Fantasy Baseball Draft: An Optimization Analysis

Submitted to Dr. S. Sebnem Ahiska King

ISE501: Introduction to Operations Research

2024 April 27

Submitted by:

Jacob Norman

Abstract

As one of the oldest fantasy sports, fantasy baseball has a rich history of using mathematics and statistics to gain a competitive edge. The foundation of any fantasy team is the players that are selected during the draft at the beginning of the season. An integer program was developed in the Julia programming language to model the behavior of a draft for a 5x5 roto fantasy league. The decision that needs to be made is simple: what player should be selected in each of the 25 rounds of the fantasy draft. The optimal team maximizes the improvement over the target value set for 10 scoring categories, including on base percentage, home runs, and strikeouts. The model was solved to optimality and determined that drafting starting pitchers, then relief pitchers, and finally, hitters was the best overall strategy given the set of constraints. In addition, five key inputs to the model were adjusted to determine if there were alternative strategies or specific players that should be considered in a real-world draft. This included changing the projection system used, adjusting the proxy for player draft value, starting at three different positions in the first round, giving up on one or more goals, and relaxing bound constraints on the goal values. It was determined that "punting" saves and holds, a common strategy in the fantasy community, is a viable option as well. Additionally, several players were identified as having superb value relative to their position in the draft and should be considered as candidates to be rostered. These players include, but are not limited to, Zack Wheeler, Kyle Schwarber, Cedric Mullins, Daulton Varsho, and Matt Chapman. It is noted that there are several key limitations of the model, mainly related to uncertainty, that restrict the utility of applying the optimal roster in its entirety. The value of the model is primarily derived from what-if analysis that uncovers common players and draft strategies that lead to success.

Section 1. Introduction

Most people are familiar with at least one professional sport, such as football, hockey, or baseball. Fans routinely choose to engage with these sports through something called fantasy leagues. These are essentially groups of individuals who compete against each other with their own teams. Even within a single type of fantasy sport, there are many different permutations of league rules; however, the performance of a team's players in real-life games is always tracked and then converted into some sort of point system. Generally, a manager's fantasy team competes against another managers's fantasy team each week to earn the most points. Winning these weekly matchups allows a team to potentially qualify for the playoffs. This is where the top teams in the league compete in a single-elimination tournament and the ultimate winner is crowned champion of the league.

Of course, there needs to be an equitable way to distribute the player pool, otherwise most teams would end up composed of the same, top performing players and the league would be homogenous. There are two primary means of selecting players: auction drafts and snake drafts. This analysis will focus on snake drafts, which involve assigning each fantasy manager a number between one and the number of teams in the league. There are then a defined number of rounds where teams select players in sequential order. The "snake" part of the draft refers to the fact that the team with the last pick in round j has the first pick in round $j + 1$. This behavior is shown in greater detail in Table F1 of Appendix F. The draft ends after the final round when all selections are made.

A limited amount of background knowledge of baseball is assumed as this model is discussed, which is shared in Appendix B. This problem concerns a 12-team fantasy baseball league that has a 25-round snake format draft. The league is standard 5x5 roto format with OBP and SOLDs as altered categories. Weekly head-to-head matchups pit two teams against each other based on the five hitting and five pitching categories tied to the performance of Major League Baseball (MLB) players. The fantasy team with the better total in a category at the end of the week wins one point, where ties are allowed. This means there are a maximum of 10 points won each week. Additionally, there are 25 roster spots on a team, but only 18 can accumulate points each day. Of those 18, there are positional requirements that mimic the makeup of an actual baseball team.

The goal of the model is to draft the optimal fantasy roster subject to a set of constraints that are based on the author's fantasy baseball league. Targets for scoring categories were determined based on the winning team in the same fantasy baseball league in the 2023 season. Specifically, the objective of the model is to maximize the amount that the drafted roster improves on the established target in all categories. This is analogous to a workforce balancing problem where a business seeks the greatest benefit subject to a set of worker position and other constraints that must be satisfied to perform operations.

By conducting this analysis, the author seeks to gain a competitive advantage in the fantasy baseball league relative to the 11 other teams and make the playoffs. While there are various statistical models and software that are designed to aid teams in constructing a competitive roster, online research suggested that it is uncommon to use an optimization model to achieve this. Since this approach is relatively novel, there is a high probability that insights gained from this analysis will result in arbitrage opportunities for undervalued players and unorthodox draft strategies that can be employed to build the optimal roster in a real-life fantasy baseball league draft.

The optimization model is an integer programming (IP) model that is an original formulation since the objective function and constraints were specific to the author's unique use case. Development went through several different platforms, starting with a small-scale model developed in Microsoft Excel. The model was then ported to GAMS as a proof-of-concept with a subset of the total data. After exceeding the maximum number of decision variables allowed on an academic license, the model was subsequently re-written in Julia, an open-source programming language, using the *JuMP.jl* library. The solver that was used was *HiGHS*, an open-source solver that specializes in linear programming and mixed-integer programming models.

Section 2. Problem Description

The IP model is built to simulate a fantasy draft and select the roster that will maximize the benefit in each statistical category. Since the model is applied to a niche situation, the components of the model will be discussed as they are presented in mathematical form. Firstly, the sets and indices of the model will be defined.

Sets

i = [1-579]; Player name

j = [1-25]; Round in draft

k = [HR, R, RBI, SB, OBP, W, SOLD, SO, ERA, WHIP]; Statistical category

p = [C, 1B, 2B, 3B, SS, INF, OF, UT, P]; Position

There are four different sets in total. The players are represented by i, while the rounds in the draft are defined by j. To allow for algebraic and compact formulations, two additional sets are needed. There are 10 scoring categories in the fantasy league, and these will be represented by k. Definitions of each item can be viewed in Appendix *C*. Additionally, the defensive positions that the fantasy roster must contain are modeled with p. Descriptions of each position and where they play on the field can be found in Appendix D and Appendix E, respectively.

