Enrique A. Areyan Viqueira

eareyan@umail.iu.edu | www.enriqueareyan.com

FIFA WORLD CUP SIMULATOR

INTRODUCTION

Founded in 1930, the FIFA World Cup is the world's most important soccer competition. Held every 4 years, teams from many countries around the world compete for as early as 2 years before the competition for one place of only 32 available. Teams are divided in 8 groups that compete in two phases: group stage and knockout. The knockout consists of a Round of 16, then quarterfinals, semifinals and the finals. The finals of the FIFA World Cup is the most widely viewed sporting event with an estimated of 715.1 million people watching the FIFA 2006 final match.

The purpose of this simulation is to explore the dynamics of the competition and how its structure affects its outcome. The input of the simulation consists of a list of 32 teams, each with a ranking assigned by FIFA (for the latest ranking as of 2014 please view [1]) from which a probability of winning a single match is determined. The simulation will take this list and randomly assigned teams to 8 groups (Groups A, B, C, …, H) just like in the real competition. Teams are ranked according to their probabilities of winning a single match and hence, the 8 best teams are those with the highest winning probability. The best 8 teams act as seeds and are placed in eight different groups while the rest of the teams are placed in any group randomly.

Next, the group stage is simulated. A team plays exactly once against teams in its group for a total of 3 games. Points are awarded as follow: 3 points for a victory, 1 point for a draw and 0 points for a defeat. After all games are played, only the first two teams on each group move on to the knockout phase starting with the Round of 16. In this and subsequent phases elimination from the competition is direct based

[1] http://www.fifa.com/fifa-world-ranking/ranking-table/men/rank=243/index.html

on one game. The setup of the Round of 16 follows the structure: 1A2B (where 1A means the first team on group A plays the second of group B), 1C2D, 1E2F, 1G2H, 1B2A, 1D2C, 1F2E and 1H2G. Subsequent phases follow a similar structure.

GOALS

The idea of this project is that, for given ranking of teams, run many simulations of the tournament and answer the following questions:

- How many times does the favorite (least) team (where the favorite (least) team is that with the highest (lowest) ranking) actually win the tournament? Mathematically, if we let X = number of times the favorite (least) team wins, then we want to estimate values for E[X] and Var[X].
- How many times does the favorite (least) team is at least third? Similarly, we can ask for expected value and variance for these random variables.
- How many tournaments should we wait until the favorite (least) team wins?

SIMPLIFYING ASSUMPTIONS

To keep the project feasible, I will make the following simplifying assumptions:

1. Simulations of a single game consist of a draw of a uniform random variable between 0 and 1 and compare this value with the probability of a team winning. The probability of a team winning is proportional to its ranking and its opponent ranking. For example, if two teams A and B compete with winning ranking 10 and 11 respectively, then the probability of A winning is 10/(10+/11) = 10/21, with the complement 11/21 being the probability of B winning. Suppose that the drawn of the uniform random variable is 0.375, then team A wins since $10/21 \approx .4761 > .375$. Similarly, suppose that in a different run these same teams meet but the draw of the uniform random variable yields 0.687, then in this case team A wins since $10/21 \approx .4761 < .687$.
2. I will assume that there are exactly 8 teams with highest ranking and 24 with lowest. However, inside the group of 8 there might be teams with same

