**1. Introduction**

The FIFA World Cup™ is the biggest single-event sporting competition in the world, where the senior men’s national teams from the 208 Member associations of FIFA contend for victory. Also, the FIFA World Cup is the world's most widely viewed sporting event; an estimated 715.1 million people watched the final match of the 2006 FIFA World Cup held in Germany and the 2010 event in South Africa was broadcast to 204 countries on 245 different channels. Inside the stadiums, a total of 3,170,856 spectators attended the 64 matches an average of 49,670 per match and the third highest aggregate attendance behind USA 1994 and Germany 2006. There were also over six million people who attended public viewing events in 16 sites across the world: ten within South Africa and a further six across the globe in Rome, Paris, Berlin, Sydney, Mexico City and Rio de Janeiro. A total of 350,000 fans attended the International FIFA Fan Fest in Berlin for the semi-final match between Germany and Spain.

The format of the tournament involves 32 teams competing for the title at venues within the host nation over a period of one month – this phase is often called the Final Competition. A qualification phase, the Preliminary Competition which currently takes place over the preceding three years, is used to determine which teams qualify for the tournament together with the host nation. For the last FIFA Word Cup, 200 teams played a total of 853 matches as 31 teams qualified for South Africa. So far, the 19 FIFA World Cup tournaments have been won by eight different national teams. Brazil has won five times, and they are the only team to have played in every tournament. The other winners are Italy, with four titles; Germany, with three wins; Argentina and inaugural winners Uruguay, with two; and England, France, and Spain, with one title each.

**2. Scope of work**

This study will develop the model to predict the results of the World Cup 2010 held in South Africa based on the previous World Cup results between 1930 and 2006.

**3. Data Description and Main random variables’ selection**

***3 – 1. Data description***

The data for this study was collected from the Wikipedia and FIFA official website, which was the publically available source. The data consists of data from the World cup held between 1930 and 2006.

***3 – 2. Random variable description***

Like the under, we can make a product density function of X and Y by Bernoulli’s distribution.

P (Xi) = Probability that a nation goes to the ith round by competition.

i = 1, 2, 3, 4, 5 ( “i” represents the number of rounds where a country survives)

e.g. P(X1) describes the probability which a nation can go to “1st” round at the initial competition by Bernoulli's distribution with previous World Cup data.

P (Y1j) = Probability that a country in the round of 32 advances to the round of 16.

j = 1, 2, …, 32 (“j” represents the sequence of the countries in the round of 32)

e.g. P(Y1j) describes the probability which “jth” nation at the initial round can go to “1st” round by Bernoulli's distribution with previous World Cup data.

P (Y2j) = Probability that a country in the round of 16 advances to the round of 8.

j = 1, 2, …, 16 (“j” represents the sequence of the countries in the round of 16)

P (Y3j) = Probability that a country in the round of 8 advances to the round of 4.

j = 1, 2, …, 8 (“j” represents the sequence of the countries in the round of 8)

P (Y4j) = Probability that a country in the round of 4 advances to the round of 2.

j = 1, 2, …, 4 (“j” represents the sequence of the countries in the round of 4)

P (Y5j) = Probability that a country in the round of 2 advances to the final.

j = 1, 2 (“j” represents the sequence of the countries in the round of 2)

***3 – 3. Data analysis***

Like false alarm example, X is the judging standard, Y is a target sample. Then 95% or 99% CI (Confidential Interval) results of X and, Y describes how much the target reaches to the judging one.

For example, there are 32 nations on 2010 World Cup. Each group has 4 nations. Thus, there are 8 groups. When we expect how 2 nations can survive from each group by our project result, we decide how a nation has a similar value from the CI of X1. We repeat this iteration until there is only one nation. That means 5th iteration will be the late one. This is because there are 16 nations after 1st round among 32 nations. Like binary system, the nations would be reduced by half when they experiences one round like the under.

32 [1st iteration]🡪 16 [2nd iteration]🡪8 [3rd iteration]🡪 4 [4th iteration]🡪2 [5th iteration]🡪the last one.

This is an initial draft of our project report. Thus, please give us some guideline to make a better direction.