Parameters

T_k = [275, 1000, 1000, 200, 4.55, 100, 1200, 80, 37, 12]; Target total for category k

W_k = [1, 1, 1, 1, 1, 1, 1, 1, -1, -1]; Weight for objective k

c_{ik} = Projected end-season total for category k for player i

ADP_i = Average draft position for player i in similar 12-team fantasy drafts

$MinPos_p$ = [1, 1, 1, 1, 1, 5, 4, 11, 7]; Minimum required players of position p

Pos_{ip} = Binary indicator if player i is eligible for position p

DP = 6; Starting draft position in round one

$Teams$ = 12; Number of teams in league

Next, the parameters, or given data, will be discussed. As mentioned previously, there are targets for all 10 categories, T_k. To keep the model linear, the three ratio targets are converted to summations. This is done by estimating that there will be 13 hitters and 10 pitchers on the final roster. As an example, the desired target for ERA is 3.7, but the number input into the model will be 37. Values of array w_k range from +1 to -1 and represent the direction of improvement for category k, with WHIP and ERA being the only negative categories. Each player i has a projected season total in category k, which is sourced from Fangraphs (2024) and stored in a text file, defined by the two-dimensional array c_{ik}. Similarly, each player i has an average draft position, or ADP, which dictates the latest pick in the draft where that player could be selected by the model, ADP_i. This information is stored in the same text file but was obtained from National Fantasy Baseball Championship (NFBC 2024) Rotowire Online 12-team leagues with drafts conducted during the month of March. Like there are targets for categories, there are also minimum thresholds for each position p that was previously defined, $MinPos_p$. Each player i also has an indicator field for position p eligibility, Pos_{ip}.

Decision Variables

$$x_{ij} = \begin{cases} 1, & if\ player\ i\ is\ selected\ in\ the\ jth\ round\ of\ the\ draft \\ 0, & otherwise \end{cases}, \forall i, j$$

obj_k = Normalized deviation of total of category k from target k, $\forall k$

The main set of decision variables, x_{ij}, is a binary variable indicating if player i was selected in round j of the draft. Given that there are 579 players and 25 rounds in the draft, that means there are 14,475 decision variables. This is why the model moved from GAMS to Julia. Additionally, obj_k is the subobjective for each statistical category k, meaning that there are 10 objectives in total.

Objective Function

$$\max \mathrm{z} = \sum_k w_k obj_k$$

The objective function maximizes the sum of each goal multiplied by the direction of improvement. This is a goal programming problem where each goal k is weighted equally since no single category is more important than any other.

Expressions and Constraints

Subobjectives) $$\frac{\sum_{i=1}^{579} \sum_{j=1}^{25} c_{ik} x_{ij} - T_k}{T_k} = obj_k\ \ for\ all\ k$$

An expression must be defined that links the decision variables to the subobjectives. In this case, the projected total in category k is normalized by its target and then the target is subtracted. This results in the deviation from the target for category k as a goal.

$$\text{RoundMax)} \qquad \sum_{i=1}^{579} x_{ij} = 1, \qquad j = 1 - 25$$

The RoundMax constraint helps model the correct behavior of a fantasy draft. It ensures that exactly one player is selected in each round j of the draft. This is because each fantasy team is only allotted one selection per round and cannot decide to forgo selecting a player.

$$\text{PlayerMax)} \qquad \sum_{j=1}^{25} x_{ij} \leq 1, \qquad i = 1 - 579$$

The PlayerMax constraint makes certain that each player i is only selected a maximum of once in the draft. This is because a player is removed from the draft board once they are chosen by a team; no other team can take them, and the same team cannot select them again later.

$$\text{ADPOddMax)} \qquad \sum_{i=1}^{579} ADP_i x_{ij} \geq Teams(j-1) + DP, \qquad j = 1,3,\ldots,25$$

$$\text{ADPEvenMax)} \qquad \sum_{i=1}^{579} ADP_i x_{ij} \geq Teams(j-2) + DP + 1, j = 2,4,\ldots,24$$

These two constraints mimic the behavior of the snake draft for a fantasy team drafting in the 12th position of the first round. Odd and even rounds follow different patterns, hence why there must be multiple expressions. To illustrate why this is necessary, please refer to the Table F1 in Appendix F.

$$\text{PositionMin)} \qquad \sum_{i=1}^{579} \sum_{j=1}^{25} Pos_{ip} x_{ij} \geq MinPos_p \text{ for all p}$$

The PositionMin constraint makes sure that the positional requirements of the final drafted roster meet the minimum requirement for position p. As an example, there must be at least one catcher on the roster, at least seven pitchers, and so on.

$$\text{Bounds)} \qquad \begin{array}{ll} x_{ij} \in \{0,1\}, & i = 1 - 579, j = 1 - 25 \\ obj_k \geq 0, & k = HR, R, RBI, SB, OBP, W, SOLD, SO \\ obj_k \leq 0, & k = WHIP, ERA \end{array}$$

Finally, the bounds for the decision variables must be set. As mentioned earlier, x_{ij} can only take the value of 0 or 1. To ensure that the resulting roster is well-balanced, the total for each category k must at least meet the target. In the case of WHIP and ERA, these means that it must be non-positive. All other categories must be non-negative. As stated earlier, the outcome of each category only has three possibilities, which means that the team needs to just do more than the opponent; there are no extra points for the winning by a large amount.

Section 3. Numerical Analysis

The base model has been solved to optimality and yields the roster in Table 1.

Table 1. Base Model Roster

Round	Player	Round	Player
1	Zack Wheeler	14	Jose Berrios
2	Luis Castillo	15	Willy Adames
3	Framber Valdez	16	Daulton Varsho
4	Zac Gallen	17	Tyler O'Neill
5	Camilo Doval	18	Starling Marte
6	Evan Phillips	19	Jeremy Pena
7	Andres Gimenez	20	Kyle Finnegan
8	Kyle Schwarber	21	Ryan Mountcastle
9	Dansby Swanson	22	Tommy Edman
10	Chris Bassitt	23	Matt Chapman
11	Merrill Kelly	24	Jose Siri
12	Cedric Mullins	25	Harrison Bader
13	Logan O'Hoppe		

The model targeted pitching early in the draft, beginning with starters and then moving to relievers after round four. This is counter to most modern draft strategies that tend to prioritize hitters in the first several rounds. The resulting projected totals by category are shown in Table 2.

Table 2. Base Model Category Statistics

Category	Target	Actual	Delta
HR	275	312	37
R	1,000	1,069	69
RBI	1,000	1,002	2
SB	200	250	50
OBP	0.350	0.319	-0.031
W	100	100	0
SOLD	75	96	21
SO	1,200	1,470	270
WHIP	1.20	1.20	0.00
ERA	3.70	3.68	-0.03

Nearly all objectives have been satisfied, apart from OBP. This is due to the conversion of ratio statistics from average to summation required to preserve linearity. Upon further inspection, the actual supplied target, 4.55, has been satisfied. Home runs and strikeouts are two categories with a robust surplus relative to the target. The strong emphasis on pitching, particularly starters, early in the draft is driving the strikeout volume. Hitters who specialize in power, like Kyle Schwarber and Logan O'Hoppe, are key players that are contributing to the home run total.

