International Journal of Statistics and Applied Mathematics 2018; 3(2): 101-106

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2018; 3(2): 101-106

www.mathsjournal.com
Received: 12-01-2018
Accepted: 15-02-2018

Joseph Roith
St. Catherine University,
St. Paul, Minnesota, USA

Rhonda Magel
North Dakota State University,
Fargo, North Dakota, U.S.

Correspondence
Rhonda Magel
North Dakota State University,
Fargo, North Dakota, U.S.

Predicting NFL football games based on simulation and modeling

Joseph Roith and Rhonda Magel

Abstract
Discriminant analysis is conducted to help determine which variables in an NFL football game are more important to the outcome. Discriminant analysis is also used to determine whether it is more productive to have a better offense or a better defense of the opposing team. A simulation technique is introduced and used to predict outcomes of NFL football games based on the three models of NFL football games developed in Roith and Magel (2017) and the discriminant analysis results derived in this research. Doing simulations with one model, we were able to predict the outcomes of 71% of the NFL football games considered correctly. Models were developed based on three years of NFL football games and predictions using simulations of these models were done for another year of NFL games.

Keywords: Least Squares Regression; Logistic Regression; Proportional Odds; Discriminant Analysis; Simulation

1. Introduction

The National Football League was founded on August 20, 1920 in Canton, Ohio. The league originally consisted of fourteen teams. It now has thirty-two teams across the United States. Currently, every season these teams play each other in three types of games; exhibition, regular season, and playoffs. Exhibition games start the season and do not count towards a team's spot in the standings. There are sixteen regular season games played by each team over the course of seventeen weeks that determine the relative position in the standings based on the win-loss record of the team. After the regular season, teams are ranked by record, and participate in a four round, single elimination playoff tournament to decide a league champion in the Super Bowl [1].

The league is divided into two conferences, the National Football Conference (NFC), and the American Football Conference (AFC). Each conference consists of sixteen teams separated into four divisions of four teams. At the end of the regular season schedule, the top team in every division secures a playoff berth. Then, from the remaining teams, the two in each conference with the best records, called wildcards, also make the playoffs. Thus, there are six playoff teams from both the NFC and AFC, and the seeding are as follows in each league [1].

Seed 1: Division winner with best record
Seed 2: Division winner with second best record
Seed 3: Division winner with third best record
Seed 4: Division winner with fourth best record
Seed 5: Wildcard team with best record
Seed 6: Wildcard team with second best record

Roith and Magel [2] developed various models based on NFL football games using data from all games played between 2011 and 2013 [3]. In developing these models, 17 differences of in-game statistics were considered in each case and stepwise selection techniques were used to help determine which of these differences were significant in each model. One model they developed was a point spread model that helped explain 82.79% of the variation in point spread of a football game based on the values of the differences of six known in-game statistics between the two teams playing. When the model was used on the testing data set (games played in 2014), it gave the correct winner about 86% of the time when these differences were known. Roith and Magel also developed logistic models that estimated the probability of the home team winning based on the differences of five in-game statistics. When this model was used on the testing data set, it predicted 86% of the games correctly as to which team would win when a home team was predicted to win if the model gave a probability greater than 0.5, or predicted to lose if the model gave a probability of less than 0.5 of winning for the home team. A third model was the proportional odds model which estimated the probability of a home team winning the football game by 10 points or more, winning by less than 10 points, losing by less than 10 points, or losing by 10 points or more. When this model was used on the testing data set with the differences of the in-game statistics given, the model had an accuracy of 70.6% of placing the game in the correct category for the home team, and had an 88.9% accuracy of correctly getting the winner. It is noted, that this high accuracy depends on knowing values of the 5 or 6 in-game statistics in which the models are based on, which is not known ahead of time. These models rather "explain" rather than "predict" the point spread or probability of a home team winning the game. The point spread model developed by Roith and Magel [2] when differences of in-game statistics are known is given by Eq 1 as an illustration of one of the models that were developed. The logistic and proportional odds models may be found in Roith and Magel[2].