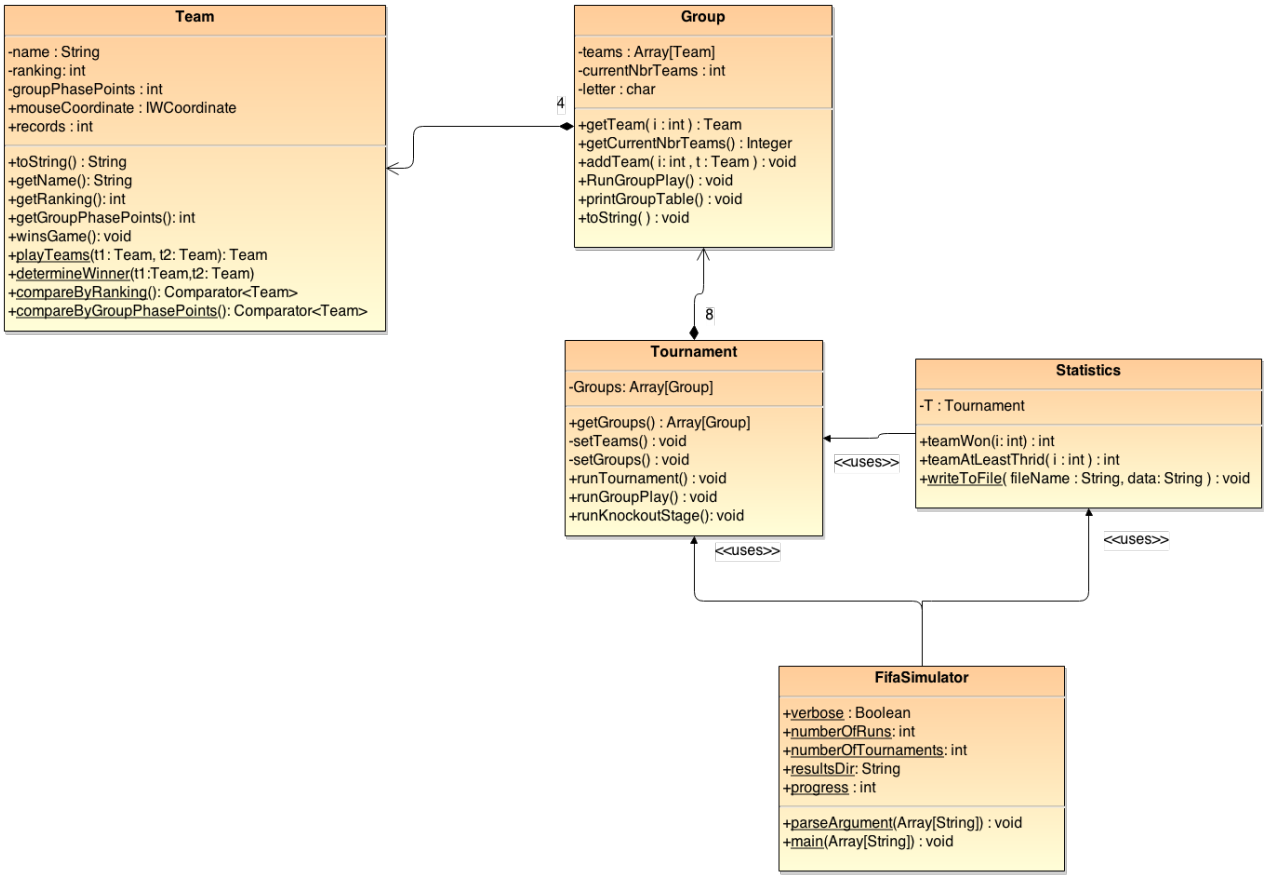
ranking as inside the group of 24. Also, there is always a favorite (least favorite) team defined as the team with the highest (lowest) ranking.

3. I will not allow for a draw on group stages. Instead, if a tie occurs for the first two places of a group, I will order teams in the group and pick the first 2.
4. I will not simulate draws on a knockout game and hence a penalty shootout. Instead, a game will always have a winner in regular time.

COMPUTER TOOLS

FifaSimulator is implemented in Java. Results of a simulation are stored in separate .CSV (comma separated values) files. Analyses of these results are conducted separately in Mathematics in which one can easily compute the Mean and Variance of data sets as well as produce histograms to display results.

The following UML diagram shows the structure of FifaSimulator



RESULTS

The following results were generated using the command:

```
java –jar FifaSimulator.jar –resultsDir ../results/
     –numberOfRuns 1000 –numberOfTournaments 100
```

This generates data for 1,000 simulations, each consisting of 100 tournaments.

