**Science & Philosophy** Volume 7(1), 2019, pp. 109-120

# Statistics for a Football Coach

Ferdinando Casolaro[•]

Mario Cristiani[*]

**Abstract**

This work presents a Decision-Making Model referring to the forecasts about the Football World Cup in Brazil (2014). This work aims to demonstrate how it is possible to approach young students to the study of Mathematics through evoking themes that are congenial to them and able to arouse their interest.

**Keywords:** Football World Cup, Motivation, Statistical Analysis, Regression line, Didactics, Students. [ł]

[•]Department of Architecture, University of Naples Federico II, ferdinando.casolaro@unina.it.

[*]Department of Law, Economics, Management and Quantitative Methods (DEMM), University of Sannio, mariocristiani31@gmail.com.

[ł] Received on December 22nd, 2018. Accepted on February 13rd, 2019. Published on June 30th, 2019. doi: 10.23756/sp.v7i1.444. ISSN 2282-7757; eISSN 2282-7765. 

# 1. Introduction

Football represents one of the most followed social phenomena at every level (F. Casolaro, 2014). The statistical analysis tells us that "being a footballer" is the first aspiration of the majority of boys!

Along with this aspiration, a strong passion for sports journalism is spreading, even among the youngest (of both sexes), therefore they are inclined to consider events with a much more critical eye. In particular, in the years in which the World Cup finals take place, the events are also followed by those who are not football fans because, through the media, the event has access to every house.

World Cup 2014: Italy, against any forecast, comes out in the first round with 3 points! Was it a case or was the statistical analysis in the forecasts wrong?

Immediately, in the various comments of technicians, experts, journalists and columnists, on the debacle of our National Team, only the responsibilities of CT Prandelli were read out and they were connected to serious errors in the tactical approach (F. Casolaro 2014). However, it is important to consider the technical resources available to Prandelli in the Serie A (2013/2014) in which 68% of the footballers in the first teams were foreign players and most of the remaining 32% were in the middle-low ranking teams.

If we do not consider the six footballers of Juventus (Buffon, Barzagli, Bonucci, Chiellini, Pirlo and Marchisio, arrived at that World Cup worn out by the efforts of three competitions – Serie A, Champion's League, Italian Cup -), the majority of 17 other players were playing in the mid-table teams, and were therefore considered, at the time of the facts, unfit to participate in the most important competitions for club teams. Below is an indication of the club teams in which the other 17 members of the Italian team (at the Brazilian World Cup) played:

- AS Roma, a team ranked second in the Serie A 2013/2014, provided only De Rossi, as 8/11 of the first team were foreigners: the other three were Totti and De Sanctis (respectively 38 and 37 years old) and Florenzi (23) who was considered too young;
- SSC Napoli, that in that league (in which it ranked third) had the first team composed almost exclusively by foreign players, provided only Insigne;
- AS Fiorentina, ranked fourth, had only three Italian in the first team. Of these the only Aquilani was called by Prandelli to join the Italian team;
- FC Inter, fifth place, had no Italian players in the first team.

From the sixth place onwards we find the other players: Cassano, Paletta and Parolo from Parma, sixth place; Darmian, Cerci and Immobile from Torino, that ranked seventh; De Sciglio, Abate and Balotelli from Milan, that ranked eighth; Candreva from Lazio (ninth); Perin from Genoa (fourteenth).

To these must be added Sirigu, Verratti and Thiago Motta who were playing abroad in Paris S. Germain, and therefore did not have the opportunity to compete in the Serie A.

It is legitimate to ask, at this point, if another coach could have done more. To answer the question, we will use a simple linear regression model.

## 2. The independent variable x: motivation, skills and experience of footballers

The World Cup is the competition that every child, every soccer lover would like to play and is, at the same time, the great goal of the overpaid professionals of the footballing universe.

The World Cup is, therefore, an event in which the motivations are extremely important and can combine in a holistic system, giving rise to sometimes extraordinary results that disavow the predictions of the best experts.

What we intend to do with our model is to provide guidelines based on pragmatic reasoning, totally unrelated to the magic of 11 men who, by throwing their hearts beyond the obstacle, exceed their limits. We will keep ourselves within those limits with an exception that we will clarify later.

The result achieved by a National Team in a World Cup obviously depends on the skill of the players that make up the team. A good indicator of players' abilities may be the number of appearances collected by them in international club events (Champion's League, Europe League, Copa Libertadores etc.).

