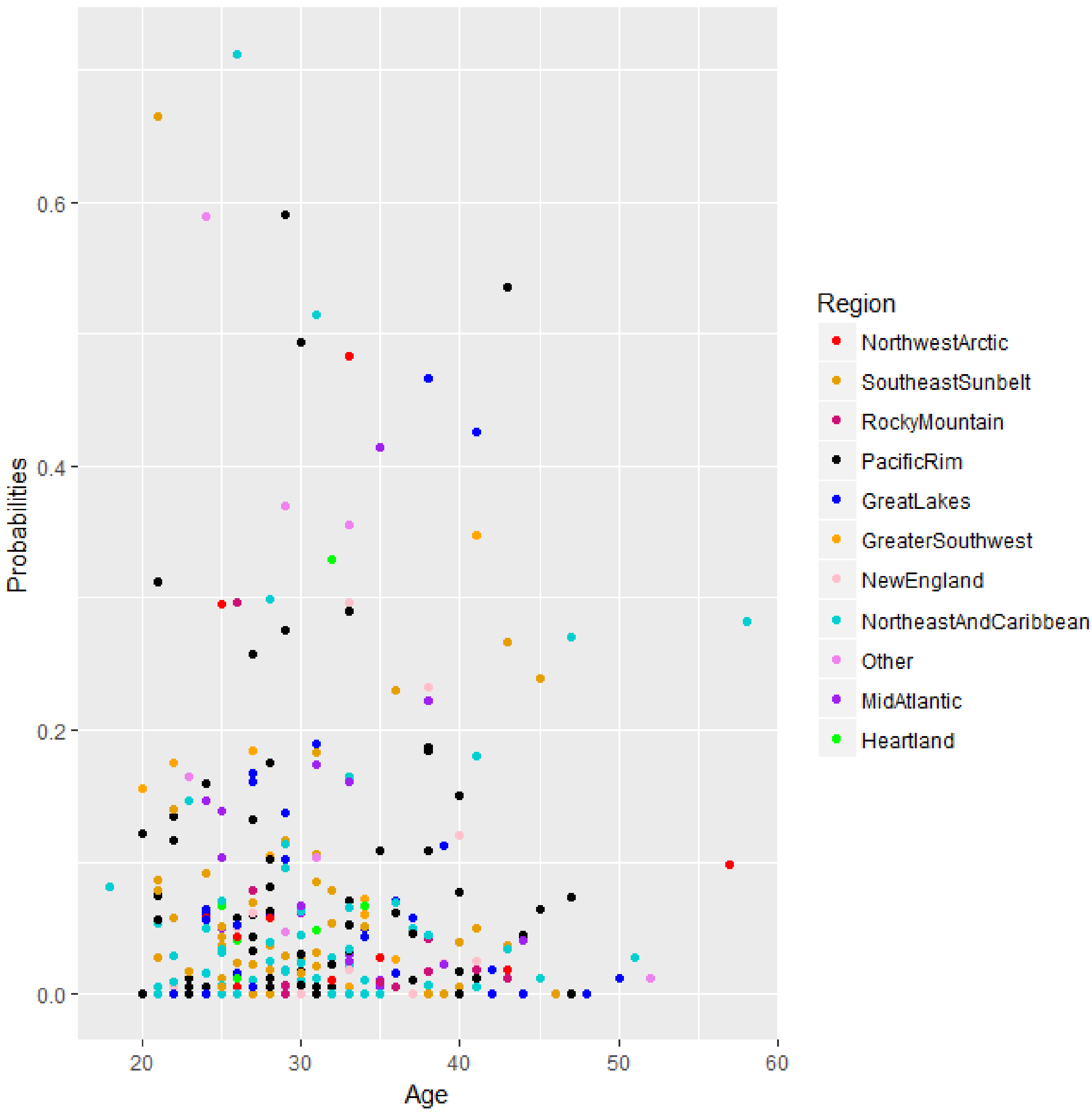


Table 4B.2 contains the average variable importance across all RF trees generated. A scatter plot of the most commonly occuring variables at t_0, Age and Region, is presented in Figure 4B.4.

Table 4B.2: Average Variable Importance

Variable	Average Variable Imporance	Variable	Average Variable Imporance
Age	2.2474	LeastBOTTOM	0.2571
Region	2.2225	LeastHIGH	0.2433
MostWIN	0.4421	MostBOTTOM	0.2273
TOP_Ep1	0.3649	AnyBOTTOM	0.2092
Gender	0.3429	AnyWIN	0.2048
LargeMetroArea	0.3224	AnyHIGH	0.1874
LeastWIN	0.3082	BOTTOM_Ep1	0.0941
MostHIGH	0.2940		

Figure 4B.3: Probability vs Age for the RF Method at t_0



4C. Comparison of Models

Figure 4C.1 contains a graph of the accuracy rates, specificity, sensitivity, and Cohen's Kappa value for both methods. Note, the Cohen's Kappa value in Figure 4C.1 compares the predicted values of a given method to the true results. A black line indicating the zero value is also indicated when applicable.

