UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE ECONOMICHE ED AZIENDALI "M.FANNO"

CORSO DI LAUREA MAGISTRALE IN

BUSINESS ADMINISTRATION

TESI DI LAUREA

"Experience as determinant of performance: elements of success and failure in the Italian football Serie B"

RELATORE:

CH.MO PROF. Diego Campagnolo

CO-RELATORE

DOTT. Luca Maestrini

LAUREANDO: Immanuel Taddei

MATRICOLA N. 1166361

ANNO ACCADEMICO 2019 – 2020

Il candidato dichiara che il presente lavoro è originale e non è già stato sottoposto, in tutto o in parte, per il conseguimento di un titolo accademico in altre Università italiane o straniere.
Il candidato dichiara altresì che tutti i materiali utilizzati durante la preparazione dell'elaborato sono stati indicati nel testo e nella sezione "Riferimenti bibliografici" e che le eventuali citazioni testuali sono individuabili attraverso l'esplicito richiamo alla pubblicazione originale.

The candidate declares that the present work is original and has not already been submitted, totally or in part, for the purposes of attaining an academic degree in other Italian or foreign universities. The candidate also declares that all the materials used during the preparation of the thesis have been explicitly indicated in the text and in the section "Bibliographical references" and that any textual citations can be identified through an explicit reference to the original publication.

Firma dello studente

Summary

EXECUTIVE SUMMARY

Among the data present in the 2019 Football Report Pwc 2019, there are those related to the table I that collects for 11 seasons the position of the winning teams of the Serie B in terms of the salaries ranking. As it can be seen, the team that usually wins is not the one that invests the most in salaries and the average position in these seasons is 5.9. In theory, the one who spends the most should be the one who gets the talent and the best players in terms of experience, performance, future development and value. As these data show, a higher investment does not always lead to better performance. This study has tried to understand if experience and which of its components influence the performance of a team in the context of Serie B, with a particular focus on players.

Season	Serie B	Winner team
2007-2008	3	Chievo Verona
2008-2009	3	Bari
2009-2010	2	Lecce
2010-2011	9	Atalanta
2011-2012	5	Pescara
2012-2013	4	Sassuolo
2013-2014	1	Palermo
2014-2015	16	Carpi
2015-2016	1	Cagliari
2016-2017	19	SPAL
2017-2018	2	Empoli
Average Placement	**5,9**	

Table I: Comparison of the ranking of the salaries of the winners of Serie B Source: Report Calcio Pwc, 2019

The question appears to be of interest for several reasons. First of all because every football team is characterized by an economic cycle in which the manager has a limited amount of financial resources that must be used to select the players to achieve the best possible sports result. The sports result is closely linked to the revenues of a team. In fact, a better result allows to increase the revenues that can then be invested to start the cycle again or, eventually, to make a profit. So the problem of selecting players within a team is fundamental to obtain not only the sports result but also a good economic gain. Besides this the literature on this particular context is practically absent, despite the extremely interesting nature of the league, which presents, from all points of view, an unexpressed potential and an exploitation not adequate to its possibilities. Finally, the

study of the experience-performance relationship in the managerial field is largely treated, the same thing is not true for the sports context, so this work is the only one that deals with this relationship in the context of the B Series.

To study the relationship, data were collected for players, coaches and the teams themselves in 5 consecutive seasons: 2018/19; 2017/18; 2016/17; 2015/16 and 2014/15. For the data collection only the teams that in each season were directly promoted to Serie A or participated in the playoffs (henceforth called HP) and the teams directly relegated to Serie C or played the playout (henceforth called LP) were considered. First of all the literature both in the managerial and sports areas of the experience-performance relationship has been reviewed. Both managerial and sports literature has yielded mixed results about the workers', managers' and organizations' experience-performance relationship.

Players have been analyzed through a different methodology from the one used for coaches and teams, being the focus of the study on players. In order to investigate the relationship between experience and performance in the dataset of coaches and teams, a simplistic model was proposed using arbitrary scores (Chapter 4.4). As far as players are concerned, instead, it has been tried to understand if in at least 4 out of 5 of the seasons analyzed there were recurring characteristics of HP experience and LP experience and if these characteristics differed. To find this out, first of all, the PCA methodology was used with a combined cluster (HP and LP players combined) and separated to see which components of the players' experience explained more variance in one and the other cluster. From the results obtained through the PCA, multiple logistical regression models were used to investigate whether there is a relationship between a characteristic of the experience and the player's probability of belonging to the HP cluster and the nature of this relationship.

Through the analysis carried out with the PCA it emerged that the HP cluster and the LP cluster have in common the variables "Age" and "NPIFLC" (number of presences in first level championships) in at least 4 out of 5 seasons while the variables that differed were "NYIT" (number of years in the team) for HP and "NPIICcc" (number of presences in international competitions (club competitions)) for LP. The results of the preliminary analysis had already showed that there were no great differences in terms of experience

characteristics between HP and LP players, which is confirmed by the multiple logistic regression models that have not shown in at least 4 years out of 5 an influence of a particular variable on the probability that a player belongs to the HP cluster. In essence, it has been shown that the players' experience does not influence the success or failure of a team.

The proposed model for the analysis of the experience-performance relationship for coaches and teams, on the other hand, provided guidance that coaches' experience does not influence performance while the teams' experience it is a better predictor for it.

The results obtained through this study have important implications for different actors in the analysed context. On the one hand, the player selection manager does not have to take into account the experience for building a successful team. This lack of a clear relationship between experience and performance can be caused by the nature of the competition itself, the lack of incentives that experienced players may have to play in a second level league and the lack of ability of the club and the coach to efficiently integrate the different backgrounds of the players. In fact, as far as coaches and teams are concerned, it has emerged that in principle for the selection of a coach one should look at his or her ability to integrate the skills and experiences of the different players, while the teams that usually perform better are those that have historically performed better, therefore a character that cannot always be changed in the short term. Maybe the team's experience is important because there are values that are fundamental and that are sedimented in the team's culture ("winning mindset"). This means that for a team is important to have a coach that is able to interpret and to transmit those values to the players. Finally, also implications at the level of the League itself can be taken into account. The fact that the context analysed is extremely fluid and it is difficult to attribute certain factors to success, as well as, the uncertainty given by the result both in the same season and in the long term allow to keep the uncertainty, that increases the appeal of the league, high and allow the league not to have to implement regulatory policies to maintain the competitive balance.

To obtain these implications the work has been structured in four chapters. The first chapter is dedicated to the analysis of sport as a business with particular attention to football and the context of Serie B. In the second chapter there is a review of the

literature both in the managerial and sports areas of the experience-performance relationship. The explanation of the criteria used to compile the datasets, the qualitative and quantitative descriptions of the variables and the preliminary analysis of the data are in the third chapter. Finally, the last chapter concerns the application of multiple logistic regression models to the players' dataset and the analysis through the proposed model of the coaches and teams dataset. In addition, in the fourth chapter the managerial implications and the limitations and possible developments of the study are present.

CHAPTER 1

THE SPORT INDUSTRY AND THE FOOTBALL BUSINESS

1.1. INTRODUCTION

The purpose of this chapter is to provide a better framework for the context in which the empirical analysis of experience as a success factor in professional football clubs in the B Series will be carried out. So the work starts from the very definition of sport and sport industry (Pitts, 1988; Pitts, Fielding & Miller, 1991). This is important to understand that the sports industry is composed of several segments of which football is a very important part being the most practised sport in the world but still constituting a small part in terms of total variety. This is essential in order to understand which direction the sports industry is moving to and how the various actors can be classified, perhaps starting from the core business within the same industry (Mancin, 2018). This industry can largely be considered as "regulated" and therefore subject to the rules of the ordinary law but also to those of the "special" law. For this reason, an attempt has been made, in the chapter, to give an overview of the organisation of sport, starting from the world level with a focus, of course, on the bodies related to football. An overview of the social and economic impact of sport in our country has also been given. Football has always been analyzed with special attention, starting from the historical vicissitudes that led to the creation of the football product as we know it today, up to the analysis of the impact in terms of value in Europe and in particular in Italy. Finally, the context in which the analysis will be carried out, focusing on the intrinsic characteristics of the championship, its economic value and its evolution over time, the Serie B have been analysed in details.

1.2. A GENERAL OVERVIEW OF SPORT BUSINESS

What is the definition of sport? Could sport be considered a business? Defining the concept of sport is very difficult and probably there is not a definitive answer. For example the Cambridge Dictionary defines sport as "*a game, competition, or activity needing physical effort and skill that is played or done according to rules, for enjoyment and/or as a job"*. The Oxford Dictionary describes it as "*An activity involving physical exertion and skill in which an individual or team competes against another or others for entertainment."*. Mancin (2018) describes it as a social and economic phenomenon with a very strong heterogeneity in terms of audience. This makes it one of the most variegated industrial sector. This is due to some peculiar features. First of all the importance of the social dimension that it assumes, the positive values that it transmits, the capacity to arouse great feelings and the fact that is a higly mediatic product. Pitts (1988) defines sport as "*any activity, experience, or enterprise for which the primary focus is fitness, recreation, athletics, and leisure related. Activity and experience are inclusive of the many athletics, fitness, recreation, and leisure-related activities of today: car racing, horse racing, boogey-boarding, knee-boarding, water skiing, golf, walking, camping, hang gliding, throw-ing the boomerang, horseback riding, participating in rodeos, sailing, and many more*". This definition is broader than those written before and it overcomes the concept of the physical activity and the competition. Considering these descriptions every person in the world, at least one time in their life, has played a sport. Every activity and sport needs products and services to satisfy the requests from the users. For this reason it is very useful to understand the nature of the sport industry better. Sport industry could be defined as: "*the market in which the products offered to its buyers are fitness, sport, recreation, and leisure related. These products include goods, services, people, places, and ideas. Sport industry products include, but are not necessarily limited to, the following: fitness activity and all fitness-related goods and services; sports activity and all sports-related goods and services; recreation and leisure activity and all recreation and leisure-related goods and services; and all related management, financial, marketing, and other administration and business goods and services*" (Pitts, 1988; Pitts, Fielding, & Miller, 1991). Starting from this point, it could be useful to understand the industry composition and segmentation following Porter (1985) better. According to him an industry is a market where there is an exchange of similar products and where products are sold to

customers. In addition to this, Porter (1985) gives the definition of industry segment. The segment could be understood like an identifiable component of the industry and could be identified by product, buyers, geographic location, channels or any mix of these variables. These theories have been applied also to the companies of the sport industry, as well as the other industry. An example of this is the study about the sport industry segment model made by Pitts, Fielding and Miller (1994). The purpose of their study was to develop a sport industry segmentation model based on Porter's industry segmentation. They chose to limit the segmentation to two of the four Porter's variables: product segments and buyer segments. The results of the study highlights three different segments, identifiable in Sport Performance Segment, the Sport Production Segment and the Sport Promotion Segment (Figure 1.1). For each of these segments they identified industrial and consumers goods buyer. Considering the product correlated to the sport performance segment, they identified two types of product that the buyer can "buy": the product for participants and the product for spectators. This main division could be divided in several varieties and they found seven types of varieties: athletics; private nonsport business, tax-supported sport organizations, membership supported sport organizations, nonprofit sport organizations, sport education and fitness and sport firms. For what concerns the buyer types they considered both industrial/commercial and consumer goods categories, for examples companies that contract for fitness/wellness programs for their employees; companies that purchase large numbers of sport event tickets; companies that buy the right to broadcast sporting events; and companies that buy materials to produce sporting goods for sale to retailers or another channel (Figure 1.1). The sport production segment is the one related to all the products that are necessary to play and practice sport or fitness. The varieties considered applying the Porter's variable relative to product are outfitting products (splitted in equipment and apparel) and performance production products. These two main varieties could be posed on a continuum based on the level of the product itself. For this reason they identified different levels for the variables like fitness trainer, medical care, sport facilities and governing bodies and officials. Also in this case the buyers are divided into consumers good buyer and commercial and industrial buyer. For what concerns the former they recognized sport participants in general and gift shoppers, the latter distinguished between high school and college athletic departments, professional sport

firms, tax-supported sport organizations, membership-supported sport organizations, nonprofit sport organizations, sport educators, fitness and sport firms, private nonsport firms offering sport as a secondary product, and private nonsport firms (Figure 1.1).

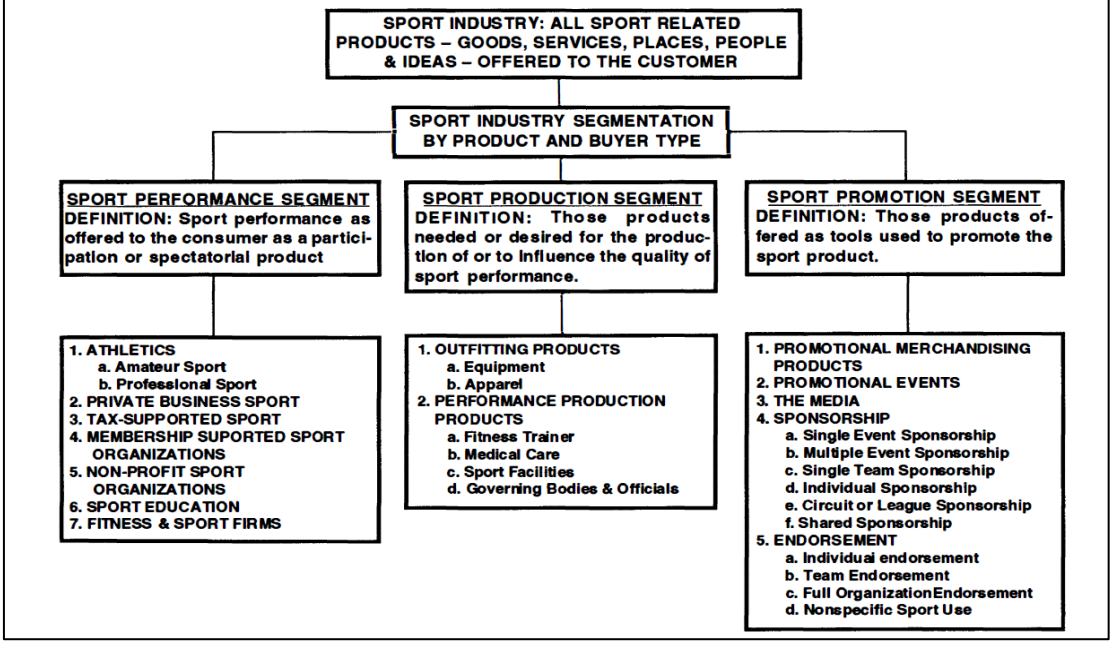

Figure 1.1: Sport industry. Source: Pitts; Fielding & Miller, 1994

Finally the sport promotion segment is about the promotion of the products and services of the sport industry. Sport exists and can be practiced also without this type of product but sales and fitness related activities are very influenced by these. The classes correlated to this segment are: promotional merchandising products, promotional events, the media (splitted in print media and audio/visual media), sponsorship (they distinguish different avenues like single event sponsorship, multiple event sponsorship, single team sponsorship, individual sponsorship, circuit or league sponsorship and shared sponsorship) and endorsement (for example individual endorsement, team endorsement, full organization, nonspecific-sport use). The division relative to the buyer is the same said for the segment before. Some examples are high school athletic departments, college athletic departments, professional sports organizations, private sport businesses, and tax-supported sport organizations for the commercial and industrial buyers, instead for the consumer good buyers there are gift buyers, collector buyers and "fan" buyers (Figure 1.1). In conclusion the authors provided some critics and recommendations about the model. First of all they considered only a part of Porter's theories. So a further development could be to make a segmentation relative

to the channels and the geographic location. Another step could be the use of different methodologies for the segmentation.

An example is Rohm's paper (2006) which proposed a mixed method approach to the market segmentation typologies. This study takes into consideration a very narrow segment of sport product, the running footwear market, but the methodology used could be applied on a broader base. It proposes an approach that includes multivariate statistical methods with qualitative data acquired with surveys (participation motivation data). The purpose is to understand better the motivations in sport activities to improve and to develop a deeper and effective segmentation of the market and of the customers. So the research proposes a segmenting approach which is not only based on a demographic background but also on a participation motivation data given by qualitative responses of an open-ended survey question. The mixed methods, according to Teddlie and Tashakkori (2003), are better than single methods in three main areas. First of all single methods cannot provide the same insights of mixed methods, moreover mixed methods provide stronger inferences and they capture a great diversity of respondent views. It is important for an industry like the sport one, to find more efficient way and methods for the segmentation. The companies which act in this industry, in order to have a real and sustainable competitive advantage, have to understand deeply which segment they are playing in and which the consumer's motivations are. In this sense the comprehension of the consumer typology become fundamental. Stewart et al. (2003) in their critical review about sport consumer typology, claim that it is almost impossible to summarize the sport consumption with homogeneous traits (Table 1.1). On the other hand there is the necessity to build sport consumption models and the market segmentation will be based on them. There have been large trials for the classification of the sport consumers. The paper collects them on the basis of three different perspectives. The first is about dualistic models that make a comparison between a type of sport consumption and its opposite, the second is about groupages of consumers ranked on the basis of their sport commitment (Table 1.1). Finally the third deals with multidimensional approaches that combine factors like motives for consuming sport products; factors that give sport meaning; indicators of loyalty; levels of emotional attachment and frequency of game attendance. An example of dualistic model is Nash's

		NATURA DEL BENE E SERVIZIO CONSUMATO	
		Natura Sportiva	Natura non sportiva
TIPOLOGIA DI CONSUMATORE	Praticante sportivo	**1** Prodotti tecnici progettati per la pratica sportiva (abbigliamento e attrezzatura sportiva)	**2** Prodotti e servizi dell'indotto correlati alla pratica sportiva
	Spettatore sportivo	**3** Servizi relativi alla fruizione diretta o mediata dello spettacolo sportivo	**4** Prodotti e servizi dell'indotto associati alla fruizione dello spettacolo sportivo

Table 1.1: Type of customer and good in the sport industry. Source: Slides from the course "Economia delle imprese sportive", Mancin (2018)

study (2000) which makes a distinction between core fans, who have grown up with the game and have a strong emotional attachment to teams and the corporate fans who have used sport only like a way to consolidate their network. Instead the model theorized by Mullin, Hardy & Sutton (1993) poses all the sport consumers on a frequency escalator that goes from the highly committed consumers to the low commitment consumers. In the middle there are the moderately committed. For what concerns the third perspective, McDonald and Milne (1997) have created a model based on the relationship marketing theory. The consumers are ranked on two dimensions: their value to the team, so the LTV (lifetime value), and their level of commitment to the club or team, so the RRS (relative relation strenght). Starting from this they created four combinations relative to the leves of consumer's commitment and financial support. In general all these models highlight the fact that sport consumption has several cognitive, effective and behavioral components. Moreover loyalty and commitment govern and influence the active consumption.

The consumption is growing according to PwC's sport survey (2019) and to the Deloitte's sport industry starting lineup (2019). The average growth in the past 3-5 years was about 7,4% and in the next years will be around 6,4%. The top ten of sports in terms

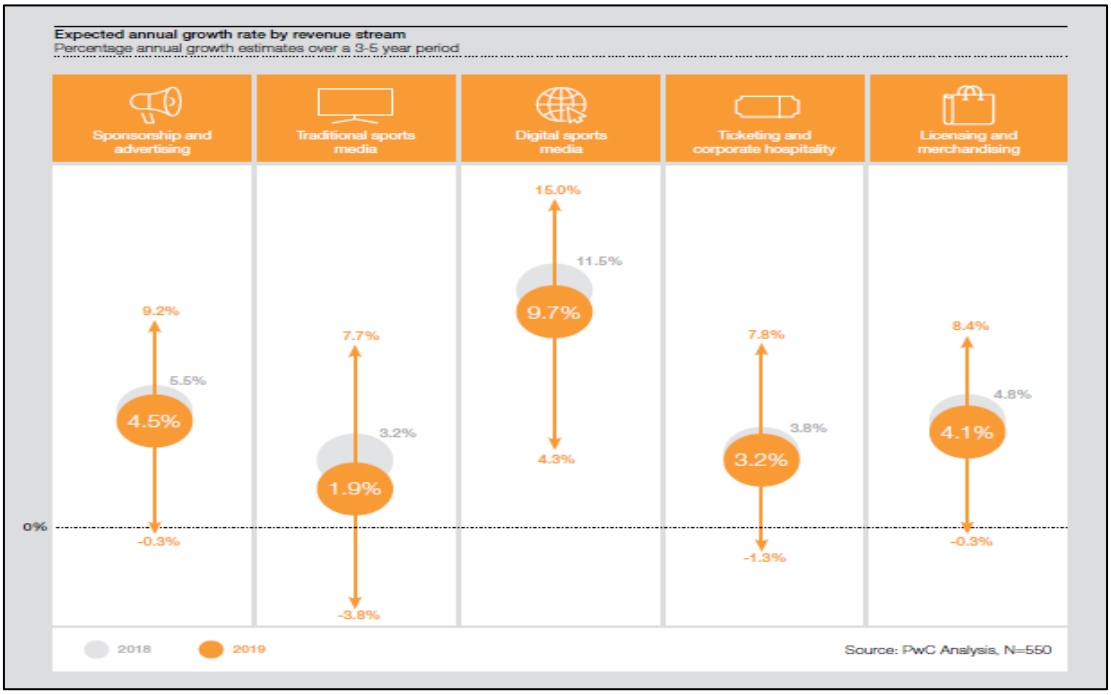

Figure 1.2: Principal revenue streams in the sport industry. Source: Pwc sport survey 2019

of potential for revenues growth is Esports in the first place, followed by football/soccer, then basketball, urban sport, tennis, rugby, golf, cycling, american football and cricket (Figure 1.2). In PwC's survey (2019), they asked to a selected group of industry leaders about the key opportunities and threats faced by the industry. Talking about growth there are three different patterns. Africa and South America are small markets in relative terms but experts are expecting the highest growth rates relative to the past 3-5 years; Europe, North America and the Austral-asia region are relatively saturated and they present slower growth rate than before. Finally Asia and Middle East are quite stable but anyway robust. In general the annual average expected growth is around 4,7%.