Point Spread = 1.00306 + 1.37997*(First Down Margin) – 0.53459*(Total Play Margin) + 1.00567*(Yards per Pass Margin) – 3.88568*(Turnover Margin) + 0.17715*(3rd Down Conversion Percent Margin) – 0.12464*(Yards Lost to Sacks Margin) (Eq 1)

Other recent literature on the NFL on predicting the outcome of football games is focused forecasting professional football games with respect to the efficiency of the betting market (see Glickman & Stern [4] and Stern [5], and Baker and McHale [6]; and Boulier and Stekler [7]). Harville [8] and Stefani [9] looked at games from the 1970's to create a predictive model and found much more success than more current research, but teams might have been a lot further apart at that point as far as ability, and therefore, games would have been easier to predict.

Long and Magel [10] considered football games at the collegiate level, analyzing games from the NCAA Division I Football Championship Subdivision. Here they used regression techniques to identify significant in game statistics and develop prediction models for the outcome of games. They concluded that six factors contribute to wins for collegiate teams, difference in turnovers, difference in the probability of pass completion, difference in the probability of a 3rd down conversion, difference in the number of sacks, difference in the number of punt returns, and difference in the number of offensive yards per play. Combining estimates of these differences along with a computer ranking of the individual teams playing, they were able to predict the correct winner of games around 73% of the time.

In other sports, Roith and Magel [11] use discriminant analysis to compare the importance of offensive versus defensive statistics in the National Hockey League. Unruh and Magel [12] considered NCAA Division I basketball games and the important influencers that determine the winners of individual games, along with predicting the winners of the championship tournament at the end of each season. Ultimately, their models performed with around 67% accuracy. So we can see that modeling individual games for most sports seems to have some limitations in predictive power regardless of the sport. In each case, the most accurate models can correctly predict the winner of the contest around 65-70% of the time.

2. Materials and Methods

In this paper, we would like to expand the work of Roith and Magel [2] by applying two additional techniques to predict or explain game outcomes; discriminant analysis, and simulation. In this case, discriminant analysis will be used to determine the most important factors of classifying an NFL football game into the two categories "home team win", and "home team lost" for regular season games based on significant differences of the in-game statistics between the two teams. Discriminant analysis creates a linear function from the significant in-game statistics differences so that the separation of groups (home team winning and home team loosing) is maximized (Rencher [13]). Stepwise selection was used with an alpha level for entry of 0.20 and an alpha level for stay equal to 0.25 to help determine the significant differences of the in-game statistics. The magnitudes of each of the in-game statistics differences were determined using standardized coefficients. Cross validation will be used to determine how well the discriminant analysis was doing in classifying a win or a loss for the home team (Rencher [13]). An attempt will also be made using discriminant analysis to determine if offensive or defensive factors make more of a difference in whether or not a home football team wins the game.

We will use simulation techniques to simulate the results of a football game and to make predictions using based on the three models given earlier in Roith and Magel [2] and also based on the results from the discriminant analysis in this paper.

Roith and Magel [2] found the in-game statistics they collected follow approximately normal distributions. We will assume that the distributions of the in-game statistics are normal when doing the simulations. The mean and variance has to be estimated for each of these normal distributions of the in-game statistics. When doing the simulations to predict winners of a football game, we considered each team that was to play in the game and computed the mean and variance of every in-game statistic that was used in the models for a team playing in the game that we wanted to predict based on the games that the team has already played. We then assumed that the values of a particular in-game statistic for each game followed a normal distribution with the mean equal to the mean for that statistic that the team had played so far and the variance equal to the variance of values of that statistic for the games that the team had so far. We then randomly generated a value of the in-game statistic being considered by simulating a sample of size one from a normal distribution with the appropriate mean and variance. This was done for each statistic and for both teams playing in the game. The differences were then taken between the two games for each of the corresponding statistics. These values were placed into each of the models that were developed and a winner predicted. This process was repeated 10,000 times.

For example, in the point spread model, there were six in-game statistics found to be significant (first down margin, total play margin, yards per pass margin, turnover margin, 3rd down conversion percent margin, and yards lost to sacks margin). We do simulation using the point spread model to predict the outcome of a game between two teams during Week 9 of the NFL season. Values will need to be simulated to replace the actual in-game statistics for both teams and then differences will be taken. We will simulate 10,000 sets of in-game statistics, and estimate the point spread 10,000 times. The team predicted to win, will be the team that had a positive point spread for more than half of the 10,000 simulations. We will do this for several games and then compare the actual game results to the results obtained from simulation.

