

**GOAL**  
**STOCK EXCHANGE**

STEPHEN KIMBROUGH · MARCO NICOLINI · DAMJAN ŠTRUS



📊 spreadsheet, goals statistics, averages, graphs, relative frequencies, coincidence, probability, quota

🖥️ mathematics, statistics, ICT

👥 15–19 years

### 1 | SUMMARY

This teaching unit will give students the opportunity to work with, make sense of and ask questions about real-life football data that is freely accessible on the Internet<sup>[1]</sup> or in daily newspapers.

### 2 | CONCEPTUAL INTRODUCTION

Football is the world's most popular sport with a global appeal that transcends national, cultural, gender and socioeconomic class boundaries. With an ever-expanding audience around the globe, football's appeal continues to grow, such that football has now become one of the most powerful businesses within the sport industry worldwide.

The European Football Market is valued at €19.4 billion<sup>[2]</sup>. Many people around the world depend on this industry to make a living, including players, coaches, referees, marketing companies, the media and, last but not least, bookmakers. The sports betting industry is worth anywhere between €606 billion and €870 billion a year. Bookmakers' work involves predicting whether a team will win or lose and calculating betting odds on this basis. A successful bookmaker needs not only luck but also acute mathematical skills in order to analyse complex sets of data, taking into consideration a variety of combinatorial factors and complex variables.

### 3 | WHAT THE STUDENTS DO

The most important skill that students first need to develop is to become familiar with designing and building a database using spreadsheets. The type of football data available online includes a wide variety of variables, including dates of games,

home and away scores, full-time and half-time scores, number of shots, corners, fouls, off-sides, yellow and red cards, and, of course, betting odds. Students can take the desired data sets from these sources and import them into their own spreadsheets.

#### 3 | 1 Data entry

First ask the students to make a spreadsheet with the results of games. An example of a spreadsheet is provided in **FIG. 1**. This spreadsheet is based on the German Bundesliga 1 in the season 2014/15.

The names of all the teams are listed in the left column (home team) and upper row (away team) in alphabetical order.

The results of each game are included in the two corresponding cells: the left cell indicates the number of goals scored by the home team, and the right cell indicates the number of goals scored by the away team. For example, when Bayern Munich played at home against Augsburg, the result was 0:1. When Augsburg were at home against Bayern Munich, the score was 0:4.

#### 3 | 2 Calculations

Ask the students to:

1. Develop a formula to calculate the number of games that were played in the whole season in the Bundesliga 1 (Hint: 18 teams all playing against each other).

Solution: Each team has 17 opponents that have all played both home and away games, so each team plays  $2 \cdot 17 = 34$  games (the Bundesliga 1 has 34 rounds). Since there are 18 teams, each round has nine matches. Therefore, in total, 306 matches were played in the season.