Projection System

A key requirement of fantasy drafts is to have an estimate of how the players will perform in the upcoming season. The naïve method might be to use the most recent season's results, but best practice is to use one of the industry standard projection systems. This includes *ZiPS*, *Steamer*, *ATC*, *THE BAT*, and *THE BAT X*, which were all obtained from Fangraphs (2024). The base model utilized *THE BAT X* projections for hitters, while *ATC* was used for pitchers. The parameter c_{ik} will be adjusted for each system based on the Fangraphs data. The optimal roster for each projection system is in Table 3.

Table 3. Roster by Projection System Scenario

Round	Base	ZiPS	Steamer	ATC	THE BAT
1	Zack Wheeler	Elly De La Cruz	Kevin Gausman	Zack Wheeler	Max Fried
2	Luis Castillo	Emmanuel Clase	Framber Valdez	Luis Castillo	Zack Wheeler
3	Framber Valdez	Yoshinobu Yamamoto	Pablo Lopez	Framber Valdez	Pablo Lopez
4	Zac Gallen	Framber Valdez	Max Fried	Zac Gallen	Framber Valdez
5	Camilo Doval	Raisel Iglesias	Camilo Doval	Logan Gilbert	Camilo Doval
6	Evan Phillips	Camilo Doval	Grayson Rodriguez	Camilo Doval	Evan Phillips
7	Andres Gimenez	Kyle Schwarber	Kyle Schwarber	Kyle Schwarber	Dylan Cease
8	Kyle Schwarber	Jackson Chourio	Pete Fairbanks	Clay Holmes	Kyle Schwarber
9	Dansby Swanson	Tanner Scott	George Springer	Chris Bassitt	Chris Bassitt
10	Chris Bassitt	Esteury Ruiz	Esteury Ruiz	Esteury Ruiz	Dansby Swanson
11	Merrill Kelly	Salvador Perez	Carlos Rodon	Salvador Perez	Hunter Brown
12	Cedric Mullins	Cedric Mullins	Cedric Mullins	Cedric Mullins	Cedric Mullins
13	Logan O'Hoppe	Gerrit Cole	Trevor Story	Jose Berrios	Logan O'Hoppe
14	Jose Berrios	Jordan Montgomery	Mitch Garver	Jose Alvarado	Christopher Morel
15	Willy Adames	Matt McLain	Mason Miller	Willy Adames	Daulton Varsho
16	Daulton Varsho	Willy Adames	Daulton Varsho	Daulton Varsho	Willy Adames
17	Tyler O'Neill	Charlie Morton	Starling Marte	Starling Marte	Nestor Cortes
18	Starling Marte	Kodai Senga	Steven Kwan	Tyler O'Neill	Starling Marte
19	Jeremy Pena	Jack Suwinski	Reid Detmers	Jonathan India	Ryan McMahon
20	Kyle Finnegan	Jake McCarthy	Jonathan India	Jack Suwinski	Jonathan India
21	Ryan Mountcastle	Tommy Edman	Ryan Mountcastle	Matt Chapman	Nathaniel Lowe
22	Tommy Edman	Ceddanne Rafaela	Anthony Rendon	Zach Neto	Tommy Edman
23	Matt Chapman	Jasson Dominguez	Jose Siri	Eugenio Suarez	Jose Siri
24	Jose Siri	Victor Scott II	Josh Bell	Jose Siri	Matt Chapman
25	Harrison Bader	Pete Crow-Armstrong	Jake Fraley	Jake Fraley	Jake Fraley

There are numerous commonalities between the different scenarios since four of the five follow the exact same draft strategy: prioritize starting pitching, then relief pitching, and, finally, hitters. In all cases, the IP model is targeting position players that excel at one or two categories in the middle to late rounds. Examples include the speedy Esteury Ruiz, whose primary contribution is stolen bases, and Cedric Mullins, who offers moderate power and speed, ultimately adding to home runs and steals, respectively. Mullins is drafted in the same round in all five scenarios, identifying him as a player of interest.

The *ZiPS* model is clearly unique and is worth exploring further. The model makes two early selections that would be considered high-risk: Elly De La Cruz and Yoshinobu Yamamoto. De La Cruz is an electric combination of power and speed who set the fantasy world on fire with his debut last season, but there were stretches where his skills were at or below replacement level. While he has had tremendous success in Japan, Yamamoto has not thrown a major league pitch prior to 2024. Investigating further, it becomes clear that there is a player profile that *ZiP*S is bullish on: young players with top prospect pedigree. In addition to those already mentioned, this includes Jackson Chourio, Ceddanne Rafaela, Jasson Dominguez, Victor Scott II, and Pete Crow-Armstrong. The other systems discount the value of these players since it is rare that a first-year player becomes immediately fantasy relevant, but *ZiPS* is gambling on the potential upside these players offer. The projected totals for each category for all scenarios are presented in Table 4.

Table 4. Category Statistics by Projection System Scenario

Category	Target	Base	ZiPS	Steamer	ATC	THE BAT
HR	275	312	275	315	312	317
R	1,000	1,069	1,138	1,151	1,023	1,061
RBI	1,000	1,002	1,091	1,032	1,002	1,001
SB	200	250	351	246	213	214
OBP	0.350	0.319	0.317	0.328	0.319	0.320
W	100	100	100	100	100	100
SOLD	75	96	139	98	97	75
SO	1,200	1,470	1,372	1,647	1,495	1,515
WHIP	1.20	1.20	1.18	1.19	1.19	1.19
ERA	3.70	3.68	3.46	3.49	3.58	3.69

Since they followed the same draft strategy, the base case, *Steamer*, *ATC*, and *THE BAT* have similar team totals. These scenarios exhibit strong categorical balance, but usually one or two statistics are favored slightly. As an example, *Steamer* has the most projected RBIs and strikeouts among the group. *ZiPS* is the outlier once again, producing a somewhat unbalanced team. This case projects for the least number of total home runs but greatly exceeds target stolen bases and SOLDs. This can partly be explained by the players on the *ZiPS* roster leaning towards speed over power, but it seems that this system is more aggressive in its projections overall.

Wins are clearly a bottleneck since no model has more than the minimum target of 100. This observation also helps explain the primary draft strategy the models are following. Wins are a difficult statistic to project since they are more prone to randomness than most of the other categories. Starting pitchers are the most reliable source of wins; therefore, the best starters need to be selected early in the draft to ensure that the 100-win threshold is met.

