Jiachun Lu
Angela Zhang

Stat 157
Course Project

Strategies in Fantasy NBA Basketball

For this project, we will discuss a popular game among basketball fans, Fantasy NBA Basketball, and analyze various statistical related strategies applicable to the game. The strategies will involve different stages of the game play, such as drafting players, selection of free agents, as well as trading players. We will use a more efficient drafting strategies based on a model that we refer to as efficient points rather than the commonly used "O-Rank". We will also use expected values and variances of the players' statistics decide how to either pick or trade free agents during the season. Our final goal is to use these techniques to optimize team performance in order to win the league.

Fantasy Basketball is essentially a free online game that allows players to manage their own team of NBA players upon creating an account. There are two types of major leagues, Rotisserie League and Head to Head League. For rotisserie League, there are several statistical categories such as rebounds, assists, or free throws that fantasy basketball teams are ranked on. Usually, ranks in each statistical category are then converted to corresponding points and are totaled to determine an overall score for the team. The team with the highest score naturally becomes the winner. On the other hand, in a head to head league, different fantasy teams created by different managers square off in a weekly match against each other to see which manager can compile the highest points in each of the different categories. For our purposes, we will only focus on Rotisserie League.

In order to create fantasy teams, managers need to draft players they want during the preseason. Usually managers base their draft on a calculated rank referred to as "O-Rank". "O-Rank" stands for original rank and is calculated based on NBA player's performance from the previous season. For example, for the data we are using, which is obtained from season 04-05, the "O-Rank" is calculated based on data from season 03-04. Usually, the higher the ranking, the better the player has performed in the previous season.