Figure 4C.1: Classification Measurements Over Time for the MNL and RF Models

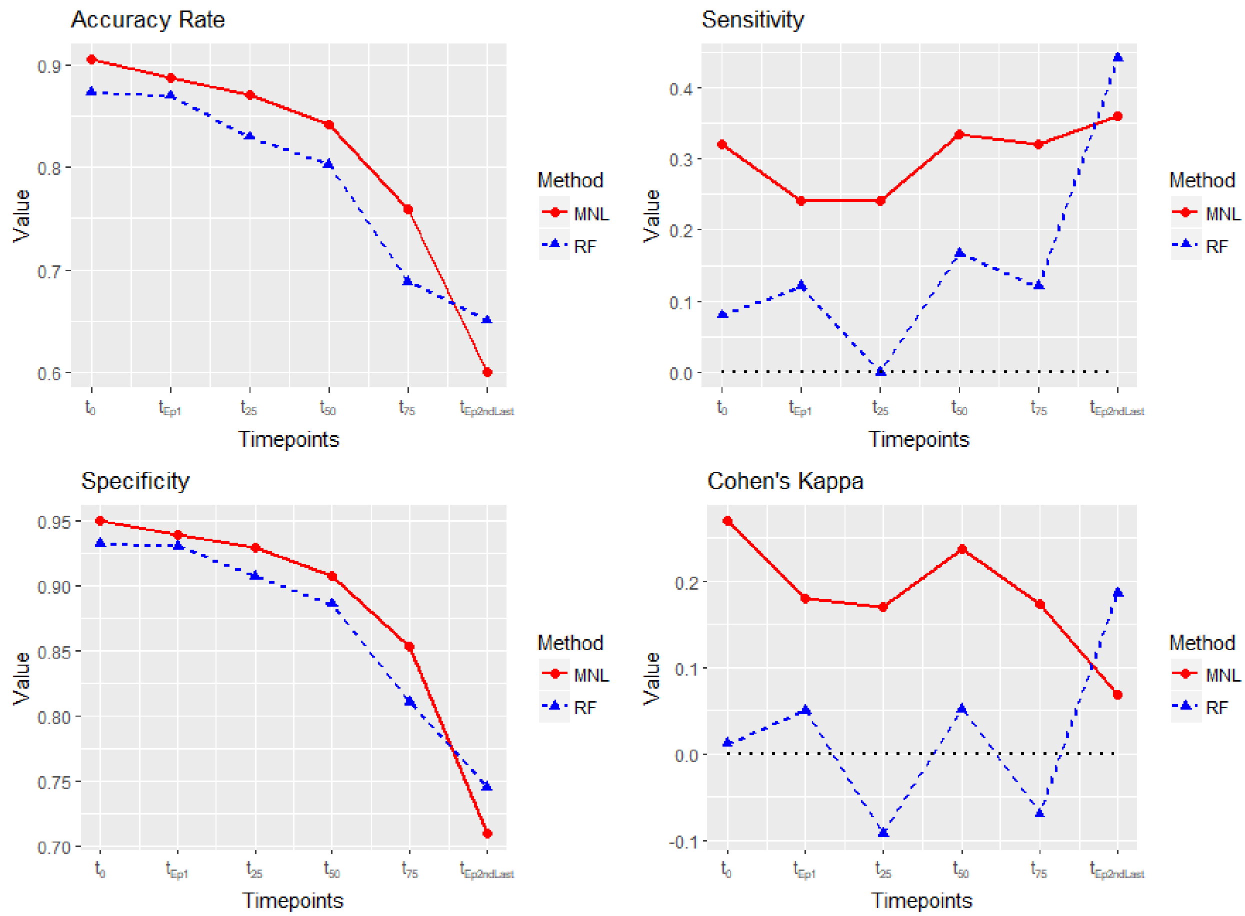


Table 4C.1 contains the Cohen’s Kappa value for the two methods over time. Recall, Cohen’s Kappa returns how similar the predicted values are when comparing the predicted results for the two methods. A negative value indicates the predictions for the two methods were more different then similar, a value of zero indicates they were essentially random, and a positive value indicates the two datasets had similiarties. A value of 1 means the datasets were equivalent.

Table 4C.1: Cohen's Kappa: MNL and RF Predicted Classes Compared

	t_0	t_{Ep1}	t_{25}	t_{50}	t_{75}	$t_{Ep2ndLast}$
Cohen’s Kappa: MNL versus RF Predicted Classes	0.0117	0.0929	-0.0055	-0.0002	-0.1668	-0.0473

5. Discussion

5A. Classification Evaluation by Method

To assess the classification performance across methods, it is useful to consider the accuracy rate, specificity, sensitivity, and Cohen's Kappa. This classification problem consists of two categories, winners and losers. At each timepoint there is a higher percentage of contestants who lost than those who won. Recall, specificity is the percentage of contestants that lost who were correctly identified, and sensitivity is the percentage of contestants who won that were correctly identified. An ideal model would have specificity and sensitivity of 1. However, most procedures do not yield perfect results.

