**2022 FIFA World Cup – Statistical Analysis**

**Mouaz Ali**

**Table of Contents**

Introduction……………………………………………….………………………………..…..….1

Content…………………………………………………………………………………………….1

Statistical Predictions…..………………………………………………………………………….2

ELO Ratings………….……………………………………………………………………………3

Questions and Solutions…………….………………………………………………………..……4

Work Cited……………………………………………………………………………………….10

Introduction

Every four years, the FIFA World Cup takes center stage and brings people together in a symphony of passion, skill, and national pride. This extraordinary event captures the hearts of millions of people worldwide, cutting over boundaries, languages, and cultural divides. The World Cup, which has been regarded as the height of international soccer competition, is more than just a competition; it's a celebration of the beautiful game and evidence of the ability of sport to bring people together. As the drama plays out on the biggest platform, nations compete for glory, players rise to the status of heroes, and everyone on the planet holds their breath.

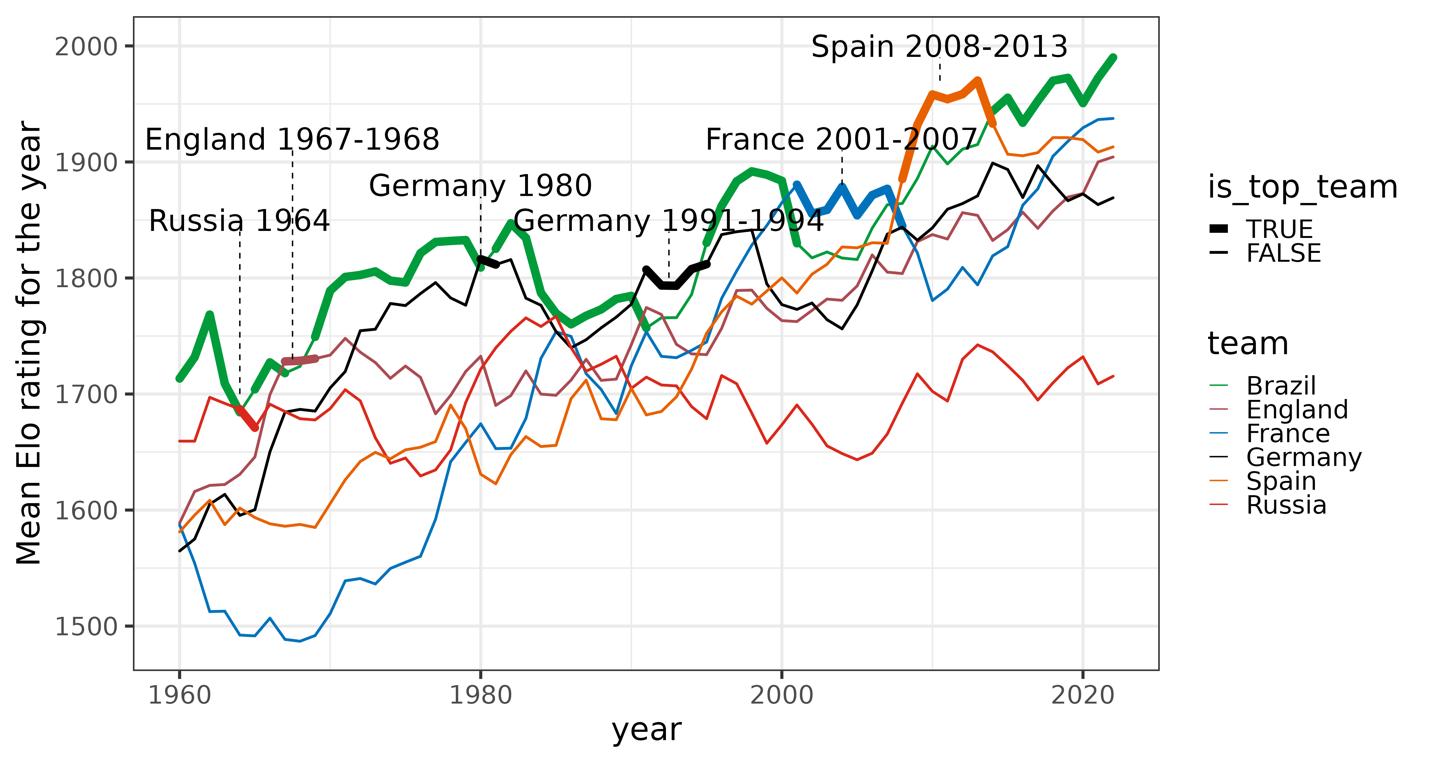
Content

The FIFA World Cup, often simply called the World Cup, is an international association football, or soccer, competition contested by the senior men's national teams of the members of the Federation International de Football Association (FIFA), the sport's global governing body. Since the tournament was started in 1930, the World Cup is held every four years, except in 1942 and 1946 due to World War II. After witnessing the events unfold last year, the current champion is Argentina, which won its third title at the 2022 tournament in Qatar.

The dataset, which was put together through web scraping from FIFA.com, contains numerical information of all matches of the 2022 FIFA World Cup. Having 88 columns, the data can be used for multiple purposes such as research, application, or simply learning. After being cleaned and discarding unnecessary information, the dataset includes various statistical data for each match played during the tournament, such as: score, assists, possession, crosses, passes, fouls, number of cards given out, offsides, and what part of the pitch certain events took place. These small pieces of information and statistics have played a major role in each one of the matches played at the tournament.

Statistical Predictions

Leading up to the 2022 FIFA World Cup, the predicted winners were Brazil based on statistical analysis. By examining all international soccer matches, including friendly games, for teams with at least 200 games, a model was created. To keep it simple, only wins, draws, and losses were taken into account. A win would give a team one point, while a draw gives half a point, and a loss will result in zero points. For example, in the 2014 World Cup final, Germany beat Argentina to become world champions with a score of 1-0, which gives Germany one point in the model and Argentina zero.



Elo Ratings