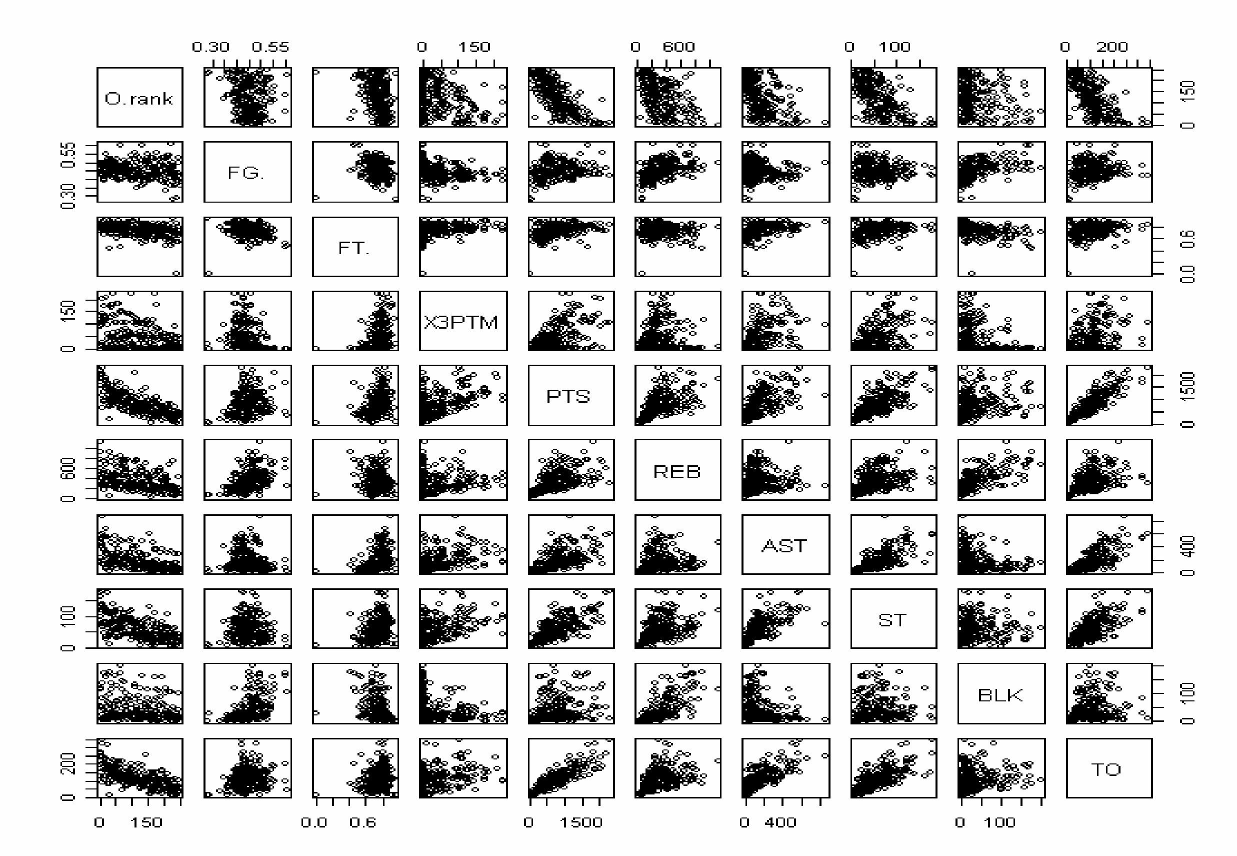
Once the players are drafted and the team is created, real-life statistics are converted to corresponding points for each NBA player as NBA games occurs. For example, if Kobe Bryant's field goal percentage is 30% during the Lakers VS Spurs game, then that percentage gets converted to points under the field goal category in Fantasy Basketball under Kobe Bryant's name and so on. As each NBA game occurs during the week, each player's points in a specific category on a particular fantasy team will be added to obtain the team's total point for that category. For example, if a fantasy team has Kobe Bryant, Tim Duncan, and Kevin Garnet and their assist counts are 3, 5, and 4 respectively, then the team's total assist count would be 3+4+5=12. This method applies to all other categories. Next, the fantasy team's total in each statistical category will be compared to other fantasy teams' totals and ranks are assigned with number 1 rank as the team with the highest number of points in that category. Finally, these ranks are converted to points according to the following method. Say if we are trying to assign points in a particular 12-team league for the category of rebounds. The team with the most rebounds will receive 12 points, and the team with the second most rebounds will receive 11 points and so on. If there is a tie, then each team involved receives 11.5 points because (12+11)/2=11.5. The last step is to sum points across categories for each team to determine the team's total points. Since rankings within the individual scoring categories are based on the cumulative stats earned by all active players during the season, the overall ranking of your team will rise and fall depending on how it performs relative to the performance of other league members. The phenomenon of "losing points" can be explained as your rank falling in one or more of the statistical categories used in your league. During the season of the game, managers are allowed to trade players with each other or pick up free agents to increase their team performance.

As mentioned earlier, the game can be played in different ways based on which league you choose to play, and the strategies behind them are very different. The following is the default settings of a public league-Rotisserie League. There are typically 12 teams in a league and a maximum of 13 players on each team. There are no limits in terms of trading. The starting roster contains 1 point guard, 1 shooting guard, 1 guard, 1 small forward, 1 power forward, 1 forward, 2 centers and 2 utility. There can be up to 3 bench players of any position and the maximum number of games played per position is 82. The statistical categories used are points score, total rebounds, assists, steals, turnovers, block shots, 3-point shots made, field goal percentage, and free throw percentage.

Before we discuss the strategies in the two stages of the game play, preseason and during the season, we need to state our assumptions. One of the most important assumptions to make is that a player's performance can be totally random. Since we can't measure the injury time for each player, we consider this as one of the random factors. In addition, player's mood, which affects how well they play greatly, is another random factor. On the same note, players' performance vary greatly at different time periods; some play better at the beginning of the season, while others play better towards the end of the season, and this is completely random as well.