The accuracy rate is one of the most important assessment tools when comparing methods. Recall, the accuracy rate is calculated using Equation 3D.1, and each method's accuracy rates are graphed in Figure 4C.1. Table 5A.1 contains a summary across all 6 timepoints of the accuracy rates for both methods, and the difference between the two methods. The mean difference between the accuracy rate of the MNL model and the RF model is 0.0252. The largest difference between the two models was at timepoint $t_{Ep2ndLast}$ with a difference of -0.05, which was the only instance where the RF method's accuracy rate was higher than the MNL method.

Table 5A.1: Summary of Accuracy Rates

	t_0	t_{Ep1}	t_{25}	t_{50}	t_{75}	$t_{Ep2ndLast}$
MNL Accuracy Rate	0.9061	0.8876	0.8707	0.8413	0.7589	0.6000
RF Accuracy Rate	0.8729	0.8698	0.8299	0.8029	0.6879	0.6500
Difference in Accuracy Rate Between Methods	0.0332	0.0178	0.0408	0.0384	0.071	-0.0500

When adjusting or changing the classification procedure specificity and sensitivity often change as well. Consider a model that classifies every contestant as a loser, specificity would be 1 and sensitivity would be 0. In contrast, if the model classified every contestant as a winner, the reverse would be true. Determining how to balance specificity and sensitivity is based on the problem at hand (Gorunescu, 2011). Thus, when evaluating classification procedures, it is often helpful to consider which type of misclassification would be a greater detriment, incorrectly classifying those who actually won or incorrectly classifying those who actually lost (Gorunescu, 2011). This classification problem is most concerned with correctly identifying the winners, sensitivity, but both measurements will be considered.

Recall, the formulas for specificity and sensitivity are Equation 3D.2 and 3D.3, respectively, and Figure 4C.1 contains a graph of the results for each measurement. Table 5A.2 contains a summary of these values. The mean difference of sensitivity between the two methods is -0.1477, with the MNL outperforming the RF method; however, the method with the highest sensitivity measurement over all timepoints was the RF model with sensitivity 0.44 at $t_{Ep2ndLast}$. The next highest sensitivity level for the RF model was 0.1667 at t_{50} . In contrast, all MNL models had a sensitivity measurement of 0.24 or higher.

Table 5A.2: Summary of Sensitivity and Specificity Values

		t_0	t_{Ep1}	t_{25}	t_{50}	t_{75}	$t_{Ep2ndLast}$
Sensitivity	MNL	0.3200	0.2400	0.2400	0.3333	0.3200	0.3600
	RF	0.0800	0.1200	0.0000	0.1667	0.1200	0.4400
	Difference	-0.2400	-0.1200	-0.2400	-0.1666	-0.2000	0.0800
Specificity	MNL	0.9496	0.9393	0.9294	0.9076	0.8535	0.7091
	RF	0.9318	0.9297	0.90706	0.88587	0.81034	0.7455
	Difference	-0.0178	-0.0096	-0.02234	-0.02173	-0.04316	0.0364

The highest specificity value was for the MNL method at 0.9496 at t_0. The mean difference in specificity across the timepoints was 0.013 in favor of the MNL model. The only instance the RF method outperformed the MNL method was at timepoint $t_{Ep2ndLast}$. In contrast to specificity, sensitivity changed more drastically across methods.

Lastly, Cohen's Kappa can also be used to assess the two methods. Recall Cohen's Kappa measures the similarity of matched data sets after accounting for random chance. The formula for Cohen's Kappa presented in Equation 3D.6. Figure 4C.1 contains a graph of the Cohen's Kappa values comparing the true results and the predicted results for the MNL and RF model. Table 5A.3 contains a summary of these Cohen Kappa values. The mean difference between methods of Cohen's Kappa values is 0.1600 in favor of the MNL method. The only instance when the RF model had a higher Cohen's Kappa value was at $t_{Ep2ndLast}$, which also had the smallest absolute difference between the two methods, 0.1164.

Table 5A.3: Summary of Cohen's Kappa Results

	t_0	t_{Ep1}	t_{25}	t_{50}	t_{75}	$t_{Ep2ndLast}$
Cohen's Kappa: Actual vs MNL Predicted	0.2696	0.1793	0.1694	0.2367	0.1734	0.0691
Cohen's Kappa: Actual vs RF Predicted	0.0118	0.0497	-0.0929	0.0516	-0.0697	0.1855
Difference	-0.2578	-0.1296	-0.2623	-0.1851	-0.2431	0.1164

Overall, the most successful classification method was the MNL approach. The only timepoint the RF method outperformed the MNL method was timepoint $t_{Ep2ndLast}$. Note, the Cohen' Kappa values comparing the predicted results of the two methods presented in Table 4C.1 indicate that the predictions of the two models were not similar for any timepoint.

