TEAM ORGANIZATION FOR SENIOR SOFTBALL (TOSS)

Submitted to the Faculty of the

GEORGE MASON UNIVERSITY

In partial fulfillment of the requirements for the

Degree of Masters of Operations Research/Systems Engineering

By:

Trevor Caskey

Lisa Kaiser

Adam Robison

Christopher Teixeira

Abstract

The teams in the Northern Virginia Senior Softball (NVSS) league are reorganized every year with the desire to make them as even as possible and promote social interactions. This project's goals were to develop a methodology to produce team assignments for players such that the teams are equal in regards to player ability and other attributes. Multiple regression was employed to obtain offensive, defensive, and team metrics. An integer program (IP) model was developed where constraints were aligned with the goals of the NVSS league to solve the assignment problem. The IP model was linked to a team assignment tool which allows the user to select settings and parameters before running the model.

Acknowledgements

Our team would like to thank the following individuals and organizations for their time, effort, and guidance throughout the project:

Dr. Andrew Loerch

Joyce Pfeffer

The Northern Virginia Senior Softball League

Table of Contents

Tables and Figures

1. Introduction

Northern Virginia Senior Softball (NVSS) league is a nonprofit, co-ed, slow pitch softball league for seniors. The teams in the league are reorganized every year with the desire to make them as even as possible and promote social interactions. NVSS is seeking assistance in developing a methodology to produce team assignments for players such that the teams are equal in regards to player ability and other attributes.

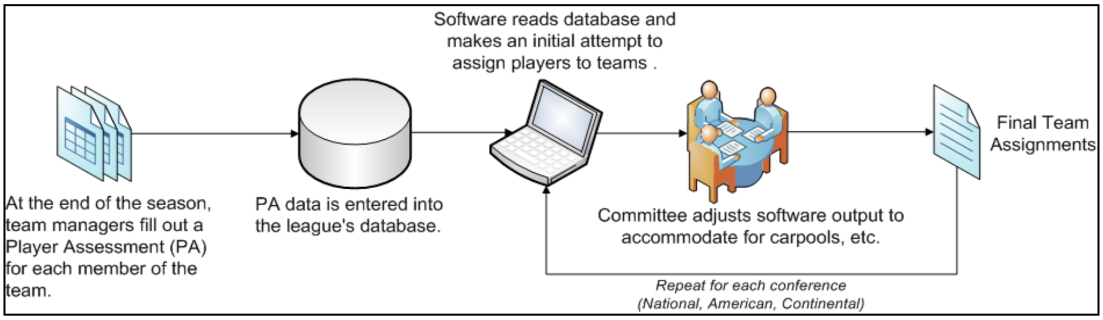
1.1 Background

NVSS began in the summer of 1980 with 47 players. As of 2010, the league has over 500 members on 25 to 30 teams. The league is sorted into three conferences based upon skill level: National, American, and Continental. The league requires men to be 50 or older and women to be 40 or older. The spring season is from mid-April to August, and the fall season is from September to October. Some players play both seasons while others may play only one season. They play two games on Tuesdays and Thursdays which are 7 innings each. The games are played with traditional softball rules with a few exceptions to improve gameplay and competitiveness within the leagues. At the end of each spring season, each team manager fills out an NVSS Player Assessment for each player on their team. The NVSS Player Assessment includes the number of games played by the player, times at bat, number of hits, total number of extra bases, primary position, secondary position, versatility score, and base running skill. These statistics are used to calculate other statistics, such as the player's on base average and extra base average. The team was also able to obtain two other data sets which contain personal information, such gender and age. All new players are given a skill level assessment before they start playing in the league.

1.2 Current NVSS Process

The current system is a two-step process: 1) NVSS player assignment software and 2) the committee. This process produces an initial list of players per team based on individual player statistics taken from the player assessments. The initial list is followed by manual updates by a committee to incorporate attributes not included in the model. This process is repeated for each conference.

Figure 1. Current NVSS Process



Players are assigned to teams using a heuristic approach starting with assigning the managers which are manually set, then players with higher speed and power, and finally defensive positions based upon a preset order. Following the initial model run, the committee makes manual updates to incorporate social requirements (carpooling), political requirements (teammate restrictions), and the distribution of new players. After many manual updates, the output statistics are computed again and the committee returns to the drawing board for more manual updates. The NVSS's current process to generate teams takes about four afternoons.

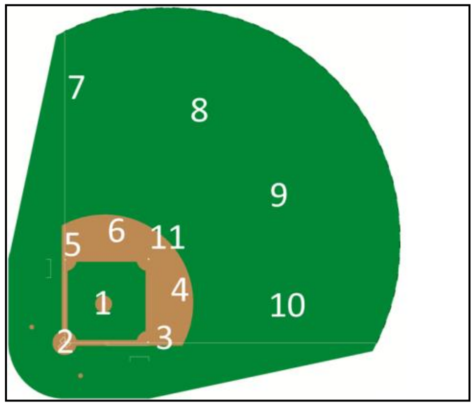
1.3 Problem Statement, Purpose, and Objectives

Given the player ability assessments, the problem is to develop a methodology to create teams of equal ability by choice of skill and position subject to carpools, managers, personality and availability. The purpose of this project is to develop a team generation tool which provides the NVSS league with equal teams. The objectives are to: 1) develop an optimization model to create balanced teams for the three different softball conferences, 2) perform statistical analysis to determine the significant metrics, if any, and 3) provide more functionality to assigning players with less manual user input.

1.4 Scope

The scope includes all three conferences of the Northern Virginia Senior Softball (NVSS) League: National, American, and Continental. The conferences are separated based on skill level, with the most talented players in the National Conference, followed by the American and Continental, respectively. The scope also includes all positions: one pitcher, one catcher, five infielders, and four outfielders for a total of eleven players on the field.

Figure 2. Player Positions



1.5 Limitations

There are several limitations to this study. The first limitation is our team is limited to the statistics which the NVSS has collected over the past 2010 spring season. The second limitation is the team is limited to free and open source software. NVSS is not interested in purchasing software to run the model. The third limitation is the methodology cannot incorporate all aspects associated with generating equal teams, due to time and resource restrictions.

1.6 NVSS Goals

NVSS has requested that the model evenly distribute players based on the following attributes:

- New Players
- Players with Pitching Experience
- Players by positions
- Base running ability
- Part – Time Players (Players that can play only Tuesdays or Thursdays.)
- Versatile Players (Players that can play multiple positions.)
- Power Hitters
- Not assigned to same team more than two years in a row
- Tournament Team players
- Match up carpools

2. Statistical Methods

2.1 Multiple Regression

Multiple regression is an extension of a simple linear regression (SLR) model. Similar assumptions are used within the models, but the multiple regression model allows for more flexibility than the SLR model. The differences lie within the predictor variables and parameters. The SLR model is represented by a function of the response variable Y, predictor variable Z, regressor variable X, parameters β_i, and random error term ε. The equation for the SLR model,

$$Y = \beta_0 + \beta_1 * X(Z) + \varepsilon$$

is a simple additive model. The multiple regression model is a more complex equation that involves the summation of the product of multiple predictors, parameters, and regressors. In general, the multiple regression model can take on several different variations of the regressors. For this work, the generalized equation

$$Y = \beta_0 + \beta_1 Z_1 + \cdots + \beta_n Z_n + \varepsilon$$

is called the additive model. This formula is the general form used in all regression models for this project.

For multiple reasons, the software R was chosen to run the statistical models for this work. R is an open source programming language and software environment. It has the capabilities of reading in data from multiple sources and providing functions with the ability to calculate the regression analysis. The majority of work was done using the glm function.