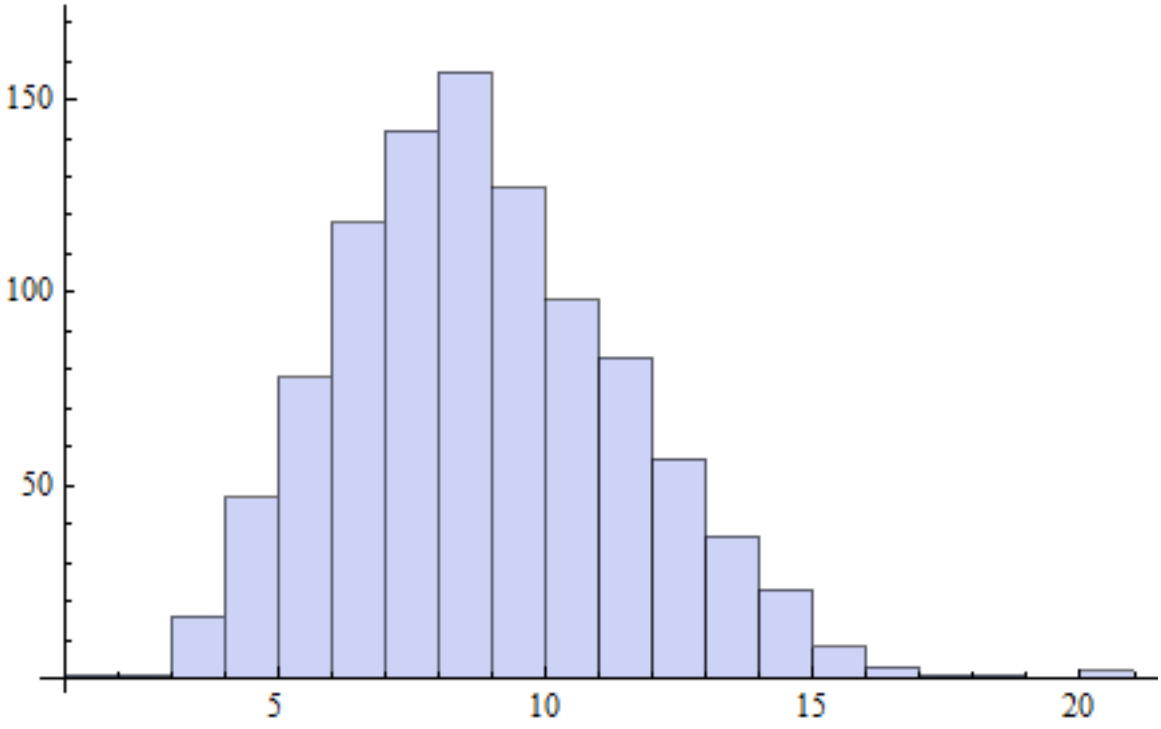
The following data and histograms summarizes the results obtained:

Favorite Team

1) Let X = number of times the favorite team won in 100 tournaments. Then:

E[X]	Var[X]
8.353	2.72393

The following is a histogram of X.