As stated earlier, we will need to simulate 10,000 sets of in-game statistics. Two methods will be used to simulate these 10,000 sets. Method 1 will use only offensive statistics from both teams. Method 2 will use both offensive and defensive statistics for both teams.

One in-game statistic needed for both teams in the point spread model is the total number of plays. We will first discuss doing simulations based on Method 1. For both teams (home and away) we can find their mean and variance of total plays from the previous eight weeks. We can take this mean and variance for the home team and use these as the actual mean and variance of the in-game statistic of total plays for the home team. We can then simulate 10,000 values from a normal distribution with that mean and variance to represent a sample of the total number of plays that the home team would have per game. We could also do this simulation for the away team based on the mean and variance of the total number of plays they had in their last 8 games. Differences of these 10,000 sets of values can be taken and placed into the point spread model in place of total play margin for 10,000 times. This could also be done to simulate the first down margin. We could find the average number of first downs that the home team got during their last 8 games as well as the variance of the number of first downs for the home team. We would then simulate 10,000 values from a normal distributions having that mean and variance to represent a sample of the number of first downs that the home team could receive. Likewise, we could do the same for the away team, and take the differences of these pairs of 10,000 values to place into the point spread model in place of first down margin. This technique could be used for all six in-game statistics in the point spread model, and 10,000 estimates of the point spread could be found. If the point spread is positive, this would indicate a win for the home team, and if it is negative, it would indicate a loss for the home team. We could then count the number of times, out of 10,000 that the point spread would be positive for the home team and divide this by the 10,000 which is the number of simulations. This value will give us the estimated probability that the home team will win the game based on using the point spread model with the generated values of the in-game statistics. In this case, only offensive statistics have been taken into consideration. We have not considered how these statistics compared each week with the teams they were playing.

For the second part of the simulations, we will also take into consideration the defensive statistics for both teams each week. For the point spread model, we need to estimate the differences of 6 in-game statistics, such as the difference between the number of total plays made by the home team and the number of plays made by the away team. Taking the defensive aspect of football into consideration also, we will average the total plays of the home team for the last 8 games (*HomeTotPly)*, and we will average the total plays of the last 8 games of the teams that they were playing (*HomeTotPlyAllowed).* We will also do this for the last 8 games that the away team has played. We will average the total plays for the last 8 games that the away team has made (*AwayTotPly)*, and we will average the total plays that each team playing the away team has made during the last 8 games (*AwayTotPlyAllowed).* We can then generate 10,000 values from four normal distributions with means equal to the sample means and standard deviations equal to the sample standard deviations based on these last games. With all of these numbers, 10,000 values of a new marginal statistic can be computed and entered into both the point spread and logistic models developed by Roith and Magel (2017) to replace Total Play Margin. Here is an example of how we will calculate the new marginal measure using the total yards gained and allowed by the home and away teams:

$$\left[\frac{(Home\ TotPly + Away\ TotPly\ Allowed)}{2}\right] - \left[\frac{(Away\ TotPly + Home\ TotPly\ Allowed)}{2}\right]$$

Then the fitted results of 10,000 simulated games can be viewed as a whole and we will be able to provide an estimated probability of the home team winning the game based on these simulations to compare with the observed results from the actual game. This procedure can also be repeated for the remaining 5 in-game statistic differences.

We will need at least four games played prior to create our values of mean and standard deviation for the six variables in the point spread model, so only games played after Week 4 will be considered. That means we will study the 193 games from the testing data set after that point in the season. We will evaluate the OLS regression, logistic, discriminant analysis, and proportional odds models for individual games only, using the simulations. The offensive statistics simulated for each team will be used first to create the marginal variables, for instance, the simulated number of passing yards for the home team minus the simulated number of passing yards for the away team. Next, we will use the simulated values for each significant variable gained by each team, and allowed by each team to create a marginal value in the manner introduced above. This will represent the presence of defensive capabilities along with offensive prowess for both teams.

We will compare our simulation results using each of the models developed in Roith and Magel [2] and the results obtained using the discriminant analysis conducted in this paper. We will also compare the simulation results to the naïve approach of selecting the home team to win. We do not expect to be able to predict games with any specific amount of certainty, but any accuracy we do find should help indicate an idea of the level of consistency of performances from week to week. Simulation also provides a way to look at our fitted responses for models such as the point spread model, and determine the probability of a range of different outcomes. Just because the model provides a fitted value, we also want to know the distribution of the outcome, this can tell us if there is a wide variation in the simulated outcome or if we can be fairly confident in our predicted value.