2. Calculate the goal statistics (goals scored and conceded) for each team for the whole season.

**FIG. 1** Spreadsheet with results of games; German Bundesliga 1, season 2014/15

	away team home team	Augsburg	Bayern	Bremen	Dortmund	Frankfurt	Freiburg	Hamburger SV	Hannover	Hertha	Hoffenheim	Köln	Leverkusen	Mainz	Mönchengladbach	Paderborn	Schalke	Stuttgart	Wolfsburg
1	Augsburg		0 4	4 2	2 3	2 2	2 0	3 1	1 2	1 0	3 1	0 0	2 2	0 2	2 1	3 0	0 0	2 1	1 0
2	Bayern	0 1		6 0	2 1	3 0	2 0	8 0	4 0	1 0	4 0	4 1	1 0	2 0	0 2	4 0	1 1	2 0	2 1
3	Bremen	3 2	0 4		2 1	1 0	1 1	1 0	3 3	2 0	1 1	0 1	2 1	0 0	0 2	4 0	0 3	2 0	3 5
4	Dortmund	0 1	0 1	3 2		2 0	3 1	0 1	0 1	2 0	1 0	0 0	0 2	4 2	2 1	0 3	0 0	2 2	2 2
5	Frankfurt	0 1	0 4	5 2	2 0		1 0	2 1	2 2	4 4	3 1	3 2	2 1	2 2	0 0	4 0	1 0	4 5	1 1
6	Freiburg	2 0	2 1	0 1	0 3	4 1		0 0	2 2	2 2	1 1	1 0	0 0	2 3	0 0	1 2	2 0	1 4	1 2
7	Hamburger SV	3 2	0 0	2 0	0 0	1 2	1 1		2 1	0 1	1 1	0 2	1 0	2 1	1 1	0 3	2 0	0 1	0 2
8	Hannover	2 0	1 3	1 1	2 3	1 0	2 1	2 0		1 1	1 2	1 0	1 3	1 1	0 3	1 2	2 1	1 1	1 3
9	Hertha	1 0	0 1	2 2	1 0	0 0	0 2	3 0	0 2		0 5	0 0	0 1	1 3	1 2	2 0	2 2	3 2	1 0
10	Hoffenheim	2 0	0 2	1 2	1 1	3 2	3 3	3 0	4 3	2 1		3 4	0 1	2 0	1 4	1 0	2 1	2 1	1 1
11	Köln	1 2	0 2	1 1	2 1	4 2	0 1	0 0	1 1	1 2	3 2		1 1	0 0	0 0	0 0	2 0	0 0	2 2
12	Leverkusen	1 0	2 0	3 3	0 0	1 1	1 0	4 0	4 0	4 2	2 0	5 1		0 0	1 1	2 2	1 0	4 0	4 5
13	Mainz	2 1	1 2	1 2	2 0	3 1	2 2	1 2	0 0	0 2	0 0	2 0	2 3		2 2	5 0	2 0	1 1	1 1
14	Mönchengladbach	1 3	0 0	4 1	2 1	3 1	3 1	0 1	0 2	0 3	2 3	1 1	0 3	0 1	1 1		2 0	4 1	1 1
15	Paderborn	2 1	0 6	2 2	2 2	3 1	1 1	0 3	2 0	3 1	0 0	0 0	0 3	2 2	1 2		1 2	1 2	1 3
16	Schalke	1 0	1 1	1 1	2 1	2 2	0 0	0 0	1 0	2 0	3 1	1 2	0 1	4 1	1 0	1 0		3 2	3 2
17	Stuttgart	0 1	0 2	3 2	2 3	3 1	2 2	2 1	1 0	0 0	0 2	0 2	3 3	2 0	0 1	0 0	0 4		0 4
18	Wolfsburg	1 0	4 1	2 1	2 1	2 2	3 0	2 0	2 2	2 1	3 0	2 1	4 1	3 0	1 0	1 1	1 1	1 3	1

FIG. 2 Graph of scored goals (green) and conceded goals (red) for each team, German Bundesliga 1, season 2014/15

