Juan Villaman

05/01/2025

**Exploring FC Barcelona’s 2014/15 Season with Probability and Statistics**

# Introduction

Futbol Club Barcelona, also known as FC Barcelona, is a football club based in Barcelona, Spain (in this paper, we will refer to soccer as football, and clubs are teams). Barcelona is arguably one of the biggest clubs in the history of the sport, as they have a rich 100+ year history that rooted deeply into the sport. Barcelona has an extensive trophy cabinet, filled with 27 league titles, 32 national cup titles, and 5 UEFA Champions League trophies.



FC Barcelona’s Badge ([Source](https://en.wikipedia.org/wiki/FC_Barcelona#/media/File:FC_Barcelona_(crest).svg))

## Overview of Competitions

These 3 competitions are the highest level of professional football that can be achieved in European football. Across countries like France, Spain, England, Germany, and Italy all have their own competitions, and teams from those competitions compete against each other in the UEFA Champions League.

A quick overview: each country has a league competition with their own set of teams. For example, Spain has LaLiga, France has Ligue 1, England has the Premier League, Germany has the Bundesliga, and Italy has the Serie A. All these leagues contain 20 teams except the Bundesliga which only has 18. Each team play each other twice per season, once home and once away. A win is worth 3 points, a draw is worth 1 point for each team, and a loss is worth 0 points. How is league title won? The team with the greatest number of points accumulated by the end of the last matchday wins the league title. With this rule, ties at the end of the season between 2 team is very possible, Usually a tiebreaker rule is included within each competition, so it may be different for each league. It is also possible to win the title **before** the last matchday. For example, Team 1 has 90 points, and 37 matches played out of 38 total. Team 2 has 85 points, and 37 matches played out of 38 total. This means that the **maximum** number of points Team 2 can get for the season is 88, making Team 1 automatic league champions before the final match is played.

The national cup competition, just like the league, is different in each of those countries. Spain has Copa del Rey, France has Coupe de France, Italy has Coppa Italia, Germany has DFB Pokal, and England has FA Cup. These cup competitions have a knockout round system instead of a point system. All these tournaments include many teams from all over their countries, even including lower division sides. It goes through the usual Round of 16 to Final rounds to determine a champion.

The UEFA Champions League, better known as the UCL, contains the top teams of each of those leagues (and a few more from countries like the Netherlands and Ukraine). These teams are then pitted against each other to see which team is the best in Europe. 36 teams gather from these leagues (usually the Top 2 to Top 4 teams) enter the first phase of the competition: the League Phase. Every team plays 8 random opponents from 4 different pools, 2 from each pool. It works with the points system explained above. At the end of the 8 games, the top 8 teams automatically qualify for the Round of 16. Places 9th-24th play a [2-legged tie](https://en.wikipedia.org/wiki/Two-legged_tie) to decide who goes through to the Round of 16. From there, the knockout rounds continue until the single-match Final is played.

### The Treble and Its Rarity

One of the most elusive feats in football is the treble —winning the domestic league, the national cup, and the UCL all in one season. To put it into perspective, the treble has only been achieved a measly 10 times by 8 teams. Since the founding of the UCL in 1955, only 14.29% of seasons produced a treble (this is for all clubs included).

FC Barcelona is one of the only two teams in history to achieve this feat **twice**. Barcelona achieved two historic trebles in the span of 6 years, in seasons 2008/2009 and 2014/2015. The other club to achieve this feat is Bayern Munchen of Germany. Due to the rarity of this feat, I will be looking at interesting statistics and calculating different probabilities of events that occurred throughout the season. This may give insight on truly how rare it is achieving this feat, and we will be doing so by exploring Barcelona’s 2014/2015 season.

#### Section 1.2: Graphical Methods

Barcelona was a dominant goal-machine in this season. Beating an opponent by a gap that is more than 2 goals is what they call in Spain a “goleada”. In this section, we focused on making frequency histograms to study the probability of certain events occurring. From the data set of the 2014/15 season, we can graph a relative frequency histogram to see how often Barcelona scored a certain number of goals, and to see if any particular one was more frequent. Graphed below:

A graph of goals per match

AI-generated content may be incorrect.

To calculate the relative frequency of goals scored in a game, it is solved this way: If Barcelona scored 3 goals in 13 games out of 60 total.