One of the main drivers of growth will be the digital transformation for example in the improvement for engagement, enhancement of media offering, improvement live event experience and synergies with gaming/exports. On the one hand digital transformation is a great opportunity, on the other one it presents threats but experts are very positive about opportunities. Another strong trend is the one about women's sport. In fact women represent the 70-80% of the consumer market and there is a very high potential for the audience of women's sports: 2019 was an incredible year in terms of audience and attendance. Also sponsorship deals are growing in terms of value. Innovation strategy, according to the leaders, is very important for the 79% of the respondents and

important for the 15%. This total high percentage is not reflected by a clear strategy. In fact the 46% of the respondents have concrete plans and the 29,8% is developing a strategy.

It is also important to try to classify the actors of the sport industry. For this aim Mancin (2018) uses two macrovariables: the business purpose divided into profit orientation and no profit orientation; and core business divided into sport and no sport. For what concerns the entities that are profit oriented with sport core business there are companies that produce goods and service for the sport activity and professional sport clubs and the new amateur companies with profit purpose. Instead, no profit entites with sport as core business, are amateur sport association, national and international sport federation and national and international olympic committees. On the contrary sponsors, official suppliers, other entities interested in visibility in sport and companies that produce goods and services for sport industries are actors that are profit oriented but their core business is not about sport. Finally, municipalities, tourist promotion agencies, onlus and foundations of various kinds are classified as no profit and with a no sport core business.

1.3. SPORT ORGANIZATION

The sport industry could be considered as a "regulated market" because it has to follow the commercial law rules common to every industry but there are also appropriate organs created for the management and the organization of the sport product (Mancin, 2018). This situation creates the sport competitive paradox. In fact in a common industry two firms compete for increasing their market share, so the competitor reduces the market penetration. In much sectors of the sport industry instead two firms collaborate to create a unique product and the competitor is essential. The product increases its value based on the strenght of the competitors (Mancin M., 2018). For example a match like Real Madrid against Barcelona is more valuable than a Serie B match.

There is a real "sport legal system" that is composed of several units that have legislative, executive and judiciary power inside the sorting. The italian law n°280/2003 confirms

the sport legal system autonomy from the repubblican one and it fixs its limits (Sandulli, 2005).

1.3.1. WORLD SPORT ORGANIZATION

The actual organization of the most important sport entities was born in 1896 with the creation of the International Olympic Committee for the first edition of the modern Olympic games, strongly desired by the Baron Pierre de Coubertin. All the governing bodies are "independent", in the sense that are elected by thier administrators and they are not nominated by politicians or other external entities. A very important feature is the absence of grants, differently from others extranational structures like ONU, UNESCO, OMS, FAO, ecc. (Colasante-Enciclopedia dello Sport, Treccani Website, 2003). All the purposes and the tasks must be realized through economic returns from their own activities. In general the world sport organization could be intended like a pyramid.

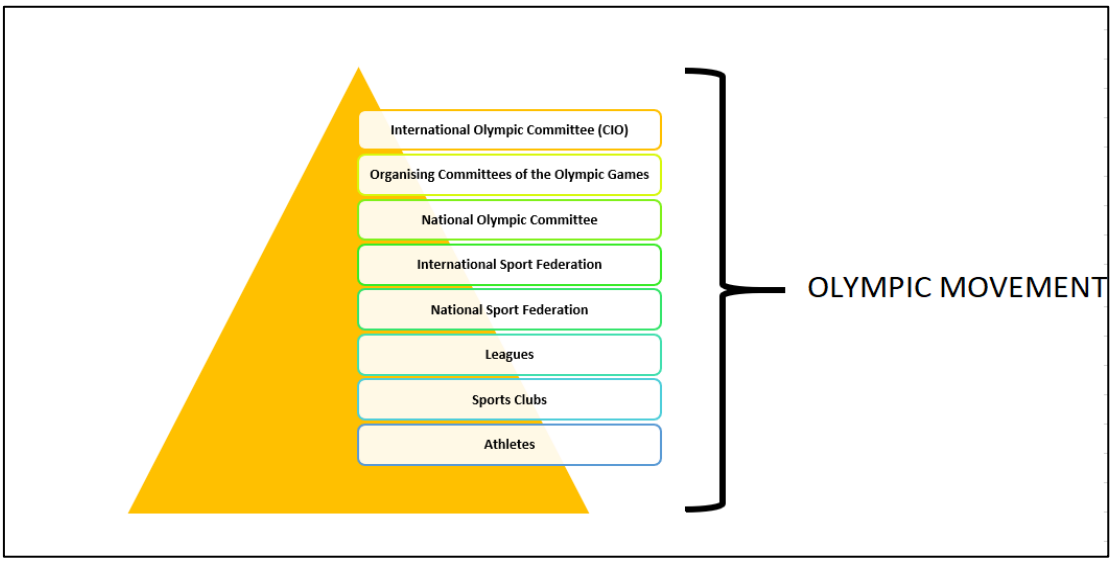

Figure 1.3: Sport Organization Pyramid. Source: personal rielaboration from Treccani website

The tip is the International Olympic Committee (CIO), then there are the Organising Committess of the Olympic Games, the National Olympic Committees, the International Sport Federation, the National Sport Federation, the Leagues, the Sports Club and the athletes. All these entities constitute the Olympic Movement (Colasante-Enciclopedia dello sport, Treccani Website, 2003; Ferrara, 2003).

The CIO was founded at the Sorbonne University in 1894. The entity is governed by a presidential office that is composed by the president, by the four vice-presidents and by ten members. For the realization of the purposes the CIO is constituted by commissions and the most important are: Athletes Commission; Olympic Games Coordination Commission; Olympic Programs Commission; Olympic Candidature Valuation Commission; Internet and TV rights Commission; Ethics Commission; Philatelic, Numismatic and "memorabilia" Commission; Marketing Commission; Medical Commission; Olympic Education and Culture Commission; Reform of the CIO Commission; Radio and Television Commission; Environment and Sport Commission; Sport for everyone Commission; Press Commission; Women and Sport workgroup and Olympic Solidarity (Olympic Games Website, 2019). The principal economic resources of the CIO are from the sell of the tv rights of the Olympic Games and income from sponsors.

The National Olympic Committees have the aim to spread the principles and the values of the Olympism in their country and to select the athletes for the Olympic Games. They are grouped in seven regional association rispectively: ANOCA for Africa; ODEPA for America; ODECABE for Centre-America; ODESUR for South-America; OCA for Asia; EOC for Europe and ONOC for Oceania (Olympic Games Website, 2019).

The International Federations are composed by all the National Federations that practise a particular sports discipline. In general they are non-governmental organizations, recognised by the CIO, that work for the promotion and the administration of a specific sports discipline. The International Federations are the entities which organize the world championship for a specific sport.

For having a better coordination, the International Federations are associated in Trade Associations. Some examples are the ASOIF (Association of Summer Olympic International Federations); AIOWF (Association of International Olympic Winter Sport Federations); ARISF (Association of IOC recognised International sports federation) and the GAISF (General Association of International Sports Federation). The CIO recognises also 27 International Federations which act outside the normal Olympic competition. Moreover there are others recognised entities, for example ICAS (International council of arbitration for sport); CIFP (International committee for fair play); IPC (International

paraolympic committee); WADA (World antidoping agency) and WOA (World Olypic Association) (Olympic Games Website, 2019).

1.3.2. SPORT ORGANIZATION IN ITALY

In Italy there is not a framework law which gives precise indications in terms of sport organization. In general CONI assumes the most central and important position. In the middle there are the national federation which are the link between CONI and sports clubs and associations (Colasante-Enciclopedia dello Sport, Treccani Website, 2003). Since the foundation of CONI with the law 426/42, it is considered as the "Federation of the Sports Federations" with the aim of "*incrementare e proteggere l'olimpismo e lo sport dilettantistico, nonché di incoraggiare e sviluppare l'educazione fisica, morale e culturale della gioventù del paese per migliorarne il carattere, la salute e il senso civico"* (L.426/42).

During the years CONI's organs have several modifications. In conclusion with the d.m. 28 December 2000 CONI was constituted by the National Council; the National Council; the President; the Secretary-General; "Comitato nazionale dello sport per tutti" and the Board of Auditors. The structure is completed by peripheral organs which are 20 regional committees; 104 provincial committees and "fiduciari locali".

Besides CONI there are the sport national federations which are constituted by sports clubs, by sports associations and by individual members. The Federations' economic resources are in part from government grants given by CONI and in part from self-financing derived from institutional activities.

Furthermore CONI recognises as sports structures also 16 "Associate disciplines". Finally the sport national scenario is completed by 14 "Sports promotion bodies" which work mainly as part of social sport and "sport for everyone" (CONI website, 2019).

1.3.3. FOOTBALL ORGANIZATION IN THE WORLD

The entities responsibles of the football system can be classified in three levels: the world level; the continental level and the national level. The most important entity is FIFA (Federation Internationale de Football Association). Inside FIFA there is an organ called IFAB which has the function of drafting, innovation and interpretation of football rules. This function is practised in exclusivity, in fact the continental and national federations don't have this task.

For what concerns the continental level there are six confederations affiliated to FIFA (IFAB cited in Wikipedia Website, 2019):

- Asian Football Confederation (AFC) for Asia (with the exception of Armenia, Azerbaigian, Georgia, Israel, Kazakistan e Turkey) and Australia. It is responsible for the organization of AFC Asian Cup, AFC Champions League, AFC Cup.
- Confédération Africaine de Football (CAF) for Africa. It is responsible for the organization of CAF: Cup of African nations, CAF Champions League, CAF Confederation's cup, CAF Supercup.
- Confederation of North, Central American and Caribbean Association Football (CONCACAF) for North and Central America, Caribbeans, Guyana, French Guyana and Suriname. It is responsible for the organization of the CONCACAF Gold Cup and CONCACAF Champions League.
- Confederación Sudamericana de Fútbol (CONMEBOL) for South America with the exception of Guyana, French Guyana and Suriname. It is responsible for the organization of Copa America, Copa Libertadores, Copa Sudamericana and Recopa Sudamericana.
- Oceania Football Confederation (OFC) for Oceania with the exception of Australia. It is responsible for the organization of Oceania Cup and OFC Champions League.
- Union of European Football Associations (UEFA) for Europe and Armenia, Azerbaigian, Georgia, Israel, Kazakistan and Turkey. It is responsible for the organization of UEFA Europian championship, UEFA Champions League, UEFA Europa League, UEFA Supercup.

Each continental federation is constituted of several national federations which organize and handle the competitions of their national territory. For what concerns national competitions for clubs, there are usually two main tournaments namely the national championship and the national cup.

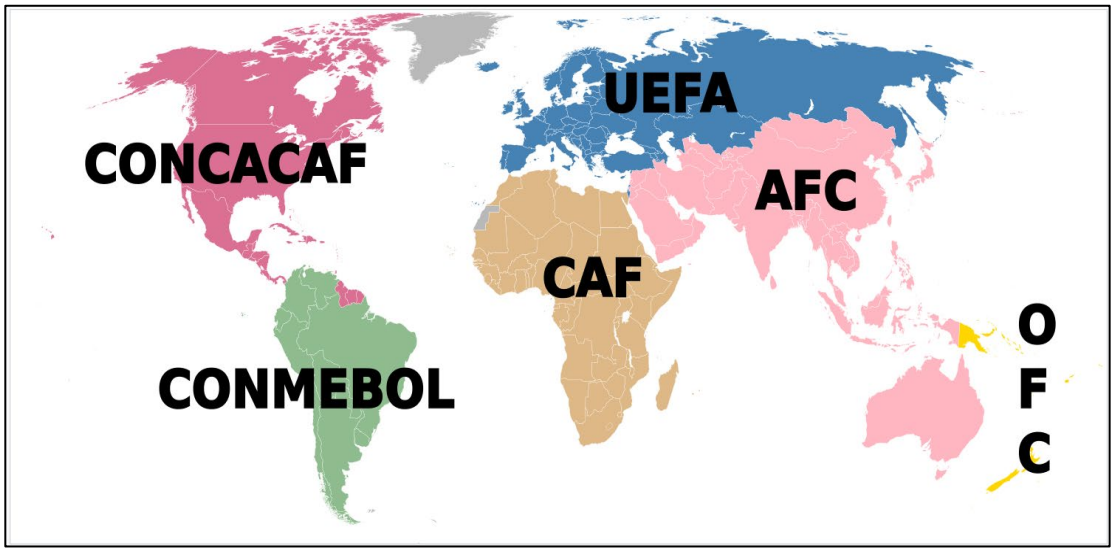

Figure 1.4: World map of football confederations. Source: Wikipedia website

1.3.4. FOOTBALL ORGANIZATION IN ITALY

The most important italian football organ is the FIGC, founded in 1898 and it was recognised by FIFA in 1905 and it is a founding member of UEFA in 1954. The second article of the FIGC's statute stipulates that: "*La FIGC è l'associazione delle società e delle associazioni sportive (le "società") che perseguono il fine di praticare il giuoco del calcio in Italia e degli altri organismi a essa affiliati che svolgono attività strumentali al perseguimento di tale fine. I regolamenti federali disciplinano il tesseramento degli atleti, dei tecnici, degli ufficiali di gara, dei dirigenti e degli altri soggetti dell'ordinamento federale.*" (Statuto Federale FIGC, 2019).

The main purposes of FIGC are:

- To promote and discipline the football game and all the aspects that are correlated, reconciling the professional and amateur dimensions through a centralized structure (FIGC Website, 2019).

- To promote the exclusion of every form of social discrimination, racism, xenophobia and violence from football (FIGC Website, 2019).

Figure 1.5: Leagues and associations that compose the FIGC. Source: FIGC website

The FIGC is composed of 4 main types of Leagues. In general Leagues have as principal aim to organize the competitive activity of their associates through the management of the events and the competitions. In addition to this they have an administrative and economic role. These two functions became more and more important with the development of sport as a "business" (Sanino, 2005). So it is possible to distinguish the Professional Championship Leagues namely "Lega Serie A"; "Lega Serie B" and "Lega Pro"; and the Amateur Championship Leagues namely "Lega nazionale dilettanti". There are the Technical Components namely "Associazione Italiana Calciatori" and "Associazione italiana allenatori calcio"; there there is AIA ("Associazione italiana arbitri") and finally the "settori", which are already included in the federal structure that are "Settore tecnico" and "Settore giovanile scolastico".

1.4. SPORT IN ITALY

Sport can be considered one of the most important form of societal involvement for larger parts of European populations. For example in Italy, sport is the sector of civil society that involves the highest number of active person (Baglioni, 2011). Sport can be

observed from two points of view: a "quantitative" one and a "qualitative" one. For what concerns the latter, sport is considered as the sphere where people learn to know each other; to develop trust and loyalty and increase cooperative behaviour (Baglioni, 2011). On several occasion italian sport has been one of Italy's greatest success, one of the few occasions in which italians could freely unite and be proud (Martin, 2011).

Sport is not something that is linked only to recreation. The Dutch historian and linguist Johan Huizinga (1938) sustained that play was the primary formative element in human culture and the sport is more ancient than culture. In Italy sport could be seen as a combination of release from everyday life and order establisher. Sport, leisure and free time were sized upon by political opportunities. During Roman era athletic culture held great relevance for the preparation for war. On the other side the construction of amphitheatres, baths and "sport places" for the masses evidence a concept of leisure and free time. After the collapse of the Roman Empire, the next thrust for italian sport comes from the Reinassance period, during the fourteenth to sixteenth century. Sport and other forms of athletic spread all over the country, local celebrations or medieval pageants were born in the territory like the Siena's Palio horse race. Events primarly considered entertainments, become enduring events in the peninsula. Unable to stop Renaissance liberalism's encouragement of pleasure seeking among the public that began to increasingly drink, gamble and enjoy blood sports, Church resistance permits events only on holy days. The Protestan Reformation puts religion in a central position and it posed work on a higher level than leisure. With this new perspective, there was a reinterpretation of sport. The conception shifts from a frivolous pursuit to a useful activity that forged disciplined bodies and and minds dedicated to labour. During the period of Italian industrialization a huge numbers of people moved into city centres and factory jobs, so the need for discipline and labour grew. Responding to social inhustice and very long work schedule, pioneers proposed reform and increasingly unionized workforce asked for time off and for time leisure. Moreover there was the creation of factory based teams and activities organized by employees and the modification of the work week. These conditions boosted sport enormously. After the unification in 1861, there was a strong increase in sporting festivals, events and activities. Sport had and has a fundamental role in the standardization of life and the forging of italians. For example

physical education was a very important piece in the creation of Fascist bodies. In fact during the Fascist period, patriotism was linked with gymnastic, men were to go, fit and disciplied, to the battlefield to defend their country (McCarthy, 2000). In general sport has always been an element of italian life. This is demonstrated also by the use of football for politics with Silvio Berlusconi (Martin, 2011).

As mentioned before, sport is a social and economic phenomenon (Mancin, 2018). So it could be useful to analyse data about the two aspects in Italy.