2.2 Multiple Regression Model for NVSS

The regression analysis was performed with a few goals in mind. The NVSS had created metrics to measure each player's ability on three different aspects: offensive, fielding, and pitching. Each team coordinated to keep track of several statistics (full list in **Appendix A**) to be used in a formula to create each metric. However, only a certain amount of these statistics are kept at the end of the season. Our team analyzed these statistics to see if a better set of metrics could be created.

The analysis was performed on the most recent data from the summer season in 2010. Data was collected for each team and each player that participated in the summer season. The actual data takes on a usable format for modeling, but required some manipulating for the regression analysis to be performed. The player assessment data collected for each player is either measured as a numeric data point (like on base average and team wins) or categorical data point (like gender, speed and versatility). Within R, the adjustment for each of the data points was made to accommodate the statistical functions. In addition to adjusting the statistics for each player, the entire database was split into one subset for each conference in the NVSS.

2.3 Offense Metric

The offensive statistics captured at the end of the season were combined and manipulated to form three main metrics: on base average points, extra base average points, and base running skills. Each of the three metrics is thought to capture the offensive profile of a player in the NVSS. The sum of these three metrics formed the offense metric reported for each player.

The regression analysis incorporated many of the same ideas to create an offensive metric for each player. In addition to the metrics for each player, the total runs scored for each team was used within the regression formula. The general form of the offensive regression formula is

Runs Scored = On base average + Extra Base Average + Base Running Skills + Bats + Intercept

The intercept included in the model is meant to measure the variation not captured in the other variables in the model. Each of the models was run against each conference dataset to get results. The results varied by conference, but concluded in very similar results. The initial look at the significance of each variable revealed that only the intercept was significant. This leads to two implications for these results. One reason for this occurrence could be that each team has their offensive abilities equally distributed among the players. Another reason could be the lack of negative statistics captured. All of the information available measures positive impacts on the offensive abilities of the game. The table below shows the estimates for each conference. The full results of the regression analysis for each conference can be found in **Appendix B**.

Figure 3. Estimates for the Offense Metric

	Conference		
Predictor	**American**	**National**	**Continental**
Intercept	619.22	485.90	401.27
On Base Average	-98.06	-56.54	21.16
Extra Base Average	12.20	95.63	-2.62
Speed = 1	-26.12	1.51	63.55
Speed = 2	-116.24	-19.39	44.15
Speed = 3	-81.35	-20.82	17.26
Speed = 4	-62.87	-19.41	40.66
Speed = 5	-79.42	-27.23	68.90
Speed = 6	-	-47.18	-
Bats Left-handed	-10.31	100.09	83.57
Bats Right-handed	-7.02	111.69	72.76

2.4 Defensive Metric

The player assessments capture defensive statistics depending on where in the field the player is positioned. The NVSS player assessment group splits positions into pitchers, infielders, and outfielders. The infielders and outfielders are measured using similar statistics whereas there is a separate measurement for pitchers. For each group of positions, those statistics are combined to form a fielding metric. Another defensive ability measured is each player's versatility in the field. This is measured as the ability to play in a variety of defensive roles in the field. These two variables are combined to create the overall defensive ability metric for each player.

In addition to the defensive statistics provided in the NVSS database, the primary position and alternate position are provided for each player. This information can be utilized in two different ways. It can help determine how important each position is using the statistics for each player. This information can also provide insight into how important having players in that position are to the team. For example, it can give perspective on whether it's more beneficial to have an extra infielder or an extra outfielder. Due to the lack of statistics provided in the database, this analysis was left out of the project.

Similar to the regression analysis done for the offensive metric, the multiple regression analysis done for the defensive metric uses the defensive statistics from the database inside the regression formula. In order to measure the estimators for each of the statistics, the runs scored against each team was used as the response variable. In general, the regression model

$$\text{Runs Against} = \text{Base Running Skills} + \text{Fielding} + \text{Versatility} + \text{Pitcher Grade} + \text{Intercept}$$

was used for each conference dataset. The results varied slightly for each conference, but resulted in the same conclusions. The conclusions from the offensive regression model are similar to the results from the defensive regression model. The consistent result in the models for each conference is the significance of the intercept. The reasons for this are the same as the offensive model. The table below shows the results from the model. Refer to **Appendix B** to see the full results of the regression analysis.

Figure 4. Estimates for the Defense Metric

	Conference		
Predictor	**American**	**National**	**Continental**
Intercept	463.73	471.87	442.50
Speed = 1	23.63	52.62	-36.21
Speed = 2	70.98	97.71	-9.14
Speed = 3	53.05	50.53	-46.34
Speed = 4	45.70	52.18	-43.57
Speed = 5	50.46	60.63	-105.46
Speed = 6	-	-26.97	-
Fielding	-0.51	0.88	-0.35
Versatility = 1	-	-	66.51
Versatility = 2	27.19	84.71	117.26
Versatility = 3	-10.65	8.95	120.49
Versatility = 4	-2.85	9.28	148.53
Versatility = 5	50.72	-29.98	93.84
Pitcher Grade	0.10	-0.07	-0.11

2.5 Team Metric

In addition to the offensive and defensive models, the team decided to analyze the contribution of each statistic to the team's ability to produce wins. Unlike the previous two models, the NVSS does not have an existing metric to compare this against. The team took this opportunity to include information on each player that was considered to have no importance to their ability before but would be considered in player assignments. This information includes statistics like age, county of residence, attendance, and other membership data. The formula below was used to evaluate the model for team wins:

Team Wins =Base Running Skills + Fielding + Versatility +Pitcher Grade
+ On base average + Extra Base Average + Bats + Attendance
+ Year of Birth + New Player + Gender + County of Residence + Intercept

It incorporates all the information found in the database for each player. The results were similar to the defensive and offensive models with a small amount of the predictors being significant. In addition, the results for this model did not follow common sense. Certain statistics and information that should have a positive influence end up with negative estimate in the model results. The table below shows the results for each conference. The details of the regression analysis for each conference can be found in **Appendix B**.

Figure 5. Estimates for the Team Metric

	Conference		
Predictor	**American**	**National**	**Continental**
Intercept	26.79	29.38	38.45
Renewal	-3.97	-	-
Attendance	0.44	-0.29	-0.11
Year of Birth	0.22	0.06	0.11
Male	3.60	-	-0.41
County = Arl	-6.15	2.90	-3.42
County = DC	1.35	9.35	-
County = Fauq	-	8.47	-10.94
County = FC	-	-	1.34
County = Fx	-6.89	2.97	-2.89
County = Loud	-10.58	7.46	-8.13
County = Md	-1.32	17.34	-
County = Orang	-	-	-5.30
County = Pw	-5.26	5.58	-1.59
County = Spots	4.27	-7.47	-
County = Staff	-	8.75	-
County = Warrn	-	-	-11.89
On Base Average	-15.95	-11.53	9.41
Extra Base Average	-7.29	14.73	-7.41
Speed = 1	-3.26	-0.07	6.88
Speed = 2	-13.89	-6.86	3.44
Speed = 3	-11.36	-2.52	6.00
Speed = 4	-9.06	-2.86	6.75
Speed = 5	-10.77	-4.77	12.11
Speed = 6	-	-4.94	-
Bats Left-handed	-6.60	10.53	1.75
Bats Right-handed	-6.53	11.52	0.90
Fielding	0.26	-0.15	0.08
Versatility = 1	-	-	-16.94
Versatility = 2	-2.89	-7.29	-17.87
Versatility = 3	2.65	0.07	-21.08
Versatility = 4	1.84	0.41	-25.13
Versatility = 5	-9.99	5.92	-19.01
Pitcher Grade	-0.01	-0.01	0.02

2.6 Regression Conclusions

The regression analysis proved to be a difficult challenge with the given statistics and metrics available. A small amount of the statistics taken throughout the season was used to create a few metrics that were entered into the NVSS database and carried over to this analysis. Using the provided statistics and metrics, the analysis broke down into three different models. The offensive metric and defensive metric models were based off of previous work done by the NVSS. It addition to those two models, the team decided to look at a model to look at each statistics contribution to the team's ability to win.