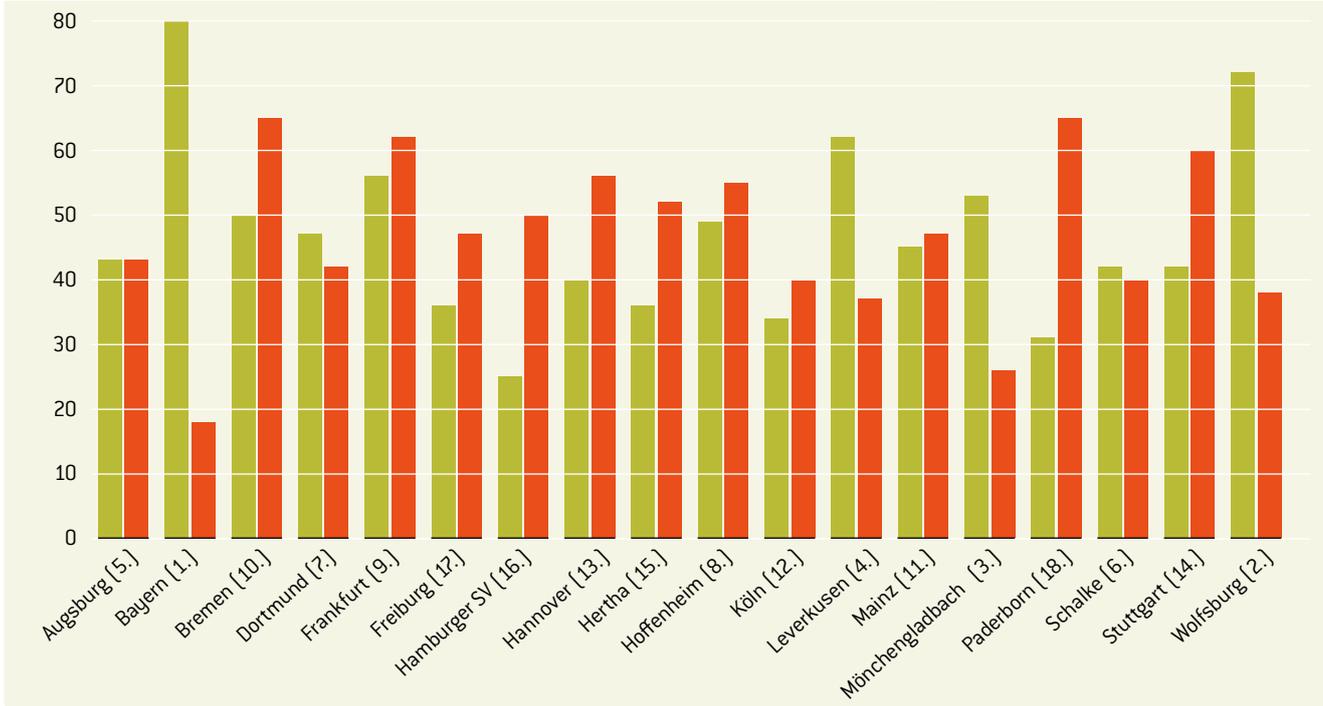


FIG. 2 shows all the goals that each team scored (highlighted in green) and all the goals that each team conceded (highlighted in red). Students can then compare the results from their spreadsheet with the real data from online databases in order to check their calculations.

3. Calculate the average number of goals per game in the whole season.

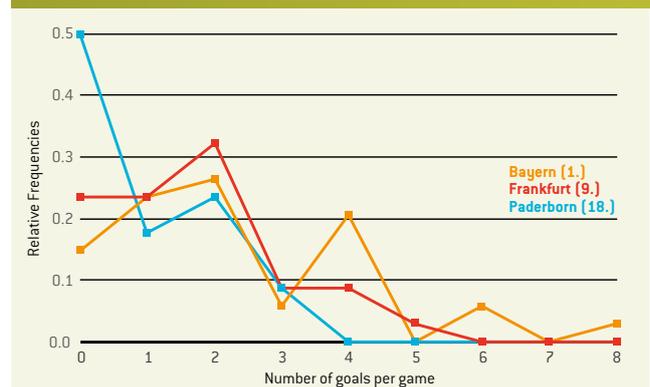
Solution: 2.75

4. Calculate the average number of goals per game that each single team scored and conceded. Students plot a graph of goals scored and conceded per game for each team. Ask the students to compare the graph with the position of each team in the final table, and allow them time to identify the connection between the shape of the graph and the ranking in the final table (in FIG. 2).
5. Calculate the relative frequency  $p(n)$  of the number of goals per game. Students can count the number of games in which each team scored 0, 1, 2, 3 goals etc. They make a spreadsheet for each team and plot a graph with relative frequencies versus number of goals per game for all teams. FIG. 3 shows that Bayern played a total of 34 matches and did not score in five matches, scored one goal in eight matches, two goals in nine matches etc. Encourage students to use formulas that the spreadsheet program offers in order to design the suggested table in FIG. 3.

FIG. 3 Relative frequencies  $p(n)$  for three teams

$n$	Relative Frequency					
	Bayern (1.)		Frankfurt (9.)		Paderborn (18.)	
	$N \cdot p(n)$	$p(n)$	$N \cdot p(n)$	$p(n)$	$N \cdot p(n)$	$p(n)$
0	5	0.15	8	0.24	17	0.50
1	8	0.24	8	0.24	6	0.18
2	9	0.26	11	0.32	8	0.24
3	2	0.06	3	0.09	3	0.09
4	7	0.21	3	0.09	0	0.00
5	0	0.00	1	0.03	0	0.00
6	2	0.06	0	0.00	0	0.00
7	0	0.00	0	0.00	0	0.00
8	1	0.03	0	0.00	0	0.00
	34	1	34	1	34	1

FIG. 4 Relative frequency versus number of goals per game for three teams



The sum of the second column is the number of games in the whole season played by one team, the sum of the third column is 1.

6. Find out what information (already calculated beforehand) students get if they multiply the number of goals  $n$  with corresponding relative frequency  $p(n)$  on each row of the table. Then summarise all products:

$$\sum_{n=0}^{\infty} n \cdot p(n).$$

Solution: They discover the average number of goals  $\bar{n}$  each team scored per season.

7. Use the average number of goals to calculate so-called coincidence in the outcome of the games. Coincidence is a relatively effective deviation and according to the Poisson distribution, it equals  $\sqrt{\frac{1}{\bar{n}}}$ .

The outcome of each game is more difficult to predict as the coincidence value increases. This is only a rough estimation; however, it can be argued that football is based on coincidence. Coincidence in real gaming can often be as high as 100%. At the same time, coincidence is higher when the football team is lower in the table.

8. Plot a graph to show how the position of each team changes in the table during the season (for each of 34 rounds). Discuss with the students some possible reasons that may have caused this rise or fall in the table.