The higher the attendance in these competitions, the greater the players' abilities and the habit of the same to face high-level competitions like a World Cup will presumably be.

Using the data provided by the "Soccerway" website, we have identified, for each player of the different National Teams, the number of appearances in the most important international club competitions in the four football seasons that preceded the World Cup in Brazil (2014):

- Champion's League (CHL) and Europe League (EUL) for Europe;
- Copa Libertadores (COL) and South American National Cup (CNS) for South America;
- Caf Champion's League (CCL) for Africa;
- AFC Champion's League (ACL) for Asia;
- Concacaf Champion's League (CCC) for North and Central America.

The appearances, however, must be appropriately weighted: an appearance in the European Champion's League must be worth more than an appearance in the Europe League or in the CONCACAF (North and Central American Champions Cup). This is because depending on the events the value of the participating teams changes. Below are our evaluations concerning the importance of the events and their weight in the model we have built:

- An appearance in the Champion's League (CHL) will be worth 1 point;
- An appearance in the Europe League (EUL) will be worth 0.7 points;
- The Copa Libertadores (COL) is the South American Champion's League. An appearance in this competition will be worth 0.8 points (0.2 less than the Champions League, as the strongest South American players play in Europe);
- The South American National Cup (CNS) is the equivalent of the Europe League in South America; an appearance in this cup will be worth 0.56 points;
- The AFC Champion's League (ACL) is the most important international club competition in the Asian Continent. An appearance in this event will be worth 0.3 points. In the 2013 edition, Guangzhou Evergrande won the competition: its coach was Marcello Lippi who in 2006 won the World Cup with Italy;
- The Concacaf Champion's League (CCC) is the equivalent of the Europe League for clubs in Central and North America. The CAF Champion's League (CCL) is the African Champions League. An appearance in these competitions will be worth 0.3 points.

With the assigned weights, we geometrically structure our model.

Imagine we have, in a National Team, players with a total number of appearances in the listed competitions as follows: 40 Champion's League (weight 1), 50 Europa League (weight 0,7), 30 Copa Libertadores (weight 0,8).

Step 1 - We multiply the appearances for the weights and we carry out the sum of the products 40x1 + 50x0,7 + 30x0,8 = 99.

Step 2 - We divide the value of the sum (99) by 23 (number of players of each National Team): $\frac{99}{23} = 4{,}3$.

What we have achieved is the Average of the weighted appearances of the national players in the most important international club events. In our model, it indicates the strength of the team and the result achieved by the National Team in the World Cup is made to depend on it. It is the independent variable and is indicated with the letter x (M. Squillante et. al 2016).

# 3. The dependent variable y: the score obtained by the National Teams in the World Cup

Each team is assigned a score based on the results obtained in the competition:

- 3 points for victory in the regular time;
- 1 point for a tie in the regular time;
- 1 point for victory in extra time;

- 0.3 points for a tie in extra time;
- 0 for defeat;
- 0.2 points for victory on penalties.

Points are reduced to one third for the final for the third and fourth place.

Thus, a National Team that has passed the group with 6 points and has been eliminated in the round of 16 on penalties will be assigned a score calculated as follows:

6 + 1 (tie in the regular time of the round of 16) + 0,3 (tie in the extra time of the round of 16) = 7,3.

This score is our dependent variable and will be indicated with the letter y.

Let us briefly summarize the course of the 30 National Teams in the 2014 World Cup:

- Holland: Overcame the group with 9 points. Overcame the round of 16 in the regular time and won the quarter-finals on penalties. It was eliminated in the semi-finals on penalties. Won the final for the third/fourth place in the regular time;
- Belgium: Overcame the group with 9 points. Overcame the round of 16 after extra time. It was eliminated in the quarterfinals in the regular time;
- Switzerland: Overcame the group with 6 points. It was eliminated in the round of 16 after extra time;
- Germany: Overcame the group with 7 points. Overcame the round of 16 after extra time. Overcame the quarterfinals in the regular time. Won the semifinals in the regular time. Won the final after extra time;
- Russia: Was eliminated in the group with 2 points;
- Bosnia: Was eliminated in the group with 3 points;
- England: Was eliminated in the group with 1 point;
- Greece: Overcame the group with 4 points. Was eliminated in the round of 16 after penalties;
- Croatia: Was eliminated in the group with 3 points;
- Portugal: Was eliminated in the group with 4 points;
- France: Overcame the group with 7 points. Overcame the round of 16 in the regular time. Was eliminated in the quarterfinals in the regular time;
- Brazil: Overcame the group with 7 points. It won the round of 16 after penalties. Overcame the quarterfinals in the regular time. Was eliminated in the semifinals in the regular time. Lost the third/fourth final in the regular time;
- Argentina: Overcame the group with 9 points. Overcame the round of 16 after extra time. Won the quarter-finals in the regular time. Overcame the semi-finals after penalties. Lost the final after extra time;
- Colombia: Won the group with 9 points. It won the round of 16 in the regular time. Was eliminated in the quarterfinals in the regular time;
- Chile: Overcame the group with 6 points. Was eliminated in the round of 16 after penalties;