Each model had mixed results for estimates of each factor. In general, very few statistics were considered significant and the estimates often contradicted the expected results. There are a couple reasons that could describe this outcome. The statistics provided positive influences to the model without a negative component to balance the model out. This would explain why certain statistics thought to have a positive impact on the metric resulted in a negative estimate for the predictor. Another reason for the insignificance of the models would be the equal balance of player's abilities on each team. The prefect spread of talent would imply that a wide range of values for each statistic occurs for each team even though certain teams had more wins at the end of the season. This implies that there is a piece of data that is not captured in the statistics provided in the database. The significance of the intercept in each model would back this conclusion.

3. Optimization Model

The team recognized this NVSS task to be a specific integer programming model known as the assignment problem. The team manipulated the integer program into a goal program to allow for constraints to be flexible. The goals of the NVSS could then be aligned directly with constraints in the model formulation.

One issue with goal programming is that once a goal is satisfied, the model continues on whether or not the current feasible solution is the optimal solution. Several means have been developed to encourage goal programs to continue searching for more efficient solutions. One is known as preemptive goal programming in which the model is solved sequentially with only the most important goal in the objective function. If the model can be solved to a feasible solution, the objective function becomes a constraint in the model and then the second most important goal becomes the new objective function. This sequential solving continues for all goals. The process is time consuming and requires manual work. The second method to encourage the goal program to find a better solution is to assign weights to the deviation variables of each goal in the objective function in order of the goals importance. This in essence simulates the sequential solving of the model by establishing harsh penalties for the most important goals and minimal penalties for the least important goals. This will force the model to be constrained on goals weighted to be the most important and will deviate on the goals weighted to be less important.

3.1 Goals to Constraints

The goals of the NVSS (listed in **Section 1.6 NVSS Goals**) are aligned with the constraints in the model formulation with a few alterations. The goal to ensure that players are assigned to a different team after two consecutive years will not be programmed as a constraint. It will be programmed heuristically, where if player *i* played on team *j* last year, then the variable x_{ij} will not be used in the model to simulate not being able to be assigned to that team this year.

The tournament team players and carpool match-ups currently have no data structure in the leagues database. These two will still need to be done manually in the same fashion that NVSS currently handles them. Current database structure has tournament players and carpool match-ups in a comment section. However, if proper updates are made to the database structure, it would be an easy extension to apply these as goals in the model formulation.

3.2 Input Variables and Indices

The following variables are read into the model either as user inputs or from the NVSS database.

Figure 6. Input Variables

$numTeams$	Number of teams to be created for the conference.
$numAssign$	Number of players to be assigned to each team.
$numPlayers$	Number of assignable players in the conference. Counted from the NVSS database.

There is a distinct difference between the variable $numAssign$ and the variable $numPlayers$. In most cases the NVSS will want to assign all players at once and go back to make minimal changes manually. In this case the variable $numPlayers$ will be used to find the average number of players to be assigned to each team. However, the NVSS expressed interest in the ability to assign only a subset of the players and manually assign the rest. Here $numAssign$ will be used.

The following are the indices for the model:

$i = 1 \ldots numPlayers$

$j = 1 \ldots numTeams$

$k = 1 \ldots$ # of Goal Constraints (13)

Figure 7. Goal Constraints and Indices

Goal	Index (*k*)
Power Hitters	1
Speed Base Runners	2
Slow Base Runners	3
Versatile Players	4
Pitching Experience	5
New Players	6
Thursday Only Players	8
Either T or R Players	9
Pos. 1 (Pitcher)	10
Pos. 2,3,5 (Corner Infielders)	11
Pos. 4,6,11 (Middle Infielders)	12
Pos. 7,8,9,10 (Outfielders)	13

3.3 Goal Weights

The weights assigned to the goals were established using the Rank Sum method. The goals were ranked based on the current NVSS heuristics approach to their goals and from information gathered from discussions with our sponsor.

Figure 8. Goal Weights

Goal	Rank	Weight
Speed Base Runners	1	0.1500
Power Hitters	2	0.1375
Pitching Experience	3	0.1250
Versatile Players	4	0.1125
Slow Base Runners	5	0.1000
Pos. 4,6,11 (Middle Infielders)	6	0.0875
Pos. 1 (Pitcher)	7	0.0750
Pos. 2,3,5 (Corner Infielders)	8	0.0625
Pos. 7,8,9,10 (Outfielders)	9	0.0500
New Players	10	0.0375
Tuesday Only Players	11	0.0250
Thursday Only Players	11	0.0250
Either T or R Players	12	0.0125

3.4 Model Variables and Variable Bounds

The following are model variables:

$$x_{ij} = \begin{cases} 1 & \text{if player } i \text{ assigned to team } j \\ 0 & \text{otherwise} \end{cases}$$

$$dOver_{jk} = \text{deviation variable for goal constraint } k \text{ being over limit for team } j$$

$$dUnder_{jk} = \text{deviation variable for goal constraint } k \text{ being under limit for team } j$$

The following are variable bounds.

$$x_{ij} = BINARY$$

$$dOver_{jk} = \mathrm{Z}^{+}$$

$$dUnder_{jk} = \mathrm{Z}^{+}$$

Note: Z^{+} represents the set of positive integers.

3.5 Data Vectors

Each goal constraint has its own data vector that is constructed from the NVSS database.

Figure 9. Data Vectors

$PH_i = \begin{cases} 1 & \text{if player } i \text{ is a Power Hitter} \\ 0 & \text{otherwise} \end{cases}$	A Power Hitter is defined as someone who has an Extra Base Average Points (XBPts) value greater than 7.
$SP_i = \begin{cases} 1 & \text{if player } i \text{ is a Speed Base Runner} \\ 0 & \text{otherwise} \end{cases}$	A Speed Base Runner is defined as someone who has been rated a 4 or 5 in their base running ability.
$SL_i = \begin{cases} 1 & \text{if player } i \text{ is a Slow Base Runner} \\ 0 & \text{otherwise} \end{cases}$	A Slow Base Runner is defined as someone who has been rated a 1 or 2 in their base running ability.
$V_i = \begin{cases} 1 & \text{if player } i \text{ is a Versatile Player} \\ 0 & \text{otherwise} \end{cases}$	A Versatile Player is defined as someone who was been rated a 4 or 5 in the overall playing versatility.
$P_i = \begin{cases} 1 & \text{if player } i \text{ has Pitching Experience} \\ 0 & \text{otherwise} \end{cases}$	All players with Pitching Experience (not necessarily someone who only plays pitcher) are identified in the NVSS database with a P.
$N_i = \begin{cases} 1 & \text{if player } i \text{ is a New Player} \\ 0 & \text{otherwise} \end{cases}$	All New Players are identified in the NVSS database with an N.
$T_i = \begin{cases} 1 & \text{if player } i \text{ can only play on Tuesdays} \\ 0 & \text{otherwise} \end{cases}$	Tuesday only players are identified in the NVSS database with 'Tues'.
$R_i = \begin{cases} 1 & \text{if player } i \text{ can only play on Thursdays} \\ 0 & \text{otherwise} \end{cases}$	Thursday only players are identified in the NVSS database with 'Thurs',
$E_i = \begin{cases} 1 & \text{if player } i \text{ can only play one day a week} \\ 0 & \text{otherwise} \end{cases}$	Players who can only play one day a week, Either Tuesday or Thursday, are indentified in the NVSS database with 'Either'.
$pPitcher_i = \begin{cases} 1 & \text{if player } i \text{ plays position 1} \\ 0 & \text{otherwise} \end{cases}$	Players whose primary position is pitcher.