Elo ratings were developed to quantify chess players' skills and deter elite players from merely playing against weaker opponents in order to rack up a lot of points. Elo ratings are named after physicist Arpad Elo and are not to be confused with the British music band ELO from the 1970s. Players with comparable Elo ratings have an equal chance of winning since Elo ratings may also be converted into probabilities.

Well, how does Elo function? The calculations are pretty straightforward actually. The winner receives points from the loser, with the amount of points awarded based on the disparity in ranks. A player with a higher ranking receives few points when they defeat a player with a lower ranking, but they receive many points when they defeat a player with a higher ranking.

In this case, we know that England lost to Croatia 1-2 in the 2018 FIFA World Cup. Before the game had began, England was favored to win with a higher Elo score. Surprisingly, Croatia was able to defeat England in extra time, gaining 12 ELO points. On the other hand, England lost 12 ELO points. Generally, ELO changes function both ways. Switzerland's victory over Spain in the 2010 World Cup group stage was a major upset. Spain, the eventual World Cup victor, had the greatest Elo rating at the time and, again excluding draws, was expected to win by 84%. As we will see in a moment, they fell short by 17 Elo points. Another example of this in the previous World Cup is Argentina being defeated by Saudi Arabia. No one had expected Argentina to lose their first game against Saudi, yet they eventually went on the win the tournament.

A screenshot of a graph

Description automatically generated

In the chart above are the ELO ratings leading up to the 2022 FIFA World Cup. As you can see, Brazil has been on top of the ratings since 2013.

Now, let’s use the data set to answer some questions.

1) Let's consider Group F for the group stage matches in the 2022 FIFA World Cup involving Belgium (BEL), Canada (CAN), Morocco (MAR), and Croatia (CRO). We'll consider each team playing two matches, and each match can result in either a win for one team or a draw.

*a. Define the sample space S using set notation:*

S={(BEL,BEL),(BEL,CAN),(BEL,MAR),(BEL,CRO),(CAN,BEL),(CAN,CAN),(CAN,MAR),(CAN,CRO),(MAR,BEL),(MAR,CAN),(MAR,MAR),(MAR,CRO),(CRO,BEL),(CRO,CAN),(CRO,MAR),(CRO,CRO)}

*2. Determine the probability of Belgium (BEL) winning at least one match.*

Assuming Belgium wins both matches, there are two outcomes where Belgium wins at least one match:

*3.. Find the probability of either Canada (CAN) or Morocco (MAR) winning their respective matches.*

*4. Calculate the probability of no team winning a match (all matches ending in a draw)*

*5. Determine the probability of Croatia (CRO) not winning any match*

6) Consider the knockout stage of the 2022 FIFA World Cup, where there are two remaining teams: Argentina (ARG) and Croatia (CRO). The probability of Argentina reaching the final is P(ARG)=0.6, and the probability of Croatia reaching the final is P(CRO)=0.4.