One of the main concerns for managers is to get prepared for drafting. Drafting is very crucial because having better selections of players at the beginning guarantees later success. A question that a typical manager would ask is "How would we pick our desired players?" We are positive that a good manager will take serious consideration to this question. But before we dive in to the strategies, one thing to keep in mind is that personal preferences will negatively affect a manager's ability to stay objective when selecting players. For example, a particular manager is a big fan of the Lakers, and Kobe Bryant happens to be his or her favorite player on the team. When such a manager encounters a situation where he or she have to make a choice between picking Kevin Garnett and picking Kobe Bryant, it is obvious that he or she will pick Kobe Bryant. However, according to previous records, it would be wiser to choose Kevin Garnett over Kobe Bryant since Kevin Garnett has performed well in many of the categories such as points, rebounds, block shots, FG%. On the other hand, although Kobe has done excellently in scoring points, he still has a comparatively weaker performance in other categories which makes him less valuable compared to Kevin Garnett. Therefore, in our analysis, we assume all managers are unbiased and unaffected by their personal preferences. In order to make the best drafting decision at the most critical time, we have developed a method called "Efficiency Point", which is calculated to help managers with selecting players during drafting.

As mentioned earlier, we use efficient points to decide which players to pick because O-rank might not be the most efficient way for drafting. O-rank is calculated based on almost every statistical category there exits and contains rank for every basketball player in NBA. While O-rank has been the most significant piece of statistics in drafting players, we argue that it is not necessarily the most efficient or appropriate in delivering information regarding the performance of players. Some of the problems that O-rank has are: 1) it contains ranks for players who may not have played very much during the last season for two possible reasons, either the player is rookie or he has been injured. In this sense, the rank is very inaccurate. The player may be a great player who has not been able to play due to injuries or a rookie who has great growing potentials, but simply looking at O-rank would tell you that they had bad performance. 2) Another problem with O-rank is that it may contain more information than we need. For example, as mentioned before, it contains data from all statistical categories. However, a particular manager may decide that only certain statistical categories are important to him or her in drafting. It could be rebounds, assists, free throw percentage, blocks, and field goal percentage. The way of using O-rank as an indicator of drafting players might not be the most efficient way, since there are other leagues which use O-rank as an indicator, too. Hence, it is important to come up with a better method for managers' tailor drafting to their own interest and improving performance of their fantasy basketball team.