Table 7. Data Vectors (Continued)

$pCInfielder_i = \begin{cases} 1 & \text{if player } i \text{ plays position } 2, 3, \text{or } 5 \\ 0 & \text{otherwise} \end{cases}$	Players whose primary position is a corner infielder: Catcher, First Base, or Third Base.
$pMInfielder_i = \begin{cases} 1 & \text{if player } i \text{ plays position } 4, 6, \text{or } 11 \\ 0 & \text{otherwise} \end{cases}$	Players whose primary position is a middle infielder: Second Base, Short Stop, and Extra Infielder.
$pOutfielder_i = \begin{cases} 1 & \text{if player } i \text{ plays position } 7, 8, 9, \text{or } 10 \\ 0 & \text{otherwise} \end{cases}$	Players whose primary position is an outfielder: Left Field, Left Center, Right Center, and Right Field.

3.6 Computed Constant Variables

To enforce spreading out the players by our goals the model must find the average number of players per team for each goal and take the ceiling (the value must be integer since we cannot assign fractions of individuals to a team).

Power Hitters:

$$\overline{PH} = \left\lceil \sum_i PH_i \Big/ numTeams \right\rceil$$

Speed Base Runners:

$$\overline{SP} = \left\lceil \sum_i SP_i \Big/ numTeams \right\rceil$$

Slow Base Runners:

$$\overline{SL} = \left\lceil \sum_i SL_i \Big/ numTeams \right\rceil$$

Versatile Players:

$$\overline{V} = \left\lceil \sum_i V_i \Big/ numTeams \right\rceil$$

Pitching Experience:

$$\overline{P} = \left\lceil \sum_i P_i \Big/ numTeams \right\rceil$$

New Players:

$$\overline{N} = \left\lceil \sum_i N_i \Big/ numTeams \right\rceil$$

Tuesday Only Players:

$$\overline{T} = \left\lceil \sum_i T_i \Big/ numTeams \right\rceil$$

Thursday Only Players:

$$\overline{R} = \left\lceil \sum_i R_i \Big/ numTeams \right\rceil$$

Either Tuesday or Thursday Players:

$$\overline{E} = \left\lceil \sum_i E_i \Big/ numTeams \right\rceil$$

Position 1 (Pitcher):

$$\overline{pPitcher} = \left\lceil \sum_i pPitcher_i \Big/ numTeams \right\rceil$$

Position 2, 3, and 5 (Corner Infielders and Catcher):

$$\overline{pCInfielder} = \left\lceil \sum_i pCInfielder_i \Big/ numTeams \right\rceil$$

Position 4, 6, and 11 (Middle Infielders):

$$\overline{pMInfielder} = \left\lceil \sum_i pMInfielder_i \Big/ numTeams \right\rceil$$

Position 7, 8, 9, and 10 (Outfielders):

$$\overline{pOutfielder} = \left\lceil \sum_i pOutfielder_i \Big/ numTeams \right\rceil$$

3.7 Objective and Constraints

3.7.1 Objective

The objective function will be the minimization of the weighted sum of the deviation variables.

$$MIN: \sum_k \sum_j W_k (dOver_{jk} + dUnder_{jk})$$

3.7.2 Hard Constraints

All players must be assigned to a team:

$$\sum_i \sum_j x_{ij} = \min(numTeams * numAssign, numPlayers)$$

Note: The right hand side will be the minimum of either the product of the number of teams and the number of players to be assigned to each team, or the total number of assignable players in the conference.

Players can only be assigned to one team:

$$\sum_j x_{ij} \le 1 \quad \forall i$$

Set upper limit for number of players per team:

$$\sum_i x_{ij} \le \min\left(numAssign, \left\lceil {numPlayers}/{numTeams} \right\rceil \right) \quad \forall j$$

Set lower limit for number of players per team:

$$\sum_i x_{ij} \ge \min\left(numAssign, \left\lfloor {numPlayers}/{numTeams} \right\rfloor \right) \quad \forall j$$

Note: If *numAssign* is less than the average number of assignable players per team, then these two constraints will be the same as an equality constraint equal to *numAssign*. Else, these two constraints will force the teams to be even in size within one player.

Pre-Assignment of Player *i* to team *j* (i.e. to assign team managers, etc):

$$x_{ij} = 1$$

3.7.3 Goal Constraints

Goal 1: Spread out Power Hitters

$$\sum_i PH_i \; x_{ij} + dUnder_{j,1} - dOver_{j,1} = \overline{PH} \quad \forall j$$

Goal 2: Spread out Speed Base Runners

$$\sum_i SP_i \; x_{ij} + dUnder_{j,2} - dOver_{j,2} = \overline{SP} \quad \forall j$$

Goal 3: Spread out Slow Base Runners

$$\sum_i SL_i \; x_{ij} + dUnder_{j,3} - dOver_{j,3} = \overline{SL} \quad \forall j$$

Goal 4: Spread out Versatile Players

$$\sum_i V_i \; x_{ij} + dUnder_{j,4} - dOver_{j,4} = \overline{V} \quad \forall j$$

Goal 5: Spread out players with Pitching Experience

$$\sum_i P_i \; x_{ij} + dUnder_{j,5} - dOver_{j,5} = \overline{P} \quad \forall j$$

Goal 6: Spread out New Players

$$\sum_i N_i \; x_{ij} + dUnder_{j,6} - dOver_{j,6} = \overline{N} \quad \forall j$$

Goal 7: Spread out Tuesday Only Players

$$\sum_i T_i \; x_{ij} + dUnder_{j,7} - dOver_{j,7} = \overline{T} \quad \forall j$$

Goal 8: Spread out Thursday Only Players

$$\sum_i R_i \; x_{ij} + dUnder_{j,8} - dOver_{j,8} = \overline{R} \quad \forall j$$

Goal 9: Spread out Either Tuesday or Thursday Players

$$\sum_i E_i \; x_{ij} + dUnder_{j,9} - dOver_{j,9} = \overline{E} \quad \forall j$$

Goal 10: Spread out Pitchers

$$\sum_i pPitcher_i \; x_{ij} + dUnder_{j,10} - dOver_{j,10} = \overline{pPitcher} \quad \forall j$$

Goal 11: Spread out Corner Infielders

$$\sum_i pCInfielder_i \; x_{ij} + dUnder_{j,11} - dOver_{j,11} = \overline{pCInfielder} \quad \forall j$$

Goal 12: Spread out Middle Infielders

$$\sum_i pMInfielder_i \; x_{ij} + dUnder_{j,12} - dOver_{j,12} = \overline{pMInfielder} \quad \forall j$$

Goal 13: Spread out Outfielders

$$\sum_i pOutfielder_i \; x_{ij} + dUnder_{j,13} - dOver_{j,13} = \overline{pOutfielder} \quad \forall j$$

4. Team Assignment Tool

To support the goals of the NVSS, a team assignment tool was created to implement the model described in section 3 and to support the automation of a process which assigns players to evenly balanced teams. The team assignment tool can be used to assign NVSS players to teams, and current NVSS tools support printing reports and making manual team adjustments.