In summary, the projection system applied to the draft pool is a key consideration when evaluating which players to draft. The fantasy manager should investigate the underlying model parameters of their projection system of choice to ensure they are aligned with the methodology.

Draft Value

In the base model, average draft position (ADP) is used as a proxy for market value, but there are other options that could be applied to the parameter ADP_i. This includes the latest pick a player was selected, MaxDP. Both values were sourced from NFBC (2024) data from similar 12-team fantasy leagues to estimate market sentiment towards players. ADP will also be adjusted 15 percent lower (LowDP) and 15 percent higher (HighDP) to account for potential randomness in the fantasy draft. The resulting roster for each case is shown in Table 5.

Table 5. Roster by Draft Value Scenario

Round	Base	LowDP	HighDP	MaxDP
1	Zack Wheeler	Zack Wheeler	Shohei Ohtani	Spencer Strider
2	Luis Castillo	Luis Castillo	Zack Wheeler	Fernando Tatis Jr.
3	Framber Valdez	Zac Gallen	Zac Gallen	Luis Castillo
4	Zac Gallen	Framber Valdez	Kevin Gausman	Randy Arozarena
5	Camilo Doval	Andres Munoz	Max Fried	Framber Valdez
6	Evan Phillips	Kyle Schwarber	Framber Valdez	Zac Gallen
7	Andres Gimenez	Ryan Helsley	Kyle Schwarber	Raisel Iglesias
8	Kyle Schwarber	Andres Gimenez	Andres Munoz	Andres Munoz
9	Dansby Swanson	Chris Bassitt	Clay Holmes	Evan Phillips
10	Chris Bassitt	Cedric Mullins	Andres Gimenez	Blake Snell
11	Merrill Kelly	Logan O'Hoppe	Esteury Ruiz	Kyle Schwarber
12	Cedric Mullins	Jose Berrios	Chris Bassitt	Andres Gimenez
13	Logan O'Hoppe	Willy Adames	Cedric Mullins	Chris Bassitt
14	Jose Berrios	Daulton Varsho	Merrill Kelly	Jarren Duran
15	Willy Adames	Brayan Bello	Logan O'Hoppe	Dansby Swanson
16	Daulton Varsho	Starling Marte	Christopher Morel	Cedric Mullins
17	Tyler O'Neill	Ryan McMahon	Daulton Varsho	Christopher Morel
18	Starling Marte	Tommy Edman	Willy Adames	Willy Adames
19	Jeremy Pena	Matt Chapman	Starling Marte	Esteury Ruiz
20	Kyle Finnegan	Jose Siri	Tyler O'Neill	Rhys Hoskins
21	Ryan Mountcastle	Jose Abreu	Kyle Finnegan	Daulton Varsho
22	Tommy Edman	Eugenio Suarez	Jeremy Pena	Matt Chapman
23	Matt Chapman	Harrison Bader	Ryan Mountcastle	Jordan Montgomery
24	Jose Siri	Andrew Nardi	Matt Chapman	Henry Davis
25	Harrison Bader	Javier Baez	Jose Siri	Starling Marte

The LowDP model essentially increases the value of each player in the draft. This model opts for many of the same players as the base model, albeit drafted several rounds sooner. This suggests that players like Matt Chapman, drafter four rounds earlier, are exceptional values at their ADP. Camilo Doval is not selected in the LowDP model because his draft value was increased to the point that it is no longer optimal to select him given his output. The HighDP and MaxDP models essentially relax the draft value constraint, resulting in all players having the same or decreased draft values. Since the model can select better players at certain spots, Shohei Ohtani, a dynamic two-way superstar, is selected in the first round of the HighDP model. The MaxDP model instead opts for Spencer Strider in the first round, widely considered the top fantasy baseball pitcher coming into the 2024 season. Otherwise, both cases have many of the same players present in the base case, just selected in later rounds of the draft. The total team statistics for each scenario are outlined in Table 6.

Table 6. Category Statistics by Draft Value Scenario

Category	Target	Base	LowDP	HighDP	MaxDP
HR	275	312	310	335	341
R	1,000	1,069	1,049	1,115	1,161
RBI	1,000	1,002	1,006	1,035	1,062
SB	200	250	222	276	311
OBP	0.350	0.319	0.316	0.322	0.327
W	100	100	100	100	100
SOLD	75	96	84	95	100
SO	1,200	1,470	1,464	1,471	1,557
WHIP	1.20	1.20	1.20	1.09	1.17
ERA	3.70	3.68	3.63	3.25	3.48

As anticipated, the HighDP and MaxDP models have similar or better statistics in all categories when compared to the base model. This is an obvious consequence of lowering the draft value of all players. While the MaxDP model is superior to the HighDP model in nearly all categories, the HighDP roster's pitching ratios have a strong advantage. Similarly, the LowDP model has improved ERA and the same WHIP as the base model. This might indicate that ratio categories are resistant to changes in draft value. The LowDP model is the most conservative and therefore, the most probable scenario, but is still able to achieve all model objectives except OBP. A shrewd fantasy manager would be wise to construct their team with this in mind.

<u>Starting Draft Position</u>

Within the fantasy community, there are differing opinions on where the optimal starting draft position is, which stem from the snake-like nature of traditional drafts. Three popular philosophies assert that starting at the beginning, middle, or end of the first round will ultimately construct the best team. This analysis will investigate the impact of starting with Pick 1, Pick 6, and Pick 12 in the first round, which means that the scalar DP and the ADPEvenMax constraint will be adjusted accordingly. The different rosters for each scenario are shown in Table 7.

Pick 1) $$\sum_{i=1}^{579} ADP_i x_{ij} \geq Teams * j, \;\; j = 2,4,\dots,24$$

Pick 6) $$\sum_{i=1}^{579} ADP_i x_{ij} \geq Teams(j-1) + DP + 1, j = 2,4,\dots,24$$