3. Results and Discussions

3.1 Results – Discriminant Analysis

When using discriminant analysis we first considered the game data, using the classes “Home Win” and “Home Loss” as our categories. Using the stepwise selection process, seven significant differences (margins) of in-game statistics were found. All were significant at 0.05.

Table 3 shows the standardized linear discriminant functions for our two groups. In both cases, turnovers contribute the most towards an observation being classified into that group. The function that produces the largest value determines the group the data point is classified into. For every increase in the turnover margin on behalf of the home team, the discriminant function will penalize classification as a home win more than for any other category. The conversion percentage for 3rd down is another variable that has a large magnitude for both groups.

The sign of the coefficient is also informative, and makes natural sense in most cases. An increase in yards per pass margin is beneficial to the home team and increases the classification value for a home win, while at the same time, decreasing the classification value for a home loss. Total number of plays margin is the only variable that seems to be counter intuitive, but this may be due to the efficiency of a team throughout the game.

Table 3: Standardized Linear Discriminant Functions for Game Model

Variable	Home Loss	Home Win
TurnoverM	0.79468	-0.82641
FirstDownM	-0.34551	0.78191
TotalPlayM	0.27263	-0.64415
3DPerM	-0.59911	0.52289
YPPassM	-0.45613	0.40363
SackedM	0.33322	-0.28138
PenaltyM	0.18074	-0.27553

Cross validation of the training data was used as a way to consider the performance of the discriminant functions. With the hold out method, 195 out of 215 home losses were correctly classified, and 247 out of 282 home wins were correctly classified. That corresponds to an 89% accuracy rate for classifying a home win using these linear discriminant functions (Table 4).

Table 4: Cross Validation Summary of Discriminant Functions for Game Model

Home Wins	*Classified Loss*	*Classified Win*	Total
Observed Loss	195	20	215
Observed Win	35	247	282
Total	230	267	497
Error Rate	0.0930	0.1241	**0.1086**

When the discriminant functions are applied to the testing data set, 219 out of the 256 games were correctly grouped, an 85.5% accuracy. This is almost identical to the other models when the testing data set was considered. Therefore, the discriminant functions can be generalized to include any game played outside of the training data set, while still played under the same set of rules and conditions.

There are a lot of similarities between the results from discriminant analysis and the two models developed by Roith and Magel [2] including the point spread and the logistic model. However, one area that may have been overlooked is 3rd down conversion percentage margin. It had a smaller coefficient in the point spread and logistic models, but it is clear here that this has a significant impact on whether or not a team will win the game.

Discriminant analysis was also used to compare offensive and defensive abilities and to determine one is more important than the other. Instead of the differences (margins) data used in the point spread and logistic models of Roith and Magel [2], we considered only the individual team totals for both the home and away team. However, now we had a designation of offensive amount gained, or defensive amount given up for each in-game statistic. In this case we used the offensive and defensive pairs for turnovers, yards per pass, 3rd down conversion percentage, sack yards, and first downs. Penalties were omitted since they could be either offensive or defensive. Table 5 displays the linear discriminant functions for the two classes.

To analyze these functions, we will need to compare the sets of variables. For instance, committing a turnover is a measurement of offense, and for every turnover committed, a team will move away from a "Win" classification by about 0.504 standardized units. However, if a team forces a turnover, which is a defensive measurement, they will move towards a Win classification by 0.713 standardized units. For turnovers, it would be more beneficial to have one more turnover forced than one less turnover committed. Defense has the clear advantage for turnovers.

Table 5: Standardized Linear Discriminant Functions for Offense vs. Defense Model

Variable	Lose	Win
TurnoverCommit	0.51450	-0.50449
TurnoverForce	-0.72730	0.71315
YPPassGain	-0.47315	0.46394
YPPassAllow	0.31672	-0.31056
Per3D	-0.34140	0.33476
Per3DAllow	0.46584	-0.45678
SackYdsLost	0.25987	-0.25481
SackYdsGain	-0.22023	0.21594
FDGain	-0.36894	0.36176
FDAllow	-0.03998	0.03920

For the yards per pass variable, we see the opposite effect, the coefficient for the "Win" class has a magnitude of 0.464 for the yards per pass gained, while the yards per pass against measure has a magnitude of 0.311. When considering yards per pass, it is advantageous to be more efficient on offense than it is to prevent your opponent from having a lot of yards per pass. Similarly, for first downs, it seems as though offense is more important, making sure your team is moving down the field.