4.1 Requirements

To design a team assignment tool, the team had to consider and balance several different factors. These factors included technical requirements, NVSS requests, and schedule/cost requirements.

Figure 10. Technical Requirements and Status

Technical Requirements	**Status**
Team assignment tool must work on a Windows PC	Supported within design
Team assignment tool must be able to use current team data from the NVSS database or from a spreadsheet version of the NVSS database	Supported within design
Team assignment tool must be support a user interface, or use technology that does not require a technical background or much training	Supported within design

Figure 11. NVSS Goals and Status

NVSS Goals	**Status**
Team assignment tool must evenly distribute new players	Supported within design
Team assignment tool must evenly distribute those with pitching experience	Supported within design
Team assignment tool must evenly distribute players on each tournament team (designated TT)	Not supported
Team assignment tool must evenly distribute base running ability	Supported within design
Team assignment tool must evenly distribute part-time players	Supported within design
Team assignment tool must evenly distribute power hitters	Supported within design
Team assignment tool must match up carpools	Not supported
Team assignment tool must account for teammate restrictions	Not supported

Figure 12. NVSS Goals and Status (Continued)

Team assignment tool must make sure players are not assigned to the same team more than 2 years in a row	Supported within design
Team assignment tool must support data security	N/A – Data maintained within NVSS database
Team assignment tool must evenly distribute players with a "versatility" rating of "5" or "4"	Supported within design
Team assignment tool must update NVSS database with resulting team assignments	Supported within design

Figure 13. Cost/Schedule Requirements and Status

Cost/Schedule Requirements	**Status**
Team assignment tool must not use software that requires expensive licenses	Supported within design
Team assignment tool must be completed and meet core requirements by the end of the course	Supported within design

4.2 Impact of Design Requirements

Based on the design requirements the following critical decisions were made:

- Use of Integer Programming – This provided a method of solving the team distribution requirements using goals and constraints.
- Use of Java – Java supported the requirement for running on Windows and being low cost. Java Swing provided an effective method to create a user interface that accepts user input for settings and assignment decisions.
- Use of lp_solve - Using lp_solve provided the solver for the integer program that works with Java and PCs running Windows. More powerful commercial solver tools are available, but can have a high license cost. lp_solve is free and is open source.

The following tools are required by the team assignment tool:

- NVSS Database – Access database that store player assessment and league renewal information.
- Java Runtime Environment (JRE) version 1.6.0_22 – The JRE is required to run the team assignment tool code. The JRE is a product of Oracle and can be downloaded from the Oracle website.

- lp_solve version 5.5.2.0 – lp_solve files are required to use the team assignment tool. The lp_solve files are available from the website http://sourceforge.net/projects/lpsolve/ .

The team assignment tool will be provided to the NVSS for their use. The design of the team assignment interface included the following steps to promote flexibility and usability:

4.2.1 Flexibility

- Allow operator selection of a NVSS database file – Player assessment data is stored in a NVSS database. The team assignment tool allows the database file to be selected, preventing the tool from being hard coded with a specific path where the database must be located.

- Support configuration file to read in default setting values– Use of a configuration file allow users of the team assignment tool to change default values without having to update the team assignment tool code.

4.2.2 Usability

- Disable control buttons based on dependencies – Some buttons on the team assignment tool are not available until another step has been completed. For example, the user is prevented for inputting the number of teams to create until the user has selected the NVSS database file.

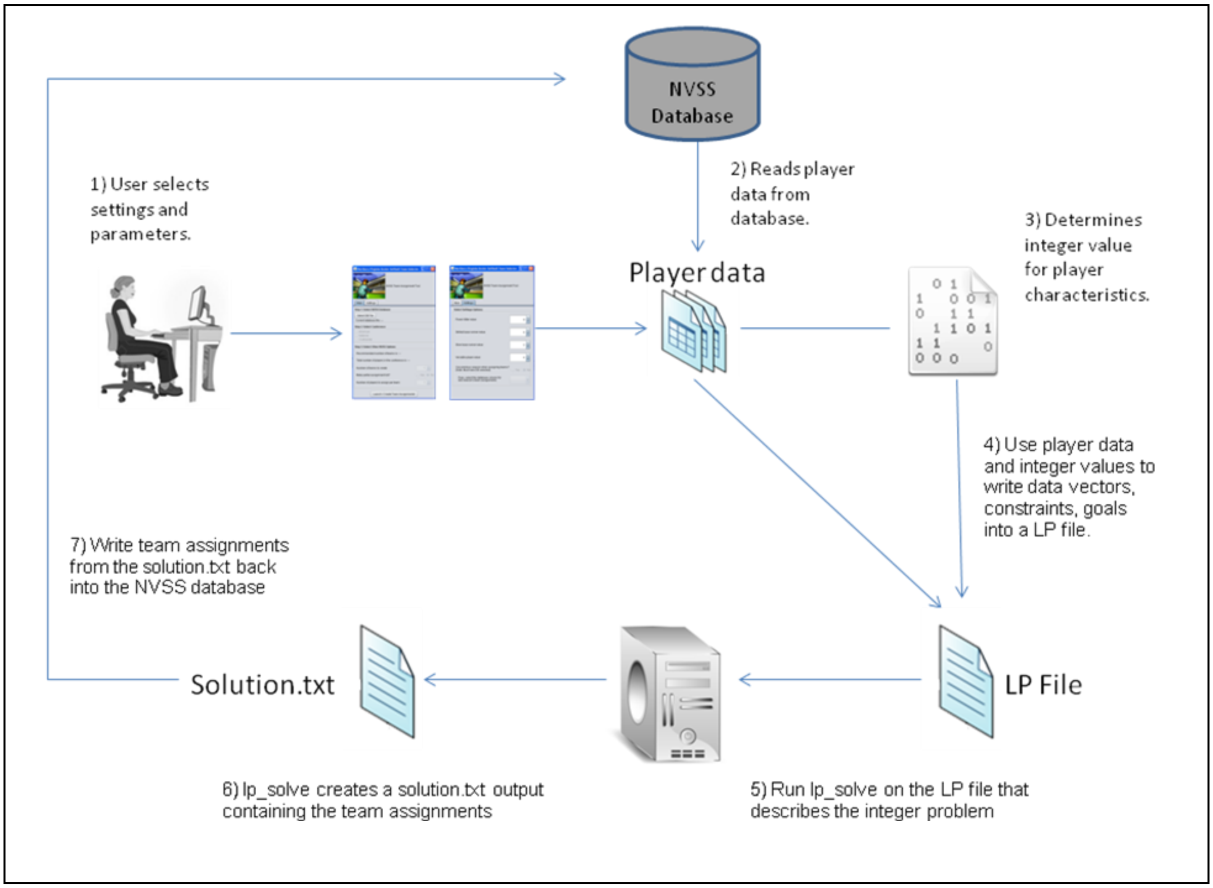
- Provide error dialog boxes as appropriate – The team assignment tool makes use of pop-up windows to report certain error types to the user. These error types include reporting for a data connection error and database update errors.

- Provide appropriate default values for settings and controls – The team assignment tool uses default values from a configuration file to populate appropriate controls. For example, the team assignment tool has a default power hitter value that is uses for determining power hitters. The team assignment tool also sets controls to default values based data from the NVSS database. For example, the team assignment tool sets a default number of teams to create based on the number of players in a conference. The default number of teams is based on creating teams of 14 players.

4.3 General Flow

The assignment tool allows the user to control the settings and parameters. After the user has selected the settings based on their preference, the tool reads player data from the NVSS Database tables. Integer values for player characteristics are determined and used in conjunction with the player data to write data vectors, constraints, and goals into a LP file. lp_solve then solves the LP file to create an output file containing the team assignments. Finally, the tool writes the team assignments from the output file back into the NVSS database.