This number is our Y-axis in our graph, and it was calculated for each number of goals scored per match. We can also reinforce the fact that Barcelona was a goal-machine by looking at the proportion of matches that Barcelona scored more than 2 goals. In the 2014/15 season, Barcelona scored more than 2 goals in 29 games out of 60. This shows us that:

In other words, Barcelona scored more than 2 goals in **48.33%** of their matches in that treble-winning season.

##### Section 1.3: Numerical Methods

In this section, we explore different aspects of our data, like the mean and standard deviation. Using the goals scored per match in the season, we can calculate the mean like this:

By using Excel, we can import our list of goals scored per match and use the STDEV.S function to accurately get our standard deviation. Excel also has an AVERAGE function which allows us to quickly calculate the mean of our data.

A close-up of a table

AI-generated content may be incorrect.

With this information, we can form a simple question about our data:

**What number of goals is exactly 1 standard deviation below the mean?**

Where is the mean, and is our standard deviation.

###### Section 1.6: Average Cards and Standard Deviation

There are other interesting events that may happen during a football match. There are permanent player warnings in football, and these are called yellow cards and red cards. A player may get a yellow card up after committing obvious and maybe dangerous foul play. It could be accidental as well, as football pitches are very slippery and may cause players to crash into each other after sliding for a challenge. A player may get a yellow card due to arguing with ref’s decision. This yellow card sticks to the player until the end of the game. If the player commits yet another yellow cad offense, he will receive a second yellow, which is a red card.

A red card is more severe, as the player is immediately ejected off the pitch, with no substitution allowed. This means that the team goes down to 10 men permanently and must play the rest of the game 10 versus 11. We’ll look at the average of cards Barcelona received by per match, as well as its standard deviation.

Using Excel’s built-in functions, we can get mean and the standard deviation of our total cards per game. As shown, Barcelona had gotten an average of 1.7 total cards per match, with a standard deviation of 1.31. The values used are below, grabbed from our data and counting the number of cards per match. The mean and standard deviation is calculated using Excel.

**Total cards per match 2014/15 FC Barcelona**

1, 2, 1, 2, 2, 1, 2, 2, 4, 1, 0, 1, 4, 0, 0, 1, 4, 3, 2, 2, 1, 1, 1, 4, 3, 4, 5, 2, 1, 2, 4 , 3, 0, 0, 1, 3, 1, 0, 1, 1, 3, 2, 0, 1, 0, 3, 1, 0, 1, 2, 3, 1, 1, 3, 2, 0, 3, 2, 1.

A close-up of a table

AI-generated content may be incorrect.

Section 2.3: Set Notation

We can use set notation to figure out more information of our data which may help us calculate probabilities later (hint: they do!). Let’s look at a situation: the game where Barcelona scored 3 or more goals but did not keep a clean sheet (not conceding a goal). We can use the set difference to find the number of games where this scenario came true.

A is the set of games where Barcelona scored 3 or more goals. B is the set of matches where Barcelona kept a clean sheet. This tells us that Barcelona had 12 matches where they had scored 3 or more goals but failed to keep a clean sheet. This could be to relaxation of the team. Having such a big gap in goals may cause players’ concentration to slip up.

Section 2.4: Discrete Cases

Barcelona also managed to have a great defensive record this season. There is a very big importance to the defense as well as the attack. A famous quote by one of the most successful English managers Sir Alex Ferguson says, “Attack wins you games, defense wins you titles.” Using probability, we can find exactly how dominant either side was during this treble season.

We can figure out the probability that Barcelona either scored 3 or more goals or kept a clean sheet (a clean sheet is 0 goals conceded). Using this data, we can figure this information rather efficiently. From our data, we can gather that Barcelona scored 3 or more goals in 32 matches, and Barcelona kept a clean sheet 33 times. Both times these conditions were true at the same time was 16 times, which is rather impressive. That’s at least a 3-goal gap win. Let’s figure that probability using this formula for **unions** below:

: Scored 3 or more goals / total matches

: Total clean sheets / total matches

: Matches where both conditions are true / total matches

The probability either of these events occurred in the 2014/15 season was 81.87% which is extremely high. This shows not only the attack, but the defense was rock solid at the back. For either event to happen is not as common as it may seem. A team needs to be achieving such ridiculous heights to reach a historical season

Section 2.5: Probability of an Event using Sample Points

We can handpick sample points from our data to make certain observations and calculate probability. Let’s look at Barcelona and them competing in the big competition, the UCL. In this historic tournament win, Barcelona knocked out English Champions Manchester City and German Champions Bayern Munich. Between these two teams, 4 games were played. If we picked a game, **what’s the probability that game was against a German team, in this case Bayern Munich**?