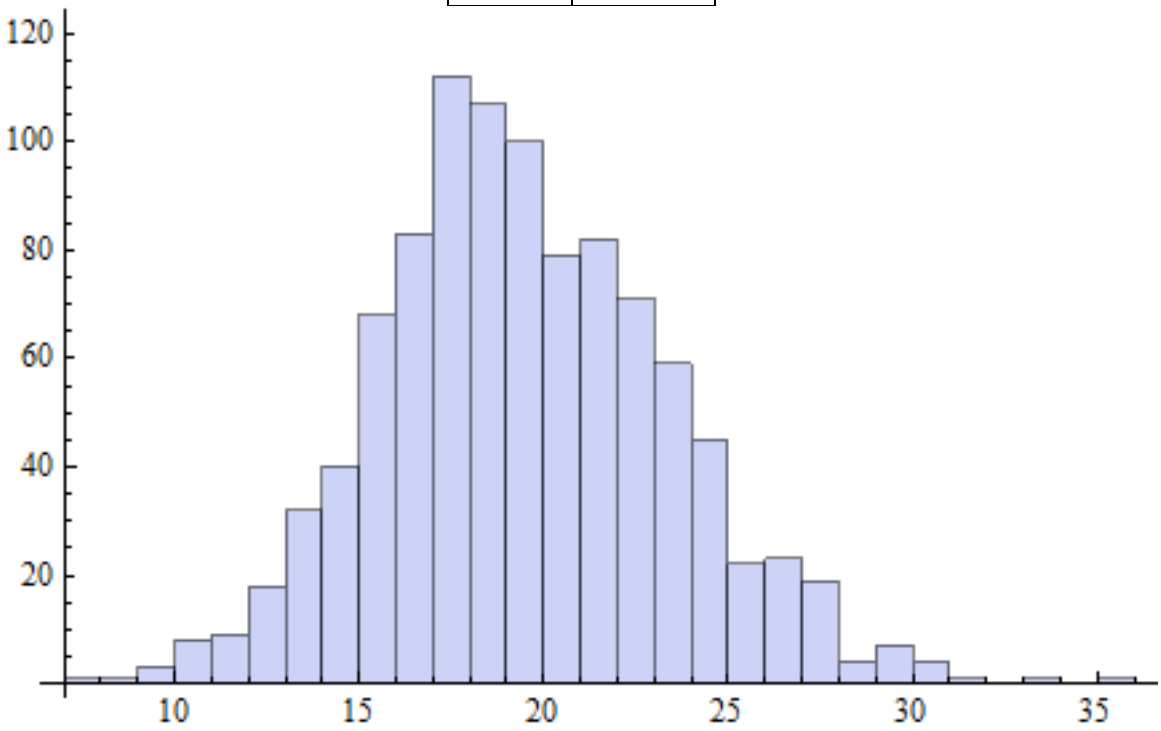


Hence, the favorite team wins only about 8.353% of the time with a standard deviation of only 2.723. This is evidence suggesting that the structure of the tournament, together with the close rankings of teams provided by FIFA, make for a very unpredictable tournament. One can further argue that, assuming that X is

normally distributed (which is a reasonable assumption given the shape of the histogram), then the favorite team will almost surely win (with probability .9973) between 8.353 – 3*sqrt(2.723) = **3.4017** and 8.353 +3*sqrt(2.723) = **13.3043** times. So the favorite team will almost surely **do not** win approximately **87 out of 100** tournaments. Betting for the favorite team as the winner in this kind of tournament carries a high degree of risk.

2) Let X = number of times the favorite team is at least third. Then:

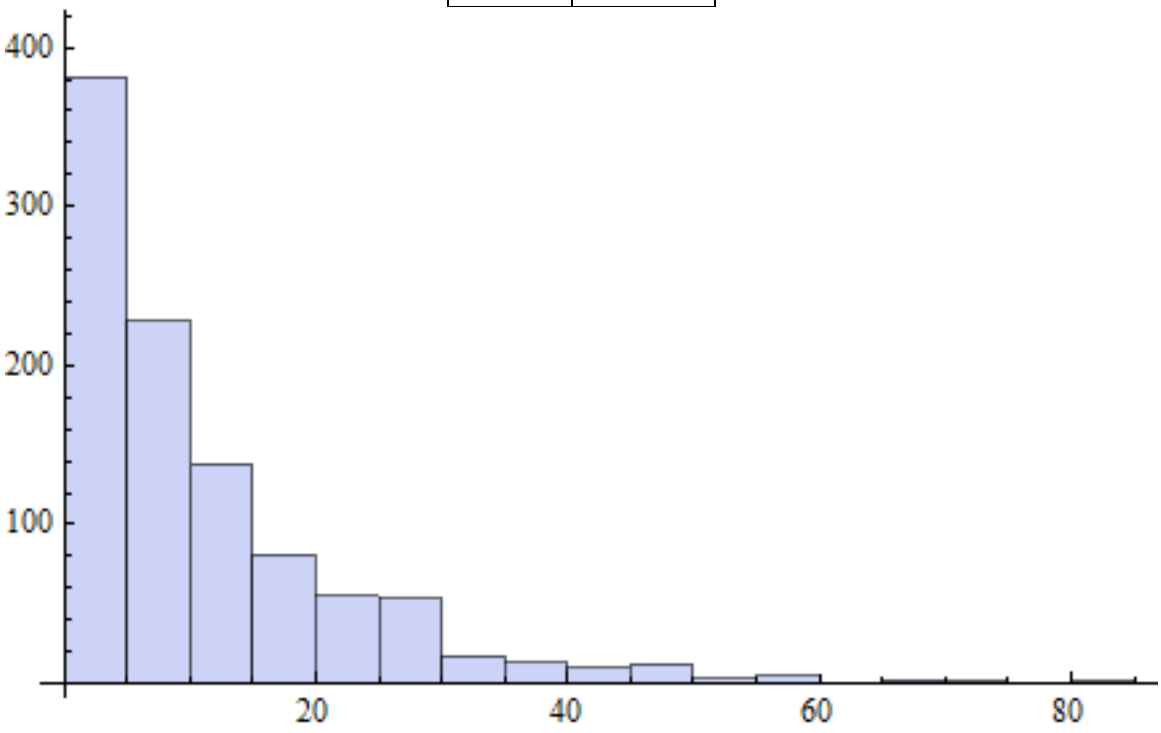
E[X]	Var[X]
18.987	3.94746



The evidence suggests that the favorite team will be at least third 18.987% of the time. This evidence seems reasonable. Interestingly enough, this data has a higher variance and the graph still provides evidence of a normal distribution but not as strong as the data for the first place.

3) Let X = number of tournaments until the favorite team wins. Then:

E[X]	Var[X]
10.599	11.4441



Evidence suggests that X ~ Geometric(p = 1/10.599 ≅ 0.1). We would expect to wait about 10 tournaments for the favorite team to win its first tournament. Again, this suggests that the dynamics of the tournament make it so that it is unlikely to be very confident about the favorite team winning a given tournament.

Least Favorite Team

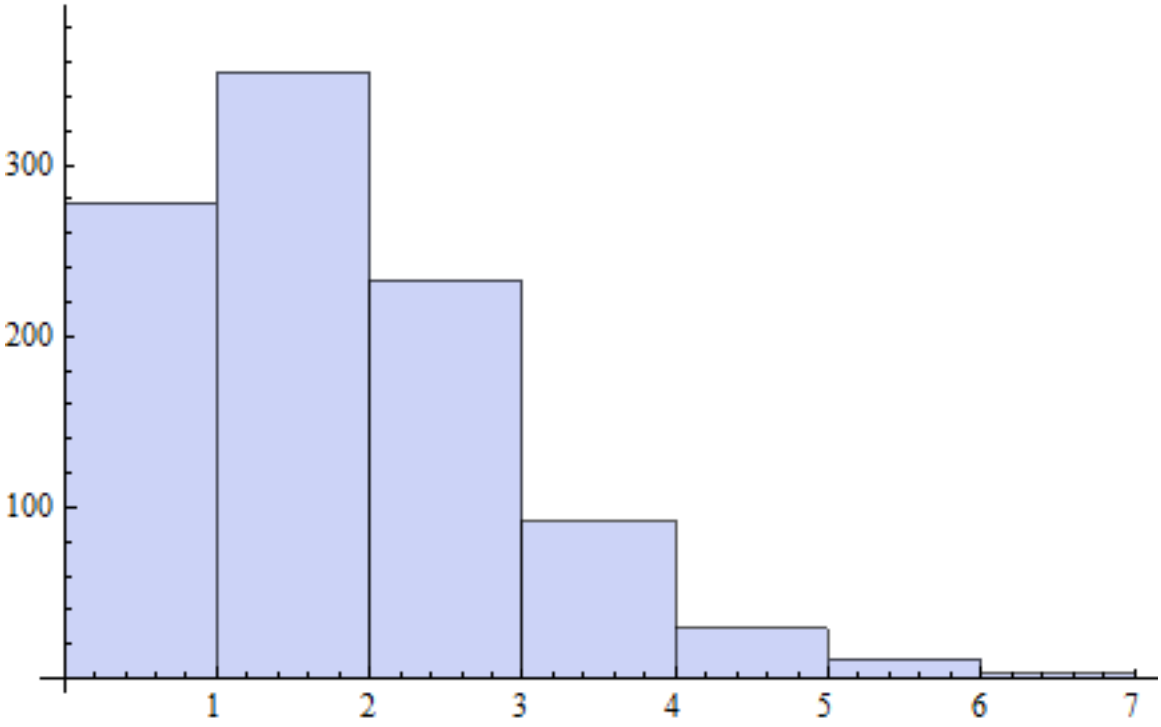
1) Let X = number of times the least favorite team won in 100 tournaments.