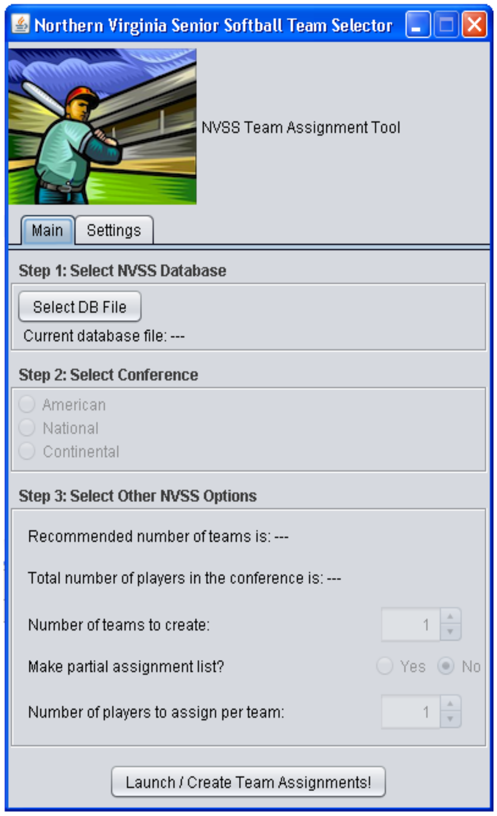
Figure 14. Assignment Tool Flow



4.4 Running the Tool

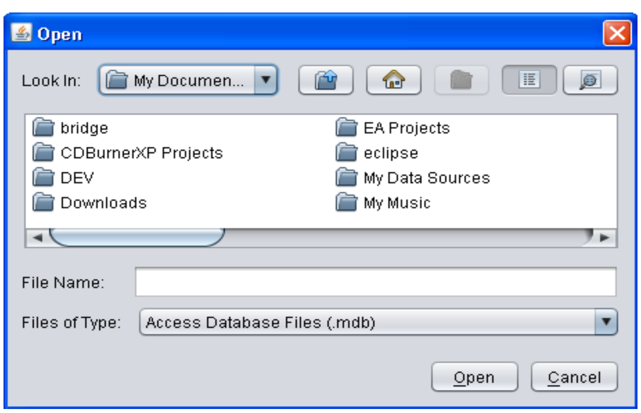
The user can launch the assignment tool by double clicking the provided executable file. The assignment tool will open as shown in Figure 4 below.

Figure 15. Main Screen Shot



First the user must select the NVSS Database File by clicking on the “Select DB File” button to open the file browser window shown below. They will use the file browser to locate the NVSS database file, select the file and click open.

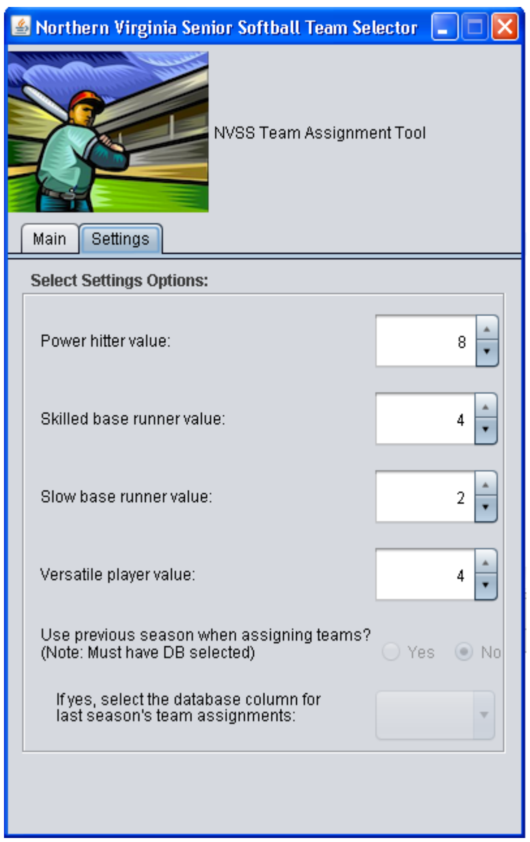
Figure 16. File Browser



Once a database is selected, the user must select a conference by clicking one of the listed radio buttons. The last necessary step is that the user selects the number of team and players by choosing the number of teams to create using the associated spin dial. The radio button associated with "make partial assignment list?" is asking if the user wants all available conference players to be assigned to their designed team. If the "No" radio button is selected, all conference players will be assigned. If "Yes" radio button is selected, the team assignment tool will create partial teams, leaving some players unassigned. These unassigned players can be manually assigned at a later time. If the "Yes" radio button was selected to create partial team assignments, the user chooses the number of players to assign to each partial team using the associated spin dial. The user selects the "Launch/Create Team Assignments" button to create the team assignments and update the database selected with the team assignments.

The user has the option of clicking on the "Settings" tab to adjust parameters using the spin dial for the power hitter, skilled based runner, slow base runner and, versatile player scores. In the "Settings" tab, the user also has the option of choosing data from a previous season.

Figure 17. Settings Screen Shot



5. Conclusions

5.1 Recommendations to NVSS

As mentioned in the **Section 2. Statistical Methods,** each team keeps track of several statistics (listed in **Appendix A**). However, only a certain amount of these statistics are kept at the end of the season. The team recommends that NVSS keep all statistics in the database. The team also recommends that NVSS create additional columns in the database for tournament team players, and carpool match-ups. If these additional columns are added, then these constraints could be added into the model and would no longer have to be done manually by the committee.

5.2 Potential Future Work

Potential future work could be done to include more pieces of data into these models. They should not be limited to the statistics used to create the existing metrics, but include additional metrics that might help explain some of the outliers and peculiarity of why certain teams end up winning. A statistic to measure how effective a manager is at managing the game would be an example. This could help determine the manager's influence on players and the strategy used in games throughout the season. Addition of these new statistics in the model will help to explain the variation of the data and create a better predictive model.

Additional future work also includes updating the default weights for each constraint in the model. This could be done by conducting a working group session with the NVSS committee members to prioritize the goals of the model. Using input from the NVSS committee would create more accurate weights.

Future work could also be done in the user interface to give the user the ability to turn constraints on or off when writing the formulation for the LP file. Added flexibility could also be added to adjust the weights for each constraint quicker than the current method. Other potential options for future work on the program could include adding in a security level to protect the database, the program and the results of the model. In addition, a reporting feature could be created to link the statistical regression results with the database and model results. This would produce a predictive model that provides additional metrics for evaluating the results of the team assignments.

5.3 Validation

The team hopes to have the NVSS review the output of the model's team assignments. The team would also like to meet with the assignment committee to compare the results of their current model with the results of the team's integer program model. The assignment committee has the ability to review the results of our model and make a mental prediction on the equality of the teams produced by the integer program. True validation won't be confirmed until the end of the 2011 spring/summer season by comparing the winning percentage across each of the conferences

References

Berkelaar, Michel, Eikland, Kjell, and Notebaert, Peter. *lp_solve: Open source (Mixed-Integer) Linear Programming system*. Version 5.5.0.0 dated May 17, 2005. URL http://lpsolve.sourceforge.net.

Casella, George, and Roger L. Berger. *Statistical Inference* . 2nd ed. Australia: Thomson Learning, 2002.

Petruccelli, Joseph D., BalgobinNandram, and Minghui Chen.*Applied Statistics for Engineers and Scientists*. Upper Saddle River, N.J.: Prentice Hall, 1999.