Since there are only 4 options, and 2 of them are German, it’s a 50% chance of the match being against a German team.

Section 2.6: Permutations and Combinations

A **permutation** is an ordered arrangement of r distinct objects. The formula is:

Barcelona played 7 matches in the knockout rounds of the UCL to be crowned Champions of Europe. If we randomly select three of 3 matches and the order matters in terms of how and when we select them, how many ways could we possibly arrange those 3 selections? Using the formula above

There are 210 different ordered ways to select 3 of the 7 matches played in that historic 2014/15 season by Barcelona

Section 2.7: Conditional Probability

Conditional Probability is simple concept to put into English. In short, what is the probability event A happens, given that event B has happened. One event may depend on one another, and some may not. That’s called *dependence* and *independence*. In the case we’ll look at, the events may or may not be related, and I’ll explain why.

First, let’s calculate **the probability that Barcelona concedes a goal, given they have a yellow card**. In 48 matches, Barcelona received yellow cards. Out of those 48 matches, they conceded a goal in 24 of them.

Given that Barcelona had a yellow card, the probability that Barcelona conceded a goal is 50%. That is high, as it shows that yellow cards may influence the way the game is played. Players may have to restraint themselves to avoid expulsion through a second yellow, and the opponent will take advantage of it. So many other factors impact these scenarios that it would be difficult to tell if they are dependent or independent of each other.

Section 2.8: Two Laws of Probability

The additive law of probability and the multiplicative law of probability are extremely useful when look at intersections and unions between sets. In the Discrete Cases section above, we used the additive law of probability to solve our problem. In this section, we’ll focus on the multiplicative law. The multiplicative law states:

We are using the exact same numbers used in the Discrete Cases section. By plugging in our numbers:

The multiplicative law gives us the probability that Barcelona keeps a clean sheet and score 3 or more goals is 0.293 or 29.3%. This number differs from our data, which is 0.317 or 31.7%. This difference in numbers indicate that scoring 3 or more goals and keeping a clean sheet are not entirely independent.

Section 2.10: Bayes’ Rule

Bayes’ Rule helps us expand upon calculating conditional probability. Bayes’ Rule states:

Where:  
: Probability of A given B occurred

: Probability of B given A occurred

: Probability of A occurring

: Probability of B occurring

Bayes’ Rule is very interesting, as it can show interesting details about goal scoring in the home football pitch and away football pitch. In the 2014/2015 season, Barcelona played 29 home matches, and 31 away matches. In 24 of the 29 home matches, Barcelona scored 2 or more goals. In 23 of their 31 away matches, Barcelona scored 2 or more goals. With this information, let’s choose a random game in which Barcelona scored 2 or more goals. The question we can ask and solve using Bayes’ Rule is:

**What is the probability that the match was a home game?**

Let’s define our variables.

: Total Home Matches / Total Matches

: Total Matches with 2 or more goals scored / total matches

: 2 goals given match is home / total home matches

We are trying to find : Given that Barcelona scored 2 or more goals, what is the probability that the match was at home. We can plug these values into Bayes’ Rule to calculate this probability

The probability that the match was home given that Barcelona scored 2 goal is 51.06%. This shows that at their home ground, Barcelona was slightly more dominant over opponent teams, as it is the usual case for sports team to have “home-ground advantage”

Section 2.11: Random Variables

A **random variable** is an outcome of a **random** experiment. Every football match is technically a random experiment, as many things can happen within those 90 minutes of game time. The random variable we’ll be looking at is the number of matches out a set-in which Barcelona failed to score a goal. Barcelona’s probability of failing to score a goal in a match is 10%.