1.4.1. SOCIAL DIMENSION OF SPORT IN ITALY

The ISTAT's investigation called "Aspetti della vita quotidiana 2017", analyze through a sample survey the information about citizens' and families' daily life in Italy. This report divides the population into 4 macrorange for what concerns sport activities. These 4 range are: people that practise sport continually; people that practise sport occasionally; people that practise sport only few times; people that don't practise sport. Considering the percentages in 2017, the 24,8% of the sample declares that the do sport continually, the 9,1% in an occasionally way, the 27,6% only few times and the 38,1% don't practise sport at all(Figure 1.6). This last percentage is at all-time low since 2001, with a

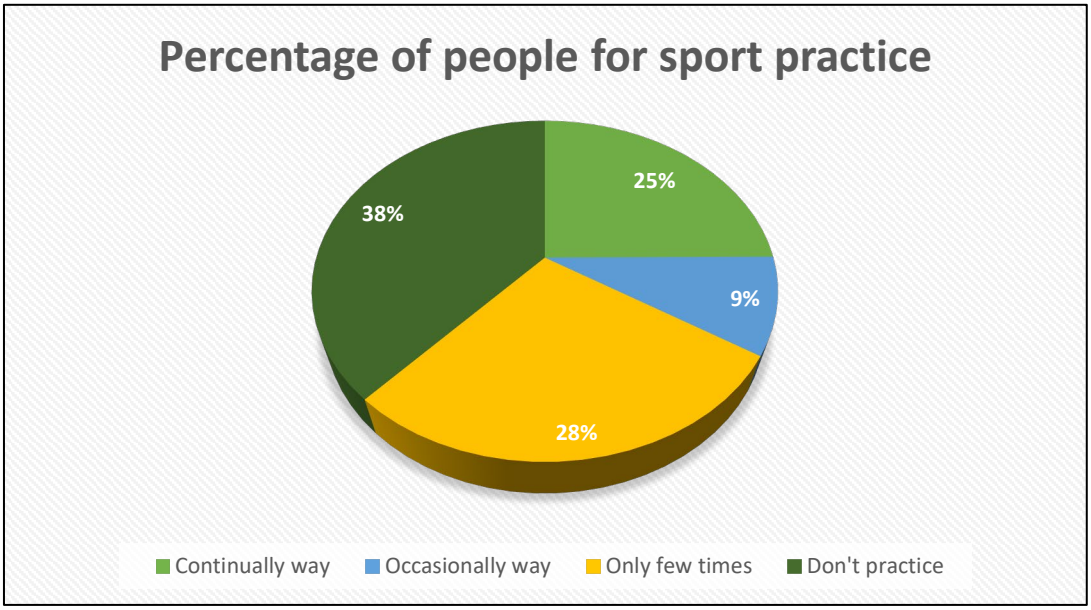

Figure 1.6: Percentage of people for sport practice. Source: "I numeri dello sport italiano", CONI 2017

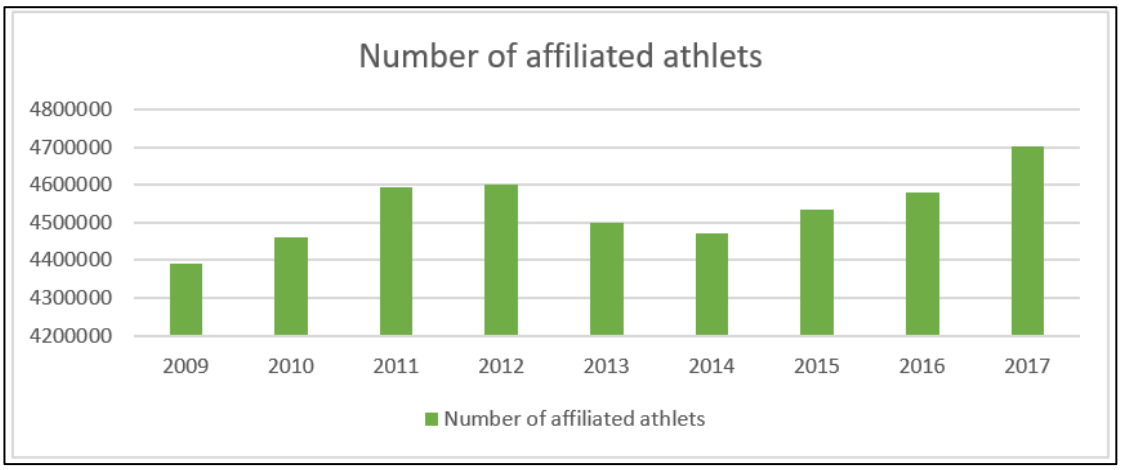

Figure 1.7: Number of affiliated athlets. Source: "I numeri dello sport italiano", CONI 2017

reduction of the 1,1% compared to 2016. To better understand the dimension of the sport phenomenon in Italy, we will take into consideration the report "I Numeri dello Sport Italiano 2017", realized by the "Centro studi e osservatoristatistici per lo sport della CONI servizi s.p.a." with the purpose to produce and spread high quality statistical information, relevant for the cuntry and for the European Union. The 2017 registered the highest number ever reached of athletes enrolled. They were 4 milion and 703 thousand spread in almost 70 thousand of associative centres (Figure 1.7). For what concerns the operators in federal and societal organizations, the number surpasses one million. In general the FSN ("Federazioni sportive nazionali") and DSA ("Discipline sportive associate") represent almost the 9,6% of the over 3 years old italian population. The affiliate athlets of FSN and DSA are increased in absolute values of 123307 units compared to the previous year, with an increase of the 2,7%. This growth is due to a continuos insertion of emerging disciplines. In this trend we can find also the sport movements generated by the new Olympic disciplines that will be present during Tokyo 2020. Some examples are climbing, karate, surfing, skateboarding and the return of baseball and softball. Moreover the projects that linked Federations and MIUR have facilitated the enrollment in federations of students that play particular sports disciplines at school. The presence of many sport events all over the country has increased the interest in sport, in particular for those disciplines that are less visible on a mediatic level. For what concerns the growth of "dirigenti societari" and technicians, they have increased in percentage for the 2,7% and the 2,2%. For the former this is due to a natural

increment of the clubs and the associations, for the latter to a growing necessity for qualified personal.

As said before there was an increase in the number of sport associations. Talking about numbers, there were 789 new affiliations. This could be correlated to economic facilitation for the clubs and with the policies of consolidatin of the more structured associations (Figure 1.8).

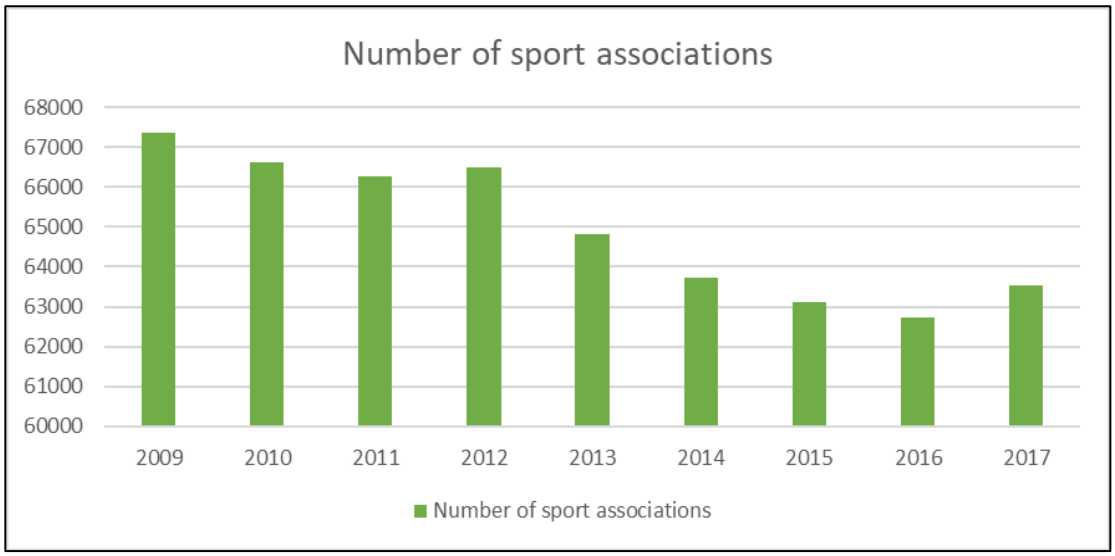

Figure 1.8: Number of sport associations. Source: "I numeri dello sport italiano", CONI 2017

Instead there is a drop about the 2% for the match officials and about the 1% for the federal managers. Probably these drops were caused by a renovation of umpire's sector for the former and by cuts to the territorial committees for the latter.

More in general if the time series are considered, it can be found a growth above 200 thousands of affiliated athletes between the 2013 and the 2017.

Instead if we take into consideration the period 2008-2017 the growth overcomes the 500 thousands units. All of this despite a geneal demographic trend that shows increasingly ageing and a diminishing birth.

Talking about specific sport disciplines, football is the most praticated sport with 1056824 athletes that correspond to the 23,8% of the total affiliated number. After football is tennis with 372964 atheles (8,4% of the total affiliated), then volleyball with 331843 players (7,5% of the total affiliated), basketball with the 7,1% and athletics with

6,1%. These five federations represent above the 50% of the affiliated while the first 10 federations represent the 70% of athletes. Nevertheless the less spread disciplines are growing in importance with an increase in the diversification of the sport offer and a decrease in the concentration of the practitioners.

About the number of sport associations, football is still the first with 12795 clubs. The second place is occupied by volleyball with 4390 clubs then cycling, basketball, tennis and martial arts with more than 3000 associations.

In Italy the sport diffusion is not omogeneous and this is influenced by the strong eterogeneity under the socio-economic point of view, the climate condition, the history and the territory conformation. Moreover the presence or the absence of adequate sport facilities influence the spread of particular disciplines. If we observe the regional ranking for the first 5 sport for numbers of athletes, it is clear that a region like Valle D'Aosta has in the first place winter sports and Liguria has sailing in the second place (Table 1.2).

In addition to this it is also important to analyse the demographic characteristics of athlets and operators. For what concerns the genre, it is evident that women are under-

Sport(N.Atleti) Regioni	1° sport	2° sport	3° sport	4° sport	5° sport
PIEMONTE	Calcio	Tennis	Atletica Leggera	Pallacanestro	Pallavolo
VALLE D'AOSTA	Sport Invernali	Calcio	Atletica Leggera	Tennis	Ciclismo
LOMBARDIA	Calcio	Pallacanestro	Pallavolo	Atletica Leggera	Pesca Sp. Att. Subacq.
LIGURIA	Calcio	Vela	Tennis	Pesca Sp. Att. Subacq.	Pallavolo
TRENTINO A.A.	Calcio	Sport Invernali	Tennis	Atletica Leggera	Pallavolo
P.A. Bolzano	*Calcio*	*Sport Invernali*	*Tennis*	*Sport del Ghiaccio*	*Atletica Leggera*
P.A. Trento	*Calcio*	*Sport Invernali*	*Pallavolo*	*Atletica Leggera*	*Tennis*
VENETO	Calcio	Pallavolo	Pallacanestro	Tennis	Atletica Leggera
FRIULIV.G.	Calcio	Pallacanestro	Vela	Tennis	Pallavolo
EMILIA ROMAGNA	Calcio	Tennis	Pallacanestro	Pallavolo	Atletica Leggera
TOSCANA	Calcio	Tennis	Pallavolo	Pallacanestro	Atletica Leggera
UMBRIA	Calcio	Tennis	Pallavolo	Pallacanestro	Atletica Leggera
MARCHE	Calcio	Pallavolo	Pallacanestro	Tennis	Ginnastica
LAZIO	Calcio	Tennis	Atletica Leggera	Nuoto	Pallavolo
ABRUZZO	Calcio	Tennis	Pallavolo	Pallacanestro	Vela
MOLISE	Calcio	Pentathlon Moderno	Badminton	Danza Sportiva	Tennis
CAMPANIA	Calcio	Pallacanestro	Pallavolo	Tennis	Atletica Leggera
PUGLIA	Calcio	Tennis	Pallavolo	Pallacanestro	Atletica Leggera
BASILICATA	Calcio	Badminton	Pallavolo	Tennis	Pallacanestro
CALABRIA	Calcio	Tennis	Badminton	Pallavolo	Pallacanestro
SICILIA	Calcio	Tennis	Pallavolo	Nuoto	Pallacanestro
SARDEGNA	Calcio	Pallavolo	Pallacanestro	Tennis	Atletica Leggera
ITALIA	Calcio	Tennis	Pallavolo	Pallacanestro	Atletica Leggera

Table 1.2: Sport diffusion in Italy. Source: "I numeri dello sport italiano", CONI 2017

represented. Considering the percentages for athletes, the 71,8% are men while the 28,2% are women (this trend a growing trend). If the data about "dirigenti societari", technicians, match officials and federals managers are observed, the percentage of women is respectively the 15,4%, the 19,8%, the 18,2% anad the 12,4 %. One of the causes for this gap can be reconciled with the less availability of free time for women that on average have 36 minutes less than men in terms of free time for the age group 3-24 and 55 minutes less for the age group 25-64. Moreover there is a general age advancement. In fact the 56,7% of affiliated is under 18 and the 33,6% is the age range 8-13. About integration, sport is an integration instrument. In fact the 2,3% of athletes are foreigners. This data will grow.

1.4.2. ECONOMIC DIMENSION OF SPORT IN ITALY

The sport industry, as said before (Pitts, 1988), is difficult to be defined and framed. Sport rapresents a very important industry the Italian economic system. In Italy there is not a statistical accounting like in other European countries. The point of reference for sport's value misurations is usually the "Vilnius definition" but it does not permit a fully fledged construction of a sport account because this will required an ATECO classification for at least six figures for more than 400 voices (CONI, 2012).

The "Vilnius definition" is articulated on three levels:

- Activities closely correlated with sport.
- Activities closely correlated with sport and all the activities that are necessary to produce sport (good and services that are necessary for the sport practice).
- Activities strictly correlated with sport and all the activities that are necessary to produce sport and all the activities that are correlated to sport but are not necessary for the production of sport.

The annual spending for sport practice is about 8,2 mld of euros, rispectively divided into clothes, shoes, tools and equipments for 3,8 mld; instructors, courses and trainers for 2,8 mld and 1,6 mld for facilities. The practitioners have a pro-capite spending about

442 euros per year. On average man spends 458 euros and a woman 424 euros (Mancin, 2018; Source: Sita Nielsen). According to Deloitte's estimates (2017) there are more than 35000 companies in the sport industry with more than 100000 workers. Talking about percentages the workers in the sport industry represent the 0,53% of the total, against an european average of 0,76%. Moreover taking into consideration the added value and the turnover, the estimates are about respectively 4,5 mld and 14 mld.

1.5. A FOCUS ON FOOTBALL BUSINESS

The study for sports and its market has grown significantly. Economists give more and more attention to sport for several recent developments. First of all the increased economic significance of sport in terms of value linked to the growth of litigations in matter relating to the organization of sport tournaments and the need for economic analysis of specific rules regulation. Moreover universities are finding a huge demand for teaching of sports economies. Finally the recognition that sports markets provide a number of natural experiments in theories of incentives and labour-market behaviour (Baroncelli, 2019).

One of the most important and popular sport in the world, both amateur and professional level, is football. Like every sport, football is characterized by three main elements that are the competition, the uncertainty of the result and the rules (Ciampaglia, 2013). For the preservation of these elements in football, as in the other sports, there are apposite organs, described previously. Football has changed deeply in the years and the growing number of matches and events caused an increase in the necessity of adequate number and quality of players and technicians. This drives the attention on the importance of the human resources and on their quality for the success and the performance (Pfeffer, 1998, cited in Ciampaglia, 2013).

This is because football clubs in general, whether large or small, are characterised by the cycle shown in Figure 1.9. In fact, through the financial resources available, the manager for the selection of players must choose the best players that can make the team obtain the best possible result in order to obtain higher revenues and increase the value of the

players in order to obtain new financial resources to restart the cycle and possibly make profit. The focus of this research is on the players and their selection to allow the cycle to be a virtuous cycle and not a failure.

Despite the great attention, clubs have failed to generate an adequate level of income. Moreover the competitive success doesn't always mean economic success.

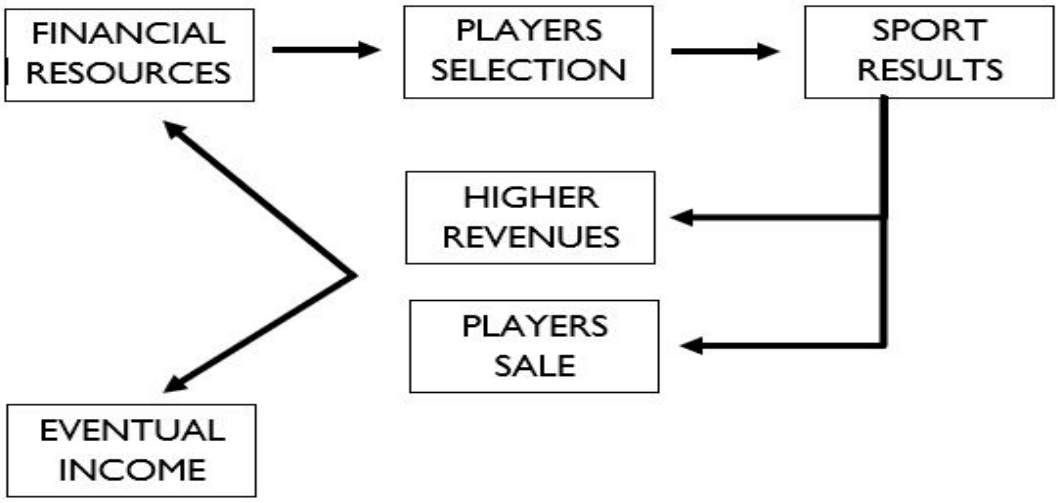

Figure 1.9: Economic virtuos cycle of professional football team. Source: personal rielaboration of Lago et al. (2004)

For what concerns football, there is not only the sport event in itself but there are income sources that have to be managed through managerial politics, business plans, the employment of the best human resources, economics resources with the aim to obtain the maximum possible with the minimum use of resources (Sorci, 2002, quoted in Ciampaglia, 2013). It is important to make managerial choices that can reach the goals using the principle of the affordability namely the idea of development and lasting life of the club. So it is necessary to assicurate the survival of the team in the short and in the long period with the aim to obtain results (Lago, Baroncelli, Szymansky, 2004).

In this environment the role of HR has grown exponentially, so every choice is not influenced only by economical reasons but there are also social purposes. This explains why football has received large investments without having an adequate flow of income from the typical items of income.

Football clubs can be divided into amateur football clubs and professional football clubs. The formers' mission is to promote the sport practice and the latters have as typical

activity the show generated by the sport events. According to Lago; Baroncelli and Szymansky (2004) the fundamental elements of the football business is the entertainment (product) sold to supporters or other spectators, provided by players, trainers, technicians, ecc., that use play fields by utilizing appropriate equipment for the creation of a competitive environment between competitors.

Taking into consideration the main types of revenues for a football club, according to Sener and Karapolatgil (2015) are from matchday (including ticket and corporate hospitality sales), broadcast rights (including distributions from participation in domestic leagues, cups and European club competitions) and commercial sources (including sponsorship, merchandising and other commercial operations). These categorization is also used by Deloitte for their report "Football Money League". We will discuss about the value and the distribution of these income flows later in this chapter.

In general the company's activity of the football clubs is composed by several subgroups that refer to different stakeholders that could be divided in: actors of the competitive system and social actors that bring interests. (Ciampaglia, 2013). Applying Porter's five forces model (already discussed in the first paragraph) to the football clubs we can distinguish as direct competitors the other football teams; as potential entrants the teams that move from a competitive environment to another for sport merits and demerits; as substitute products all the leisure and entertainments activities; as clients the spectators, the people that buy merchandising items, television networks and sponsors; as suppliers the other clubs that sell players, players' managers, technical equipment providers and services suppliers. All these stakeholders can be posed on four different dimensions (Ciampaglia, 2013):

- The sport dimension
- The economic dimension
- The social dimension
- The communication dimension

The nature of the differentiation of the revenues streams and the stakeholders is linked to the transformations that football clubs have had in the years. We can distinguish various phases in the transformations (Mancin M, 2018):

- '60-'70: Patronage → Entrepreneur's investments for personal image returns
- '80- half '90: Sponsorship
- Half '90-'00: Sponsorship and TV rights/Broadcasting
- '00: Less public resources and Broadcasting revenues boom

We will see the legislative transformation of football club later in this chapter.

Business Model Typologies: Expected and Resulting Value Creation and Appropriation

Internal Accumulated Shared Experience	External Market Value of Talent: Low	External Market Value of Talent: High
Low	**Business Model 1:** 'B' Team Talent Low Shared Team Experience **Performance Expectations:** Theory • Low value creation • Low value appropriation Results: • Low value creation • Moderate value appropriation	**Business Model 4:** 'A' Team Talent Low Shared Team Experience **Performance Expectations:** Theory • Moderate value creation • Low value appropriation Results: • Moderate value creation • Very low value appropriation
High	**Business Model 2:** 'B' Team Talent High Shared Team Experience **Performance Expectations:** Theory • Moderate value creation • High value appropriation Results: • Low value creation • High value appropriation	**Business Model 3:** 'A' Team Talent High Shared Team Experience **Performance Expectations:** Theory • High value creation • High value appropriation Results: • High value creation • High value appropriation

Table 1.3: Business model typologies. Source: McNamara; Peck; Sasson, 2011

In general the professional sport companies have to face a very complex business model with the management of the different types of stakeholders that are private, public and institutional. Four recurrent types can be recognized, all characterized by the competitive results used to gain the profit result. These are the clubs as a passion which is characterized by the presence of a patron that invests in the club for passion or for philanthropy; the club as an image vehicle where the shareholders invest in the club with the purpose to obtain greater visibility; the club as a place where enhance new talents where the shareholders search for young talents with the purpose to sell them to other clubs and searching for profit; the club as entertainment company where the shareholders try to reach an economic balance between revenues and costs (Mancin, 2018). Football clubs' business models can be analyzed from different points of view. An

example could be the study conducted by McNamara, Peck and Sasson (2011) where they analyzed the English Premier League environment for the definition of four business models based on two variables: the talent of the team's players and the shared team experience accumulated (Table 1.3). In this research the empirical question is whether more than one stable business model configuration can exist within an industry. They consider the stability of a business model in terms of its ability to generate both value for the customer and adequate financial returns for the firm. They demonstrated that some resource and capability configurations of talent and accumulated shared team experience are associated with both success on the field of play (which in this case can be rapresented by value creation for customers) and higher financial performance relative to the industry for the firm (i.e., value capture).