E[X]	Var[X]
1.285	1.14503

In this case we can see that, as one would expected, it is considerably more likely that the least favorite team **do not win**. In fact, it only wins 1.285% of the time with a very tight standard deviation of 1.14504. In other words, a safer bet is that of the least team not winning. This is reflected in the big payoff proposed by bookies for

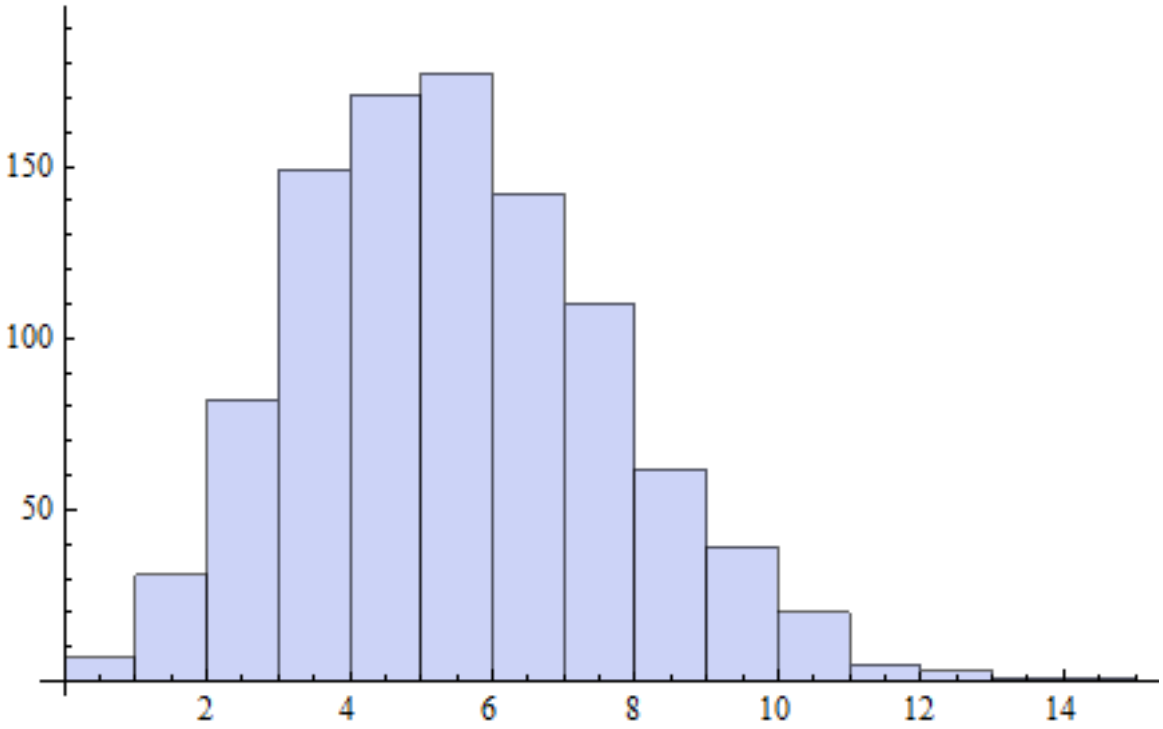
teams that are very unlikely to win a tournament of this nature. For example, Greece was the winner of the EuroCup 2004 a tournament with similar structure as the FifaWorld cup. They were one of the least favorite team and their odds arriving the tournament were 150:1 of winning. However, they did end up winning, providing huge returns for whoever placed a beat for them at the beginning of the tournament. For more on this read http://news.bbc.co.uk/sport2/hi/football/4021041.stm.

The following is a histogram of X.