Defining the following events:

A: Argentina wins the semifinal match

B: Croatia wins the semifinal match.

*a. Write the expression for P(ARG∣A) in terms of conditional probability.*

*b. Calculate the conditional probability P(ARG∣A) using the information provided.*

*c. Calculate the conditional probability P(CRO∣B) using the information provided.*

7) Consider the group stage of the 2022 FIFA World Cup, where four teams (England, Iran, USA, Wales) compete. The probability of each team advancing to the knockout stage is as follows: P(England)=0.3, P(Iran)=0.2, P(USA)=0.4, P(Wales)=0.1

Additionally, the probability of a team winning a match in the knockout stage is given as follows:

P(Win|England)=0.5

P(Win|Iran)=0.6

P(Win|USA)=0.7

P(Win|Wales)=0.4

Write the expression for the law of total probability for the event W in terms of the events K (advancing to the knockout stage) and the conditional probabilities of winning given reaching the knockout stage.

8) Now, suppose a team has won their knockout stage match. What is the probability that this team is England? Formulate the problem using Bayes' Rule and then solve it.

A: England wins the knockout stage match.

B: A team wins their knockout stage match.

Bayes’ Rule:

Therefore,

To calculator probability,

The probability that England wins their knockout stage match given that a team wins a knockout stage match is 25.4%.

9) In the group stage of the 2022 FIFA World Cup, Portugal plays three matches. The probability of Portugal winning any individual match is 0.4, and the probability of losing is 0.6. Let's define the following events:

X: The number of matches Portugal wins in the group stage.

*Assuming each match is independent, calculate the probabilities for each possible value of X (0, 1, 2, 3) using the binomial probability distribution.*

PMF:

Where:

n is the number of trials (matches), �

x is the number of successes (matches won),

p is the probability of success (winning a match),

(1−p) is the probability of failure (losing a match).

* Calculate P(X=0):
* Calculate P(X=1):
* Calculate P(X=2):
* Calculate P(X=3):

10) In the 2022 FIFA World Cup, there are 32 teams in total, including Portugal (P), Argentina (A), and France (F). Let's assume that a draw is conducted, randomly selecting 8 teams to form Group A. The hypergeometric distribution can be used to model the probability of specific outcomes based on this draw.

Define the following events:

K: The number of selected teams in Group A that are from Portugal.

N: The total number of teams in the draw (32).

n: The number of teams selected in the draw to form Group A (8).

k: The specific number of teams from Portugal in Group A.

11. Write the hypergeometric probability mass function (PMF) for K in terms of N, n, and k.

12. Calculate the probability that exactly 3 teams from Portugal are selected in Group A

13. In the 2022 FIFA World Cup, consider a scenario where a fan wants to create a personalized playlist to represent the group stage matches. The fan selects 5 specific teams: Portugal (P), Argentina (A), France (F), Brazil (B), and Germany (G).

*a. How many different ways can the fan arrange these 5 teams in a sequence to represent the order of matches?*

14. Problem: In the 2022 FIFA World Cup, a group of friends plans to attend three matches together. The matches involve five teams: Portugal (P), Argentina (A), France (F), Brazil (B), and Germany (G). The friends want to select two teams to support in each match.

*How many different ways can the group of friends select two teams out of the five (P, A, F, B, G) to support in the first match?*

Work Cited

Iron486. (2022, December 18). *FIFA World Cup 2022: Complete dataset*. Kaggle. https://www.kaggle.com/datasets/die9origephit/fifa-world-cup-2022-complete-dataset

FOX Sports. (n.d.). *FIFA Men’s World Cup Stats & Leaders*. FIFA Men’s World Cup Stats & Leaders | FOX Sports. https://www.foxsports.com/soccer/fifa-world-cup-men/stats