Surely the English Premier League is the most important league in the world and this is not occured by chance. In fact the English football system has experienced a long process of reorganization and modernation. All started with the document called "Taylor Report" which contains 43 reccomendations, proposed by Lord Taylor, for the stadia safety and enhance the economic value of the sport event (Tirrito, 2005). In particular, the focus of the English reformation was the rebuild of stadia with the purpose to having an aggregation place for everyone not only for the football event but for every leisure activity that produce flows of revenue. For this reason the footbal clubs started to be the owners of the stadia. So the english clubs can rely on income from the management of all the stadia's related activities which add to the broadcasting rights, the sponsors and merchandising. One example of this virtuous control is Manchester United which excellently exploit the economic possibilities derived from the ownership of the stadium (Old Trafford) and the brand awareness (Tirrito, 2005). The Manchester United case underlines the importance of the brand for a football club and for their strategy in the football industry.

For this reason Sener and Karapolatgil (2015) indentified among the 50 global football club brands the strategic groups and the common main strategies of each group. They found three different strategic groups with the members of each group following similar strategies. Moreover they demonstrated that brand value forms a mobility barrier between the movement from a strategic group to another. So the football industry

shows eterogeneity and as a result of the analysis, 50 football clubs were classified as industry-leaders, runner-ups and weak clubs. When the clubs within each cluster are ranked according to their brand value, there exists a clear distinction among three groups. There are 7 football clubs classified as industry leaders. The total brand value of these clubs account for 45% of the whole sample. The main strategy followed by these clubs is identified as offensive strategy. In general, the main aim is to win as much as possible in each field. The clubs generate high revenues from broadcast and matchdays, their stadiums are fully crowded and they also earn from special tournaments and training camps all around the world. Besides, they have high valued contracts with sponsors and shirt manufacturers, many companies compete with each other to be a sponsor for these clubs or for an advertisement at the stadium; these clubs have easily recognizable logos, symbols and nicknames, and museums as well that defined as touristic spots. Then there are 12 football clubs classified as runner-up clubs. These clubs as a common strategy try to reach to the levels of industry-leaders and strengthen themselves in order not to be weakened. All of the runner-up clubs follow distinctive image strategy. Finally Most of the sample (31 football clubs) is classified as weak clubs. These clubs as a common strategy follow defense strategy. Another interesting finding of this study is relative to the fact that being a domestic industrial leader does not make sense globally if the country does not have a high profitable football industry. Even a weak brand in a successful system has more brand value than industrial leaders of countries that have a weak football industry.

1.5.1. HISTORY OF FOOTBALL BUSINESS

The steps that bring to the actual nature of the football clubs see the constant presence of a financial-economic precarious situation with a ethic crysis of the actors and deep changes in the law about the relationship between the professional sportman and the club (Mancin, 2018). The development of the historical analysis of the football club could be divided in three main periods (Mancin, 2018):

- First period: '60-'70 → the growth of the economical dimension of the sport activity

- Second period: '80-half '90 → the first law on the professional sport
- Third period: half '90-actual → Bosman sentence, tv rights and financial fair play

1.5.1.1. FIRST PERIOD

Before this period the last law about sport topic was the law n. 426/42 which created CONI and defined the sport companies ad association without profit making purpose (Barbieri, 2005). So the football clubs were simply associations. This period was characterized by the growth of the technical quality of the competitions and by the birth of international competitions (Mancin, 2018). Moreover the media started to be strongly interested in football and this encouraged the public's interest (Ranieri, 2005). The football associations saw a strong presence of patronages that increase investments searching for the sport success to have an image return. With the growth of the economic dimension there was a transformation of the professional sportman. In fact the practitioner-affiliated became a professional athlet (de facto but not de iure). In the same time this transformation was accompanied by financial deficits and growing indebtness (Mancin, 2018). In this context the associative form appeared inappropriate. For this reason the FIGC in 1966 adopted two deliberies (Barbieri, 2005; Carta, 2005):

- 16 September 1966 → dissolution of football associations and nomination of a extraordinary administrator for each team
- 21 December → approvement and compulsory nature of a "standard statute"

De facto there has been a transformation from football association to S.P.A. with some specific limitations. For what concerns the social object there was the mandatory sport purpose, there was the imposition of the absence subject profit-making, the presence between the associates of external representatives and external control by management company (Barbieri, 2005; Mancin, 2018).

The aim of these procedures was to impose the football clubs control systems adequate to the size of the business.

1.5.1.2. SECOND PERIOD

Despite the above indications, the world of professional football was experiencing serious legal uncertainty (Carta, 2005). In the meantime, the dynamics above listed continued to develop (increasing economic value and indebtedness). A change in this direction occurred in 1981 with the law n. 91/81. This law deals in particular with:

- the form of professional football clubs and corporate purpose
- the profit-making aim
- working relations with professional sportsmen
- federal control system

This law required that the legal form of professional football clubs be the S.P.A or the S.R.L. and that the transformation should take place within one year. The 1966 directives of the FIGC were transposed, requiring teams to carry out a series of checks that are even more intrusive than those intended for companies with share capital. The rule also provides for the definition of a professional football club, i.e. a club that has contractual relations with professional sportsmen. In addition, the memorandum of association must stipulate that the club may only carry out sporting activities and related or instrumental activities.

As regards profit-making, law n. 91/81 fully reiterates the prohibition on the distribution of profits made. The latter are to be entirely reinvested by the clubs in the exclusive pursuit of sporting activity.

Another important change concerns working relations with professional sportsmen. First of all, the nature of the employment relationship was outlined as an employment contract. The contract may provide for a termination period not exceeding 5 years with the possibility of successive contracts. The contract may be transferred before it expires between several companies. Certainly one of the most important changes made by the law was the abolition of the "vincolo sportivo" through which it meant the exclusive relationship that linked the player to the club. In fact, Article 4(6) prohibits any clause of non-competition or restriction of professional freedom. Moreover art. 6 of the law established the so-called “indennità di preparazione e promozione” (“IPP”). When a

player moved from one club to another as a result of the expiration of a contract, the former was entitled to a consideration calculated by averaging the gross annual emoluments of the last two seasons multiplied by a coefficient influenced by age and the category of destination. The right to recognition of the "IPP" could be exercised until the end of the second year following the expiry of the contract. In practice, the IPP was a continuation of the "vincolo sportivo" but with a limit in terms of duration and maximum limit of the value (Mancin, 2009). In general, as a result of this rule, there has been an increase in the contractual power of players, which has also led to an increase in labour costs, which has been one of the major problems for the financial stability of clubs (Mancin, 2009).

Finally there was an extension of controls to investment transactions about real estate and extraordinary administration and the federation, for serious management irregularities, can ask the court to put the company into liquidation.

1.5.1.3. THIRD PERIOD

This period is characterized by three elements that have profoundly changed the business of football:

- The Bosman sentence
- The advent of pay-per-view
- Financial Fair Play (FFP)

The first two elements are located further back in time than the FFP. Marc Bosman was a footballer who, once concluded his contract with the "Royal Football Club de Liege", wanted to move to the French team Dunkerque. His old team had forced him to choose between stopping playing or continuing to play for them under unfavourable conditions. For this reason Bosman appealed to the court involving the club itself, the Federation and UEFA. On 15 December 1995, the European Court of Justice in Luxembourg abolished the transfer system and the 3+2 rule (Késsene, 2006). All this with the sole interpretation of Article 48 of the EEC Treaty. The Bosman judgment calls into question all the rules relating to the possibility for one team to demand from another, a payment

for the services of a player with an expired contract and all the rules imposed by the federations to limit the participation of foreign players in EU competitions (3+2 rule) (Mancin, 2009). Obviously this ruling has had serious effects on the entire football system. First of all, it has severely unbalanced the contractual power between players and clubs in favour of the former (Lago; Baroncelli; Szymanski, 2004).

In addition, very often in Europe many football teams marked the transfer value of their players as an asset in their balance sheets and used those assets as collateral with banks for loans (Késsene, 2006). For this reason many clubs lost a lot of money for their previous investments. Another effect was the lengthening of the contracts, this is because the clubs were allowed to buy and sell players before the end of the contract. The most dramatic impact of the Bosman verdict was the increase in international mobility of free agent players in Europe (Maguire & Stead, 1998, cited in Késsene, 2006).

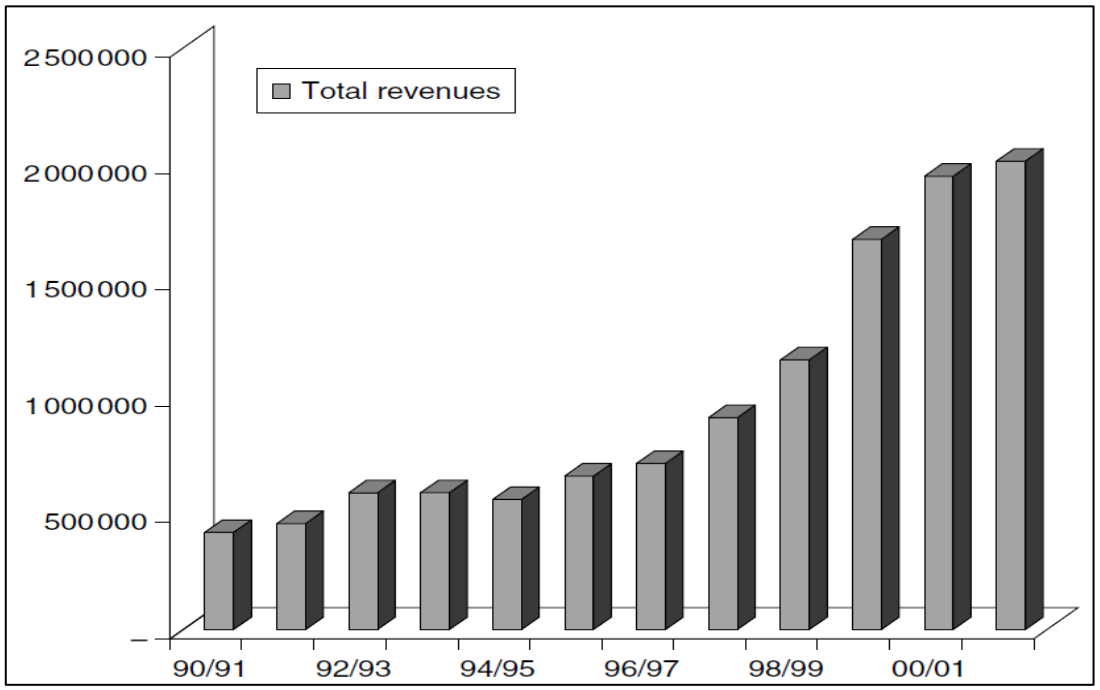

Figure 1.10: Revenues of Serie A football clubs, 1990/91 to 2001/02 (€000). Source: Lago, 2006

Finally, all these elements led to a trend towards higher wages and increasing competition for the best players. The second characteristic element of this period, the advent of pay-TV, also contributed to this. In the period 1990-2000 the revenues of football clubs grew at a rate never seen before thanks to the advent of pay-TV (Lago, 2006. Figure 1.10). For the reasons explained above, there is also a significant increase

in costs during the same period, in particular from the 1996/97 season, the season following the Bosman judgment. As regards Italy, there was a significant increase with Law 78/99, which allowed the subjective negotiation of television rights. Thanks to this circle of increase in revenues and consequent increase in investment costs, the Italian teams hoped to start a virtuous circle of sporting success accompanied by economic success, but this was not the case (Lago, 2006).

	95–96	96–97	97–98	98–99	99–00	00–01
England	47	47	51	58	62	60
Italy	57	58	64	72	62	75
Spain	53	44	53	56	–	–
Germany	–	46	51	51	54	50
France	58	60	69	69	57	64

Table 1.4: Ratio wages/turnover in major Europeans Leagues (€m, %). Source: Lago, 2006

During the 2001/02 season, the so-called "bursting" of the speculative bubble led to the enactment of two emergency measures, the "legge anti-insolvenza" and the "decreto salva-calcio". To mention is the failure of Fiorentina in 2001 that could not enroll in the championship of Serie B. As regards the relevant Italian laws, reference should be made to Law No 586/96. This law transposes Decree-Law No 485/96 and amends certain points of it. An important change that led to this law was the elimination of the ban on subjective profit with only the obligation to reinvest 10% in schools and youth sectors upgrade (Carta, 2005). In addition, taking up the suggestion made by the European legislator in the Bosman judgment, it repealed the "IPP" by effectively eliminating a claim in the companies' balance sheets and creating a negative effect on the profit and loss account (Carta, 2006).

Finally, the most recent legislative revolution in the world of football took place with the Financial Fair Play (FFP). Its main function is to implement measures to ensure the long-term economic and financial stability of football (Benoldi and Sottoriva, 2011). It is based on well-defined quality standards to be met in order to gain access to UEFA competitions. The specific criteria are 36 and can be divided into 5 macro-areas:

- Sports area

- Infrastructural area
- Of personnel area
- Legal area
- Financial area

Each criterion is distinguished by a letter (A, B or C) according to the level of nature that the constraint imposes. In general, the FFP requires clubs not to spend more than they earn, no old liabilities during the season are allowed, and greater financial transparency on the part of the clubs (Benoldi and Sottoriva, 2011).

1.6. FOOTBALL IN EUROPE

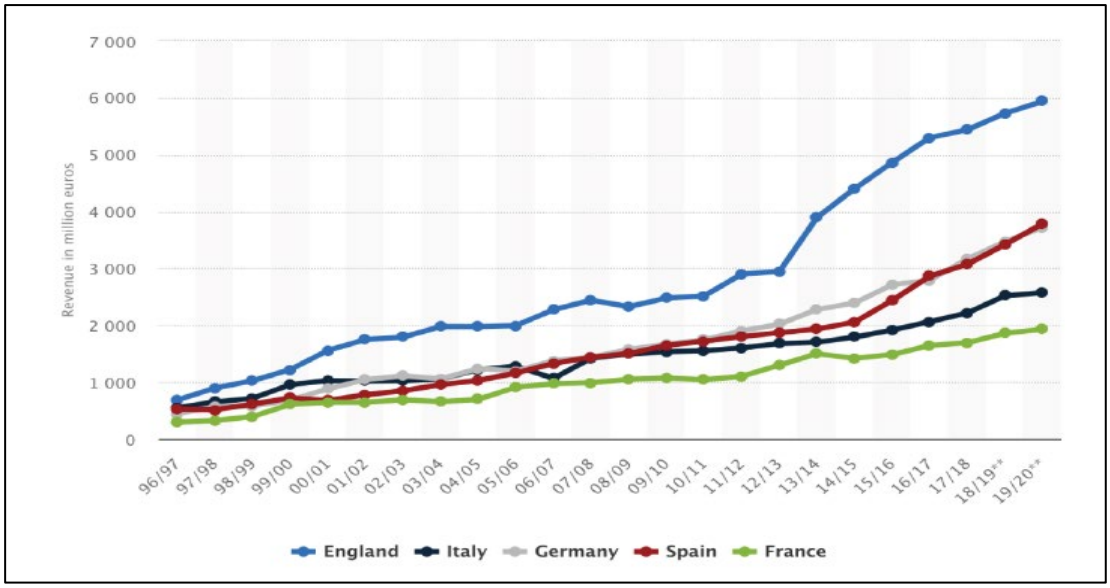

Figure 1.11: Revenues of "Big Five" football leagues from 1996/97 to 2019/20, Source: Statista 2019

In Europe there are 36 professional leagues and associations of clubs with 990 professional football clubs representing 29 European countries in which there are 58 football divisons (European Leagues Website, 2019). The overall size of European football continues to grow in terms of revenues, reaching 28.4 billion in 2017/18 (Deloitte, 2019). European leagues have seen strong growth in terms of revenues in the last 10 years mainly due to an increase in the value of broadcasting agreements (Teodor, 2015; Deloitte, 2019). European leagues have seen strong growth in terms of revenues

in the last 10 years mainly due to an increase in the value of broadcasting agreements (Teodor, 2015; Deloitte, 2019. Figure 1.11).

The leagues present in Europe can be divided into two groups, the so-called "Big Five" composed of England, Germany, Spain, Italy and France and a second group composed of all the others. Looking at the european football market size it can be seen how between the season 2016/17 and 2017/18 there was an aggregate increase of 2.9 billion euros (Figure 1.12). As for the "Big Five", the leading nation is England (5440 million), followed by Germany (3168 million), Spain (3073 million), Italy (2217 million) and finally France (1692 million) (Figure 1.13). The 2017/18 season saw an increase in revenues for

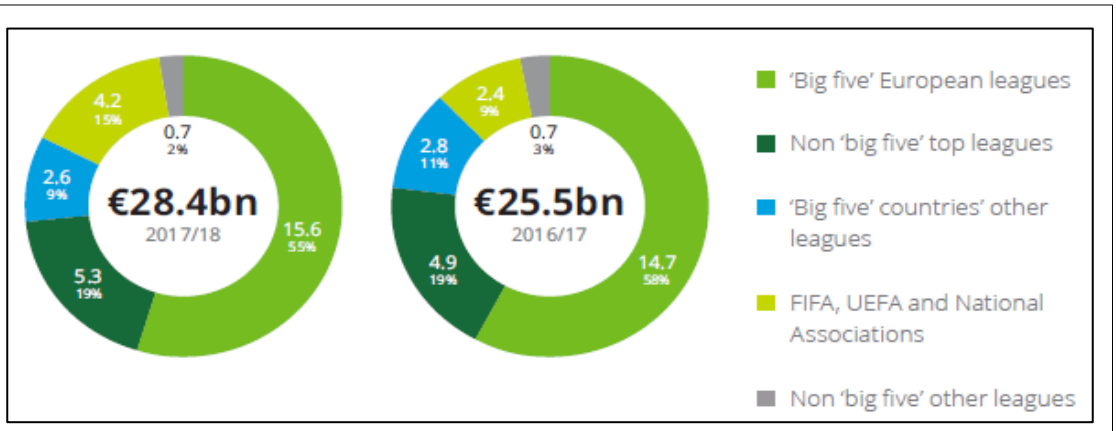

Figure 1.12: European football market size- 2016/17 -2018/19 season. Source: Annual Review of Footbal Finance, Deloitte, 2019

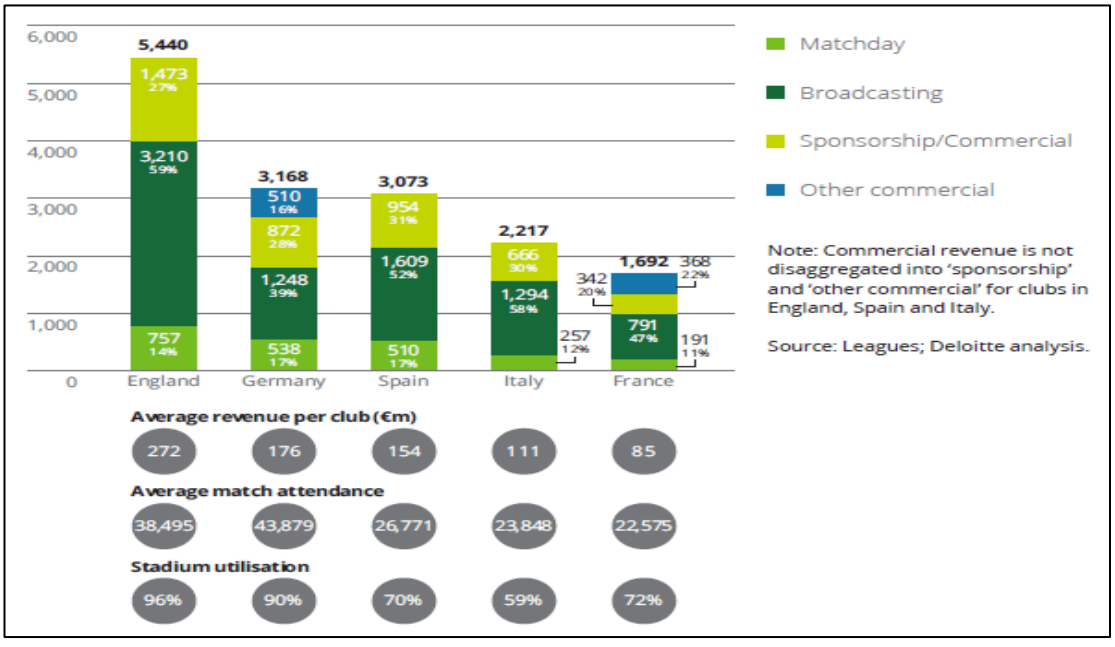

Figure 1.13: "Big Five" European league clubs revenues-2017/18 (€mln). Source: Annual Review of Football Finance Deloitte, 2019

La Liga of 7% and for the Bundesliga of 13% against 6% in the Premier League. This allowed the two leagues to narrow the gap with Premier. The Bundesliga has seen such strong growth thanks to a new agreement for television rights that has also allowed it to overtake the Spanish league in terms of revenue. This is despite the fact that, according to Deloitte Football Money League (2019), Real Madrid and Barcelona are the clubs respectively in first and second place in the ranking of the top clubs in terms of revenues. The growth of the Italian league was around 8%, although for the first time in the history of the Deloitte Football Money League Report, no Italian club is in the top 10. This growth is due, among other things, to a strong increase in commercial revenues for Internazionale and to the increase in revenues from television rights of AS Roma following the achievement of the Champions League semi-final. Another element that contributed was the growth in matchday revenues driven by the Milanese clubs and AS Roma. Finally, France is the last of the "Big Five", with a decrease in terms of revenues from broadcast and sponsorship but an increase in revenues for matchday and other commercial income.