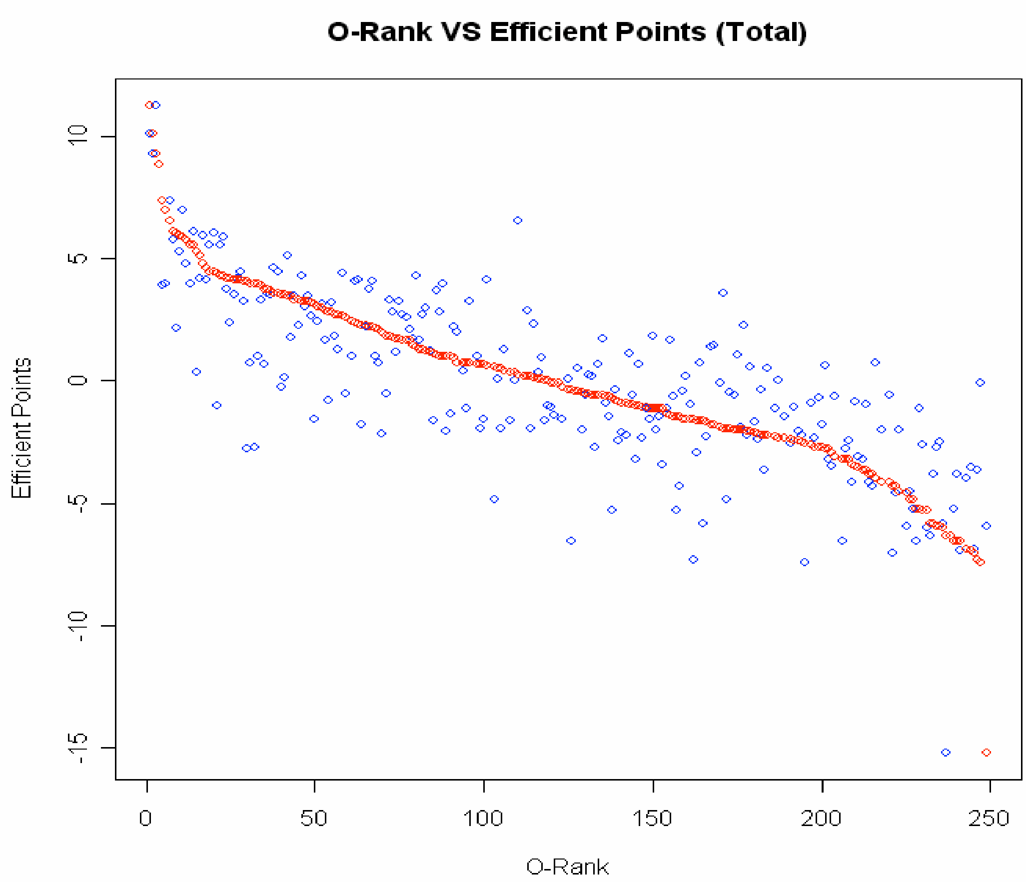
This is a graph plotting O-rank against all the other explanatory variables, such as FG%, FT%, points, rebounds, assists, 3 points made, steals, block shots, and turnovers. From the plot, we noticed that FG%, FT%, 3 points made, block shots does not really have any relationship with O-Rank. Only points, rebounds, assists, steals, and turnovers appear to have a negative relationship with O-rank which indicates that in general, a low number of O-rank will score more points, grab more rebounds, have more assists, steal more balls, and turn over balls more often.

	O.rank	FG%	FT%	X3PTM	PTS	REB	AST	ST	BLK	TO
O.rank	1.00	-0.15	-0.34	-0.37	-0.75	-0.45	-0.49	-0.61	-0.23	-0.67
FG%	-0.15	1.00	-0.22	-0.34	0.18	0.48	-0.14	-0.05	0.51	0.15
FT%	-0.34	-0.22	1.00	0.38	0.31	-0.14	0.35	0.26	-0.37	0.22
X3PTM	-0.37	-0.34	0.38	1.00	0.46	-0.12	0.46	0.44	-0.32	0.28
PTS	-0.75	0.18	0.31	0.46	1.00	0.51	0.59	0.68	0.17	0.85
REB	-0.45	0.48	-0.14	-0.12	0.51	1.00	0.01	0.30	0.68	0.40
AST	-0.49	-0.14	0.35	0.46	0.59	0.01	1.00	0.70	-0.24	0.73
ST	-0.61	-0.05	0.26	0.44	0.68	0.30	0.70	1.00	-0.04	0.69
BLK	-0.23	0.51	-0.37	-0.32	0.17	0.68	-0.24	-0.04	1.00	0.12
TO	-0.67	0.15	0.22	0.28	0.85	0.40	0.73	0.69	0.12	1.00

Here is a table of correlation of how response variable O.rank interacts with its explanatory variables. We notice that O.rank is negatively related to all the explanatory variables which indicate that a lower O.rank will generally have more of everything else. Also, some of the variables have more prediction power than the other variables; a regression model of predicting O.rank might only contain points, rebounds assists, steals and turnovers. In addition, we know that more turnovers will result in more points which make a player less valuable; however, neither the graph nor the correlation table demonstrates this result.