Let Y represent this random variable. Let’s select 3 random matches and calculate the probability that Barcelona did not score a single goal. This problem fits a binomial probability distribution where we need the formula:

We need to calculate Y = 0, 1, 2, 3. Our results are below:

**Barcelona scored in all 3 matches**

**Barcelona failed to score in exactly 1 match**

**Barcelona failed to score in exactly 2 matches**

**Barcelona failed to score in all 3 matches**

This model does a good job at showing how consistent Barcelona was at scoring, with a 72.9% chance that they scored at least one goal in 3 randomly selected matches from the 2014/15 season.

Section 3.2: Probability Distribution of a Discrete Random Variable

To further explore how consistent Barcelona was as a goal-machine, we can use a discrete random variable Y to represent the number of goals in a randomly selected match during this treble season. Since Y is a discrete random variable, we can compute the probability distribution since it is countable. First, let’s make a frequency table with goals scored.

|  |  |
| --- | --- |
| Goals Scored (Y) | Frequency |
| 0 | 5 |
| 1 | 8 |
| 2 | 15 |
| 3 | 14 |
| 4 | 5 |
| 5 | 6 |
| 6 | 5 |
| 8 | 2 |

Now we can compute the probability distribution for our variable Y by diving each frequency by the total number of matches

|  |  |
| --- | --- |
| Goals Scored (Y) | P(Y) |
| 0 | 0.083 |
| 1 | 0.133 |
| 2 | 0.25 |
| 3 | 0.233 |
| 4 | 0.083 |
| 5 | 0.10 |
| 6 | 0.083 |
| 8 | 0.033 |

From this table, we can see the probability distribution for Y. We can make some quick observations, like how Barcelona is most likely to score 2 goals maximum in any random match selected from this season

Section 3.3: Expected Value of a Random Variable

This section covers the **expected value** of the random variable we just looked at. The **expected value** is better known as the average outcome of an experiment. Since Y is a **discrete random variable**, there is a specific way to calculate its expected value. This formula below lets us calculate it:

We already have a table from the previous section that contains P(Y). Plugging in the numbers to the formula:

This simple formula shows us that expected value of goals scored in a randomly selected match is 2.9 goals, which shows Barcelona’s incredible consistency of scoring enough goals to beat their opponents comfortably.

Section 3.4: Binomial Probability Distribution

When looking at binomial experiments and their properties, we can relate them to football match. Directly from the Mathematical Statistics 7th Edition textbook, it states that a binomial experiment has these properties:

1. **The experiment consists of a fixed number, n, of identical trials.**
2. **Each trial results in one of two outcomes: success, S, or failure, F.**
3. **The probability of success on a single trial is equal to some value p and remains the same from trial to trial. The probability of a failure is equal to q= (1− p).**
4. **The trials are independent.**
5. **The random variable of interest is Y , the number of successes observed during the n trials.**

Using these properties, we will relate them to a football match. The **fixed number of identical trials** is our sampled data, which is all 60 matches from the 2014/15 for Barcelona. We can use an event in a match to determine if it is a success or a failure. The probability of each trial, in our case, scoring in match, is the same. Each match is independent of each other, and we can look at the number of success or failures of events happening in the matches.

First, let’s estimate the probability of Barcelona scoring at least one goal. We need to take count of matches in which Barcelona scored **at least 1 goal.** Divide that by the total matches played, and we have the probability of Barcelona scoring in a match.

This number helps us to solve a problem that requires calculating the probability of a random variable with binomial distribution. Let’s look at this specific scenario: **What is the probability that exactly 58 matches had at least one goal?** We can solve this using the binomial distribution formula:

In the formula, n is the total amount of matches, y is the amount of matches Barcelona scored at least one goal, p is the probability of success, and q is the probability of failure. Stated by the properties above, q is calculated by:

Therefore, our probability of failure is 0. 10, that is Barcelona failing to score at least one goal in the 2014/15 season.

This binomial distribution tells us that the probability that Barcelona scores at least one goal in exactly 58 matches is 6.43%.

Section 3.5: Geometric Probability Distribution

The random variable that has geometric probability distribution shares some binomial properties, but instead of finding how many out how many successes (binomial), the trials run until the first success is encountered. This leaves us with the formula for geometric distribution:

Where is the probability of failure, p is the probability of success, and y is the number of trails until the first success. We can see the chances of a Barcelona committing a yellow card offense before a certain number of games. We first need to estimate probability for Barcelona getting at least one yellow card in the game. We first count the number of games that Barcelona had a yellow card and divide it by the total number of matches played.