Taking into consideration other European championships than those belonging to the "Big Five", the Annual Review of Football Finance Deloitte (2019) places Russia, Turkey, the Netherlands, Portugal, Scotland, Denmark, Austria, Norway, Sweden and Poland in the top ten, respectively. The "first place" of Russia is certainly to be found in the effect that the World Cup organized at home have had on the entire Russian football system. This has enabled it to overcome Turkey, despite the fact that Turkey has negotiated a new agreement on television rights. The growth of Portugal and Austria thanks to the results obtained by their teams in the Champions League and Europa League are to be mentioned. It is precisely the international competitions in European football that play a fundamental role and have a huge influence on national championships. In fact, UEFA competitions have an impact on club budgets and competitive league balances (Bellia, 2019). This is because those who participate to the Champions League or the Europa League receive more money than those who do not participate to these competitions, causing some teams to become increasingly powerful economically. For example, for the 2018/19 season, a total of €2 billion of euros was distributed to clubs participating in the Champions League and €504 million of euros to those of the Europa League. In

particular, the UEFA awards (2018/19) for performances in the Champions League are as follows:

- The 32 teams in the group stage will receive a fixed payment of 15,3 € and 2,7 € mln per win and 900000 € per draw
- The teams competing in the round of 16 will receive 9.5 € mln
- The quarter-finalists will receive 10,5 € mln
- The semi-finalists will receive 12 € mln
- The winner will receive 19.5 € mln + 3.5 € mln Super Cup

Competitive balance is crucial for sport. For example, from a commercial point of view, it is established that the uncertainty of the result favours the show and the increase in revenues. For this reason the CIES in his 40th Monthly Report has analyzed 24 European competitions over a period of ten seasons between 2008/09 and 2017/18. To measure

	Total	2009-13	2014-18	Evolution
POR1	84.4%	85.1%	83.7%	-1.4%
ESP1	82.7%	84.6%	80.9%	-3.7%
SCO1	79.8%	77.0%	82.6%	+5.6%
GER1	79.3%	75.7%	82.9%	+7.2%
NED1	78.3%	76.7%	80.0%	+3.3%
ITA1	77.6%	73.5%	81.8%	+8.3%
ENG1	77.3%	76.1%	78.4%	+2.3%
FRA1	75.0%	70.0%	80.0%	+10.0%
TUR1	74.1%	73.4%	74.9%	+1.5%
BEL1	73.4%	75.7%	71.1%	-4.6%
NOR1	72.8%	71.1%	74.5%	+3.4%
SUI1	72.7%	70.8%	74.6%	+3.8%
DEN1	72.5%	71.9%	73.0%	+1.1%
RUS1	71.7%	71.1%	72.2%	+1.1%
SWE1	70.8%	69.3%	72.2%	+2.9%
FIN1	70.4%	70.7%	70.1%	-0.6%
AUT1	70.1%	67.8%	72.4%	+4.6%
GER2	67.7%	69.6%	65.9%	-3.7%
ENG2	67.7%	66.1%	69.3%	+3.2%
POL1	66.8%	68.4%	65.1%	-3.3%
ITA2	64.6%	63.8%	65.4%	+1.6%
FRA2	64.4%	61.6%	67.2%	+5.6%

Table 1.5: Average % of points of champions by league. Source: CIES

	Total	2009-13	2014-18	Evolution
UEFA CL	1.578	1.470	1.686	+0.216
NED1	1.554	1.616	1.492	-0.124
AUT1	1.502	1.483	1.521	+0.038
SUI1	1.452	1.473	1.432	-0.041
ESP1	1.452	1.429	1.474	+0.045
GER1	1.433	1.436	1.430	-0.006
ENG1	1.377	1.349	1.406	+0.057
BEL1	1.376	1.362	1.388	+0.026
SCO1	1.369	1.317	1.422	+0.105
NOR1	1.361	1.344	1.378	+0.034
SWE1	1.358	1.330	1.387	+0.057
UEFA EL	1.338	1.333	1.343	+0.010
DEN1	1.332	1.318	1.344	+0.026
FIN1	1.322	1.380	1.267	-0.113
POR1	1.320	1.260	1.370	+0.110
TUR1	1.310	1.275	1.344	+0.069
ITA1	1.299	1.247	1.351	+0.104
GER2	1.288	1.355	1.221	-0.134
RUS1	1.287	1.292	1.283	-0.009
POL1	1.266	1.241	1.286	+0.045
FRA1	1.263	1.184	1.343	+0.159
ENG2	1.227	1.221	1.233	+0.012
FRA2	1.123	1.094	1.151	+0.057
ITA2	1.098	1.112	1.084	-0.028

Table 1.6: Average goal difference per game, by competition. Source: CIES

the level of competitive balance they considered the percentage of points obtained by the club that ended the season in first place and the percentage of points obtained by the teams in the top three at the end of the championship. The results showed an increase in the level of imbalances over the period considered. The growth of the imbalance level is even more evident in the "Big Five" leagues. This can be caused by a number of factors, including in particular the above-mentioned increase in revenues from participation in international competitions. Perhaps it is precisely for this reason that within the sample studied the cadet alloys are those that generally show the lowest Imbalance. The same result was obtained through an analysis of the competitive balance for matches instead of for the final result of the championship. It should be noted that the context to be analyzed, that of Serie B, presents a situation that shows a low level of imbalance compared to that of the other leagues, making the context to be analyzed extremely interesting.

1.7. A FOCUS ON ITALIAN FOOTBALL

Every year AREL and PWC draw up the report of Italian football in which they analyze the Italian phenomenon from different points of view. In this paragraph we will not deal with the data relating to the census of Italian football and amateur youth football, as it has already been dealt with in the previous paragraphs. Rather, this section will analyse the economic-financial profile of professional football, its tax and social security implications and its socio-economic impact. In Italy, the aggregate value of the production of the professional football system in the 2017-18 season was around 3551 million, registering a variation of +6% compared to the previous season.

This number differs from that reported by Deloitte because in the PWC are also included the capital gains from the transfer of players. The aggregate EBITDA of the football system amounted to around 780 million euros (Table 1.7), with a positive change of 6.2%. On the other hand, the aggregate net result worsened by 37.8% compared to the previous year, with an aggregate loss of 215 million euros (Table 1.7).

Campione analizzato	20 su 20
Dati chiave/Indici	17-18
Valore della produzione	€ 3.071m
Costi operativi	(€ 2.256m)
Ebitda	€ 815m
Ammortamenti e svalutazioni	(€ 713m)
Ebit	€ 103m
Proventi (oneri) straordinari e finanziari	(€ 106m)
Ebt	(€ 4m)
Imposte	(€ 94m)
Risultato netto	(€ 98m)
Costo del lavoro p.tesserato/ricavi di vendita	*55%*
Patrimonio netto/totale attività	*9%*
Debiti finanziari/totale debiti	*30%*

Campione analizzato	19 su 22
Dati chiave/Indici	17-18
Valore della produzione	€ 353m
Costi operativi	(€ 344m)
Ebitda	€ 10m
Ammortamenti e svalutazioni	(€ 52m)
Ebit	(€ 42m)
Proventi (oneri) straordinari e finanziari	(€ 1m)
Ebt	(€ 43m)
Imposte	(€ 13m)
Risultato netto	(€ 56m)
Costo del lavoro p.tesserato/ricavi di vendita	*69%*
Patrimonio netto/totale attività	*19%*
Debiti finanziari/totale debiti	*21%*

Campione analizzato	46 su 57
Dati chiave/Indici	17-18
Valore della produzione	€ 126m
Costi operativi	(€ 172m)
Ebitda	(€ 45m)
Ammortamenti e svalutazioni	(€ 12m)
Ebit	(€ 57m)
Proventi (oneri) straordinari e finanziari	(€ 2m)
Ebt	(€ 60m)
Imposte	(€ 1m)
Risultato netto	(€ 60m)
Costo del lavoro p.tesserato/valore della produzione	*83%*
Patrimonio netto/totale attività	*0%*
Debiti finanziari/totale debiti	*n.a*

Table 1.7: Summary results for 2017/18 season. Source: Report calcio 2019, Arel & Pwc

In the period analysed by the report, i.e. from 2013 to 2018, the value of production had a positive CAGR of 6.8% accompanied by a CAGR relating to the increase in costs of 4.3%. To complete the picture of the economic data, the net result had a CAGR of 9.3%. The balance sheet figures, on the other hand, generally show a positive CAGR for total fixed assets of 5,2%, 3,7% for debits and 15,7% for shareholders' equity (Table 1.8).

As far as the tax and social security contribution of professional football is concerned, there was in general a positive CAGR in the period 2006-16 of +4.3%, with a CAGR of +25.7% for IRES in particular. As far as betting is concerned, on the other hand, there is a negative CAGR of -2.6%. In general, in the historical series most of the contribution comes from Serie A, followed by Serie B and then by C. This breakdown can be seen for all the taxes analysed, i.e. VAT, IRES, IRAP and IRPEF.

Finally, the socio-economic impact is estimated at around 3.01 billion euros. The direct contribution to the national economy is 742.1 million, with estimates of 1.215 billion of

health expenditure savings from football and the economic savings generated by social benefits is around 1051.4 million.

	13-14	14-15	15-16	16-17	17-18
Serie A	€ 3.612,9m	€ 3.358,6m	€ 3.513,2m	€ 4.269,3m	€ 4.718,2m
Serie B	€ 567,9m	€ 404,4m	€ 446,3m	€ 345,4m	€ 354,8m
Lega Pro 1° Div.	€ 127,3m	€ 134,6m	€ 167,9m	€ 161,7m	€ 148,4m
Lega Pro 2° Div.	€ 28,8m				
TOTALE ATTIVITÀ (TA)	€ 4.336,8m	€ 3.897,6m	€ 4.127,4m	€ 4.776,4m	€ 5.221,4m
Serie A	€ 197,9m	-€ 12,8m	€ 75,2m	€ 301,1m	€ 428,7m
Serie B	€ 59,4m	€ 47,6m	€ 64,8m	€ 58,6m	€ 68,6m
Lega Pro 1° Div.	€ 13,6m	€ 2,4m	€ 10,6m	-€ 1,6m	-€ 7,1m
Lega Pro 2° Div.	€ 2,5m				
Totale Patrimonio netto (PN)	€ 273,4m	€ 37,2m	€ 150,5m	€ 358,1m	€ 490,3m
Serie A	€ 3.093,3m	€ 2.974,2m	€ 3.066,2m	€ 3.624,5m	€ 3.883,0m
Serie B	€ 465,0m	€ 296,0m	€ 302,9m	€ 238,5m	€ 246,1m
Lega Pro 1° Div.	€ 103,3m	€ 115,7m	€ 134,5m	€ 145,8m	€ 136,5m
Lega Pro 2° Div.	€ 24,7m				
Totale Debiti (TD)	€ 3.686,3m	€ 3.385,9m	€ 3.503,5m	€ 4.008,8m	€ 4.265,6m
Serie A	€ 321,8m	€ 397,2m	€ 371,9m	€ 343,7m	€ 406,5m
Serie B	€ 43,5m	€ 60,9m	€ 78,7m	€ 48,2m	€ 40,1m
Lega Pro 1° Div.	€ 10,3m	€ 16,4m	€ 22,8m	€ 17,5m	€ 19,0m
Lega Pro 2° Div.	€ 1,6m				
Totale Altre Passività (TAP)	€ 377,1m	€ 474,4m	€ 473,3m	€ 409,4m	€ 465,6m
TOTALE PASSIVITÀ	€ 4.336,8m	€ 3.897,6m	€ 4.127,4m	€ 4.776,4m	€ 5.221,4m
Equity ratio (PN/TA)	*6,3%*	*1,0%*	*3,6%*	*7,5%*	*9,4%*
Debiti su totale attività (TD/TA)	*85,0%*	*86,9%*	*84,9%*	*83,9%*	*81,7%*

Table 1.8: Shareholders' equity, payables, total assets and liabilities 2013-18. Source: Report Calcio 2019, Arel & Pwc

The direct contribution to the economy can be divided into infrastructure investments for a value of about 135.5 million and in the consumption of players, i.e. the expenditure related to the expenses of players such as travel, sports equipment, registration fees, etc... for a figure that stands at around 606.6 million. The analysis of sociality, on the other hand, is characterized by 4 indicators concerning the reduction of crimes for a saved expenditure of 0.37 million, 182.5 million for the reduction of the risk of NEET, 55.9 million for the improvement of training and a value produced by voluntary sports for 812.7 million. Finally, its impact on people's health is around 1175.4 million in terms of perceived well-being, with a decrease of 42.1 million for the cost of accidents and a positive influence on the costs of cardiovascular disease, type 2 diabetes, mental illness and colon and breast cancer.

1.8. ENVIRONMENT EMPIRICAL ANALYSIS: SERIE B

Serie B is the second professional level of the Italian football championship, organised and managed by the Lega Nazionale Professionisti di Serie B. Series B is the context in which the empirical analysis of experience as a success factor in professional football clubs will be carried out. Following the Charter of Viareggio of 1926, in 1928 the Bolognese fascist hierarch Leandro Arpinati, new president of the FIGC, imposed the creation of a cadet championship no longer interregional but with a single round the same as that of the major championship. So in the 1929/30 season there were for the first time the Serie A and Serie B as we know them today. Initially 18 teams were registered and the number remained so until the 1933/34 season when it tried to divide the championship into two groups. The experience failed and two seasons later the old organization returned until the outbreak of World War II. The resumption of the single group championship took place in the 1948/49 season with 22 teams then reduced again to 18 in 1952/53. During the 1958/59 season it was decided to have a 20 team championship that remained unchanged for 35 years, except in the 1967/68 season following the decrease in the number of participants in the Serie A from 18 to 16 teams. In 2003/04 there was the absolute record of 24 teams then fell to 22 in the following season. In the 2018/19 season, due to the lack of registration of Avellino, Bari and Cesena and the inability to find teams in good time, the championship went from 22 to 19 teams. In the 2019/20 season the championship returns to have 20 teams (Serie B Website, 2019; Wikipedia Website, 2019). The formula of the championship provides for a single round with Italian races outward and return. According to the rules 3 points to the winner of the match are assigned, 1 point for the draw and 0 for the loser. The championship provides for an annual replacement of 7 teams with three promotions in Serie A and 4 relegation to Serie C. The first two teams of the championship were promoted directly while the third promotion is decreed through playoffs between the teams classified between the third and eighth place. There is no playoff if there is a gap of 14 points or more between third and fourth place. As for relegation, teams ranked between eighteenth and twentieth place were directly relegated while sixteenth and seventeenth play a playout unless there was a gap between the two of 5 points or more. In the period taken into account for the compilation of the dataset this thing occurred

PERIOD	PARTICIPANTS
1929-1933	18
1933-1934	26 in two groups
1934-1935	32 in two groups
1935-1936	18
1936-1937	16
1937-1938	17
1938-1943	18
1946-1947	60 in three groups
1947-1948	54 in three groups
1948-1950	22
1950-1951	21
1951-1952	20
1952-1958	18
1958-1967	20
1967-1968	21
1968-2003	20
2003-2004	24
2004-2018	22
2018-2019	19
2019-	20

Table 1.9: History of participants in Serie B. Source: Wikipedia website

only once, in the year 2015/16. The playoff formula has changed during the time window taken into account for the construction of the dataset. In fact, until the 2016/17 season, the regulations provided for what was previously said, while from the 2017/18 season, the points gap between the third and fourth is no longer considered. The preliminary round includes only one home game of the team that has obtained the best result in the standings. In case of a tie, the match goes to overtime and in case of a further tie, the home team enters the semifinals. The semifinals are composed of two games, outward and return, at the end of which, in case of a draw at the ninetieth minute of the return match, the team that had done better in the league enters the final. The final consists of two games, in case of overall parity then wins the team best placed in the standings. Only in the event that the two teams have finished the championship with equal points, the overtime and possibly penalties will be disputed. The away goal rule does not apply to

both the final and the semi-final. For playouts, however, is still in force the rule that does not provide for the dispute in the event that there are 5 or more points between the fifth and fourth last. In general they are composed of two games, one outward and one return where in the event of a tie at the end of the 180 minutes, the fifth last in the standings remains in Serie B, while in the event of a tie in points in the standings are disputed overtime and penalty kicks. As for the playoffs, the rule of away goals does not apply.

1.8.1. SERIE B'S PROFIT AND PATRIMONIAL PROFILE

If the economic results are compared to sports results it can be seen that in Italy, taking into account the seasons between 2007/08 and 2017/18, the Serie A is regular if the ranking of salaries and the championship victory are compared, in fact, the winner of the championship, usually is the team that has the highest total of fees. The average ranking is 1.2, while for the comparison between the ranking of the value of production with the victory of the Serie A championship, the average ranking is 1.4. Very different situation for the Serie B where the above mentioned comparisons have an average position for the first d 5,9 and for the second d 7,8. If we look at the Serie C instead, we have an even more varied situation with a positioning for the first of 9.7 and for the second of 11.4 (Report Calcio Pwc, 2019. Table 1.10-1.11).

As far as average investments or divestments are concerned, the Report Calcio 2019 distinguishes 4 clusters for the Serie B. The data analyzed concern the season 2017/18. The four clusters are Investments/Disinvestments for direct promotion for an average of 9.3 million where 49% of the fixed assets of the clubs that have been directly promoted have been invested in players' registration rights; Investments/Disinvestments for clubs that have participated in the playoffs for an average value of 2.6 million, Investments/Disinvestments for clubs that have remained in Serie B for 1.1 million and finally Investments/Disinvestments for clubs that have participated in the playouts or have been directly relegated by 5.34 million.

	Serie A	Serie B	Serie C
2007-2008	1	3	4
2008-2009	1	3	23
2009-2010	1	2	3
2010-2011	1	9	25
2011-2012	3	5	3
2012-2013	1	4	14
2013-2014	1	1	7
2014-2015	1	16	1
2015-2016	1	1	12
2016-2017	1	19	8
2017-2018	1	2	7
Posizionamento medio	1,2	5,9	9,7

	Serie A	Serie B	Serie C
2007-2008	2	3	1
2008-2009	2	16	28
2009-2010	1	9	1
2010-2011	1	5	19
2011-2012	3	10	13
2012-2013	1	10	18
2013-2014	1	1	9
2014-2015	1	11	1
2015-2016	1	1	15
2016-2017	1	19	11
2017-2018	1	1	9
Posizionamento medio	1,4	7,8	11,4

Table 1.10: Comparison of the ranking of the salaries of the winners of Serie A, Serie B and Serie C. Source: Report Calcio Pwc, 2019

Table 1.11: Comparison of ranking positioning in the value of production of the winners of Serie A, Serie B and Serie C. Source: Report Calcio Pwc, 2019

Thus, the largest investments in average terms are made by the best and worst in the class. This is probably dictated by the same cause but with different purposes. Both clusters usually invest more than the others in the repair market on the one hand to ensure direct promotion and on the other to avoid relegation or playouts. If the income and equity data of the Serie B are considered, it can be seen that the value of production in the 2017/18 season has returned to growth after the fall of 2016/17. The increase was +13.5%, reaching an aggregate figure of 353 million euro. There are also positive figures for EBITDA, which grew by 2% to €10 million (Report Calcio 2019, Pwc). The item "contributi in conto esercizio", which represent an important parachute for the teams relegating from Serie A, is becoming increasingly important. It increased by 25.3% in the 2017/18 season, accounting for 35% of total production value (Report Calcio 2019, Pwc).

The value of production is growing, but also the value of costs, which is up 11.9%, mainly due to the cost of labour, which alone represents more than half of the total. The net result of the companies worsened compared to 2016/17 by around 14%. On the other hand, there is an increase of 17% in net assets, which makes it possible to control the evolution of debts, which account for 69% of total assets. (Report Calcio 2019, Pwc).