Next, we will present a different method that we came up with, which uses efficient points as the deciding indicator. The following steps will allow us to obtain efficiency point for each player based on the given dataset. First, we need to remove players who do not have previous records for reasons mentioned earlier. Next, since we cannot compare data in different units, we normalize the dataset across all categories by subtracting its mean and then dividing by standard deviation. A normalized table shows us how much better one player is performing relative to the others in any particular category. According to the default setting of the game, we noticed that only nine categories are being played: points, rebounds, assists, steals, turnovers, block shots, 3 points made, FG%, and FT%. Since these categories are equally important, they have equal weights when calculating the efficient points. Therefore, efficiency point can be calculated by summing up the players' total points, rebounds, assists, steals, block shots, 3 points made, FG% and FT% minus turnover, because more turnover results in less points while all other categories are positively related to points.

There are two types of efficient points: efficient points for total value across each category and efficient points for average value across each category. We understand that a player may have more total points because he plays more games, and that is reasonable. However, we know that some of the players do not have a lot of total points, but on average he is performing well. For example, Yao Ming has only scored 1271 points for Rockets during the 05-06 season because of his injury; however, on average he can score 22.3 points per game, which is really good. If we only evaluate players based either on total points or on average points, it might not give us enough information to thoroughly judge a player's value. Therefore, the existence of two types of efficient points seems necessary, and thus we can use these two indicators to determine which players we should pick for the team.



Here is a graph of plotting efficient points against O-rank based on total. The blue dots are where you plot efficient points against O-rank without sorting, and the red dots are where you sorted the players according to their efficient points in a decreasing order, and then plot it against O-rank.

From the graph, we notice that some of the players with lower efficient points are actually in the top rank, and a play with high efficient point is actually in the back. According to the O-rank, those players are No 40, No 41, and No 105.

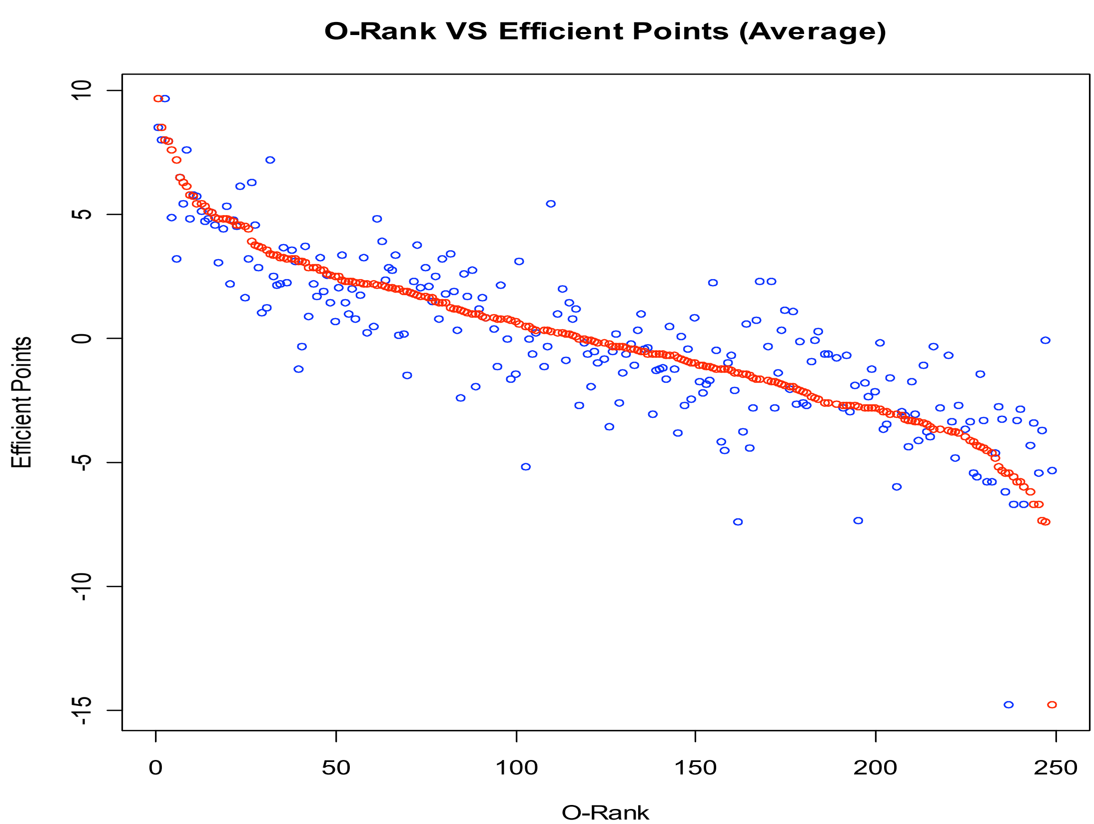
We notice that the graph shows a negative relationship between O-rank and efficient point which indicates that in general, lower value of O-Rank will have a high efficient point.
However there are a few outliers such as players with O-rank number 40, 41, and 105.