In the 2014/2015 season, Barcelona had an 80% chance of getting a yellow card in a match. Now that we have this number, we can use the geometric distribution formula to set up interesting results and show how often yellow cards were present in this season.

We can ask ourselves two interesting questions: **What is the probability that Barcelona receive their first yellow card in the 5th match of the season? And what is the probability that Barcelona do not receive any yellow cards in the first 3 games?**

For the first question, we can setup the formula this way:

This shows that the probability that Barcelona received their yellow card in the 5th match of the season is 0.128%. This means that it doesn’t take a long season to commit the first yellow. These yellows could’ve also been tactical yellows, like deliberately stopping the play without causing harm (pulling a shirt to stop a counterattack).

To solve the second question, we can setup the formula this way:

The chance of Barcelona **not** committing a yellow card offense in the first 3 games is 0.032%. That is extremely low, which means the first yellow card of the season most likely happened before the third match, which is true, as Barcelona did get yellow carded in the first game of the 14/15 season.

Section 3.6: Negative Binomial Probability Distribution

Negative Binomial Distribution mainly focuses on the getting the probability that a success occurs on trial. The conditions of the trials stay like binomial distribution, causing the formulas to look almost alike. This is the formula below:

In terms of football, we can use negative binomial to determine if Barcelona was a consistent goal scoring machine. Let’s say scoring at least one goal is a success for our trial. This means the probability is 0.90, as calculated in earlier sections. **What is the probability that Barcelona’s 5th goal scoring match happens on the 6th match?** A little clarity: if Barcelona have scored at least one goal in the first 4/5 matches, what is the probability the next match they score at least once happens on the 6th one.

This result is interesting when comparing it to the real-world data. 2.95% is rather low for a treble winning team. When looking at the first 6 games, the scores are:

One thing this model fails to consider is the level of the rivals. Some bottom-table teams are easier to score against than top-table teams who are direct rivals for the league title. It’s an interesting conclusion that the real world may have unprecedented factors that some models like these can’t handle.

Section 3.7: Hypergeometric Probability Distribution

Hypergeometric Geometric distribution allows us to observe a certain number of successes in a chosen sample, without replacement. This means that the next trial is directly affected by the last one. Let’s continue the line of 2 or more goals being scored at the home field. Barcelona played 29 homes matches, and in 24 of them scored 2 or more goals. From this sample of home games, let’s pick 5 random matches. Out of those 5 matches, **what is the probability that 3 of them had Barcelona scoring 2 or more goals**. For questions like this, hypergeometric distribution helps us answer this question. The formula for Hypergeometric is below:

To solve the problem above, we’re going to have to use this formula above with information given.

: Total Matches

: Home matches were Barcelona scored 2 or more goals

: Sample of matches

: Number of successes

The results show that the probability that exactly 3 out of 5 randomly selected matches were home games where Barcelona scored 2 or more goals is 23.35%. This shows that Barcelona was more likely to be high scoring in all 5 of the games. This further reinforces their dominant status in this incredible season

Section 3.8: Poisson Probability

Poisson Probability allows us to calculate the probability of certain events occurring within a fixed interval of time or space. With our football data, we can use Poisson to calculate the probability of events happening in the game like goals and cards. Conveniently, our data has events listed with the minute it occurred in. With this, we can find a 15-minute interval which Barcelona scored the most goals.

After looking through all the intervals, the 15-minute interval of 31st-45th minutes was the most fruitful for Barcelona, as they scored 22 goals in that time frame in their games this season. Now that we have a time interval, we can solve a Poisson Probability problem. The formula for Poisson probability goes as follows:

Lambda () can be calculated by taking the goals scored in that interval divided by the total number of matches (60).

The question is: Assume that goals in this interval follow a Poisson distribution (for education purposes), what is the probability that Barcelona scores at least one goal in the 31st-45th minute interval of a randomly selected match.

The probability of Barcelona scoring at least one goal in the 15-minute interval between the 31st-45th minute is 30.75%. This show that it took Barcelona a little bit to grow into their game and be able to get past the opponent. This calmness allowed Barcelona to control the tempo of the game and hold on to important scores.