In general, if the time series from the 2013/14 season to the 2017/18 season are considered for the main income items (value of production, operating costs, EBITDA, etc. ...) and balance sheet items (total debt, shareholders' equity, total assets ...) there is not a clear upward or downward trend. (Report Calcio 2019, Pwc). This can be caused by many factors including the large turnover that characterizes the championship and therefore makes each season very different from the others in terms of participating teams both for the regulation of the championship itself and for the events related to failures and irregularities.

1.8.2. COMPARISON WITH OTHER CADET CHAMPIONSHIPS

League	Number of promoted clubs that were relegated after one season	Percentage of promoted clubs that avoided relegation in the first season	Best performance by newly promoted clubs
Premier League	11/30	63%	Wolverhampton 7th in 2018/19 Leicester 14th in 2014/15 and champions next year
LaLiga	8/30	73%	Villarreal 6th in 2013/14
SERIE A	13/30	57%	Parma 8th in 2009/10
BUNDESLIGA *	6/22	73%	RB Leipzig 2nd in 2016/17
LIGUE 1 *	10/29	66%	Monaco 2nd in 2013/14

Table 1.12: Newly promoted clubs' performance in the past ten season (2009/10-2018/19). Source: Football Benchmark, KPMG

Every "top" "Big Five" league has its own cadet league. As previously explained by each maximum championship a total of teams are relegated and a total of teams are promoted from the cadet championship. Considering the seasons ranging from 2009/10 to 2018/19 you can see that in the top 5 European championships the percentage of clubs promoted that remained in the highest series in the season following the promotion, varies (KPMG-Football Benchmark Website, 2019). For example, for the Premier League there is a percentage of 63%, La Liga 73% as well as the Bundesliga and Ligue 1 66%. Serie A has the lowest figure at 57% (Table 1.12). These figures are also demonstrated by the best results achieved by the newly promoted clubs in the major leagues in the period considered. This figure for Serie A probably indicates a difficulty for the newly promoted clubs that is greater than the other championships due to structural, economic and managerial issues that characterize the entire Italian context as previously highlighted in this chapter.

Finally, one last feature that concerns the Serie B championship is related to the competitive balance of the championship. The study proposed by CIES presents, for both indicators previously discussed, low values of imbalance for the championship cadet (Monthly Report 40, CIES).

1.9. CONCLUSIONS

The review and the data collected in this chapter allow to understand how professional sport is now an increasingly important economic phenomenon in terms of turnover and need for managerial attention. It has been underlined the importance and the concrete possibility to observe a managerial management of sport in order to have a competitive advantage both in terms of business and sport. This is related to the purpose of this study, i.e. to try to provide tools to ensure that the manager for the selection of the team can choose the best players in order to obtain the set sports result and start the virtuous circle that characterizes professional football clubs. The focus of the analysis on football and Serie B has allowed to understand in depth the context in which it will occur if the players' experience and what elements of it are important for the achievement of success. The context of Serie B has proved to be very fluid, with a constant turnover of

players and extremely competitive, where usually those who spend the most on engagements or have a higher production value do not win. In this environment it is interesting to analyze the possible relationship between a not strictly financial and asset element such as experience and related performance. In the next chapter, on the other hand, there will be an in-depth discussion of the literature concerning the relationship under study in different contexts.

CHAPTER 2

EXPERIENCE-PERFORMANCE RELATIONSHIP

2.1 INTRODUCTION

In the first chapter of this work, the context of the data collection was analyzed in detail, stressing the nature of the sports business and the peculiarities of the football business with a magnifying glass on Serie B compared to the major leagues and other cadet championships. In this chapter, however, it will be analyzed how the experience and what comes out of it, affect the performance and what moderates this relationship. Despite the context, i.e. sports, it is considered important also to examine the managerial literature on other more common and present topics in literature. We started from the influence of the workers' experience and performance that can be understood as "players" in our context, then moved on to the analysis of the same relationship for managers and entrepreneurs who can be compared to coaches and, finally, to what links the experience of the organization itself to performance. The last area that was reviewed was that of experience as a determinant of success in the sporting context. The literature on this subject is not as extensive as that of the other areas taken into consideration and, in any case, the contexts are different from those of Serie B. The present study analyzes a little treated environment such as the Italian Serie B trying to understand which elements of the accumulated experience of players characterized the teams that perform well (playoffs and promotions) and those that perform badly (playouts and relegation). The focus of this study will be on the players and the methodologies which will be used on that dataset. However the team's dataset and the coach's dataset will be studied for what concerns the descriptive statistics and a model for their analysis will be proposed.

2.2. A GENERAL OVERVIEW OF THE EXPERIENCE-PERFORMANCE RELATIONSHIP

There are several factors that can influence the performance of a worker, a manager or a company and one of these is definitely experience. In fact, in a given field, it is necessary to have a repeated and continuous experience to achieve high levels of performance. Despite this, the extensive experience does not always lead to the same level of performance (Ericsson, 2006).

In general, it happens that people who approach a new job for the first time, are supported by more experienced individuals who guide them on their path to professional growth. Once the person has reached a sufficient level of experience, he or she is ready to act independently. At this point the subject can develop in two distinct ways, either to achieve a certain level of performance and maintain it for the rest of their working life or to continue to improve to an even higher level (Ericsson, 2006. Figure 2.1). The improvement that a person can have is not only dictated by experience but also by factors that cannot be modified through training and practice. In fact, performance is also influenced by the subject's innate abilities, skills and talents (Krampe & Charness, 2006).

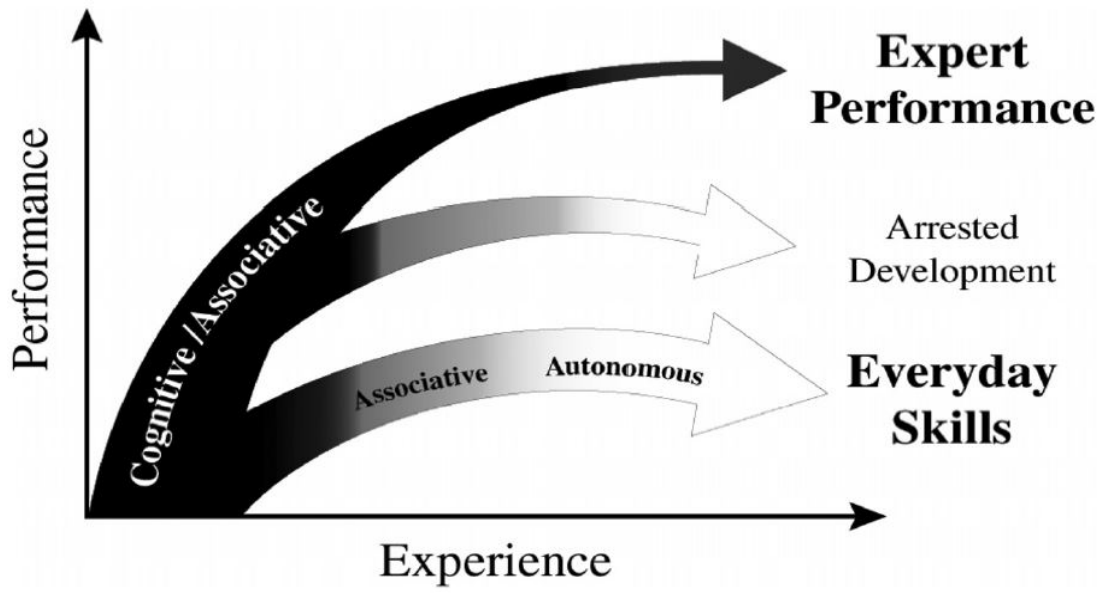

Figure 2.1: Qualitative difference between the course of improvement of expert performance and of everyday activities. Source: Ericsson, 2006

Beyond these issues, several studies have identified the presence of different types of experience. For example, it may be direct or vicarious as described in the paper of Levitt & March (1988), may vary in terms of peace and timing (Levinthal & March, 1981), as well as in its novelty (Lampel et. al., 2009), heterogeneity (Haunschild & Sullivan, 2002) and ambiguity (Bohn, 1995). Also experience can be gained from co-located or geographically dispersed units (Argote, 1999) as well as from a failure or success (Sitkin, 1996).

When an individual faces a new activity that requires skills for the first time, their main objective is to reach a level that can perform the activity in a functional way (Ericsson, 2006), then they will try to act without making big and obvious mistakes. This can be understood as the first phase of learning (Fitts & Posner, 1967). The second phase of learning eventually involves the individual having learned more about the task and the performance seems more fluid and there is a strong reduction of serious errors. So after a period of experience and training an acceptable level of performance can be reached. The third phase of learning sees an automation of the actions carried out where the subject takes little effort to achieve an acceptable performance and in this phase the individual loses the ability to control his skills and the performance reaches a stable level in which there are no significant improvements (Ericsson, 2006). In fact if it takes week or months to get an acceptable performance, it may take years or decades of experience to get a better one, because everyone needs to accumulate patterns and knowledge on how to react to situations by remembering similar situations in the past (Simon & Chase, 1973).

From three laboratory tasks in the field of chess (selecting the best move given a given situation), typing (writing as much as possible of a text given in one minute) and music (playing the same piece twice in the same way) Ericsson & Lehman (1999) have come to define how an extensive experience is necessary to achieve superior expert performance and how only some types of domain related experience show to be connected to an increase in performance and that this must be accompanied by thousands of hours of training and practice.

Particular importance is given to the characteristics concerning the specific field of experience, in fact the development of an expert performance can be understood as a function of age and years of experience, in particular all individuals improve gradually (Ericsson, 2006). The peak of performance is usually achieved in adulthood after several years of exposure and practice. Normally in sport the highest level of performance is reached between 25 and 30 years (Ericsson, 2006) while for what concerns art and science, the peak is between 30 and 40 years (Shulz & Curnow, 1988). Obviously each context has its own peculiarities and peak performance after several years of practice and at different ages. To deepen the topic, Ericsson (2006) argues that the increase in performance does not happen automatically as a link with experience but that the improvements are caused by a change in the cognitive mechanisms through which the brain and nervous system control the performance and by the degree of physiological system of the body, so there must be specific and stable changes that allow an incremental increase in performance.

Experience Measure	No. of Estimates Across Literature
Management Experience	45
Ownership/Entrepreneurial Experience	43
Traditional Experience	40
Start-up Experience	35
Related Activities Experience	34
Industry Experience	18
Experience squared	16
Wage Experience	6
Marketing Experience	6
Finance Experience	4
Supervisory Experience	3
Tenure	3
Tenure Squared	3
R&D Experience	2
Manufacturing Experience	2
Joint Experience (Team)	2
Total	262

Table 1.1: Experience measures tested across the entrepreneurship experience-performance literature and the number of estimates for each measure. Source: Peake & Marshall, 2009

In general, the relationship between experience and performance has been studied in many areas but the results are varied and in some cases opposite. This is certainly due to the difference in the measurement of experience and performance (Cooper & Gimeno-Gascon, 1992). In this regard, in their study Peake & Marshall (2009. Table 2.1-2.2) made a review of the types of experience of the entrepreneur analyzed with the number of estimates of each type and also the effects derived from these studies.

	Experience	Management Experience	Ownership/ Entrepreneurial Experience	Related Activities Experience	Start-up Experience
Negative	4	1	1	2	2
Insignifican	18	26	29	15	25
Positive	18	18	13	17	8

Table 2.2: Estimated effects by experience measure. Source: Peake & Marshall, 2009

From table 2.1 it can be seen that the 5 most studied experiences regarding the entrepreneur are the experience in management followed by the experience in entrepreneurship, traditional experience, start-up experience and finally in related activities experience. Table 2.2 shows how in four of these types the results obtained are very different and contrasting. This is just one example of the numerous studies made on the relationship between experience and performance from many points of view. In the following paragraphs there will be description studies concerning the above mentioned relationship as far as the employees, the managers and the company itself are concerned.

2.3. EXPERIENCE-PERFORMANCE RELATIONSHIP OF WORKERS

As far as individual job performance is concerned, many studies have verified that it is dynamic and changes over time (Sturman, 2003) but not exactly what influences it (Ployhart & Hakel, 1998). In a study carried out by Sturman (2003), the relationship between job experience, organizational tenure and age and performance is analyzed. The result of the study shows an inverted U-shaped relationship between time and

performance with significant effects of influence between the method of measuring performance and the complexity of the work. For the definition of the three variables the author has based his study on the previous literature where he defines the job experience as the one obtained through the accumulation of job specific knowledge from action, practice and perception of the tasks and duties associated with a specific job (Quinones et. al., 1995 cited in Sturman, 2003). Ehrenberg & Smith (2000) argue that workers invest in themselves in terms of experience to increase their skills and this has an influence on job performance. Thus changes over time due to the accumulation of relevant knowledge, skills and abilities, implying that the work experience has a positive effect on performance. For example, studies by McDaniel, Schmidt and Hunter (1988, cited in Sturman, 2003) and Schmidt, Hunter and Outerbridge (1986, cited in Sturman, 2003) have shown that a year's work advantage has a significantly greater impact at low levels of job experience than at higher one.

As far as organizational tenure is concerned, it is important in relation to the fact that the experience gained in a given context is different from the one gained in another context (Tesluk & Jacobs, 1998). This is related to the fact that the experience and performance of the organization is extremely linked to that of the individuals who make it up. This is extremely true for fooball. In fact, individuals constitute the means by which the organization creates knowledge, that is maintained in a supra individual repository that keeps it even if the individual himself leaves the organization (Argote, 2011). If this is true, however, it is also confirmed the reciprocal of this relationship, i.e. the transfer of knowledge from the organization to the individual through the experience within it (Argote, 2011).

Continuing with the analysis of variables related to the study by Sturman (2003), if on the one hand the job experience increases as well as the organizational tenure, the same applies to the age of the individual. In the paper mentioned above several studies about the negative relationship between age and performance are cited, but in addition to the worsening of cognitive, psychological and physical characteristics, certainly this effect is dictated by a dimming of the motivation of the older worker (Wright & Hamilton, 1978). This fact may be evident in the context of the Serie B. For example, older and experienced players may, despite their experiential background, no longer have any

interest in doing their best or performing extremely well. This may confirm the theory of Sturman's study (2003) stating that an inverted U relationship is observed. Considering these conditions, the study by M.C. Sturman elaborates three hypotheses, namely that there is an inverted U-shaped relationship between temporal variables and job performance, that this relationship is moderated by the type of performance measurement and by the complexity of the work. These hypotheses are all confirmed for jobs characterized by low complexity while for the ones characterized by high complexity the relationship is not linear but not even inverted U-shaped. This suggests, for example, that experience in a job can be taken into account for selection processes in low complexity jobs while for high complexity jobs the choice becomes more complicated.

Another study on non-managerial jobs was conducted by Avolio, Waldman & McDaniel in 1990. The authors sought to understand whether the age and number of years of experience in a given job helped explain the degree of work performance showed by supervisory ratings. They found out that experience predicted performance levels better than age. In their paper, they also tried to define the elements that make up the determinants of a worker's performance. The studies of Blumberg & Pringle (1982, cited in Avolio et. al. 1990) present a work performance model in which the latter is the product of skill, motivation and context. As in the previous study, also in this case the authors have theorized hypotheses that the lenght of experience will contribute to the prediction of work performance beyond the contribution of age and that the type of employment moderates the level of linear relationship and not linear relationship between performance and age or experience. Beyond the type of employment, the study result shows that experience is more age-related in terms of its influence on performance. Nevertheless, a difference between employment types was noted. The analyses of this study were conducted on a sample of 24129 individuals working in the U.S., collected between 1970 and 1984, with ages between 18 and 74, classified into 5 different occupation classes.

Schmidt, Hunter & Outerbridge (1986) also discuss supervisory ratings on job performance in their study. In particular, their paper discusses the impact that the job experience can have on job knowledge, work sample performance and the

aforementioned assessment of job performance. The examined dataset in this case collects data from 4 independent studies, creating a total sample of 1474. The dataset analysed showed that the job experience has a strong direct impact on job knowledge and a smaller direct impact on performance capabilities. In addition, the job experience has a strong indirect effect on the performance of the work sample thanks to the effect it has on job knowledge, which has a strong impact on the performance of the work sample. The results of this study were then confirmed by the paper by McDaniel, Schmidt & Hunter (1988) which studied the relationship between job experience and job performance on a sample of 16058 individuals. The job performance is dictated by two variables: the length of the experience and the complexity of the work. In this case the highest correlation was found between the population with a low value of the average work experience and jobs with a low level of complexity.

In the paper by Kotur & Anbazhagan (2014), the authors have studied how education and work experience influence performance. To get this, they have sought through the performance of workers in an Indian sugar factory. Again, a direct effect was found between the two variables and performance, verifying that performance is influenced by their level of education and that experience initially has a very positive effect on performance and that after 20 years of experience performance decreases, demonstrating an effect to inverted U-shaped, like in the already cited studies. This study probably shows that experience alone is not enough to achieve good performance.

If, on the other hand, team familiarity and role experience are considered as experience, by the study of Huckman, Staats & Upton (2009) it results that the environment and the team in which an individual works are not stable and change over time. Starting from the assumption that the team is something fluid, they found that the level of team familiarity, i.e. the average number of times that each member of the team has worked with each other, has a positive effect on performance but that the number of years of experience in the same company does not affect performance. In contrast to this, what positively affects performance is the role experience, i.e. the years in a given role in the team. This study is very interesting for the present work, it will be discussed later in this paragraph. Although the variable collected for team familiarity is the number of years the player has been in the team and therefore it is more similar to the consideration of the number of

years since the worker is in the company, in this study football clubs do not have many groups within them and therefore the variable "NYIT" (number of years in the team) can be considered as a good proxy for team familiarity. Looking again at what determines the productivity of the worker and his gain, Holzer (1990) shows that previous experience and the length of this in a given job, have a significant and positive effect on the salary and productivity of the worker, as well as a positive effect is given by the hours of practice and training. In this paper the literature on what determines the salary and value of players in football will be considered for the construction of the variables that make up the player's experience (Chapter 3).

Finally, there are also cases where experience does not translate into superior performance. As already mentioned, the human capital theory argues that the passage of time makes workers increase their skills. Certainly the on-the-job experience is fundamental, especially in the early part of one's career, but external factors can cause a worker's ability to complete a job to diminish (Medoff & Abraham,1980). This can happen during periods of strong technological change. The study by Medoff & Abraham (1980) on the correlation between experience, performance and earnings, carried out on a sample of managers and professional employees of two large American companies, shows that, within grade levels, there is a strong and positive association between experience and earnings but there is no negative association or association between the experience and its performance.

Table 2.3 summarises the results of the literature review carried out so far and this shows that experience has rarely been analysed on its own and that the results are not always the same. As already mentioned, this depends largely on the other variables analysed together with experience, context and performance indicators. As it will be explained later, in this paragraph and in the next chapters, performance in sport is easily measurable (ranking points) and experience and its relationship to success is very little covered in the literature. The works seen so far are useful to give a general idea of what has been found by relating workers and performance but in many cases the reference is made to the performance of the individual and not to that of the group.