name	o.rank	Gp	FG%	FT%	X3ptm	Pts	Reb	Ast	St	Blk	To
B.Gordon	40	82	0.41	0.86	134	1235	215	164	53	10	186
C.Anthony	41	75	0.43	0.80	42	1558	426	194	68	30	224
A.stoudemire	110	80	0.56	0.73	3	2080	713	131	77	130	189

By looking at the table, we notice that A.stoundemire ranked 110 based on O-rank, but has a total of 2080 points, 713 rebounds, 77 steals, 130 block shots, which completely outperforms both B. Gordon, and C.Anothony. Hence, if managers only base drafting on O-Rank, they probably miss the opportunity of getting a great player such as A.stoudemire.

Next, we sort the efficient points and rank them based on their efficient points. A player with highest efficient point will be noted as No. 1 and so on. Red dots essentially show that the higher the efficient points, the higher the Rank. Also, we notice that the variances of the red dots are relatively stable. This helps managers to distinguish players. Simply by looking at the efficient points, managers can tell which player is better and which they should pick for the team during drafting.

However, will average method work the same way?



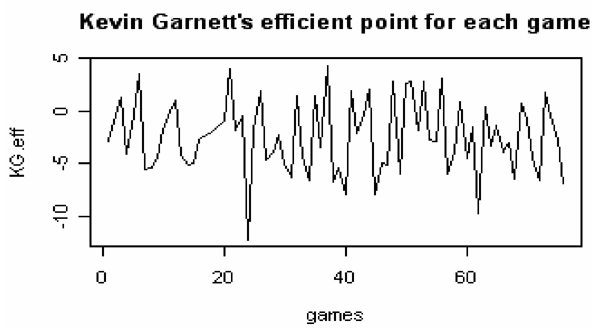
Comparing the two graphs (Average VS Total), the only difference is that blue dots on the second graph converge to red dots better than the blue dots on the first graph. For example, as we mentioned earlier, B. Gordon and C.Anthony with O-rank of 40 and 41 respectively are pretty significant in the first graph, but not as significant in the second graph because the vertical distance between blue dots and red dots has been decreased.

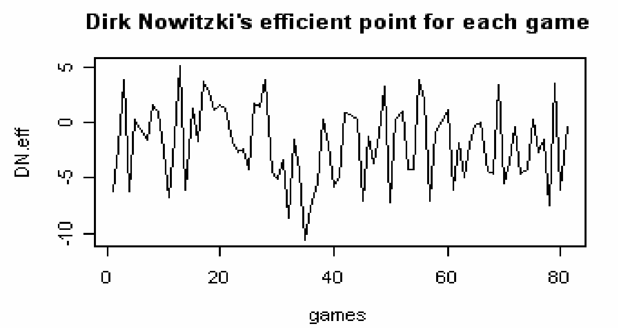
Comparing the order of the efficient points with two methods, we still find there are a lot of differences in ranking them. For example, the order of the players based on total are ranked as following: 3 1 2 4 7 11 105 14 20 17. On the other hand, the orders of the players based on average are ranked: 3 1 2 4 9 32 7 27 24 11. With the exception that the top four players maintain the same seats, the rest of the seats are filled with different players. Therefore, we conclude that best players are always the best no matter how you rank them. A detailed ranking based on total and on average is included in the appendix.