5B. Classification Evaluation by Time

When assessing the best timepoint, it is useful to consider the same three measures sensitivity, specificity, and Cohen's Kappa value. The highest sensitivity value for both methods occurred at $t_{Ep2ndLast}$, followed by t_{50}. The MNL model's sensitivity value consistently stayed within a range of [0.24, 0.36], but the RF model range was nearly four times larger, [0, 0.44].

The highest specificity value for both methods was at timepoint t_0, each steadily decreasing. However, the MNL method outperformed the RF method until timepoint $t_{Ep2ndLast}$.

The highest Cohen's Kappa value for the MNL model was 0.2696 at timepoint t_0, followed by 0.2367 at timepoint t_{50}. No other models had a Cohen's Kappa value over 0.2000. The largest Cohen's Kappa value for the RF model was 0.1855 at timepoint $t_{Ep2ndLast}$.

In general, there does not appear to be a single best timepoint; however, the MNL model preformed the best at timepoints t_0 and t_{50}, and the RF model preformed best at $t_{Ep2ndLast}$. An advantage of $t_{Ep2ndLast}$ is that there is a more balanced data set between winners and losers, and the game show variables would more accurately represent a contestant's abilities, because the variables consider more elimination challenges. In contrast, t_0 considers more observations and would be more useful to potential contestants and the public. The timepoint t_{50} is less practical then t_0 but has the same advantage of considering game information, as opposed to timepoint t_0.

5C. Variable Assessment

For the MNL model the percentage of models that selected a variable was recorded at six specific timepoints because assessing the individual coefficients in every model was both limiting and computationally expensive. Similarly, the mean Gini index decrease, or variable importance, for a variable at the same six timepoints were recorded for the RF model. These measurements can be compared from timepoint to timepoint but have no practical meaning when comparing across methods. In addition, each measurement can be used to rank the variables in order of most influential. The proportion of models that select a given variable is a measurement between [0, 1], and each nominal variable considered was evaluated as a series of dummy variables. In contrast, the variable importance measurement considers a nominal predictor as a single variable. Thus, although each measurement can be used to rank variables, the ranking comparisons are limited.

Consider the MNL variables in Figures 4A.2 and 4A.3, and Table 4A.2. Only two variables occur at every timepoint, Age and RegionNorthwestArtic. Age is a continuous variable and RegionNorthwestArtic is a dummy variable indicating a contestant's current region of residence. RegionNorthwestArctic is the only variable that occurred in a majority of models for each timepoint; it was also selected more times than any other variable. The next most common region variable was RegionNewEngland which was selected at least once for every timepoint except t_0. The remaining region variables were selected sporadically, occurring in less than 0.10 of the models overall.

Age, occurred in a majority of models for each timepoint except t_{50}, and was the second most selected variable overall. Of the remaining non-region variables LeastWIN occurred most frequently, selected for the majority of the timepoints for most models. Although a few of the remaining non-region variables were selected by the forward stepwise procedure, their selection rate was inconsistent.

Consider the RF variables in Figure 4C.3, and Table 4B.2. Clearly, two variables occurred more frequently than any other, Age and Region. Early in the season, Age carried the highest variable importance at t_0 and t_{Ep1}. For the remaining timepoints, Region had the highest variable importance value. After t_{Ep1}, the variable importance remained somewhat stationary for Age and Region. The remaining variables had a relatively small variable importance value.

Although variable assessment for both methods is limited, Age and Region clearly played a large role for both methods. Looking at the scatter plot of the MNL model's probability at t_0 versus age, the contestants in the Northwest Artic region clearly have a higher probability then the other regions.

Although there are a few other colors that appear near the top of the graph, they are inconsistent. There is also a clear decrease in probability as age increases. In contrast, Figure 4B.4 is the same graph using the RF generated probabilities. The probability still decreases with age, but it is a weaker relationship. There does not appear to be a clear best region when assessing the RF results.

5D. Limitations

There were several limitations to this study, the most evident being the lack of personal contestant information. Personal contestant information varied greatly between programs. Some programs included education level, some included experience, and others included current occupation. In addition, demographic information such as race, sexual orientation, and marital status was nearly nonexistent. The other large limitation was gathering competition results. The process is time consuming, and several aspects were not considered. For example, some competitions offer immunity or other advantages to contestants, which could have caused another contestant to be eliminated prematurely. In addition, there are often mini challenges that happen before the elimination challenge, and team challenges that could affect a contestant's scores.

In future studies, more predictor variables should be considered for both demographic information and game statistic information. In addition, reality-competition programs could be further limited to a single industry. For example, only considering competitions regarding food, fashion, or entertainment. Lastly, there are other classification methods and tools that can be used to evaluate classification performance which could be considered.