JOB EXPERIENCE-PERFORMANCE	TYPE OF RELATIONSHIP	RESULTS
Sturman (2003)	Relationship between job experience, organizational tenure and age with performancce	Inverted U-shaped relationship influenced by the performance measurement and job complexity
Ehrenberg & Smith (2000)	Relationship between experience intended as knowledge, skills and abilities and job performance	Positive effect of experience on job performance
McDaniel et al. (1988)	Relationship between job experience and performance	A year's work advantage shows a significantly greater impact at low levels of job experience
Tesluk et al. (1998)	Relationship between job performance and experience gained in different contexts	The experience differs depending on the context, so the experience in one context may not affect performance in the same way as in another.
Avolio et al. (1990)	Relationship between age and number of years of job experience with work performance explained by supervisory ratings	Positive relationshisps between age and job performance and experience and job performance but experience is a better predictor
Blumberg et al. (1982)	Work performance as a product of skills (gained with exxperience and age), motivations and context	Experience as better predictor of performance than age and the type of employment moderates the relationship
Schmidt et al. (1986)	Impact of job experience on job knowledge, work sample performance and job performance	Job experience has a strong and direct impact on job knowledge and a smaller direct impact on performance. Moreover job experience has strong and indirect effect on work sample thanks to job knowledge which shows a strong effect on work sample performance
Kotur et al. (2014)	Influence of education and work experience on job performance	Positive relationship between education and job performance, inverted U-shaped relationship between work experience and performance
Huckman et al. (2009)	Relationship between team familiarity and role experience with performance in fluid teams	Positive relationship between team familiarity and role experience with team performance. Number of years in the same company does not affect performance
Holzer (1990)	Determinants of productivity of the workers and their gain	Positive relationships between lenght of experience in a job and productivity and level of salary
Medoff et al. (1980)	Correlation between experience, performance and earnings	Strong association between experience and earnings but no association between experience and its performance

Table 2.3: Resume of literature on job experience and performance

Since the focus of this study is on the experience of the players and what elements of their experience are important to achieve high performance, it is important to focus further on the dynamics that make an individual within a team successful. This is because the debated task concerns the performance that the team had in the seasons analysed and not the performance of the individual player. As previously pointed out, an important part of the work within an organisation is conducted by groups that can be understood as fluid, i.e. groups not always formed by the same individuals. This can also be considered true in the analyzed area since the composition of the teams varies over time and even during the same season. So in the same way as groups and teams work within companies, football teams can be considered fluid systems. The aforementioned study by Huckman et al. (2009) argues the importance of considering the familiarity of team members with each other and the role changes that individuals in the group have seen in the accumulation of the fluid team experience. As far as the present study is concerned, the factor that takes into account the familiarity of the team instead of the role experience is significant because it is difficult for a player to change roles in a football team many times within the group. Their measure of team familiarity was evaluated by the average number of times each member worked with another member of the team and a positive relationship between the increase of this and the team's performance was found. The variable that was counted as team familiarity in this study was "NYIT" i.e. the number of years the player has been part of the team, so in this case more a measurement of membership of the organization rather than an estimate of the number of interactions between individuals in the same group. More generally, some of the studies mentioned above highlight how cumulative experience is important to develop routines that allow you to move on learning curves while other elements such as experience reveals to be a rigidity or impediment to improvement when the context changes. For example, the context of Serie B is a very dynamic one because it employed new players and new teams practically every year, with a turnover of at least 7 teams per year. In contexts that see fluidity in group composition it is particularly important to understand how familiarity can lead to superior performance. Reagans et al. identify this through two elements, coordination and willingness to engage in relationship. Especially if individuals find themselves having to perform a task only achievable jointly and the knowledge to be shared is mostly tacit (as in the case of football), then team familiarity

can make the group more coordinated and thus achieve the result more efficiently. The second factor, the willingness of individuals to engage in a relationship, shows that the psychological safety team has a significant impact on learning and performance, as well as increasing trust within the team (Huckman et al., 2009). In addition to the positive effect found on the familiarity team, a U-inverted relationship was also found as very long-lived teams tend to isolate themselves and counterbalance the benefits listed above (Berman et al., 2002). The study described had as a context the projects carried out within the Wipro company, an Indian software company, therefore a context that foresees a high speed of change dictated by technology. Also the context examined in this thesis, as repeatedly said, foresees a high rate of change and a high speed even if not so much in terms of technology as in terms of actors. If considering a particular football competition as a project, then analogies can be made between the study conducted by Huckman and the sports field. For example, the team work is made up of the team of players, who have had different experiences and in some cases have already played among themselves and the project manager can be understood as the coach. The study by Ruigrok et al. (2011) is based on similarities of this kind. The paper deals with the theme of international experiential diversity and performance within projects by analysing the FIFA World Cup in Germany 2006 as a context. Team performance in this case is understood as a result of the interdependence of team characteristics and the skills, knowledge and incentives of individuals. One of the advantages of studying a sports context is to have a uniformity in individuals' incentives and unambiguous measures of performance. The aim pursued by the authors is to find out if there is a combination in the characteristics of each player's experience that makes national teams perform well. The results of the study above also have implications for the selection practices of team members in professional sports teams. This is the same objective that this study has set, namely to understand what elements of a player's experience can make the team he belongs successful, i.e. reaching the playoff or promotion to Serie A. As a framework (Figure 2.2) for their study he authors proposed two distinct sources between player capabilities, i.e. the combination of past experiences of team members (as far as this thesis is concerned the number of appearances in every professional competition both at national and international level) and the entent of shared experiences between team

members (in the present case expressed by the variable "NYIT" i.e. number of years in the team").

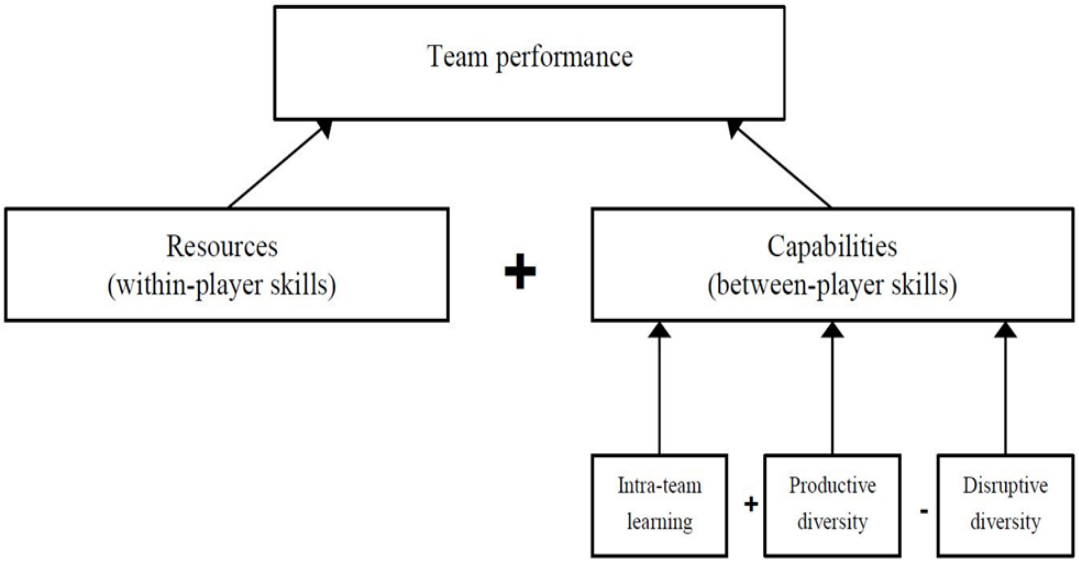

Figure 2.2: Research framework of Ruigrok et al. paper, Source Ruigrok et al. 2011

So the characteristics of a player's experience can be translated into a competitive advantage for the team if the accumulation of experience to the entire team or the combination of previous experience of each individual creates a base of tacit knowledge at the team level. This one produces value generating activities, so the fact that a team that has obtained a poor result but it turns out to have players with a lot of experience, both within the team and in other contexts, may be explained by the inability to integrate different experiences to obtain a result, given the time frame and the available resources. Obviously the context addressed in the Ruigrok et al. (2001) study concerns an international one with a very short time frame, while the one addressed in this study foresees longer periods and talks about a national second level competition. Among the results of this study (Ruigrok et al., 2011) it has emerged that a significant impact between team longevity and performance was not found in that particular context and that, on the contrary, in environment such as the national team, high levels of shared team experience may be related to a lack of ability to develop new young talent, but this is only a hypothesis. With regard to the above mentioned implications for the composition and selection of players for a team, the results of the study (Ruigrok et al., 2011) show that a manager who considers two players equally capable of holding a

certain position on the field should consider the experiential background of the two to make a choice. In that particular context the international experience of the players is important and, given the limited time available for players, it is also important to take into account the overlapping experiential background. The aim of this thesis is precisely to provide the player selection manager with an additional tool to build a successful roster. The examined context is very different from the one presented in the study described above, in fact the championship is national, it lasts much longer and therefore there is more time for players to develop coordination and integrate different experiences and styles of play matured in different contexts.

2.4. EXPERIENCE-PERFORMANCE RELATIONSHIP OF MANAGERS AND ENTREPRENEURS

In the first paragraph of this chapter it was cited an article by Peake & Marshall (2009) in which the authors tested whether there was a relationship between entrepreneur experience and performance. In this study, the role of the entrepreneur or manager can be compared, in a similar way, to that of the coach. Their research summarized the previous literature and found results of various kinds, i.e. experience as a positively correlated factor for success, experience as an insignificant factor for performance and finally experience as a negatively correlated factor for performance. From their review, through the use of exploratory and ordered probit analyses, it emerged that the start date of data collection and the type of experience tested have a strong impact on the probability of obtaining a positive relationship between performance and experience as a result of the study. To go into more detail, the authors of the paper considered all studies from 1980 to 2007 concerning the effect of experience on the entrepreneur's performance. They found that, despite the numerous studies on the subject, it was difficult to make comparisons and draw definitive conclusions on it, but the studies relating to specific experiential characteristics seem to justify a large proportion of variation in the results of the studies. In addition to this, they found that the quality of the publication outlet, when represented by journal impact factor, indicated that a higher ranked outlet increases the probability of obtaining a positive result concerning the

impact of the experience on performance. This obviously indicates a publication bias in favour of positive results. On the other hand, with regard to the influence of the starting date of data collection, it was found that the analyses that used an average of 8 years between the beginning and the end of the collection, show positive results while insignificant results are around 2 years of difference and negative results at 3. This can be given by the fact that in the early years, a company may have problems in terms of operations and the experience of the entrepreneur in the early stages may not be significant. So these results indicate that experience can improve performance after a sufficient period of time. The second factor of influence found by the authors is that of moderators, which represent the measure of experience, being in line with the research of Reuber & Fischer (1994; 1999) according to which the difference in the measurement of experience causes most of the variation of results across the literature. Based on this, it has been seen that experience in the industry is the one that most influences a positive estimate of the relationship between experience and performance with a 54% probability of achieving it, followed by management experience with a 40% probability of achieving it.

Particular mention should be made of the technology industry, where experience in this area increases the likelihood of obtaining a negative estimate of the impact of experience on performance. This may be due to the fact that a large number of entrepreneurs in the technology industry are from unrelated industries.

Taking into account Reuber and Fischer's (1994) study more in detail, it can be learnt also in this case how the previous researches about the entrepreneur's experience and the company's performance show a mixed result and that this is due to the entrepreneur's ability to acquire expertise. Their study focuses on the fact that expertise mediates between experience and firm performance and the relationship between distinct types of experience and distinct forms of expertise. In essence, the authors argue that the mixed results are due to the fact that little attention is paid to the explanation of how experience affects performance. In addition to this, the authors identify among the main reasons for this, that there is a wide variety of experience and performance measurements and that the same type of experience is defined differently in different studies, that there are other factors unrelated to experience that influence performance,

and that there may be a little variance among entrepreneurs regarding the variability that determines experience. Their results show that expertise is more strongly linked to performance than experience and that experience is an inadequate substitute for expertise in terms of firm performance. Finally, different types of expertise are linked to different types of experience. So taking experience and not expertise as a performance-related parameter can be wrong because it is understood that everyone learns in the same way and at the same speed. This could justify a possible lack of success by teams coached by coaches with higher cumulative experience. Another point touched on by the study concerns the depth and breadth of the experience. In fact, in many studies the different types of experience are evaluated independently, without considering the synergies that may exist between different experiences even if "quantitatively" scarce compared to a single and independent variety.

Considering precisely the type of experience, the study of Whitler, Krause & Lehmann (2018) analyzes the marketing experience of board members. They dealt with the biography of 64086 board members from 1500 companies and found that only 2.6% of them have marketing experience but that the presence of these members means that the growth of the company had an improvement due to the fact that growth is a strategic objective and that through their expertise the efficiency of their revenue growth strategy could be improved. This relationship is strengthened or weakened by events inside and outside the company, but the rule that there are no experienced marketers within the board members leads to a competitive disadvantage.

Experience in marketing is just one example of how the experience of a certain type can affect performance. For example Mion & Opromolla (2014) have shown that the export experience gained by managers in previous work experiences lead the company to a higher export performance and managers also got a higher salary. It has also been verified that experience is more effective when it is market specific. These researches are interesting for this study because it is possible to make a parallelism between the experience gained by managers in particular areas and the experience accumulated by coaches in particular types of competition.

On the other hand, if we consider the projects and their managers, the study of Rubin and Seeling (1967) examined experience as a factor in the selection and performance of

project managers. Initially, the research analyzed the background characteristics of 40 managers and the relative characteristics of the projects they were supposed to manage. Starting from this, the performance of the relative project was related to the experience of the project manager and the characteristics of the project itself. The result was that none of the measures related to the project manager's experience, with the exception of "growth in responsibility", showed a direct influence on project performance. In addition, it was found that organizations choose senior and experienced project managers to manage and maintain projects that are considered more important and with a higher priority. It is precisely the fact that the priority is higher that influences performance and not the fact that the manager has some experience. Again with regard to a company's projects, the 1987 study by Pinto and Slevin discusses the critical factors that lead to the success of a project. In this case it was found that managers with previous project experience have a higher probability of success in terms of successful project implementation.

Turning to the success in terms of performance of small and medium enterprises, it is difficult many times to categorize these businesses but above all to define the parameters for success (Simpson, Tuck & Bellamy, 2004). In this case the study by Felicio, Couto & Caicedo (2014) conducted on a sample of 199 Portuguese small and medium enterprises between 3 and 15 years of age from 5 different business sectors, investigated the relationship between human capital and social capital of managers with the performance of these companies. In their revision of the definition of human capital they quote Writh et al. (1995) who argued that the characteristics of human capital are education, experience and knowledge. Taking experience in particular into account, in this case the work experience does not have a significant impact on growth (Bruderl & Preisendorfer, 2000) but the previous experience in management or entrepreneurship positively influences the performance of new companies (Gimeno et al., 1997). In the present study this result could be understood as the fact that the coach's experience as a player does not influence the team's performance but his experience as a coach does. Moreover, knowledge is the result of explicit and implicit knowledge acquired through experience in certain fields (Cohen & Levinthal,1990). This study found a correlation and a positive

influence between the relationship between human capital and social capital that then influences performance and outcome.

Also the study by Simpson, Tuck & Bellamy (2004) talks about the success factors that characterize small businesses, trying to understand which factors inside and outside the company are the most critical to explain the success of the company and to verify if their results are in line with previous research. Starting from the theorization of four categories of manager/entrepreneur they found that depending on the category, there may not be clarity as to whether past experience or training and education was more important to achieve success in some cases, in others previous knowledge and experience was more important than training and in others that business and market experience was more important than training, education and previous knowledge for achieving success.

Turning to the establishment phase of a manager career, the study by McEnrue (1988) assumes that a manager who at the beginning of a career has more experience in a given role, who has been in the current location for longer and who has been in a company for longer, can achieve better results and performance than a manager with less experience in a given role and less time in the same company. In addition to this, according to the author the length of experience is more reliable in predicting performance. Analyses carried out on 89 restaurant managers of an international restaurant company showed that there is a strong relationship between the length of their role as managers and their results, also finding a lack of significant results regarding the relationship between time spent in the same location and performance and time spent within the organization and performance.

Speaking of performance, on many occasions it is necessary for a manager to perform well immediately and it is essential to understand if the experience of a manager influences early performance or not and that is what the study conducted by Stuart & Abetti (1990) analyzed data obtained through interviews with chief executives of 52 new technical ventures in NY and New England area. Their research showed that experience as an entrepreneur understood as the number of previous new venture involvements and the level of management role held in that one is the most significant factor. Other elements of experience considered such as age, years of business, management, technical

experience, etc... showed an insignificant relationship with performance. Another interesting element of the study is the negative relationship between higher bachelor degree levels of education and performance, this is due to the fact that in this particular context it is better to learn by doing and having experience in other companies than to deepen the issues with courses or experience in large companies already stable. The effect of learning by doing in the context of Serie B could be the same as the one analyzed in this study with regard to coaches. Always remaining within companies at the beginning of their lives, Wise and Vallerie (2014) deal with the topic of management experience on the performance of start-ups within accelerators. Once defined the way to evaluate the performance of an accelerator, it is certainly influenced by those who manage it in two ways, i.e. they can contribute through their knowledge and skills developed through direct experience in start-ups and knowledge gained indirectly by having access to the experience of other start-ups. The results show that the direct start-up experience of managers counts more than their connection to the network and their indirect experience, so that a manager with more experience in the role of start-up founder has a significant impact on the success of the companies he assists within the accelerator, mitigating the risk of failure of the start-up.

Kehler, Wemberg & Kim (2013) take a further step forward in the study of the relationship between experience and performance, also considering barriers to learning. In fact, if on the one hand the literature argues that entrepreneurs and managers with more experience are able to generalize their knowledge and apply it to new situations, experience cannot be understood as something that positively influences performance because it can incur in situations where previous experiences have been wrongly coded. From these assumptions the authors have shown how the positive relationship between performance and experience appears only when it comes to experienced entrepreneurs while less experienced entrepreneurs may not be able to apply the knowledge gained from their previous experiences to new contexts. For their study they based the framework on possible barriers to learning based on three context-domain differences between the entrepreneur's previous and current venture, i.e. the context related to industry, geography and time. High context similarities in the presence of low or moderate levels of experience weaken the direct negative relationship between

experience and performance, while in the presence of high levels of experience, high context similarities reinforce the positive and direct relationship between experience and performance. In general, this study confirms the research related to the theory of experience curve, i.e. by repeating several times and having more opportunities to practice, barriers to learning are overcome and performance is improved. A parallel can be drawn between the subjects analysed in this study and the coaches of a team. In fact a coach with little experience but with high similarities of context (little experience as a coach but had in Serie B) can make the negative relationship between experience and performance weaker, while a coach with a lot of experience and high similarities of context can strengthen this relationship.

Also in this paragraph it has been important to analyse the previous managerial literature on the relationship between manager/entrepreneur experience and performance to understand how there have been several studies on the subject over the years that have studied the relationship from different points of view and in different contexts. Although the focus of this study is on the players, it has been considered significant to also dwell on the dataset of the coaches and in fact in chapter 4 a model to analyze the relationship between their experience and the success of the team will be proposed. None of the reported studies analyzes a context that can be compared to that of this study but it is still important to have an overview of management examinations on the subject. In fact, as already said several times, the role of the entrepreneur or manager in a company can be compared to that of the coach within a team. Table 2.4 summarises the literature analysed in this paragraph.

JOB EXPERIENCE-PERFORMANCE	TYPE OF RELATIONSHIP	RESULTS
Peake & Marshall (2009)	Literature review about relationship between entrepreneur's experience and performance	Various king of results, positive, insignificant and negative relationship. The nature of the relationship is particularly influenced by the start date of data collection and the type of experience tested
Reuber & Fischer (1994)	Relationship between entrepreneur's experience and performance	Mixed results influenced by the ability of entrepreneur's ability to acquire expertise

Whitler et al. (2018)	Relationship between precise type of experience and performance	Positive relationship between marketing experience of board members and company performance
Mion & Opromolla (2014)	Relationship between manager's work experience and company export performance and salary	Positive relationship
Rubin & Seeling (1967)	Experience as a factor in the selection and performance of project managers	No relationship between experience and performance but experience is a tool for the selection
Pinto & Slevin (1987)	Critical factors for the success of a project	Previous project experience leads to a higher probability of success
Felicio et al. (2014)	Relationship between human and social capital of managers with performance in a small-medium portuguese enterprise context	Insignificant relationship between work experience and performance and positive relationship between managers experience and performance. Positive relationship between knowledge and performance.
Simpson et al. (2004)	Internal and external success factors that characterize small businesses	They theorized 4 manager/entrepreneur categories and depending on the category they found mixed results
McEnrue (1988)	Relationship between manager's experience in a role, in a location and in a company in the establishment phase of their career and performance	Very strong and positive relationship between role experience and performance, insignificant relationship between time in a location and in a company and performance
Stuart & Abetti (1990)	Relationship between entrepreneur's experience and early performance in new ventures	The number of previous new venture involvements and the level of management role are the most influential factors on performance. Negative relationship between higher level of education and performance due to the context
Wise & Vallerie (2014)	Experience of management and relationship with performance in start-ups within accelerators	Direct positiv influence of management previous experience in start-ups on performance
Kehler et al. (2013)	Relationship between experience and performance considering barriers to learning that are caused by the context	Confirmation of the theory of learning curve

Table 2.4: Resume of literature on manager/entrepreneur experience and performance

2.5. EXPERIENCE-PERFORMANCE RELATIONSHIP IN ORGANIZATIONS

The experience of the organization is a subject extensively discussed in the managerial literature and in many cases it is considered in terms of how much it can help the company to learn how best to deal with future situations. Organizational experience includes "historical memory of routines" that the company or organization repeats when it believes it is facing a similar situation (Perkins, 2014). This view can also be understood as being of interest to football teams. While the organization's experience can reduce uncertainty and mitigate negative performance, it can also solidify routines and negative choices for the company when it is under the illusion that it is facing a similar situation or when it has miscoded a past event (Wen & Lin, 2010).

Starting from these assumptions, Tropzynski & Banalieva (2016) studied how institutional distance changes the performance of foreign affiliates and how experience moderates this relationship. For this reason they distinguished between two types of organizational experience, i.e. the similar experience operating in institutional environments similar to those of the host country and dissimilar experience, i.e. experience in institutional environments different from those of the host country. From their analysis on a sample of 14712 Polish companies, the results showed how organisational experience can bring a competitive advantage or disadvantage. More precisely, this is influenced by the type of experience and institutional difference. In fact dissimilar experiences are not easy to transfer to more advanced markets and this can bring a strong disadvantage for the company, so an organization with a strong experience in a particular institutional context should evaluate the situation very well before entering into a different context because knowledge transfer and adaptation to a specific new market can be very expensive.