Table 7. Roster by Starting Draft Position Scenario

Round	Pick 1	Pick 6	Base (Pick 12)
1	Ronald Acuna Jr.	Fernando Tatis Jr.	Zack Wheeler
2	Zac Gallen	Luis Castillo	Luis Castillo
3	Luis Castillo	Zac Gallen	Framber Valdez
4	Max Fried	Max Fried	Zac Gallen
5	Framber Valdez	Framber Valdez	Camilo Doval
6	Andres Munoz	Andres Munoz	Evan Phillips
7	Evan Phillips	Evan Phillips	Andres Gimenez
8	Dansby Swanson	Andres Gimenez	Kyle Schwarber
9	Andres Gimenez	Chris Bassitt	Dansby Swanson
10	Esteury Ruiz	Esteury Ruiz	Chris Bassitt
11	Chris Bassitt	Dansby Swanson	Merrill Kelly
12	Jose Berrios	Cedric Mullins	Cedric Mullins
13	Cedric Mullins	Jose Berrios	Logan O'Hoppe
14	Willy Adames	Willy Adames	Jose Berrios
15	Christopher Morel	Christopher Morel	Willy Adames
16	Daulton Varsho	Daulton Varsho	Daulton Varsho
17	Brayan Bello	Henry Davis	Tyler O'Neill
18	Jack Suwinski	Brayan Bello	Starling Marte
19	Jeremy Pena	Ryan McMahon	Jeremy Pena
20	Ryan Mountcastle	Jeremy Pena	Kyle Finnegan
21	Matt Chapman	Ryan Mountcastle	Ryan Mountcastle
22	Tommy Edman	Tommy Edman	Tommy Edman
23	Shea Langeliers	Matt Chapman	Matt Chapman
24	Jose Siri	Jose Siri	Jose Siri
25	Andrew Nardi	Andrew Nardi	Harrison Bader

In very rare cases, there is a consensus top pick in a fantasy draft. Coming off a 50-home run and 70-stolen base season, Ronald Acuna Jr. was that player. The Pick 1 model capitalizes on his strong projections in all five offensive categories and selects him in the first round. Beyond that single deviation, the strategy employed by the Pick 1 model is nearly identical to the Pick 12 model. The roster of the Pick 6 model only differs from the Pick 1 model by three players. This is likely because Fernando Tatis Jr. is selected in the first round instead of Acuna. While his profile is comparable to Acuna, he has a lower floor and weaker projections in certain categories, like steals. The remaining two player exchanges attempt to close this gap. The statistical differences between the models is given in Table 8.

Table 8. Category Statistics by Starting Draft Position Scenario

Category	Target	Pick 1	Pick 6	Base (Pick 12)
HR	275	315	306	312
R	1,000	1,104	1,100	1,069
RBI	1,000	1,001	1,004	1,002
SB	200	305	289	250
OBP	0.350	0.319	0.320	0.319
W	100	100	100	100
SOLD	75	87	87	96
SO	1,200	1,417	1,417	1,470
WHIP	1.20	1.20	1.20	1.20
ERA	3.70	3.62	3.62	3.68

Overall, this analysis does not clearly indicate that starting at any of these positions is markedly better than the others. All three scenarios, even Pick 1 and Pick 12, have very similar categorical outputs, with the only notable statistic that differs being steals. The presence of either Acuna or Tatis is the reason for this variance, which indicates that their skillsets are rare. If anything, these findings imply that the snake format draft is reasonably effective at distributing player value. The best strategy might be to identify a specific player of interest in the first round and attempt to draft in a position where they will be available.

Goal Reduction

It can be difficult to construct a roster than is balanced enough to excel at all 10 categories, so a common approach is to purposefully give up, or "punt," a category (or two) to strengthen all others. This is akin to reducing the number of goals in the model, which will be achieved by dropping one or more k categories. The most common strategy is to punt saves and holds since they depend on several factors beyond the player's skill level, but there are two additional strategies that will be explored. The resulting rosters are in Table 9.

Table 9. Roster by Goal Reduction Scenario

Round	Base	Punt SOLDs	Punt Wins	Punt SOLDs & Wins
1	Zack Wheeler	Zack Wheeler	Kevin Gausman	Aaron Judge
2	Luis Castillo	Luis Castillo	Jose Ramirez	Jose Ramirez
3	Framber Valdez	Framber Valdez	Josh Hader	Jazz Chisholm Jr.
4	Zac Gallen	Zac Gallen	Edwin Diaz	Randy Arozarena
5	Camilo Doval	Logan Gilbert	Blake Snell	Blake Snell
6	Evan Phillips	Bobby Miller	Andres Munoz	J.T. Realmuto
7	Andres Gimenez	Andres Gimenez	Pete Fairbanks	Kyle Schwarber

Table 9. Roster by Goal Reduction Scenario

Round	Base	Punt SOLDs	Punt Wins	Punt SOLDs & Wins
8	Kyle Schwarber	Kyle Schwarber	Dylan Cease	Dylan Cease
9	Dansby Swanson	Esteury Ruiz	Dansby Swanson	Dansby Swanson
10	Chris Bassitt	Chris Bassitt	Tanner Scott	Esteury Ruiz
11	Merrill Kelly	Jarren Duran	Jarren Duran	Jarren Duran
12	Cedric Mullins	Cedric Mullins	Cedric Mullins	Cedric Mullins
13	Logan O'Hoppe	Logan O'Hoppe	Logan O'Hoppe	Christopher Morel
14	Jose Berrios	Jose Berrios	Jose Alvarado	Nick Pivetta
15	Willy Adames	Daulton Varsho	Daulton Varsho	Rhys Hoskins
16	Daulton Varsho	Willy Adames	Willy Adames	Daulton Varsho
17	Tyler O'Neill	Starling Marte	Tyler O'Neill	Tyler O'Neill
18	Starling Marte	Tyler O'Neill	Starling Marte	Starling Marte
19	Jeremy Pena	Jeremy Pena	Jeremy Pena	Reid Detmers
20	Kyle Finnegan	Jack Suwinski	Jack Suwinski	Yusei Kikuchi
21	Ryan Mountcastle	Tommy Edman	Tommy Edman	Griffin Canning
22	Tommy Edman	Ryan Mountcastle	Ryan Mountcastle	Tommy Edman
23	Matt Chapman	Jose Siri	Jose Siri	Jose Siri
24	Jose Siri	Matt Chapman	Matt Chapman	Matt Chapman
25	Harrison Bader	Harrison Bader	Yuki Matsui	JP Sears

By removing the SOLD goal, the model forgoes drafting any relief pitchers and instead replaces them with starters. The strategy for drafting the remaining players remains unchanged. By punting wins, the model chooses to take only two starters and fill the rest of the pitching rotation with a bounty of premium relievers. This could be an effective strategy since relief pitchers generally have better ratios than starting pitchers. Additionally, it is no mistake that the model took Kevin Gausman and Blake Snell as the only starting pitchers; both of those players excel at generating strikeouts, a category a roster built entirely of relievers would struggle to accumulate. Punting both categories results in the roster prioritizing hitters in the first dozen rounds and entirely avoiding relievers. The first starter, Blake Snell, is not taken until round five and the bulk of the pitchers are not taken until the end of the draft. This is sure to affect the projected category totals, shown in Table 10.