6. Conclusion

The aim of this study was to answer the following: Is it possible to predict the winner of a reality competition program? At what point in the competition can a winner be determined? What are the primary factors used to determine the winner? Two methods, the MNL and RF procedures, over six timepoints were used to make predictions.

The MNL preformed the best overall, but it did not successfully classify the majority of the winners, sensitivity ≥ 0.50, for any timepoint. Thus, on average, the correct winner was not selected. However, as illustrated by the Cohen's Kappa values graphed in Figure 4C.1, the predictions were typically better than random guessing. In short, to answer the first question posed, these models help provide insight, but the predictions should not be relied on heavily.

Three different timepoints stood out as the most successful $t_0,\ t_{50},\ and\ t_{Ep2ndLast}$. However, it is typically more advantageous to be able to predict the outcome of a competition as early as possible. For these competitions, a contestant's predicted probability of winning may sway their decision to participate. Timepoint t_0 is the earliest and most practical timepoint at which to predict the winner.

To answer the third question, of the variables considered, Age and Region consistently played a large role in determining a contestant's probability of winning. For the MNL model, the Northwest Artic was the most distinctive region, indicating a contestant having a high chance of winning. For the RF model, a region's influence on probability seemed to vary greatly. In both models, probability decreased as age increased.

To summarize, between these methods the MNL model tended to be most effective overall, indicating that there is a linear relationship between the variables considered. Of the timepoints considered, $t_{Ep2ndLast}$, was the most effective in correctly classifying the winners using the RF method, but timepoint t_0 was the most effective classifier overall, and most practical, using the MNL method. For both models and all timepoints, Region and Age are the most important variables when making predictions. There is some indication that the number of WINs also plays a role, but the results are inconsistent. Overall, the MNL at timepoint t_0 is the best classifier with predictors Region and Age.

7. References

7A. References for Analysis

Agresti, A. (2013). *Categorical Data Analysis* (3rd ed.). Hoboken, NJ: Wiley.

Bonato, A. (2017, December 21). *Survivor: Mathematical predictions of the big win.* Retrieved March 14, 2018, from http://theconversation.com/survivor-mathematical-predictions-of-the-big-win-89162

Breiman, L. (2001). *Random Forests* (Rep.). Berkeley, CA: University of California Berkeley.

Crockett, Z. (2016, August 17). The Man Who Got No Whammies. *Priceconomics*. Retrieved February 20, 2018, from https://priceonomics.com/the-man-who-got-no-whammies/

Gorunescu, F. (2011). Data Mining Concepts, Models and Techniques. Berlin: Springer.

GSA Regions [Digital image]. (n.d.). Retrieved March 14, 2018, from https://www.gsa.gov/about-us/gsa-regions

Haghighat, M., Rastegari, H., & Nourafza, N. (2013). A Review of Data Mining Techniques for Result Prediction in Sports. *Advances in Computer Science: An International Journal*, 2(5), 7-12.

Hoerschelmann, O. (n.d.). Quiz and Game Shows. Retrieved February 20, 2018, from http://www.museum.tv/eotv/quizandgame.htm

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2017). *An Introduction to Statistical Learning: with Applications in R*. New York: Springer.

Kaplan, D. (2006). And the Oscar Goes to: A Logistic Regression Model for Predicting Academy Award Results. *Journal of Applied Economics & Policy*, 25(1).

Karpievitch, Y. V., Hill, E. G., Leclerc, A. P., Dabney, A. R., & Almeida, J. S. (2009). An Introspective Comparison of Random Forest-Based Classifiers for the Analysis of Cluster-Correlated Data by Way of RF. *PLoS ONE*, 4(9). doi:10.1371/journal.pone.000708

Lessmann, S., Sung, M., & Johnson, J. E. (2010). Alternative methods of predicting competitive events: An application in horserace betting markets. *International Journal of Forecasting*, 26(3), 518-536. doi:10.1016/j.ijforecast.2009.12.013

Liaw, A., & Wiener, M. (2002). *Classification and Regression by randomForest*. R News, 2(3), 18-22.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics* (pp. 105-142). New York, NY: Academic Press.

Nelder, J. A., & Wedderburn, R. W. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society*, 135(3), a, 370-384.

Neter, J., Kutner, M. H., Nachtsheim, C. J., & Li, W. (2005). *Applied Linear Statistical Models* (5th ed.). New York, NY: McGraw-Hill Companies Inc.

Pardoe, I., & Simonton, D. K. (2008). Applying discrete choice models to predict Academy Award winners. *Journal of the Royal Statistical Society*: Series A (Statistics in Society), 171(2), a, 375-394. doi:10.1111/j.1467-985x.2007.00518.x

Podlas, K. (2007). Primetime Crimes: Are Reality Television Programs "Illegal Contests" In Violation of Federal Law. *Cardozo Arts & Entertainment Law Journal*, 25(141), 141- 172.