Rardin, Ronald. *Optimization in Operations Research*. Englewood Cliffs: Prentice Hall, 1998.

R Development Core Team (2010). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. URL http://www.R-project.org.

Winston, Wayne and Jeffrey Goldberg. *Operations Research*. Australia: Thomson Brooks/Cole, 2004.

Wolsey, Laurence. *Integer Programming*. Chichester: John Wiley & Sons, 1998.

Yoon, K. Paul, and Ching Hwang. *Multiple Attribute Decision Making: Introduction*. London: Sage, 1995.

Appendix A. NVSS Player Assessment Data

This appendix lists all of the data points taken for each player on the player assessment. Statistics in bold are kept after the season. Statistics in italics are numeric variables.

Offensive Statistics and Information:

- All Players:
    - **Base Running Skills**
    - ***Number of Games Played***
    - *Trips to the Plate*
    - *Singles*
    - *Doubles*
    - *Triples*
    - *Home Runs*
    - *Sacrifice Flies*
    - *Base on Balls*
    - *On Base by Fielders Error*

Defensive Statistics and Information:

- All Positions:
    - Catching Flyballs
    - Fielding Grounders
    - Fielding Range
    - Positioning
    - Throwing Strength
    - Throwing Accuracy
    - Alertness
    - **Versatility**
- Infielders:
    - Covers Base Properly
    - Catches Throws
- Outfielders
    - Moves In and Out on Flies
    - Backs Up Other Fielders
- Pitchers
    - Control
    - Mastery
    - Fielding
    - Throwing Strength
    - Throwing Accuracy
    - Covers/Backs Up Plays
    - Alertness

Appendix B. Regression Results

Offensive Regression: American, National, and Continental Conferences

American Conference

```
Call:
glm(formula = RS ~ BA + XBBA + SPD + dBATS)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-137.74   -72.11    26.39    67.57   113.23

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  619.220     90.109   6.872 4.97e-10 ***
BA           -98.055    117.097  -0.837   0.4043
XBBA          12.199    120.795   0.101   0.9198
SPD1         -26.115     58.772  -0.444   0.6577
SPD2        -116.237     51.303  -2.266   0.0256 *
SPD3         -81.346     47.902  -1.698   0.0925 .
SPD4         -62.871     49.217  -1.277   0.2043
SPD5         -79.418     51.615  -1.539   0.1270
dBATS2       -10.313     61.839  -0.167   0.8679
dBATS3        -7.015     58.068  -0.121   0.9041
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 6332.603)

    Null deviance: 712862  on 112  degrees of freedom
Residual deviance: 652258  on 103  degrees of freedom
AIC: 1321.4

Number of Fisher Scoring iterations: 2
```

National Conference

```
Call:
glm(formula = RS ~ BA + XBBA + SPD + dBATS)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-113.27   -21.45    16.66    32.97    76.30

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  485.904     72.921   6.663 4.14e-09 ***
BA           -56.540     82.382  -0.686   0.4947
XBBA          95.625     63.187   1.513   0.1344
SPD1           1.514     36.781   0.041   0.9673
SPD2         -19.388     31.319  -0.619   0.5378
SPD3         -20.822     28.693  -0.726   0.4703
SPD4         -19.411     27.819  -0.698   0.4875
SPD5         -27.231     32.655  -0.834   0.4070
SPD6         -47.175     62.365  -0.756   0.4518
dBATS2       100.091     61.483   1.628   0.1078
dBATS3       111.687     59.253   1.885   0.0634 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 3205.625)

    Null deviance: 260418  on 84  degrees of freedom
Residual deviance: 237216  on 74  degrees of freedom
AIC: 939.62

Number of Fisher Scoring iterations: 2
```

Continental Conference

```
Call:
glm(formula = RS ~ BA + XBBA + SPD + dBATS)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-141.051   -39.812     0.556    46.492   140.494

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  401.272     50.313   7.976 1.88e-12 ***
BA            21.157     66.668   0.317   0.7516
XBBA          -2.623     88.831  -0.030   0.9765
SPD1          63.550     28.703   2.214   0.0290 *
SPD2          44.149     22.901   1.928   0.0566 .
SPD3          17.260     21.404   0.806   0.4218
SPD4          40.661     21.689   1.875   0.0636 .
SPD5          68.902     29.485   2.337   0.0213 *
dBATS2        83.570     42.222   1.979   0.0504 .
dBATS3        72.759     39.277   1.852   0.0667 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 4283.408)

    Null deviance: 512909  on 115  degrees of freedom
Residual deviance: 454041  on 106  degrees of freedom
AIC: 1310.8

Number of Fisher Scoring iterations: 2
```

Defensive Regression: American, National, and Continental Conferences

American Conference

```
Call:
glm(formula = RA ~ SPD + FLD + VER + PITCHER_GRADE)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-96.611  -52.740    9.096   41.857  105.857

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   463.73385   45.68636  10.150   <2e-16 ***
SPD1           23.62523   43.78233   0.540   0.5907
SPD2           70.97622   39.20993   1.810   0.0732 .
SPD3           53.04643   38.22820   1.388   0.1683
SPD4           45.69632   39.29600   1.163   0.2476
SPD5           50.46494   42.43152   1.189   0.2371
FLD            -0.51020    0.86168  -0.592   0.5551
VER2           27.19019   31.00286   0.877   0.3826
VER3          -10.64753   25.64083  -0.415   0.6788
VER4           -2.85374   28.78582  -0.099   0.9212
VER5           50.71696   37.37298   1.357   0.1778
PITCHER_GRADE   0.09987    0.22602   0.442   0.6595
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 3277.863)

    Null deviance: 382183  on 112  degrees of freedom
Residual deviance: 331064  on 101  degrees of freedom
AIC: 1248.7

Number of Fisher Scoring iterations: 2
```

National Conference

```
Call:
glm(formula = RA ~ SPD + FLD + VER + PITCHER_GRADE)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-126.28   -67.48   -16.80    77.17   137.59

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   471.87284   51.11445   9.232 7.75e-14 ***
SPD1           52.61902   55.29958   0.952   0.3445
SPD2           97.71396   45.32733   2.156   0.0344 *
SPD3           50.52581   44.17267   1.144   0.2565
SPD4           52.18126   45.78589   1.140   0.2582
SPD5           60.63128   51.20080   1.184   0.2402
SPD6          -26.96651   92.98020  -0.290   0.7726
FLD             0.88167    1.31116   0.672   0.5035
VER2           84.70869   46.01276   1.841   0.0697 .
VER3            8.94506   32.35862   0.276   0.7830
VER4            9.28245   37.59084   0.247   0.8057
VER5          -29.98152   47.99937  -0.625   0.5342
PITCHER_GRADE  -0.07176    0.31714  -0.226   0.8216
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 6427.388)

    Null deviance: 532149  on 84  degrees of freedom
Residual deviance: 462772  on 72  degrees of freedom
AIC: 1000.4

Number of Fisher Scoring iterations: 2
```

Continental Conference

```
Call:
glm(formula = RA ~ SPD + FLD + VER + PITCHER_GRADE)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-164.24   -51.98   -11.61    52.61   159.96

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)    442.4979    80.7336   5.481 3.02e-07 ***
SPD1           -36.2126    35.8857  -1.009    0.315
SPD2            -9.1375    30.6867  -0.298    0.766
SPD3           -46.3445    29.8829  -1.551    0.124
SPD4           -43.5745    31.0521  -1.403    0.164
SPD5          -105.4566    40.7276  -2.589    0.011 *
FLD             -0.3498     1.1914  -0.294    0.770
VER1            66.5087    87.4253   0.761    0.449
VER2           117.2565    86.1303   1.361    0.176
VER3           120.4864    88.6840   1.359    0.177
VER4           148.5274    91.4916   1.623    0.108
VER5            93.8419   104.5992   0.897    0.372
PITCHER_GRADE   -0.1085     0.3083  -0.352    0.726
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 6375.96)

    Null deviance: 733134  on 115  degrees of freedom
Residual deviance: 656724  on 103  degrees of freedom
AIC: 1359.6

Number of Fisher Scoring iterations: 2
```