Table 10. Category Statistics by Goal Reduction Scenario

Category	Target	Base	Punt SOLDs	Punt Wins	Punt Solds & Wins
HR	275	312	331	307	414
R	1,000	1,069	1,185	1,076	1,373
RBI	1,000	1,002	1,067	1,004	1,259
SB	200	250	323	259	380
OBP	0.350	0.319	0.317	0.321	0.327
W	100	100	100	56	64
SOLD	75	96	0	219	0
SO	1,200	1,470	1,415	1,201	1,201
WHIP	1.20	1.20	1.17	1.18	1.26
ERA	3.70	3.68	3.74	3.21	4.07

Punting saves and holds dramatically increases all offensive output, other than OBP, with very little decrease in projected strikeouts. While WHIP improved, ERA deteriorates due to the dearth of relievers on the roster. Even though wins appeared to be a binding constraint in the model, punting wins is not an objectively better strategy than punting SOLDs. The strongest aspect of this model is that an all-reliever roster has made SOLDs, ERA, and WHIP elite categories that the fantasy team should win almost every week. The data suggests that punting SOLDs and wins is tantamount to giving up on all pitching and only focusing on offense. Even with the strength of the offensive categories, this is not a strategy to win most weeks. Of the three options, punting saves and holds appears to be the only sustainable strategy to stay competitive in all remaining fantasy categories.

Relaxed Goal Bounds

The final scenario will remove the bounds on obj_k for every k, which will allow the model the flexibility to not meet all goals. While this is similar to punting a category, this approach allows the IP model to choose which categories to limit. The resulting roster is shown in Table 11.

Table 11. Roster with Strict vs Relaxed Goal Bounds

Round	Base (Strict)	Relaxed
1	Zack Wheeler	Aaron Judge
2	Luis Castillo	Jose Ramirez
3	Framber Valdez	Edwin Diaz
4	Zac Gallen	Josh Hader
5	Camilo Doval	Camilo Doval
6	Evan Phillips	Andres Munoz
7	Andres Gimenez	Kyle Schwarber
8	Kyle Schwarber	Pete Fairbanks
9	Dansby Swanson	Esteury Ruiz
10	Chris Bassitt	Dansby Swanson
11	Merrill Kelly	Jarren Duran
12	Cedric Mullins	Cedric Mullins
13	Logan O'Hoppe	Logan O'Hoppe
14	Jose Berrios	Jose Alvarado
15	Willy Adames	Willy Adames
16	Daulton Varsho	Daulton Varsho
17	Tyler O'Neill	Starling Marte
18	Starling Marte	Tyler O'Neill
19	Jeremy Pena	Kyle Finnegan
20	Kyle Finnegan	Jeremy Pena
21	Ryan Mountcastle	Ryan Mountcastle
22	Tommy Edman	Tommy Edman
23	Matt Chapman	Matt Chapman
24	Jose Siri	Jose Siri
25	Harrison Bader	Harrison Bader

Since the Relaxed model chooses to only draft relief pitching instead of starting pitching, the model has essentially chosen to concede wins. This is not surprising given that wins are difficult to accumulate, as we have not seen any permutation of this model have a win total above the 100-win threshold. Early in the draft, the model chooses a pair of prolific hitters, Aaron Judge and Jose Ramirez, and two of the best relievers in the entire draft, Edwin Diaz and Josh Hader. The Relaxed model takes the minimum required number of pitchers to maximize offensive statistics. Table 12 displays these values.

Table 12. Category Statistics with Strict vs Relaxed Goal Bounds

Category	Target	Base (Strict)	Relaxed
HR	275	312	389
R	1,000	1,069	1,319
RBI	1,000	1,002	1,223
SB	200	250	333
OBP	0.350	0.319	0.324
W	100	100	25
SOLD	75	96	230
SO	1,200	1,470	567
WHIP	1.20	1.20	1.17
ERA	3.70	3.68	3.14

As predicted, all offensive team totals have been greatly improved in the Relaxed model; this is heavily driven by Judge, Ramirez, and the additional three hitters that the model is able to select. A further consequence of not having any starters is that the strikeout total has fallen quite short of the target. The projected total wins and strikeouts are not enough to be competitive most weeks, but since they are non-zero, there is the potential to steal a point in certain matchups. As was the case when wins was punted, a roster built entirely of relievers means that the pitching ratios are a major point of strength for this roster.

Section 4. Conclusion

This analysis affirms that a fantasy baseball draft is well-suited as an optimization problem and can be successfully modelled using integer programming. Several generalizations must be made in order to satisfy the certainty and linearity assumptions, including deterministic player draft values and summation of ratio statistics. Fantasy baseball leagues are famous for their seemingly endless set of rules that can be customized, and a model such as this could be adapted to nearly any use case to conduct similar analyses. The real power of the IP model comes with stress testing different parameters that are input into the model. Normally, different draft strategies take an entire season to play out. By formulating the problem as a mathematical model, it becomes easy to quickly determine the effects of various inputs on the resulting fantasy roster.

Through numerical experimentation, there were several important discoveries that came to light. The primary theme that kept recurring in nearly all scenarios was that starting pitching should be targeted in the early rounds of the draft. This is surprising because almost all current draft strategies favor hitters at the top of the draft. This finding underscores the impact wins has on fantasy baseball leagues and helps explain why many leagues are beginning to phase out that category in favor of one that occurs more frequently, like quality starts, to allow for more flexible roster construction. Additionally, the scenario analysis confirmed that punting saves and holds is a valid strategy, although it is not the optimal strategy. The optimal strategy avoids starting pitchers entirely to prioritize offense, SOLDs, and pitching ratios. Next, snake-format fantasy drafts are an equitable means of distributing players and a successful fantasy team can be selected regardless of the starting draft position.

Of course, following the recommendations of the optimization model does not mean that the drafted team will be the best in the league. There is uncertainty in all aspects of any fantasy league. Fantasy managers behave irrationally and unexpectedly during drafts, which can entirely derail a planned strategy. Projection systems are essentially educated guesses as to what each player will do in the upcoming season; there are always overperformers and underperformers. A major risk that this analysis completely ignores is player injuries. It is an unavoidable part of the game and impacts every team throughout the course of the season. As an example, Spencer Strider, a consensus first round pick, had season-ending elbow surgery after only two starts in the 2024 season. Depending on how the roster was constructed, that could be a debilitating loss for a fantasy team. These effects compound and essentially render the exact roster that the model selects as entirely theoretical. This does not mean that the analysis is useless, in fact, quite the opposite. The draft strategies, player valuations, and other what-if analysis can lead to insights that can be applied in real-world situations. Based on this analysis, the author followed a punt saves strategy and targeted many of the same players that were continually drafted in all scenarios, like Cedric Mullins and Daulton Varsho, in an actual fantasy draft. The recurrence of these players posits that their ratios of expected performance to market value are quite high.