Primetime Emmy Awards Rules and Procedures (Tech.). (2017, April 28). Retrieved February 20, 2018, from Television Academy website: https://www.emmys.com/sites/default/files/Downloads/2017-rules-procedures-v3.pdf

Reilly, N. (2016, September 9). It's been 15 years since Charles Ingram cheated his way to victory on Who Wants To Be A Millionaire. *Metro*. Retrieved February 20, 2018, from http://metro.co.uk/2016/09/09/its-been-15-years-since-charles-ingram-cheated-his-way-to-victory-on-who-wants-to-be-a-millionaire-6119305/

Selvin, S. (1975). On the Monty Hall Problem. *The American Statistician*, 9(3), 67. Retrieved February 20, 2018, from http://www.montyhallproblem.com/as.html

Tournament formats | Match Play Events. (n.d.). Retrieved March 16, 2018, from https://matchplay.events/tournaments#golf_bracket

Validating Nominal Data Sensitivity, Specificity and related measures. (n.d.). Retrieved February 28, 2018, from http://influentialpoints.com/Training/measures_of_validity_for_binary_and_nominal_variables.htm

VanDerWerff, T. (2016, January 07). 750 reality TV shows aired on cable in 2015. Yes, 750. Retrieved February 25, 2018, from https://www.vox.com/2016/1/7/10728206/reality-shows-how-many-peak-tv

Yahr, E., Moore, C., & Chow, E. (2015, May 29). How we went from 'Survivor' to more than 300 reality shows: A complete guide. Retrieved February 20, 2018, from https://www.washingtonpost.com/graphics/entertainment/reality-tv-shows/

7B. References to Build Dataset

A Force to Be Reckoned With. (2014, December 16). Retrieved March 22, 2018, from http://www.syfy.com/faceoff/episodes/season/3/episode/1/a-force-to-be-reckoned-with

Alice in Zombieland. (2014, December 16). Retrieved March 22, 2018, from http://www.syfy.com/faceoff/episodes/season/3/episode/4/alice-in-zombieland

Beyond The Expanse. (2015, October 07). Retrieved March 22, 2018, from http://www.syfy.com/faceoff/episodes/season/9/episode/11/beyond-the-expanse

Boardmen, M. (2017, December 07). RuPaul's Drag Race Season 7 Cast Introduced: Meet the Competitors. Retrieved March 24, 2018, from https://www.usmagazine.com/entertainment/news/rupauls-drag-race-season-7-cast-introduced-meet-the-competitors-2014812/

Cast and Info. (n.d.). Retrieved March 17, 2018, from http://www.syfy.com/faceoff/cast/1

Cast and Info. (n.d.). Retrieved March 17, 2018, from http://www.syfy.com/faceoff/cast/2

Cast and Info. (n.d.). Retrieved March 17, 2018, from http://www.syfy.com/faceoff/cast/3

Cast and Info. (n.d.). Retrieved March 17, 2018, from http://www.syfy.com/faceoff/cast4

Cast and Info. (n.d.). Retrieved March 17, 2018, from http://www.syfy.com/faceoff/cast/7

Cast and Info. (n.d.). Retrieved March 17, 2018, from http://www.syfy.com/faceoff/cast/9

Cast and Info. (n.d.). Retrieved March 17, 2018, from http://www.syfy.com/faceoff/cast/10

Coughland, M. (2012, November 19). RuPaul's Drag Race Season 5 Contestants Announced. Retrieved March 24, 2018, from http://people.com/tv/rupauls-drag-race-season-5-contestants-revealed/

Creature Carnage. (2014, December 16). Retrieved March 22, 2018, from http://www.syfy.com/faceoff/episodes/season/7/episode/13/creature-carnage

Cutforth, D., Lipsitz, J., Minoprio, S., Berwick, F., Cohen, A., ... Cook, E. (Producers). (2008). *Top Chef* [Television series]. New York, NY: Bravo.

Cutforth, D., Lipsitz, J., Minoprio, S., Berwick, F., Cohen, A., ... Leffler, K. (Producers). (2009). *Top Chef* [Television series]. Las Vegas, NV: Bravo.

Cutforth, D., Lipsitz, J., Cook, E., Kriley, C., Cohen, A., ... Cassidy, J. (Producers). (2010). *Top Chef* [Television series]. Washington, DC: Bravo.

Cutforth, D., Lipsitz, J., Strait, N., Kriley, C., Cohen, A., ... Davis, B. (Producers). (2013). *Top Chef* [Television series]. Los Angeles, CA: Bravo.