- Ecuador: Was eliminated in the group with 4 points;
- Mexico: Won the group with 7 points. Was eliminated in the round of 16 in the regular time;
- Uruguay: Overcame the group with 6 points. Was eliminated in the second round in the regular time;
- Honduras: Was eliminated in the group with 0 points;
- United States: It overcame the round with 4 points. Was eliminated in the round of 16 after extra time;
- Costa Rica: Won the group with 7 points. It won the round of 16 after penalties. Was eliminated in the quarters after penalties;
- Ivory Coast: Was eliminated in the group with 3 points;
- Nigeria: Overcame the group with 4 points. Was eliminated in the round of 16 in the regular time;
- Cameroon: Was eliminated in the group with 0 points;
- Algeria: Overcame the group with 4 points. Was eliminated in the round of 16 after extra time;
- Ghana: Was eliminated in the group with 1 point;
- Japan: Was eliminated in the group with 1 point;
- Iran: Was eliminated in the group with 1 point;
- South Korea: Was eliminated in the group with 1 point;
- Australia: Was eliminated in the group with 0 points.

# 4. From the set of observations to the regression line

Table 1 shows, for each of the 30 National Teams, the scores assigned to the teams, the players' appearances in the most important international club competitions and the Average of weighted appearances:

| National Team | Score (y) | Appearances | Average of weighted appearances (x) |
|---|---|---|---|
| HOLLAND | 15,8 | CHL=167; EUL=199 | 13,32 |
| BELGIUM | 11 | CHL=255; EUL=224 | 17,9 |
| SWITZERLAND | 7 | CHL=184; EUL=166 | 13,05 |
| GERMANY | 17 | CHL=546; EUL=105 | 26,93 |
| RUSSIA | 2 | CHL=169; EUL=207 | 13,65 |
| BOSNIA | 3 | CHL=61; EUL=141; ACL=7 | 7,03 |
| ENGLAND | 1 | CHL=226; EUL=94 | 12,7 |
| GREECE | 5,3 | CHL=162; EUL=133 | 11,09 |
| CROATIA | 3 | CHL=273; EUL=159 | 16,7 |
| PORTUGAL | 4 | CHL=282; EUL=250 | 19,87 |
| FRANCE | 10 | CHL=316; EUL=110 | 17,087 |
| BRAZIL | 11,5 | CHL = 370; EUL=96; COL=162; CNS = 13 | 24,96 |
| ARGENTINA | 16,5 | CHL = 260; EUL = 131; COL = 96; CNS =18; CCC = 24 | 19,38 |
| COLOMBIA | 12 | CHL= 74; EUL= 115; COL= 111; CNS=41 | 11,57 |
| CHILE | 7,3 | CHL= 71; EUL= 90; COL =151; CNS= 155 | 14,85 |
| ECUADOR | 4 | CHL = 36; EUL =31; COL = 206; CNS = 149; ACL =14; CCC=54 | 14,19 |
| MEXICO | 7 | CHL =37; EUL = 38; COL = 77; CCC = 111 | 6,89 |
| URUGUAY | 6 | CHL = 117; EUL = 189; COL = 41; CNS = 17; ACL = 6 | 12,76 |
| HONDURAS | 0 | CHL =36; EUL = 18; CCC = 112 | 3,57 |
| UNITED STATES | 5 | CHL = 19; EUL = 80; CCC = 72 | 4,2 |
| COSTA RICA | 9,8 | CHL = 39; EUL = 82; CCC = 88 | 5,34 |
| IVORY COAST | 3 | CHL = 104; EUL = 147; CCL = 6 | 9,07 |
| NIGERIA | 4 | CHL = 81; EUL = 53; CCL = 5 | 5,2 |
| CAMEROON | 0 | CHL= 176; EUL =97; CCL = 8 | 10,71 |
| ALGERIA | 5 | CHL 38; EUL = 62; CCL = 13; ACL = 21 | 3,98 |
| GHANA | 1 | CHL = 138; EUL =81; CCL=3; ACL=12 | 8,66 |
| JAPAN | 1 | CHL = 58; EUL = 72; ACL = 98 | 5,99 |
| IRAN | 1 | ACL = 217; CCC = 1; EUL =7 | 3,056 |
| SOUTH COREA | 1 | CHL = 24; EUL = 19; ACL = 167 | 3,8 |
| AUSTRALIA | 0 | CHL = 11; EUL = 13; ACL = 97 | 2,14 |

Table 1. The scores assigned to the 30 National Teams, the players' appearances in the major international club competitions and the Average of weighted appearances.