The remaining variables all favor the defensive side. A team should focus more on stopping their opponent on 3rd downs than converting their own, if they want to maximize their probability of winning the game. Sacks are also more significant from the defensive side, giving up sack yardage is not as critical as is sacking the opposing quarterback. To be clear, all of these variables will affect the game significantly, but the discriminant functions let us see that those effects are not necessarily equal.

In summary, for areas such as turnovers, and 3rd down percentage, a team would be better off with a defensive mindset. As for yards per pass, first downs, and preventing sacks, an offensive strategy will help a team take full advantage of their chance to win the game

3.2 Simulation Results

We will now consider simulation results of games. In our testing data set, all games after Week 4 in 2014 were selected, 193 in total. For each of these games, the two teams involved had the means and standard deviations calculated for each of the variables in the models based on all of the games they played leading up to the game of interest. Then, from these statistics, 10,000 game simulations were created based on the marginal statistics of offensive performance, and also a combination of offense and defense. The results for each model were compared to those of the actual game results.

One example of simulations used to forecast a game outcome is presented here, using Method 2. On November 3rd 2013, during Week 9 of the NFL season, the Kansas City Chiefs played at the Buffalo Bills. We will illustrate using the point spread regression model on the simulated data. Table 6 shows the average values and standard deviations for all of the variables used, including the average values allowed by each team over the previous eight weeks. These are the numbers that were used to simulate 10,000 games being played, and the marginal variables were calculated for each one. Next, the marginal values are entered into the model and the average point spread for all the simulations is calculated. For this case, the average point spread was -3.7 using only the offensive numbers for each team, and -6.1 using the combination of both offensive and defensive statistics. This represents that we would expect the away team, the Chiefs, to win the game by somewhere around four to six points. The actual outcome of the game saw the Chiefs win by a margin of ten points. So we were correct in selecting the winner using both marginal statistics, and it appears that using the combination of offense and defense provided a fitted point spread that was a little closer to the actual point spread.

Table 6: Averages and Standard Deviations Through Week 8

Variable	Chiefs	Chiefs - Allowed	Bills	Bills - Allowed
First Downs	Mean = 19.0 SD = 1.69	16.0 3.93	18.88 2.85	21.38 3.96
Total Plays	Mean = 67.75 SD = 5.036	62.25 6.84	70.75 5.83	71.88 10.08
Yards per Pass	Mean = 5.78 SD = 1.24	5.9 1.94	5.8 0.89	7.10 2.33
Turnovers	Mean = 1.0 SD = 1.19	2.5 1.31	1.63 1.06	1.88 1.73
3rd Down Percentage	Mean = 36.02 SD = 15.47	25.54 7.87	35.97 9.35	37.88 11.99
Yards Lost to Sack	Mean = 15.63 SD = 11.38	31.75 20.38	19.5 10.60	21.63 12.74

The overall percentage of simulated games won by the Chiefs was 76%, and while the average point spread value was -6.1, clearly there were many simulated games that resulted in a point spread close to -10, that of the actual game.

This process was repeated for 193 games, and for each of the individual game models. Table 7 summarizes the results and accuracy of each model when forecasting the outcome of games using historical team performance. Included in the table are the percentages of each model correctly choosing the winner based on the actual in game statistics observed.

For projecting games without any knowledge from the event itself, we start with the naïve method of choosing a winner of the game by selecting the home team to win. With this technique, 57% of the games were correctly selected. For predicting the point spread, using a combination of both offensive and defensive statistics for each team resulted in finding the actual winner 67% of the time. The mean absolute deviation of the fitted point spread to the observed point spread was just under ten points, meaning that is how close our fitted value was to the observed value on average.

For the models we developed that only try to classify the game as a Win or a Loss, the best performing models were the proportional odds models. While they were only around 35% in predicting the actual category the final point spread fell into, the correct winner was selected for 71% of the games. Overall, for each of the models, it seems that simulating games using both offensive and defensive statistics provides a more accurate forecast. This would seem to make intuitive sense, since when we consider offensive values, we are only looking at half of the picture. Simulation seemed to create a lot of conservative average point spreads, with few games being classified as blowouts. But those models were just as successful as the others when we considered only if they produced the correct winner.