### 3 | 3 Probability

9. Students already calculated the average number of goals that each single team scored per match. Let  $r_1$  be the average number of goals that the first team scores per match and  $r_2$  the average number of goals that the second team scores per match. We define  $R$  as the quotient:  $R = \frac{r_1}{r_2}$ .

The probability that the first team will score the next goal is calculated using  $p_1 = \frac{R}{R+1}$  and the probability that the second team will score the next goal is calculated using  $p_2 = 1 - p_1 = \frac{1}{R+1}$ .

Obviously, the averages change with each goal scored. However, we shall not consider this but rather use the previous averages for the whole match. Ask the students to calculate probabilities  $p_1$  and  $p_2$  for each team using the data obtained in 33 rounds so as to compare the theoretical calculations with the actual football results in the 34<sup>th</sup> round of the Bundesliga 1, 2014/15.

10. If, at a particular moment in the match, both teams have together scored  $n$  goals, the probability that all of the goals have been scored by the first team is  $p_1^n$ , and the probability that all of the goals have been scored by the second team is  $p_2^n$ . The probability that the first team has scored  $k$  of  $n$  goals is  $\binom{n}{k} p_1^k p_2^{n-k}$ .

11. The probability that the team which scores  $r$  goals per match will score  $n$  goals in time  $t$  (between 0 = start and 1 = end of the match) equals  $p = \frac{(rt)^n}{n!} e^{-rt}$ .

Ask students to plot the graph of the probability of scoring  $n$  (0, 1, 2, 3 or 4) goals during 90 minutes of a football game for each team. Use the data obtained in 33 rounds to compare the theoretical calculations with the actual football results in the 34<sup>th</sup> round of the Bundesliga 1, 2014/15.

12. Students can also be asked to check the probability of result  $n:m$ . The theory states that this probability produces the equation

$$p_{n,m} = \frac{(r_1 t)^n (r_2 t)^m}{n! m!} e^{-(r_1+r_2)t}.$$

This equation presumes that the number of goals for each team is independent of each other, which is obviously not true, but it can be used for a first approximation. Students should compare the theoretical calculations with the actual football results in the 34<sup>th</sup> round of the Bundesliga 1, 2014/15 (FIG. 5).

FIG. 5 Football results of the 34<sup>th</sup> round of the Bundesliga 1 in the season 2014/15 [3]

Bayern	Mainz	2 : 0
Dortmund	Bremen	3 : 2
Frankfurt	Leverkusen	2 : 1
Hamburger SV	Schalke	2 : 0
Hannover	Freiburg	2 : 1
Hoffenheim	Hertha	2 : 1
Köln	Wolfsburg	2 : 2
Mönchengladbach	Augsburg	1 : 3
Paderborn	Stuttgart	1 : 2

### 4 | CONCLUSION

Continuous examination and analysis of data sets can most certainly help to predict the results of football games. However, to predict the correct score for certain games, many other parameters, apart from goals, must be taken into consideration (i.e. injuries, players' form, condition of the pitch, weather conditions etc.). If there was a magic formula there would be a lot more betting millionaires. Similarly, the work for odds compilers could be considered more art than science.

However, the aim of this teaching unit has not been to talk about football betting, so let us conclude here.

### 5 | COOPERATION OPTIONS

Students from different countries can collect the results of all games of their top national football league. They should then calculate the goal statistics (score and conceded) for each team for the whole season, calculate the average number of goals per game in the whole season and calculate the average number of goals per game that each single team scored and conceded.

Finally, they should compare the results of their calculations and analyse their national league. Are all the teams more or less equal, or are there a few very strong teams, a few very weak teams and a larger number of average teams? Maybe students will even discover a third, fourth or fifth option ...

### REFERENCES

[1] [www.football-data.co.uk/](http://www.football-data.co.uk/)

[2] [www.soccerex.com/about/what-soccerex/football-industry](http://www.soccerex.com/about/what-soccerex/football-industry) (08/11/2015)

[3] [www.rezultati.com/nogomet/njemacka/bundesliga-2014-2015/](http://www.rezultati.com/nogomet/njemacka/bundesliga-2014-2015/) (12/11/2015)

- ALI JE NOGOMET IGRA NA SREČO, Janez Strnad, Presek, ISSN 0351-6652, year 13 (1985/1986), number 1, pp. 9–15
- Matematika i nogomet (<http://pptfilesearch.com/single/79931/nogomet-i-matematika>), Franka Miriam Brückler, Osijek, 1.6.2006 (08/03/2016)