Team Regression: American, National, and Continental Conferences

American Conference

```
Call:
glm(formula = TEAM_WINS ~ dTYP + ATT + YOB + dGENDER + dCOUNTY +
    BA + XBBA + SPD + dBATS + FLD + VER)

Deviance Residuals:
       Min          1Q      Median          3Q         Max
-1.931e+01  -5.848e+00   4.974e-14   6.344e+00   1.732e+01

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  26.0908    21.3947   1.219   0.2260
dTYP1        -3.6870     9.7298  -0.379   0.7057
ATT           0.4335     0.3365   1.289   0.2010
YOB           0.2231     0.2010   1.110   0.2700
dGENDER2      3.5810     6.5202   0.549   0.5843
dCOUNTY1     -6.1325     5.8388  -1.050   0.2965
dCOUNTY2      1.3038    10.8616   0.120   0.9047
dCOUNTY5     -6.9273     4.4409  -1.560   0.1224
dCOUNTY6    -10.6313     5.2583  -2.022   0.0463 *
dCOUNTY7     -1.0185    11.2814  -0.090   0.9283
dCOUNTY9     -5.3809     4.9088  -1.096   0.2760
dCOUNTY10     4.3557    10.0695   0.433   0.6664
BA          -15.7214    14.8061  -1.062   0.2913
XBBA         -7.0926    15.4357  -0.459   0.6470
SPD1         -2.9530     7.1623  -0.412   0.6811
SPD2        -13.5207     6.3271  -2.137   0.0354 *
SPD3        -10.9559     6.1213  -1.790   0.0770 .
SPD4         -8.6257     6.3572  -1.357   0.1783
SPD5        -10.0998     6.8586  -1.473   0.1445
dBATS2       -6.5864     7.9378  -0.830   0.4090
dBATS3       -6.6286     7.4098  -0.895   0.3735
FLD           0.2456     0.1424   1.725   0.0881 .
VER2         -2.6312     5.5163  -0.477   0.6346
VER3          2.9874     4.3263   0.691   0.4917
VER4          2.0446     5.1623   0.396   0.6930
VER5        -10.3259     6.2542  -1.651   0.1023
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 81.34593)

    Null deviance: 9954.9  on 112  degrees of freedom
Residual deviance: 7077.1  on  87  degrees of freedom
AIC: 842.19

Number of Fisher Scoring iterations: 2
```

National Conference

```
Call:
glm(formula = TEAM_WINS ~ ATT + YOB + dCOUNTY + BA + XBBA + SPD +
    dBATS + FLD + VER + PITCHER_GRADE)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-17.650   -4.746    1.210    4.398   12.295

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)    29.38458   18.23493   1.611    0.113
ATT            -0.28883    0.28660  -1.008    0.318
YOB             0.06297    0.22107   0.285    0.777
dCOUNTY1        2.90343    8.01044   0.362    0.718
dCOUNTY2        9.34959    9.90568   0.944    0.349
dCOUNTY3        8.47386   10.04677   0.843    0.403
dCOUNTY5        2.97318    5.81655   0.511    0.611
dCOUNTY6        7.46355    6.45856   1.156    0.253
dCOUNTY7       17.33862   11.51694   1.505    0.138
dCOUNTY9        5.58340    6.78109   0.823    0.414
dCOUNTY10      -7.46796   10.15383  -0.735    0.465
dCOUNTY11       8.74981   10.63896   0.822    0.414
BA            -11.53243   12.79700  -0.901    0.371
XBBA           14.72528    9.86717   1.492    0.141
SPD1           -0.07081    5.96966  -0.012    0.991
SPD2           -6.85682    5.01306  -1.368    0.177
SPD3           -2.51602    4.70069  -0.535    0.595
SPD4           -2.86290    4.77453  -0.600    0.551
SPD5           -4.77056    5.56475  -0.857    0.395
SPD6           -4.93937   10.29294  -0.480    0.633
dBATS2         10.53082    9.34762   1.127    0.265
dBATS3         11.51561    8.65471   1.331    0.189
FLD            -0.15209    0.13925  -1.092    0.279
VER2           -7.29168    4.96829  -1.468    0.148
VER3            0.06805    3.63194   0.019    0.985
VER4            0.41234    4.40787   0.094    0.926
VER5            5.91787    5.25943   1.125    0.265
PITCHER_GRADE  -0.01191    0.03407  -0.350    0.728

(Dispersion parameter for gaussian family taken to be 58.25698)

    Null deviance: 4573.6  on 84  degrees of freedom
Residual deviance: 3320.6  on 57  degrees of freedom
AIC: 610.77

Number of Fisher Scoring iterations: 2
```

Continental Conference

```
Call:
glm(formula = TEAM_WINS ~ ATT + YOB + dGENDER + dCOUNTY + BA +
    XBBA + SPD + dBATS + FLD + VER + PITCHER_GRADE)

Deviance Residuals:
       Min          1Q      Median          3Q         Max
-1.384e+01  -4.800e+00   3.197e-14   4.428e+00   1.339e+01

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)    38.45211   16.08302   2.391  0.01894 *
ATT            -0.10519    0.20011  -0.526  0.60045
YOB             0.10519    0.13517   0.778  0.43854
dGENDER2       -0.40700    3.10096  -0.131  0.89588
dCOUNTY1       -3.42235   10.45266  -0.327  0.74413
dCOUNTY3      -10.94315   12.55911  -0.871  0.38594
dCOUNTY4        1.34078   11.51955   0.116  0.90761
dCOUNTY5       -2.88535   10.13430  -0.285  0.77653
dCOUNTY6       -8.12814   10.61265  -0.766  0.44579
dCOUNTY8       -5.29577   12.62266  -0.420  0.67584
dCOUNTY9       -1.59115   10.82285  -0.147  0.88345
dCOUNTY12     -11.89076   12.62021  -0.942  0.34867
BA              9.41218   10.11785   0.930  0.35478
XBBA           -7.40806   11.07583  -0.669  0.50534
SPD1            6.87697    3.46454   1.985  0.05026 .
SPD2            3.44214    2.88885   1.192  0.23665
SPD3            6.00344    2.90058   2.070  0.04141 *
SPD4            6.74919    3.04748   2.215  0.02937 *
SPD5           12.10629    3.88697   3.115  0.00249 **
dBATS2          1.74551    5.89819   0.296  0.76797
dBATS3          0.90220    5.56697   0.162  0.87163
FLD             0.07768    0.12570   0.618  0.53821
VER1          -16.93796    8.14911  -2.079  0.04057 *
VER2          -17.86840    7.95745  -2.245  0.02724 *
VER3          -21.08040    8.31087  -2.536  0.01296 *
VER4          -25.12708    8.56453  -2.934  0.00427 **
VER5          -19.00625   11.20960  -1.696  0.09351 .
PITCHER_GRADE   0.02459    0.03180   0.773  0.44150
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 52.84607)

    Null deviance: 6204.7  on 115  degrees of freedom
Residual deviance: 4650.5  on  88  degrees of freedom
AIC: 815.36

Number of Fisher Scoring iterations: 2
```