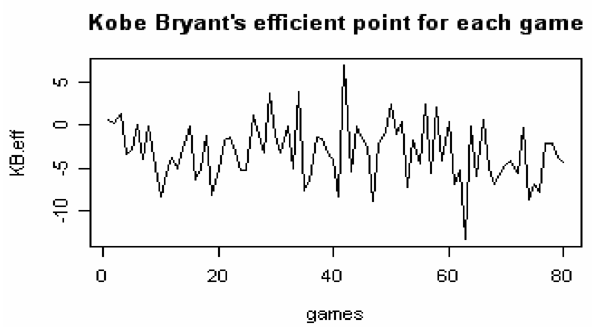
Drafting players simply by looking at efficient points is not enough. For example, if a manager picks D.Wade and G.Arenas, two of the best guards in the NBA league, the team will be very strong in scoring, assisting, and steals since they have shown the same strengths in these fields. They represent what we call overlapping players, and managers should avoid keeping them together. If a manager gets E. Brand to replace one of them, the team will be much more balanced across different categories. As a result, we can achieve more overall points and have a better chance of wining the league. We have come to the conclusion that the best way of drafting is to draft each position with the most efficient players one could possibly get. Managers should avoid drafting too many players on a single position such as either guards or forwards because it will create imbalances later in the season.

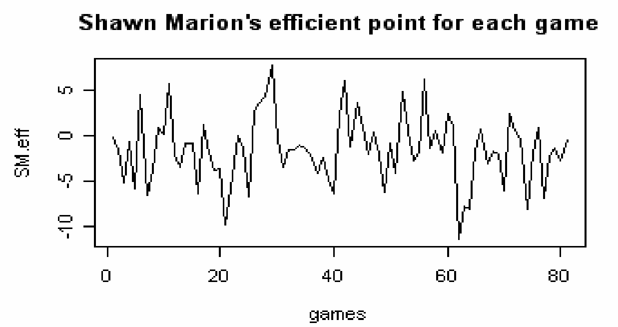
Besides drafting, during the season, a manager can increase team performance either by trading players with other team managers or picking up free agents from the player pool. Let's assume currently there is a player with a poor performance, and the manager would like to get rid of him by trading with another player or picking another player in the pool to replace him. Through my experience, it's relatively hard to trade players with other managers, since a bad performance player generally would not be accepted by other managers. The most likely condition is that a team may want to trade their uncut players to improve team performance in a desired category. An uncut player is a player whose performance is typically the best in the team and as a result, a logical manager would be reluctant to cut him. For example, if a team has a lot of rebounds, the manager may consider trading one or more of his or her forwards with players that are good at assists or 3 points, whichever is a weaker category for the team, so that it can improve overall in order to gain more total points. On the other hand, another good way to improve team performance is to drop bad performing players and pick up free agent players that could improve the team's total points. A common question to ask is which player and should I drop and which players should I pick up. To answer the first part of the question, which player should one drop, we need to look at a summary chart known as "team log" for players that can potentially be cut. A team log shows a player's recent game records. If a player is not on the uncut list and he is performing way below his expected performance, then this is a strong sign indicating that he needs to be substituted. The way of finding the right substitute can be executed in two steps: 1) we need to determine who the possible candidates are and 2) make a selection from these potential candidates. For the first step, we sort unpicked players according to their "Rank" in a decreasing order. The first few are the possible candidates since they have the highest ranks. But how will we decide which candidate is the best substitute? In order to make the right choice, managers need to realize the team's current condition as to which categories they should improve the most. Let's assume the rebound category needs to be improved; then managers should look for a player who is good at rebounding among the candidates sorted earlier. However, if there is not one particular category that needs to be improved, then a well-rounded player would be the best choice.