The Dancing Dead. (2014, December 16). Retrieved March 22, 2018, from http://www.syfy.com/faceoff/episodes/season/1/episode/6/the-dancing-dead

Dangerous Beauty. (2014, December 16). Retrieved March 22, 2018, from http://www.syfy.com/faceoff/episodes/season/2/episode/5/dangerous-beauty

Death's Doorstep. (2016, February 18). Retrieved March 22, 2018, from http://www.syfy.com/faceoff/episodes/season/10/episode/6/deaths-doorstep

Face Off (Season 1). (n.d.). Retrieved March 22, 2018, from http://syfyfaceoff.wikia.com/wiki/Face_Off_(Season_1)

Face Off (Season 2). (n.d.). Retrieved March 22, 2018, from http://syfyfaceoff.wikia.com/wiki/Face_Off_(Season_2)

Face Off (Season 3). (n.d.). Retrieved March 22, 2018, from http://syfyfaceoff.wikia.com/wiki/Face_Off_(Season_3)

Face Off (Season 4). (n.d.). Retrieved March 22, 2018, from http://syfyfaceoff.wikia.com/wiki/Face_Off_(Season_4)

Face Off (Season 7). (n.d.). Retrieved March 22, 2018, from http://syfyfaceoff.wikia.com/wiki/Face_Off_(Season_7)

Face Off (Season 4). (n.d.). Retrieved March 22, 2018, from http://syfyfaceoff.wikia.com/wiki/Face_Off_(Season_4)

Face Off (Season 10). (n.d.). Retrieved March 22, 2018, from http://syfyfaceoff.wikia.com/wiki/Face_Off_(Season_10)

Frightful Fiction. (2015, August 19). Retrieved March 22, 2018, from http://www.syfy.com/faceoff/episodes/season/9/episode/4/frightful-fiction

Intergalactic Zoo. (2015, July 29). Retrieved March 22, 2018, from http://www.syfy.com/faceoff/episodes/season/episode/1/intergalactic-zoo

Keep One Eye Open. (2016, March 17). Retrieved March 22, 2018, from http://www.syfy.com/faceoff/episodes/season/10/episode/10/keep-one-eye-open

Killer Instinct. (2014, December 16). Retrieved March 22, 2018, from http://www.syfy.com/faceoff/episodes/season/7/episode/7/killer-instinct

Lee, S. (2016, February 1). 'RuPaul's Drag Race' season 8 queens and premiere date revealed. Retrieved March 24, 2018, from http://ew.com/article/2016/02/01/rupauls-drag-race-season-8-queens-premiere-date/

Levitt, M., Goularte, J., & Slonina, R. (Producers). (2014). *Skin Wars* [Television series episode]. Los Angeles, CA: Game Show Network.

Levitt, M., Goularte, J., & Slonina, R. (Producers). (2015). *Skin Wars* [Television series episode]. Los Angeles, CA: Game Show Network.

Levitt, M., Goularte, J., & Slonina, R. (Producers). (2016). *Skin Wars* [Television series episode]. Los Angeles, CA: Game Show Network.

Life and Death. (2014, December 16). Retrieved March 22, 2018, from http://www.syfy.com/faceoff/episodes/season/7/episode/1/life-and-death

Make It Reign. (2014, December 16). Retrieved March 22, 2018, from http://www.syfy.com/faceoff/episodes/season/4/episode/1/make-it-reign

Meet the Queens of 'RuPaul's Drag Race' Season 6. (2013, December 10). Retrieved March 24, 2018, from https://www.out.com/entertainment/2013/12/10/meet-queens-rupauls-drag-race-season-6

Overview of the Atlanta Area, Georgia (Metro Area). (n.d.). Retrieved April 04, 2018, from https://statisticalatlas.com/metro-area/Georgia/Atlanta/Overview

Overview of the Boston Area (Metro Area). (n.d.). Retrieved April 04, 2018, from https://statisticalatlas.com/metro-area/Massachusetts/Boston/Overview

Overview of the Chicago Area (Metro Area). (n.d.). Retrieved April 04, 2018, from https://statisticalatlas.com/metro-area/Illinois/Chicago/Overview

Overview of the Dallas Area, Texas (Metro Area). (n.d.). Retrieved April 04, 2018, from https://statisticalatlas.com/metro-area/Texas/Dallas/Overview

Overview of the Detroit Area, Michigan (Metro Area). (n.d.). Retrieved April 04, 2018, from https://statisticalatlas.com/metro-area/Michigan/Detroit/Overview

Overview of the Houston Area, Texas (Metro Area). (n.d.). Retrieved April 04, 2018, from https://statisticalatlas.com/metro-area/Texas/Houston/Overview

Overview of the Los Angeles Area, California (Metro Area). (n.d.). Retrieved April 04, 2018, from https://statisticalatlas.com/metro-area/California/Los-Angeles/Overview

Overview of the Miami Area, Florida (Metro Area). (n.d.). Retrieved April 04, 2018, from https://statisticalatlas.com/metro-area/Florida/Miami/Overview