In future iterations, there are numerous extensions that can be made to the model to make it more robust. Introducing random variables and factoring in additional inputs, like a proxy for injury risk, would make the specific roster selected by the model more actionable. Another idea is to adjust the model to be quickly used live during a fantasy draft. This would allow it to effectively function as an assistant and make recommendations based on the current available pool of players, while also accounting for the random behavior of other teams in the draft.

References

Fangraphs (2024). *2024 Projections* [Data set]. Fangraphs.
Retrieved from https://www.fangraphs.com/projections

National Fantasy Baseball Championship (2024). *ADP / Standings* [Data set]. National Fantasy Baseball Championship. Retrieved from https://nfc.shgn.com/adp/baseball

Appendix A

Supplemental Files

There are several additional files attached to this report that were fundamental to creating the analysis:

- *base.jl*: This is the Pluto notebook that models the base case for the fantasy baseball draft problem. This will only be executable if Julia and the Pluto package are installed on the system.
- *base.html*: This an html export of the *base.jl* file and can be viewed on any computer regardless of the installation of Julia or Pluto. As a combination text and code file, this should read very similar to a blog post. All code is thoroughly documented and discussed so the reader can follow what is happening and why.
- *scenarios.jl*: This is the Pluto notebook that models numerical experimentation for all scenarios in the fantasy baseball draft problem. This will only be executable if Julia and the Pluto package are installed on the system.
- *scenarios.html*: This an html export of the *scenarios.jl* file and can be viewed on any computer regardless of the installation of Julia or Pluto. As a combination text and code file, this should read very similar to a blog post. All code is thoroughly documented and discussed so the reader can follow what is happening and why.

Appendix B

Essential Baseball Concepts

Baseball is widely considered to be America's pastime and for good reason: it is one of the oldest competitive sports in the nation's history. Throughout the decades, there have been significant changes to the game; however, the fundamental rules have remained largely unchanged. The primary action revolves around hitting a baseball; a batter uses a wood bat to attempt to hit an orange-sized ball thrown to them by the opposing team's pitcher. Therefore, batters, or hitters, are the primary means of generating offense, which is known as scoring runs. This occurs when the ball is batted into the field of play and a player touches home plate after rounding three bases. Pitchers must throw the baseball to batters, but their objective is for the ball to be turned into an out. This can happen if the batter earns three strikes by swinging and missing or failing to swing on balls thrown in the strike zone. There are other means of generating outs, where defensive players in the field of play catch a ball before it hits the ground. They might also tag a player with the ball when he is not touching one of the bases. These defensive players play nine positions on the baseball field and use a baseball mitt, or glove, as their primary tool to catch the ball. After three outs are recorded, the teams switch sides and the process repeats. This goes on for at least nine iterations, called innings, after which the team with the most runs wins the game.

The MLB season is comprised of 162 games for each of the 30 teams. These teams belong to one of two leagues, the American League or the National League. Within each league, there are three divisions that contain five teams each. After the completion of the regular season, the teams with the best record in each division, plus the remaining three teams in each league with the best records, meet in a postseason tournament. The winner of the World Series, the final series in the tournament, determines the ultimate champion of the season. While these are arguably the most meaningful games of the season for actual MLB players, they do not have fantasy baseball implications. The fantasy baseball season, including its own playoff structure, is usually concluded at the end of the regular season.

Appendix C

Statistics Glossary

- Home run (HR): A batted ball leaves the fair field of play before touching the ground. Essentially, the ball is launched in the air over the outfield fence. Hitting statistic that can take integer values only.
- Runs scored (R): A player rounds all three bases and touches home plate without being declared out. Hitting statistic that can take integer values only.
- Runs batted in (RBI): A batter hits a ball in play that scores a run, including their own. Hitting statistic that can take integer values only.
- Stolen base (SB): A baserunner advances a base without a ball hit in play. This can also be referred to as a "steal." Hitting statistic that can take integer values only.
- On base percentage (OBP): The percentage of time a batter successfully reaches a base per at bat. Hitting statistic that ranges from 0 to 1.
- Win (W): A pitcher is the active pitcher when the winning team took the lead and ultimately won the game. Pitching statistic that can take integer values only.
- Saves plus holds (SOLD): A pitcher successfully maintains a lead or closed a game out when their team was winning. Pitching statistic that can take integer values only.
- Strikeout (SO): A pitcher records three strikes from a batter during an at bat. Pitching statistic that can take integer values only.
- Earned run average (ERA). The number of runs attributed to a pitcher divided by the number of innings pitched multiplied by nine. Pitching statistic that can take floating point values.
- Walks plus hits divided by innings pitched (WHIP): The number of walks and hits a pitcher gives up in an outing divided by the number of innings pitched in the same outing. Pitching statistic that can take floating point values.
- Quality Start (QS): A starting pitcher records at least six innings pitched and less than or equal to three earned runs in a single outing. Pitching statistic that can take integer values only. Not included in model.