Section 3.11: Tchebysheff’s Theorem

Football data, as most sports data is, is not normally distributed. Tchebysheff’s Theorem helps us estimate a certain number of values that will at least fall within range of the average, or within standard deviations of the mean. This theorem directly helps us visualize the spread of the data being looked at, even if the data is not normally distributed. Tchebysheff’s Theorem is like this:

If we are looking *within* k standard deviations, we use the first formula. If we are looking for *outside* k standard deviations, we use the second formula. For our problem, we will be looking *within* k standard deviations. Let’s say . In earlier sections, we calculated the mean () and the standard deviation (). First, we must calculate our intervals using this formula:

3.72 comes from . Since we can’t have negative goals in a football match, we’ll just set that interval to 0. Using Tchebysheff’s, we can finish the problem:

This equation tells us that at least 75% of matches in the 2014/15 season had between 0 and 6.68 goals scored. This shows that watching Barcelona was a spectacle, as many their matches were high scoring. The team carried a great rhythm which brought them historic success.

Section 4.2: Probability Distribution for a Continuous Random Variable

When it comes to probability distribution for a continuous random variable, it becomes difficult, or outright misleading if you were to use the football data for these kinds of problems. Since all the football data is discrete by nature, there are simply factors that cannot be fractional, like goals. Therefore, problems that involve continuous probability distribution are simply for educational purposes to show the different formulas.

Let Y be a continuous variable that represents the minute in which Barcelona scores a goal during a match. Assume Y is uniformly distributed across the 90-minute time (for educational purposes). **What is the probability that a goal is scored between the 30th and 60th minute of the match?**

Since our distribution is uniform, that means there is an equal chance of scoring a goal in every minute. So, we solve:

Despite the data not being continuous, the result looks like the result we got in the above sections where we used discrete methods to find out this probability. Interesting that when assuming data is continuous for experimentation, it came close.

Section 4.3: Expected Values for Continuous Random Variables

The **expected value** is the expected average outcome for a continuous random variable. This section gets more complicated, as calculus with integrals is now included. Again, we are assuming the football data is continuous for educational purposes and may not represent real life results. The expected value of a continuous random variable Y is:

Let Y represent the time during the first half of a match where Barcelona scores their first goal of the game. We can use this probability density function:

This states that is expected of Barcelona to score their first goal 22.5 minutes into the first half of game. A goal this early builds momentum and sets the team up for a successful game.

Section 4.10: Revisiting Tchebysheff’s Theorem

In this section, we revisit Tchebysheff’s Theorem. The formula hasn’t changed at all, but the context in which we use it is different. Tchebysheff’s can help us make a statement about how spread out the values are around mean, or the standard deviations. It also helps since some data isn’t normally distributed when calculating them could be extremely tedious. Let’s use an example of finding these intervals using our football data.

We can use Tchebysheff’s to find the interval around the mean that contains at least 85% of the matches in terms of goals scored. Our mean is 2.9, and standard deviation is 1.89.

We can now apply this to our interval

We can’t have negative goals, so we’ll just set that bound to 0.

This means that in 85% of Barcelona matches in the 2014/15 season, Barcelona scored between 0 and 7.78 goals. We can round that to 7 since fractional goals do not exists. This shows that Tchebysheff’s can tell us things about the data even if it is not normally distributed.

Chapter 5

The way the football data is structured, analyzing data with Chapter 5 becomes extremely experimental, and I would have to create problems that I cannot plan out (due to integrals). Having to come up with numbers that naturally solve into a readable answer but also relating it to the discrete football is rather difficult. I would have love to explore these complicated realms of probability, but the data I have gathered does not give me the tools to give adequate explanation or conclusions about the results.

Conclusion

Using this interesting data, we can calculate interesting probabilities using FC Barcelona’s 2014/2015 treble-winning season. We got to see how good the team was in both attack and defense, as they had a great record of achieving either feat, sometimes achieving both in the same match. Data like this can give us interesting insight on one of the hottest trends in recent times: sports gambling. The amount of data collection that is collected across thousands of games is used to calculate the odds of bets, and that was a conclusion I arrived to while calculating my own probabilities.

Source of Data:

[Kaggle](https://www.kaggle.com/datasets/davidcariboo/player-scores?resource=download&select=games.csv)