Table 7: Summary of Model Forecasting Accuracy

Model	Using In Game Statistics	Offense Only Forecast	Offense and Defense Forecast
Naïve (Home Team Wins)	57%	57%	57%
Point Spread	85.9%	65%	67%
Logistic	85.9%	63%	66%
Discriminant Functions	85.5%	63%	67%
Proportional Odds (Correct Category)	68.8%	37%	35%
Proportional Odds (Correct Winner)	86.3%	64%	71%

Overall, every one of the models we developed had an accuracy of around 63% or higher, much better than just choosing the home team to win. This is further evidence that those variables we have identified as significant in explaining success in the NFL are indeed important. Also, this illustrates the capabilities that simulation has when looking forward to games that have not yet occurred.

4. Conclusions

It is also important to consider the amount that each of these significant variables contributes. The question of offense versus defense, or the old axiom that defense wins championships, is absolutely something that should be confirmed or discredited. The truth is, there are no universal rules that say either offense or defense is better than the other. It needs to be examined on a statistic by statistic basis, and once the value of each is quantified, it can be exploited to benefit of the teams that are willing to do so.

When it comes to turnovers, it is better for your team to create them, than try to prevent committing them. Either way, both will affect the outcome of the game, but you can expect a larger return for creating a turnover then preventing one. This can lead to different strategies, such as being more aggressive on defense by trying to intercept more passes, or trying to cause more fumbles.

On the other hand, when you consider passing productivity, it is more important to have an offense that produces more yards per pass than it is to have a defense that prevents a larger yard per pass value. So this, along with the fact that offensive turnovers do not hurt as much, would suggest an approach that is more aggressive in passing offense. An example would be calling more plays that will result in greater passing yardage, even if the risk of an interception is slightly increased.

With all of this information available to each team and the public, it is important that NFL teams use it effectively. There is still some resistance throughout the league to use analytics for improved decision making off the field, and enhanced performance on it. Hopefully, the continued analysis and application of statistics in sports will convince those who are in the position of making decisions that soon they will be at a disadvantage if they fail to make the most of these underlying tendencies in the game of football.

5. References

1. Pro Football Reference. (2016). Retrieved November 2014 from https://www.pro-football-reference.com/years/2016/#all_team_stats.
2. Roith J, Magel R. Modeling NFL Outcomes, International Journal of Mathematics & Statistics, 2017; 3:12.
3. NFL Scores. [2009-2013]. Retrieved July 2014, from ESPN.com: http://scores.espn.go.com/nfl/scoreboard
4. Glickman ME, Stern HS. A State-Space Model for National Football League Scores. Journal of the American Statistical Association, 1998; 93:441, 25-35.
5. Stern HS. On the Probability of Winning a Football Game. The American Statistician, 1991; 45(3):179-183.
6. Baker RD, McHale IG. Forecasting exact scores in National Football League games. International Journal of Forecasting, 2013; 29:122–130.
7. Boulier BL, Stekler HO. Predicting the outcomes of National Football League games. International Journal of Forecasting, 2003; 19:257–270.
8. Harville D. Predictions for National Football League Games Via Linear-Model Methodology. Journal of the American Statistical Association, 1980; 75:371, 516-524.
9. Stefani RT. Improved Least Squares Football, Basketball, and Soccer Predictions. IEEE Transactions on Systems, Man and Cybernetics. 1980, 10(2):116-123.
10. Long J, Magel R. Identifying Significant In-Game Statistics and Developing Prediction Models for Outcomes of NCAA Division 1 Football Championship Subdivision (FCS) Games. Journal of Statistical Science and Application. 2013; 1(1):51-62.
11. Roith J, Magel R. An Analysis of Factors Contributing to Wins in the National Hockey League. International Journal of Sports Science, 2014; 4(3):84-90.
12. Unruh S, Magel R. Determining Factors Influencing the Outcome of College Basketball Games. Open Journal of Statistics. 2013; 3:225-230.
13. Rencher AC. Methods of Multivariate Analysis (2nd ed.) New York: Wiley. Predictions. IEEE Transactions on Systems, Man and Cybernetics. 2002; 10(2):116-123.