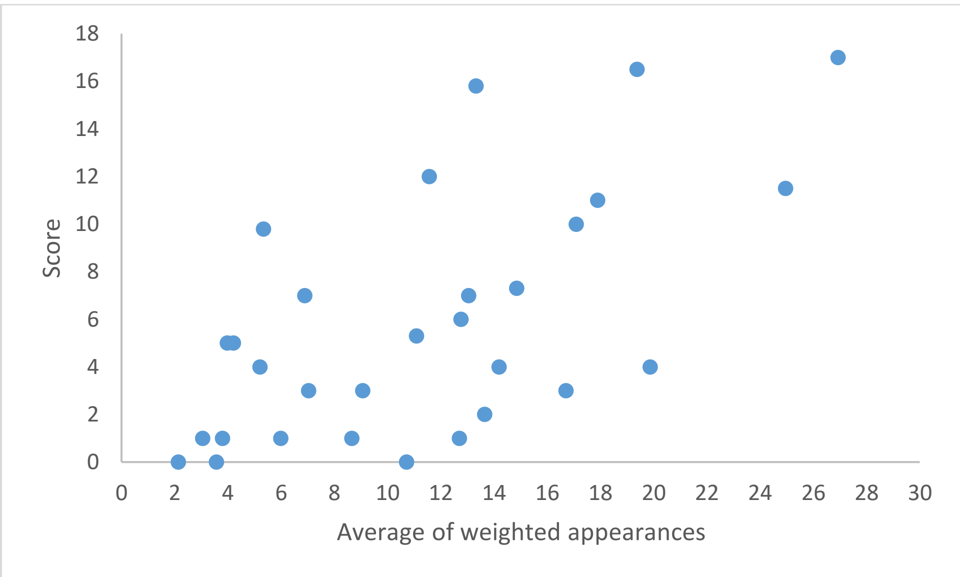
From the observations we have eliminated Spain which appeared in Brazil as the defending champion. From the 2002 World Cup to the 2014 World Cup three times out of four the defending champion team was eliminated in the first round: France in Korea and Japan (2002), Italy in South Africa (2010), and Spain in Brazil (2014). At these must be added Germany presented at the World Championship played in Russia in 2018 as World Champion: it was eliminated in the first round.

The trend of the motivational variable is, therefore, in this case, and limited to this aspect, easily predictable and it is even more so for the Spain which in addition to the 2010 World Cup had won the European Championships of 2008 and 2012.

Reported in a system of Cartesian axes, the set of observations takes the form of a point cloud: the dispersion diagram (figure 1). Each point represents one of the 30 National Teams: the abscissa is the Average of weighted appearances, the ordinate is the score obtained.

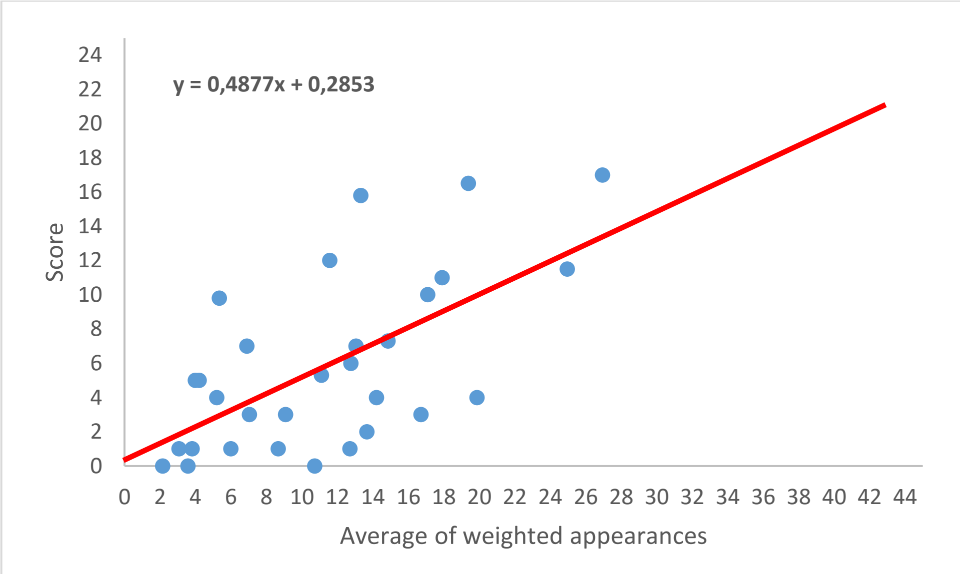
The diagram is a photograph of what actually happened:

Figure 1: Dispersion Diagram




Our regression line has equation: $y = 0{,}28 + 0{,}48x$.

Figure 2: The regression line



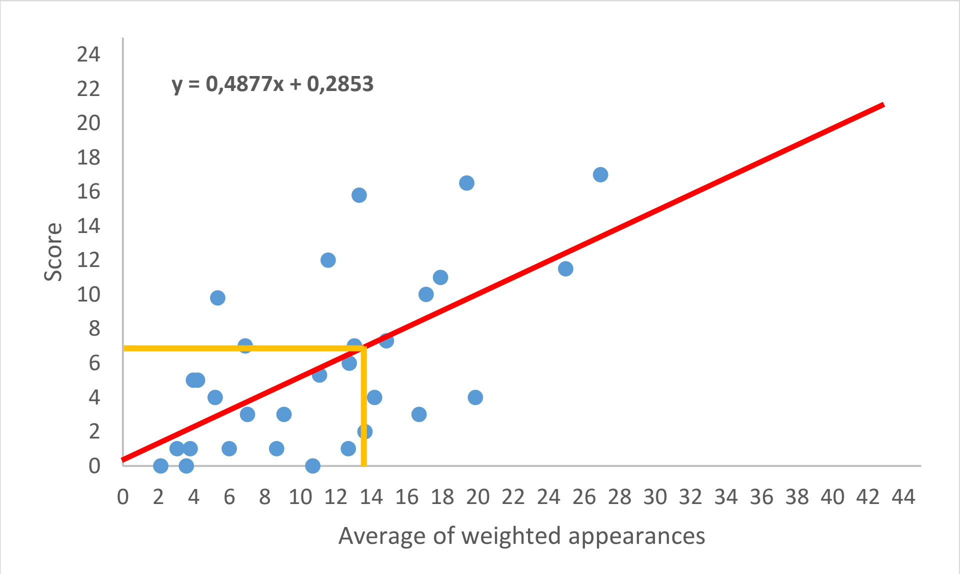

The regression line (figure 2) is the one that best interpolates the points of the dispersion diagram. It is the one that minimizes the sum of the squares of the distances of the points from the line (F. Casolaro, A. Fontana 2018).

The maximum score attainable by a National Team is 21 (victory of all matches in the regular time). The score will be obtained, according to our line, by a team with an x (Average of weighted appearances) of 42,5. With an x equal to or greater than 42,5 the score obtained will always be 21. The x can never take negative values. The angular coefficient of 0,49 indicates that for an increase in the Average of weighted appearances of 1 there is a growth of the score of 0,49. The vertical intercept of 0,28 indicates that a National Team with the Average of weighted appearances in major events for Club of 0 in our model reaches a score of 0,28 points.

At this point, by inserting the value of the Italian National Team (13,65) in the equation of the line, we calculate the score that according to our model it should have obtained (figure 3):

$$y = 0{,}28 + (0{,}48)(13{,}65) \Longrightarrow y = 6{,}94$$

Figure 3: The case of Italy




So Italy should have obtained a score of 7 (rounded up). This score is associated with combinations of different results, certainly better than those obtained by our National Team, but which in no case go beyond a complicated landing in the quarter-finals.

Only in one circumstance, the Azzurri could have overcome the quarter of finals, but this case turns out to be so fortunate and daring that we believe it is right, considering also the rounding up, to exclude it.

It may be useful to consider the score reached by the National Teams that had an Average of weighted appearances similar to that of Italy. Next to Holland (x = 13,32) which reached a truly remarkable result being eliminated in the semifinals on penalties and then beating Brazil in the final for 3rd / 4th place (y = 15.8) and to Russia (x = 13,65) which instead was surprisingly eliminated in the groups with 2 points (y = 2) we find Switzerland (x = 13,05) which exceeded the groups with 7 points and then was eliminated in the round of 16 from Argentina in the regular time (y = 7). The score of Switzerland is exactly the same as the one that, in our model, should have reached Italy and wanting to make an arithmetic average of the scores of the three National Teams, the score obtained is 8.2, so very close to the 7 estimated by us for the National Prandelli Team.