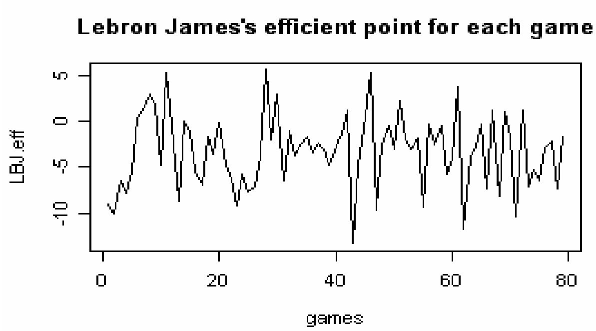
Beyond drafting players based on areas that need improvement, there is another aspect that managers need to consider when it comes to picking players: whether the players are consistent players or inconsistent players. The reason for this classification is to make the substitution more efficient. In general, a consistent player would be better in a sense that they are more predictable. If a player delivers similar statistics for a long period of time with only little variation, we would consider him as a consistent player. On the other hand, an inconsistent player would have a relatively large variation in each game he played. For example, given the same expected value of total performance, consistent players would have a much smaller variance than inconsistent players. Here are examples of the first five best players according to O-Rank. We actually calculate each player's efficient points for each game they played and use both visual and numeric way to assess whether or not they are consistent.











	Kobe Bryant	Kevin Garnett	Lebron James	Shawn Marion	Dirk.Nowitzki
Mean	-3.125	-2.368421	-3.291139	-1.54321	-1.925926
Variance	12.06678	11.97508	16.55358	13.74872	12.28599

From the graph, we can clearly see that people have better games (high efficiency points) at different time period. Game 1 simply means that that's the most recent game this particular player has played. With the same logic, the last game shows how well the player has performed in the first game of NBA basketball season 05-06. For Kevin Garnett, with the exception of the poor performance in game 22 due to injury, his performance is relatively consistent. In addition, by looking at the mean and variance of efficient points for Kevin Garnett of each game he played, he has the lowest comparative variance, which indicates that he is the most consistent player among the top five players. On the other hand, Lebron James has shown a lot of inconsistencies in his efficient point graph. He started off pretty consistent at the beginning of the season; however, as the season progressed, he begin to play a lot of good games followed by bad games as shown by the ups and downs of the efficient point graph during the middle part of the season. Lebron James did not perform at the end of the season since his efficient point hit as low as -10. Looking at the mean and variance of efficient points for Lebron James, he has the most variance of efficient points compared to the others. This implies that he is the most inconsistent player among these top five players. By using the same technique, we can assess the consistencies of all other players.

Comparatively speaking, making a decision between picking consistent and inconsistent players is like deciding whether to buy bonds or stocks. Generally, buying a bond is a safe call; however you will not gain as much as buying a risky stock. Therefore, if a consistent player performs on the same level as an inconsistent player, in other words, if they display the same range of stats, then we would prefer to pick the consistent player since it is a safer play. Such players have a long term value. For inconsistent players, we suggest keeping an eye on them so that managers can pick them up at the right timing. If it's too early, you will not benefit a lot from them. If it's too late, then the player may already be picked up by other teams. A typical usage period of an inconsistent play can be between 2 games to 5 games, and it varies depending on the player. By using consistent players and inconsistent players wisely, one's team can achieve a better performance.

As shown from the above analysis, there are many aspects of statistics that are applicable to the game of Fantasy Basketball. We have demonstrated that rather than simply using "O-Rank" as the sole decision factor, there are many other strategies that can be applied to the game to improve performance using simple statistic knowledge. On a different note, the analysis of a game such as Fantasy Basketball can also offer practical insights in other field such as stock investments. However, one fault in the analysis is that some of the assumptions may not be completely valid in real life situations. Therefore, future analysis can modify these assumptions accordingly to fit the real world data better.