Overview of the Philadelphia Area (Metro Area). (n.d.). Retrieved April 04, 2018, from https://statisticalatlas.com/metro-area/Pennsylvania/Philadelphia/Overview

Overview of the Phoenix Area, Arizona (Metro Area). (n.d.). Retrieved April 04, 2018, from https://statisticalatlas.com/metro-area/Arizona/Phoenix/Overview

Overview of the San Francisco Area, California (Metro Area). (n.d.). Retrieved April 04, 2018, from https://statisticalatlas.com/metro-area/California/San-Francisco/Overview

Overview of the Washington Area (Metro Area). (n.d.). Retrieved April 04, 2018, from https://statisticalatlas.com/metro-area/District-of-Columbia/Washington/Overview

Population of the New York Area (Metro Area). (n.d.). Retrieved April 04, 2018, from https://statisticalatlas.com/metro-area/New-York/New-York/Population

Return to Oz. (2014, December 16). Retrieved March 22, 2018, from http://www.syfy.com/faceoff/episodes/season/2/episode/1/return-to-oz

RuPaul's Drag Race Season 4: Meet the Cast. (2012, January 5). Retrieved March 24, 2018, from http://www.wetpaint.com/rupauls-drag-race-season-4-meet-the-cast-694196/

RuPaul's Drag Race (Season 4). (n.d.). Retrieved March 22, 2018, from http://rupaulsdragrace.wikia.com/wiki/RuPaul's_Drag_Race_(Season_4)

RuPaul's Drag Race (Season 5). (n.d.). Retrieved March 22, 2018, from http://rupaulsdragrace.wikia.com/wiki/RuPaul's_Drag_Race_(Season_5)

RuPaul's Drag Race (Season 6). (n.d.). Retrieved March 22, 2018, from http://rupaulsdragrace.wikia.com/wiki/RuPaul's_Drag_Race_(Season_6)

RuPaul's Drag Race (Season 7). (n.d.). Retrieved March 22, 2018, from http://rupaulsdragrace.wikia.com/wiki/RuPaul's_Drag_Race_(Season_7)

RuPaul's Drag Race (Season 8). (n.d.). Retrieved March 22, 2018, from http://rupaulsdragrace.wikia.com/wiki/RuPaul's_Drag_Race_(Season_8)

RuPaul's Drag Race (Season 9). (n.d.). Retrieved March 22, 2018, from http://rupaulsdragrace.wikia.com/wiki/RuPaul's_Drag_Race_(Season_9)

Scene of the Crime. (2014, December 16). Retrieved March 22, 2018, from http://www.syfy.com/faceoff/episodes/season/3/episode/10/scene-of-the-crime

The 150 Largest Cities in the World. (2018, January 17). Retrieved March 24, 2018, from http://www.worldatlas.com/citypops.htm

Twisted Trees. (2014, December 16). Retrieved March 22, 2018, from http://www.syfy.com/faceoff/episodes/season/7/episode/4/twisted-trees

Wanted Dead or Alive. (2016, January 14). Retrieved March 22, 2018, from http://www.syfy.com/faceoff/episodes/season/10/episode/1/wanted-dead-or-alive

Weinstein, H., Weinstein, B., Cutler, J., Gruber, D., Klum, H., ... Fletcher, L. (Producers). (2010). *Project Runway* [Television series]. New York, NY: Lifetime.

Weinstein, H., Weinstein, B., Murray, J., Gruber, D., Klum, H., ... Norton, T. (Producers). (2011). *Project Runway* [Television series]. New York, NY: Lifetime.

Weinstein, H., Weinstein, B., Murray, J., Gruber, D., Klum, H., ... McCarthy, G. (Producers). (2012). *Project Runway* [Television series]. New York, NY: Lifetime.

Weinstein, H., Weinstein, B., Murray, J., Gruber, D., Klum, H., ... McCarthy, G. (Producers). (2013). *Project Runway* [Television series]. New York, NY: Lifetime.

Weinstein, H., Weinstein, B., Murray, J., Gruber, D., Klum, H., ... McCarthy, G. (Producers). (2014). *Project Runway* [Television series]. New York, NY: Lifetime.

Weinstein, H., Weinstein, B., Murray, J., Gruber, D., Klum, H., ... McCarthy, G. (Producers). (2014). *Project Runway* [Television series]. New York, NY: Lifetime.

Welcome to the Jungle. (2014, December 16). Retrieved March 22, 2018, from http://www.syfy.com/faceoff/episodes/season/1/episode/1/welcome-to-the-jungle

When Hell Freezes Over. (2014, December 16). Retrieved March 22, 2018, from http://www.syfy.com/faceoff/episodes/season/4/episode/3/when-hell-freezes-over

The Ultimate Spotlight Challenge. (2014, December 16). Retrieved March 22, 2018, from http://www.syfy.com/faceoff/episodes/season/2/episode/10/the-ultimate-spotlight-challenge