The choice of the line as a function is completely discretionary.

The line is the simplest and most intuitive model and it is, above all for this reason, the most used in educational applications. It is not absolutely certain, however, that it is the best to approximate the link between the variables taken into consideration in our analysis (A. Maturo, R.M. Contini 2010).

The coefficient of determination:

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = 0{,}4$$

in our opinion, indicates a good adaptation of the linear function to the observed data.

# 5. Conclusions

As can be seen, the results of the statistical analysis differ only insignificantly from the opinion expressed in the introduction, according to which Italy could not have done more.

With reference to the motivational aspect it is not superfluous to underline how, in continuity with the results of the World Cups from 2002 to 2014 where on three occasions out of four the defending champion team had been eliminated in the first round, also in the 2018 event, after the one taken into consideration in the analysis, Germany (winning team in 2014 and one of the favorites in the World Cup in Russia) was eliminated in the group stage.

As clarified in the work, the pragmatism of our model limits the consideration of the motivations to this single aspect that results however fundamental in the attainment of our statistical result; therefore we consider it appropriate to add sociological reasons among the pre-requisites of the decision analysis. (F. Casolaro, L. Paladino 2012).

And the social and psychological aspects were fundamental in the presentation of the model in some conferences concerning the teaching of Mathematics: 4° Convegno Nazionale *La Matematica nel 1° Ciclo: aspetti didattici, sociologici e interdisciplinary* Chieti, 06-09 aprile 2016 Università di Chieti-Pescara; X edizione Convegno di Geometria *Giochi matematici per la scuola premio "Aldo Morelli"*, Castellammare di Stabia 20-22 maggio 2016; 10° Convegno di matematica, Montesarchio 12 e 13 maggio 2017, Auditorium "E. De Filippo" Liceo "E. Fermi"- Ambito BN5.

Also, the rhetorical implications that are always identified in sporting events were fundamental in attracting the attention of students and teachers. While the students were fascinated by the possibility of using Mathematics for the analysis of interesting phenomena and integral parts of their passions, the teachers saw in the study the possibility of presenting even complex topics through more engaging tools.

Model improvements are achievable by considering other independent variables on the basis of which to determine the National Team's score value. This could be made to depend, for example, not only on the Average of weighted appearances in international Club events but also on the results achieved by the National Teams in the most recent events (we could consider the previous World Cup or the last Continental cup to which the team has taken part). However, it would end up in the field of multi-varied regression with all the complications related to a study of this type. Having the model purely didactic aims and not being oriented to the elaboration of a statistic with forecasting purposes, improvements of this type should be made only later and following a discussion activity that actively involves the students who, once learned the rudiments of bivariate regression, could subsequently approach multivariate analysis.

# References

[1] F. Casolaro (2014). “Prandelli non poteva fare di più”, Il Crotonese 12-08-2014: “I numeri non aiutano l'Italia”.

[2] F. Casolaro (2014). “La Matematica del calcio: un corso made in Sannio”, Il Mattino 18-06-2014

[3] F. Casolaro, L. Paladino (2012). "Didactics of Statistics in Sociology" - First International Conference on Recent Trends in Social Sciences: Qualitative Theories and Quantitative Models (RTSS) - Iaşi (Romania), 23-25 September, 2012.

[4] M. Squillante et. al, (2016). “The Logic of Probability: A Trip through Uncertainty”. Science & Philosophy - Vol. 4(2), 2016,

[5] Mirray R. Spiegel (1994). “Statistica - 975 problemi risolti”. Ed. McGraw-Hill.

[6] A. Maturo, R.M. Contini (2010), "Statistical and mathematical models for the analysis of educational processes in the intercultural school", WCLTA-2010 CAIRO, Procedia Social and Behavioral Sciences, 9, (2010).

[7] F. Casolaro, A. Fontana (2018). “I pre-requisiti essenziali per lo studio della Probabilità e della Statistica” – Quaderni dell'APAV n. 2 - 2018, pag. 45-52.

[8] F. Casolaro (2018). “Un approccio didattico all'insegnamento della Statistica” - Quaderni dell'APAV n. 2 - 2018.