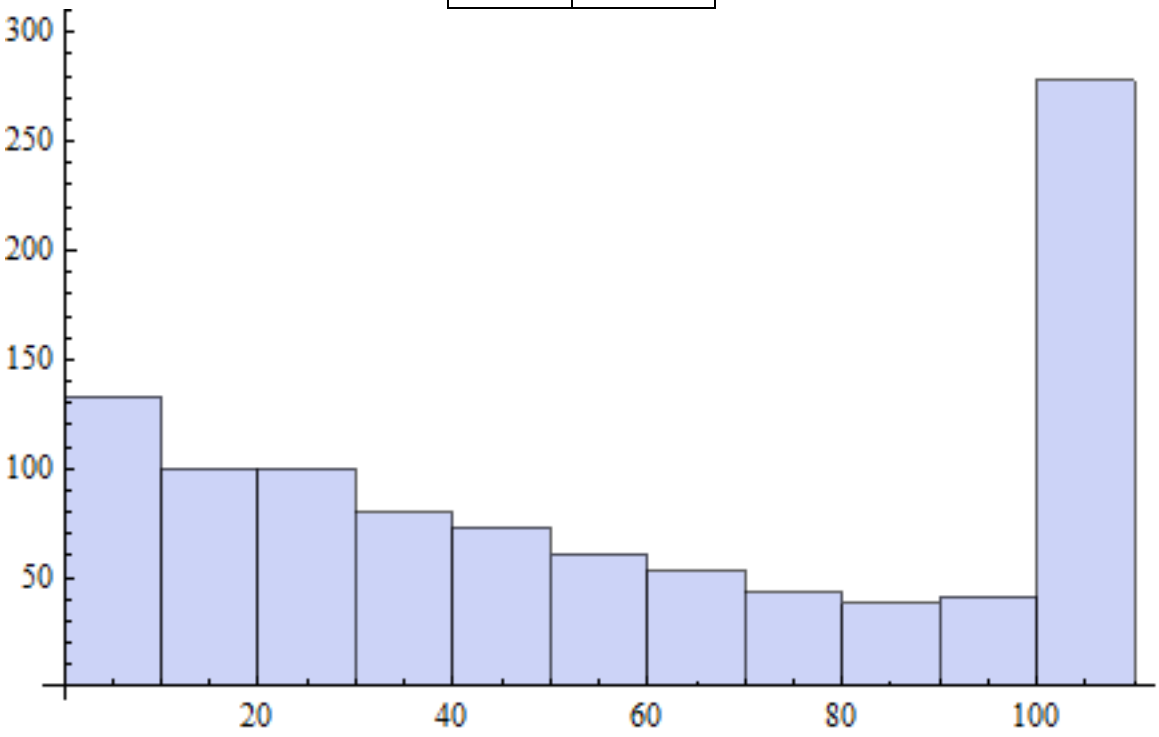
2) Let X = number of times the least favorite team is at least third. Then:

E[X]	Var[X]
4.998	2.21741



3) Let X = number of tournaments until the least favorite team wins. Then:

E[X]	Var[X]
55.244	36.4861



Sample Run

Please see the directory sample/ for a sample of one run of FifaSimulator.

Extensions and Improvements

As currently programmed, FifaSimulator is capable of recording (with very little extra programming) more data to answer more questions about the tournament. Some interesting questions would be:

- How many times does the favorite team moves to the knout stage?
- How many times does the least favorite team moves to the knockout stage?
- How many times does the favorite team ends first in their group?
- How many times does the least favorite team end first in their group?
- All of the questions so far could be asked for different teams, for example the second and third favorite teams.

Moreover, FifaSimulator is an Object Oriented simulation software package and as such, it is readily set for extensions. An obvious extension or improvement would consist on lifting some of the simplifying assumptions described at the beginning of this document. The key assumption I would like to remove is that of no draws in the group stage. Additionally, I would like to account for a game going to penalty shootout and have teams with different probabilities of winning a penalty shootout. These probabilities could be divided into two categories: 1) probability of a goalkeeper stopping a penalty and 2) probability of a shooter making the penalty.

On the technical side of the simulation, I would like to extend the software so that a user can input teams with custom rankings. This data could be input manually or via an XML file. Also, one could provide a graphical interface to more easily see the development of a tournament in real time.