Appendix D

Position Glossary

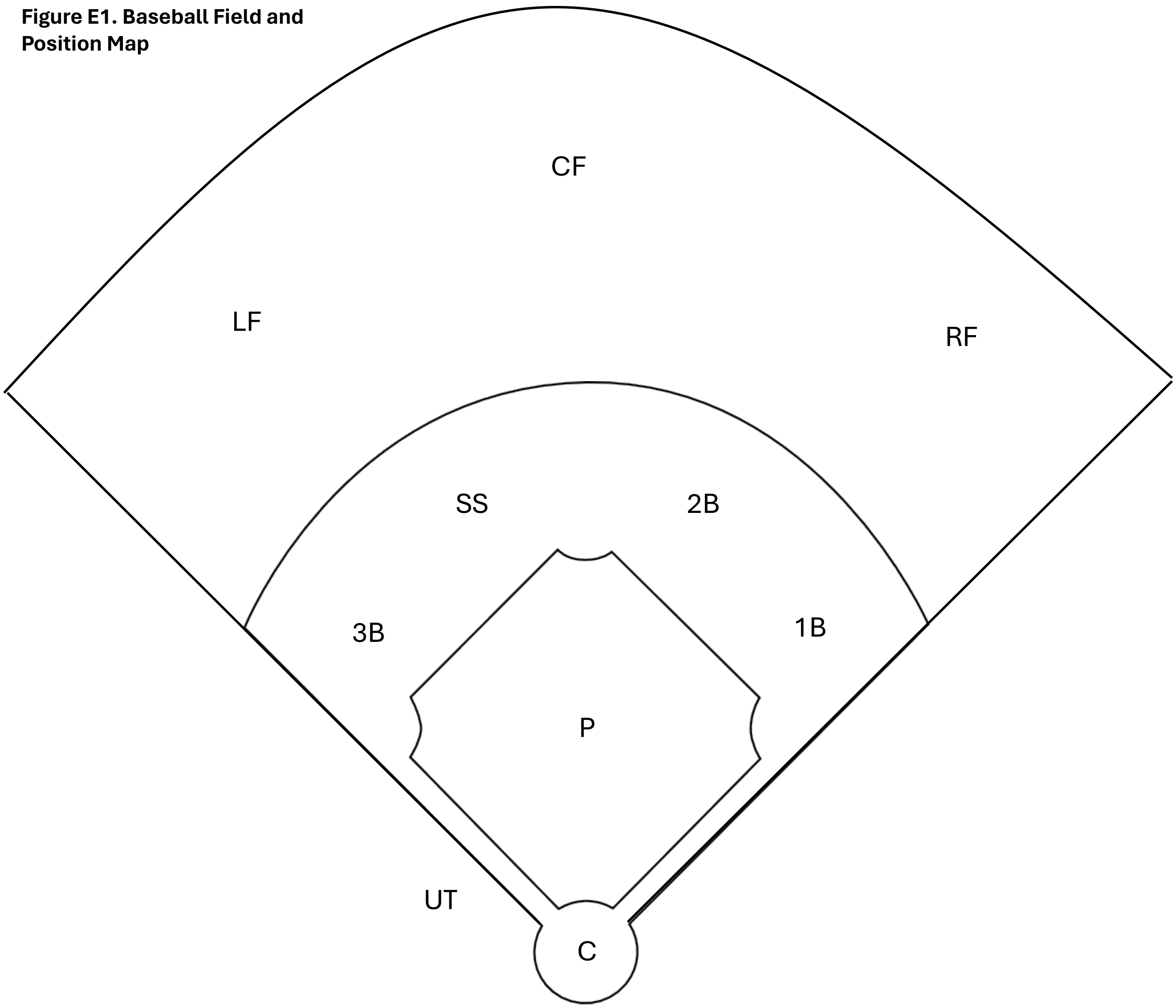
- Catcher (C): Sits behind home plate and catches the ball throw by the pitcher. Responsible for communicating which pitch type the pitcher should throw, often leading to them being known as "game-callers."
- First Baseman (1B): Positioned near first base and is responsible for fielding balls in play and managing baserunners. Typically considered the least skilled infield position.
- Second Baseman (2B): Positioned between first and second base and is responsible for fielding balls in play and managing baserunners.
- Third Baseman (3B): Positioned close to third base and is responsible for fielding balls in play and managing baserunners.
- Shortstop (SS): Positioned between second and third base and is responsible for fielding balls in play and managing baserunners. Typically considered the most difficult infield position.
- Infielder (INF): Grouping of players including 1B, 2B, 3B, and SS.
- Left Fielder (LF): Positioned behind third base in the outfield and fields balls in play. Generally considered one of the least skilled defensive positions on the entire field.
- Center Fielder (CF): Positioned behind second base in the outfield and fields balls in play. A premium defensive position as it requires athleticism and range to successfully cover the massive territory they are responsible for.
- Right Fielder (RF): Positioned behind first base in the outfield and fields balls in play. Requires superior arm strength to make high leverage throws to successfully contest runners at third base and home plate.
- Outfielder (OF): Grouping of players including LF, CF, and RF.
- Batter: Any non-pitcher, but generally any player who registers a plate appearance. Can also be referred to as a "hitter."
- Utility (UT): Any hitter; Grouping of players including C, 1B, 2B, 3B, SS, INF, OF.
- Pitcher (P): Positioned on the pitcher's mound in the infield and throws the ball to the batter standing at home plate. Arguably the most important position on the field.
- Starting pitcher (SP): A subset of pitcher that starts a ball game and usually attempts to pitch as deep into the contest as possible. Can also be referred to as a "starter."

- Relief pitcher (RP): A subset of pitcher that replaces a starter in the middle of a baseball game. Can also be referred to as a “reliever” or “closer.”

Appendix E

Baseball Field and Position Map

Figure E1. Baseball Field and Position Map



Appendix F

Snake Draft Example

Table F1. Draft Pick by Round

Initial Pick	1	2	3	4	5	6	7	8	9	10	11	12
Round 1	1	2	3	4	5	6	7	8	9	10	11	12
Round 2	24	23	22	21	20	19	18	17	16	15	14	13
Round 3	25	26	27	28	29	30	31	32	33	34	35	36
Round 4	48	47	46	45	44	43	42	41	40	39	38	37
Round 5	49	50	51	52	53	54	55	56	57	58	59	60
Round 6	72	71	70	69	68	67	66	65	64	63	62	61
Round 7	73	74	75	76	77	78	79	80	81	82	83	84
Round 8	96	95	94	93	92	91	90	89	88	87	86	85
Round 9	97	98	99	100	101	102	103	104	105	106	107	108
Round 10	120	119	118	117	116	115	114	113	112	111	110	109
Round 11	121	122	123	124	125	126	127	128	129	130	131	132
Round 12	144	143	142	141	140	139	138	137	136	135	134	133
Round 13	145	146	147	148	149	150	151	152	153	154	155	156
Round 14	168	167	166	165	164	163	162	161	160	159	158	157
Round 15	169	170	171	172	173	174	175	176	177	178	179	180
Round 16	192	191	190	189	188	187	186	185	184	183	182	181
Round 17	193	194	195	196	197	198	199	200	201	202	203	204
Round 18	216	215	214	213	212	211	210	209	208	207	206	205
Round 19	217	218	219	220	221	222	223	224	225	226	227	228
Round 20	240	239	238	237	236	235	234	233	232	231	230	229
Round 21	241	242	243	244	245	246	247	248	249	250	251	252
Round 22	264	263	262	261	260	259	258	257	256	255	254	253
Round 23	265	266	267	268	269	270	271	272	273	274	275	276
Round 24	288	287	286	285	284	283	282	281	280	279	278	277
Round 25	289	290	291	292	293	294	295	296	297	298	299	300