Speaking of knowledge, Fiol & Lyles (1985) define organizational learning as a change in organizational knowledge due to increased experience. Knowledge can be manifested in many ways, for example explicit or tacit, and can change behaviors, beliefs and conceptions of things within the organization (Argote, 2011). This path taken by the organization is possible because individuals belonging to the organization make sure that individual knowledge is kept in a supra individual repository even if the individual leaves

the organization (Argote, 2011). This process can take place through three sub-processes: creation, retention and transfer of knowledge. The studies of Gimeno et al. (1997) and Colombo & Grilli (2005) also argue that the human capital is critical to achieving high business performance. Therefore organizational learning is central to the success or failure of organizations, so it is important to fully understand the phenomenon and how it is related to the accumulation of experience. Precisely for this reason the study of the relationship between experience and performance has been applied to numerous corporate development activities such as the introduction of new products, diversification, international expansion, alliances and acquisitions (Anand, Mulotte & Ren, 2016).

The studies of Anand & Khanna (2000), Hayward (2002), Nerkar & Roberts (2004), Sampson (2005) & Shaver, Mitchell & Yeung (1997) cited in Anand, Mulotte & Ren (2016) find a positive relationship between experience and performance and attribute it to the ability to transform that experience into learning.

In particular, the field of CDAs differs from that of operations because it is not a continuous and routine semi-automatic pattern but is the result of a selection made by the company in a context where the type of decision implies a lower level of similarity and frequency but higher levels of causal ambiguity and outcome ambiguity (Anand, Mulotte & Ren, 2016). In fact, companies most likely repeat and accumulate experience in activities that have had a positive performance in the past believing that this can make them perform better in the future. The study by Anand, Mulotte & Ren (2016) in this sense argues that the experience within CDAs is influenced by both a learning effect and a selection effect, confirming the positive relationship between performance and experience, so the experience in that area may not result from random and exogenous decisions but from endogenous decisions driven by expectations of higher performance.

If attention is drawn to particular aspects of CDAs, the studies of Michael and Palandijian (2004) and Haleblian & Finkelstein (1999), respectively, analyze the role that experience and organizational learning have on the introduction of new products and the success of an acquisition. In the first case the authors identified a "competency trap" that is caused by an initial success that then causes the company to get stuck in certain routines giving little importance to subsequent experience, demonstrating through an analysis on the

introduction of new shampoos in the U.S. during the period 1974-1987, as the brands with more experience were those those less successful. This is a demonstration of how companies can find it difficult to learn from experience when it increases. The second case shows how data from 449 acquisitions show an inverted U-shaped relationship between the organization's experience in acquisitions and performance. It was also verified that the more similar the target is to the previous targets, the better the performance.

Always considering the theme that combines experience and knowledge, Argote & Spektor (2011-Figure 2.3) argue that organizational experience and context interact to create knowledge. Therefore they defined organizationl learning as a process that occurs over time where the task performance experience converts into knowledge that then transforms the context of the organization and influences future performance. If you draw a parallel with football teams, then you can say that the more competitions a team

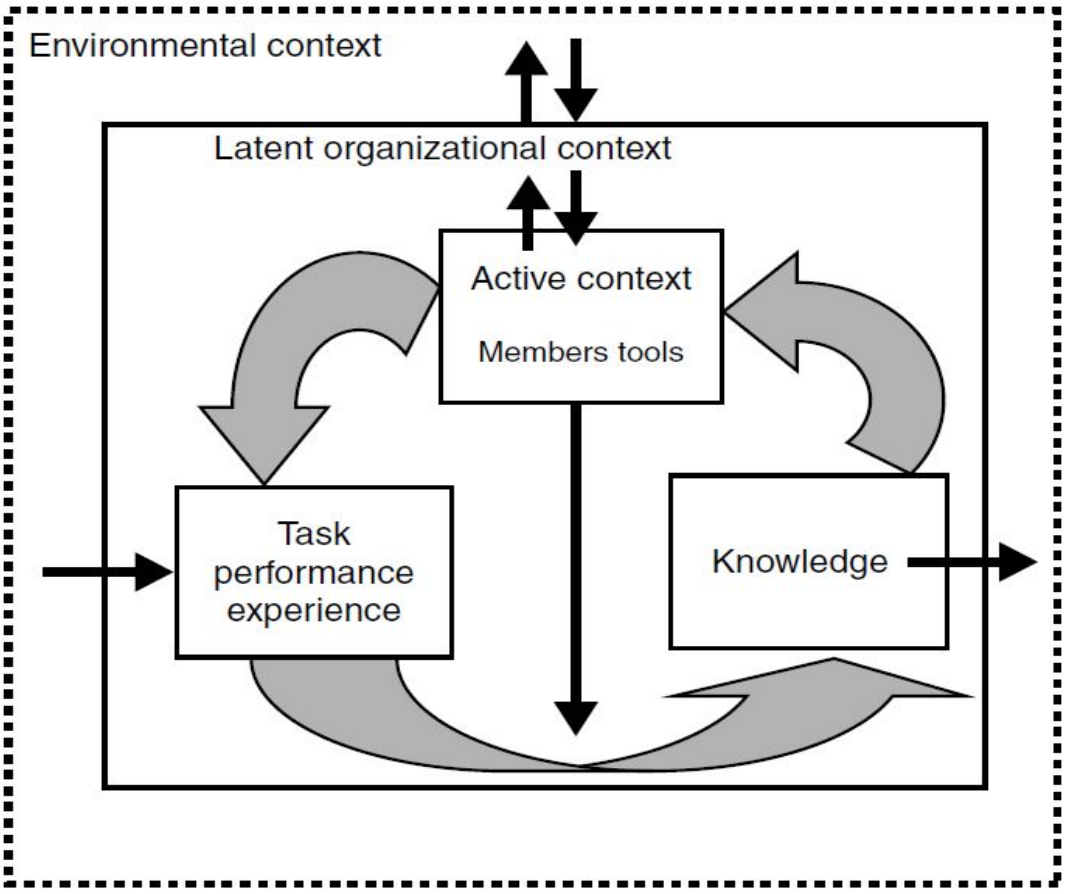

Figure 2.3: Theoretical framework for organizational learning. Source: Argote & Spektor, 2011

plays, the more this experience can turn into knowledge about the competition and how best to compete. Knowledge can then transform the context through the actions of the team that has acquired it, influencing its future performance and that of other teams. By context the authors mean something that includes the organization itself and the environment within which it acts. In turn, the experience, which can be seen in different dimensions, interacts with the context to create knowledge that once acquired transforms the context itself as the organization is part of it. So organizational memory is particularly important to understand how the organization encodes, stores and learns from the past despite staff turnover and the passage of time (Levitt & March, 1988).

It is also important to understand if an organization can benefit more in terms of learning from failures or successes and it is precisely this topic that Madsen & Desai's study (2010) deals with. Starting with previous research, the authors found that it was unclear whether organizational performance depended more on success, failure, or a combination of these. To clarify this, they considered the context of the global orbital launch vehicle industry and their analysis showed that the organization learned most effectively from failure and that the knowledge created by this type of experience was more enduring over time and that the amount of previous experience and the magnitude of failure regulated the organization's learning. In particular, this study provided evidence that organizations learn by observing and learning from their own failures and others' mistakes and failures while they found no evidence of the same thing with regard to success. Despite this, the authors did not claim that organizations cannot learn from their successes but that learning from them, in this particular context, is not an automatic process.

Another analysis concerning failures is the one conducted by Hunschild & Sullivan (2002) which deals with how the effects of previous incidents affect commercial airlines' learning. They examined the variation in learning by analysing whether companies were learning more from mistakes given by homogeneous or heterogeneous causes. Learning was measured as the reduction in accidents and incident rates over the period 1983-1997. In this period it turned out that heterogeneity was generally better for learning, which resulted in a greater and more detailed search for the causes that led to the accident, allowing a reduction in these. Accidents can be understood as rare and unlikely

events but still represent opportunities for the organization to learn. Rare events can be seen in two different ways, i.e. defined according to their probability of occurrence calculated taking into account the frequency of the event and as unique occasions to be interpreted in order to better understand and react tosimilar events in the future (Lampel, Shamsie & Shapira, 2005).

Also Perkins (2014) dealt with the role of the previous experience in a context and the related learning to explain the performance, in particular in terms of internationalization. In fact, the paper examines how the previous internationalisation experience leads to performance improvements for multinational companies that have had different experiences in different countries with different regulatory environments. The author hypothesized and demonstrated how a multinational company with high experience in institutional similarities with the target country will be successful while companies with institutional experience unrelated to the regulatory system of the target country will be more likely to fail. The context analyzed by the author is that of telecommunications and the dataset was created through interviews with managers of the sector in Brazil, Spain, Portugal, Canada and the USA and data on 96 foreign direct investments in Brazilian telecommunications between 1997 and 2004. The results of this research suggested that variations in the context influence the organization's learning curve.

The learning curve describes learning from experience and can be understood as the continuous process by which the relationship between positive and negative outcomes increases rapidly with experience before becoming subject to diminishing returns (Musaji, Schulze and De Castro, 2019). Understanding how much time or effort is required to obtain a reliable learning curve can be very useful in strategic terms for the company. The study by Musaji et al. (2019) tries to find, over a period of 14 years for a large global remittances firm in terms of choice of franchisee applicants, the point after which the performance shows a positive trend, in this case the number of choices to be made to achieve it. In fact, a limited experience causes both a learning difficulty and an increased risk of type I (false positive) and type II (false negative) errors (Dahlin et al., 2018, cited in Musaji et al., 2019). Another important aspect is the understanding of how learning is influenced by the rhythm of experience accumulation and context differences. These assumptions are the basis of the above mentioned study where the focus was on

the initial phase of experience accumulation and on how long or how much experience it takes to reach the positive portion of learning curve and in what way differences in the process of experience accumulation modify learning outcomes. Analyses have shown that a lower pace in terms of decision making shortens the time to reach the positive portion of the learning curve, making it necessary to have a sufficient interval between decisions, providing the time needed to process and accumulate information.

The same applies to the literature on the experience of the organization and the performance as to the coaches. In this paragraph it is clear that the experience-performance relationship is, as far as organizations are concerned, focused on the creation of knowledge that then influences performance. It has been seen how experience has been analysed in different contexts and in a direct and indirect way with performance. A parallelism can also be made between a normal company and a football team, in fact the accumulation of experience that then becomes knowledge can make a team learn over time to compete efficiently in a given context. In chapter 4 a model will be presented to analyse the relationship between a team's accumulated experience and its performance. In table 2.5 a review of the literature analyzed in this paragraph is provided.

JOB EXPERIENCE-PERFORMANCE	TYPE OF RELATIONSHIP	RESULTS
Tropzynski & Banalieva (2016)	Institutional distances and performance of foreign affiliates with experience as moderator	Organizational experience can bring competitive advantage or disadvantage depending on type of experience and institutional differences
Fiol & Lyles (1985)	Organizational learning as a change in organizational knowledge due to the increased experience	Positive relationship between increased experience and knowledge and then organizational learning
Anand et al. (2016)	Relationship between experience and organizational performance in CDA activities	Positive relationship between experience and performance influenced by learning and selection effects
Michael & Palandijian (2004)	Relationship between experience and organizational learning with performance in new product introduction	Negative relationship between experience and success caused by "competency trap"
Halebian & Finkelstein (1999)	Relationship between experience and acquisition success	Inverted U-shaped relationship

Argote & Spektor (2011)	Influence of organizational experience and context on knowledge	Organizational experience converts into knowledge that then transforms the context of the organization and influences future performance
Madsen & Desai (2010)	Relationship between success and failure with learning	Learning from failures is more important and more enduring than learning from success
Hunschild & Sullivan (2002)	How previous incidents affect commercial airlines' learning	Heterogeneity causes are better predictors for learning
Perkins (2014)	Relationship between previous experience in a context and learning to explain performance (internationalization)	High experience in institutional similarities brings success while low experience in institutional similarities is more linked with failure
Musaiji et al. (2019)	-How much time is required to obtain a reliable learning curve in the context of the choice of franchisee applicants. -How learning is influenced by the rhythm of experience accumulation and context differences	A lower oace in terms of decision making shortens the time to reach the positive portion of the learning curve and the pace is dictated by the context

Table 2.5: Resume of literature on organizational experience and performance

2.6. EXPERIENCE-PERFORMANCE RELATIONSHIP IN SPORT

In the three areas we have dealt with so far on the relationship between experience and performance, it has been seen that there are great differences between the studies with regard to the analysed contexts, the measure of experience and the relationships found. We have also seen how the context on numerous occasions is fundamental in defining a positive or negative relationship between experience and performance. If in the previous paragraphs a review of the managerial literature on the relationship linking experience and performance for workers, managers/entrepreneurs and organisations has been made, in this paragraph the relationship that sees experience as the cause of success in sport will be debate directly.

One of the advantages of the sport industry, but more particularly of competitions organised by leagues, is that they have a reliable and clear measure of performance. In fact, in the study of Aversa et al. (2015) on Formula 1, the final ranking was used as a measure of performance because, in general, every team aims to win as many races/matches as possible. This statement can clearly also apply to the context analysed

in this study. Each team aims to win as many games as possible and score as many points in the standings as possible. Also in the same article, as a second context-related advantage, the fact that competitors in the Formula 1 championships have comparable dimensions as well as organisational structures and focus in the production of a single technological product, i.e. the single-seater, is considered as a second advantage. As far as the study of the Serie B the same elements can be considered, obviously considering as a focus of the companies the constitution of a team led by a coach able to reach the highest possible ranking. Finally, the third advantage of the context is related to the fact of having a large amount of data available, given the great interest in Formula 1. Obviously this statement also applies to the present study. Once the performance measurement is defined for this study, it is important to define the experience as well. Data on accumulated experience have been collected for what can be understood as the worker in this context, i.e. the player, what can be understood as the manager i.e. the coach and what can be understood as the company i.e. the team. It should be remembered that the main focus of the study is on the players, but in chapter 4 a proposal has been made for the analysis of the experience-performance relationship also for coaches and teams. The method of data collection and the elements that make up experience, will be dealt with in depth in the next chapter. From the review of the literature made it has not emerged the existence of a study that included these three types of experience and especially that dealt with the context of Serie B. In particular, not much material has been found regarding the relationship between experience and performance, but experience is more analysed in terms of how it affects a player's value or how a team, through its experience, tends to be more like a win maximiser or a profit maximiser (more in depth in chapter 3.1). For coaches there are instead studies related to this relationship with mixed results but those have considered fewer elements of experience and different contexts (More in chapter 3.1). For example, the study conducted by Rodriguez et al. (2018) on the uncovering value drivers of high performer players, understood as the players with the highest value, among the determinants of the value were the age and the number of presences in national U-21 and major national teams.

Just about the age Poli, Ravanel & Besson (2018) have searched on the teams of 31 European top divisions if there were a good age to win. The study considered a period

between 2009 and 2017 and they saw that between leagues and clubs there were strong differences in the age of team members, noting that the best teams and leagues do not have very low average ages but are not even among the highest. In fact 8 out of 10 teams in the top 10 of the oldest teams are leased in Italy, Cyprus, Turkey and Greece. Instead the youngest teams are located in Serbia, Croatia and Finland, other uncompetitive leagues. The study also revealed a significant negative correlation between age and UEFA's league ranking. Even the most competitive clubs appear to be those with the average "old" clubs, which is usually between 26 and 27 years of age. From this study it can be seen how a balanced age structure allows younger players to grow well alongside more experienced players who will then be replaced.

The same authors investigated whether or not the stability of a team was important to achieve success. Again they analyzed the teams in the 31 best divisions in the period between 2009 and 2017. As an indicator of stability was considered the percentage of new players in the team for each year. In our study, instead, as a stability variable we have considered instead the number of years since the player has been playing in the team. The data analyzed by the above mentioned study showed that a high percentage of new players is a symptom of poor team management and sports difficulties. The percentage of new signings decreases as the level of the league increases, showing a strong relationship between stability and performance. These statements are also confirmed by the study by Rico et al. (2008) on the subject of team implicit coordination. From their paper it emerges how implicit coordination contributes through anticipation and dynamic adjustment of behaviors to team performance. This approach goes beyond the concept of performance caused only by explicit communication or planning interventions. In this sense Williams & Davids (1995) examined whether the skilled sport performers' enhanced declarative knowledge base is a by-product of experience or a characteristic of expertise. To do this they tested the skills on soccer recall, recognition and anticipation on three groups of equal number of experienced high skilled and low skilled players and physically disabled spectators. This showed how the high skilled soccer players demonstrated a more elaborate task-specific knowledge base, demonstrating that declarative knowledge is a characteristic of expertise rather than experience itself. Another interesting result is that declarative knowledge is formed

more effectively through performance rather than observation. On the other hand, tacit knowledge in a team can constitute a sustainable competitive advantage (Berman et al., 2002). Also in the case examined by Berman, Down & Hill (2002) using NBA data, a positive relationship between shared team experience and team performance was found. Their research also showed that although the report is positive, it is subject to diminishing returns. This is justified by the fact that the positive effect of shared experience can become negative due to the effect of knowledge ossification that reached a certain limit can outweight the benefits of collective knowledge accumulation, as a result of some rigidities that are formed with time. Another important result of the analysis shows how the coach experience in terms of tenure in the same team has a more important positive and significant influence when there are low shared level of shared experience.

The opposite result to those mentioned so far on team tenure is the study by W. Ruigrok et al. (2011) which finds an insignificant relationship between longevity of team mebership and high team performance. The considered context was that of diversity configuration in the 2006 FIFA World Cup.

A review of the literature so far on the link between experience and performance in sport as far as players are concerned has shown that in these studies the context of Serie B has not been addressed to and that experience has not been treated in the same way as it has been codified in this study (Chapter 3.1). The only study in which the experience was considered to be composed of several elements was that of Ruigrok et al. (2011), while the others were more diverted to the importance of team stability and what that entails. In the present study (deepened in the following chapters), it will be tried to understand which elements of experience characterize and diversify the components of HP teams from LP teams in order to allow the manager to better choose, given his limited financial resources, the team to compete during the season. The topic will be studied in depth in the following chapters. Table 2.6 summarizes the articles reviewed so far on players' experience and performance.

JOB EXPERIENCE-PERFORMANCE	TYPE OF RELATIONSHIP	RESULTS
Poli et al. (2018)	Perfect age to win in football	Average age between 26 and 27 years old
Poli et al. (2018)	Importance of team stability for achieving success	A strong positive relationship between team stability and performance
Rico et al. (2008)	Team stability and team implicit coordination	Implicit coordination contributes to team performance
Williams & Davids (1995)	Enhanced declarative knowledge of sport performers from experience or expertise	Declarative knowledge as a product of expertise. Declarative knowledge is formed more effectively through performance than observation
Berman et al. (2002)	Relationship between shared team experience and team performance	Positive relationship with diminishing returns because of knowledge ossification
Ruigrok et al. (2011)	Team longevity and performance	Insignificant relationship between team longevity and performance

Table 2.6: Resume of literature on players experience type and performance

Considering once again the role of the coach's experience in determining performance, the study by Hall & R. Pedace (2016) analyzes in the context of Major League Baseball whether the experience the manager had before becoming a manager translates into success and what is the relative importance of the measurable and non-measurable performance characteristics of the manager. It has been seen that managers have an influence on the team winning percentage but that the characteristics which explain this influence are not measurable for the most part. In addition, the manager's player experience has an insignificant effect on the team's performance. This result is also confirmed by Muehlheusser et al. (2016) where even managers who have been professional players perform on average worse than those who have not had a career as a player.

When considering the determinants of a team's success in certain situations, such as the playoffs, Pitts' study (2014) considered all the National Football League playoffs (1996-2012) and the likelihood that past playoff experience of quarterbacks and coaches will or will not influence a positive playoff performance. The analysis conducted showed that previous playoff experience has little evidence in influencing performance in future playoffs. The authors justified this result by arguing that playoff matches are exactly the same as in the regular season and that the matches are extremely similar in their mode

to those played by the same players in college or high school for several years. Tarlow (2012) also came up with a similar solution with his NBA study in which he considered the coach's experience as a player and a coach in playoff games to see if there was a correlation with the percentage of playoff games won.

2.7. CONCLUSIONS

This chapter the aim was to give as wide an overview as possible of the studies that have dealt with the relationship between experience and performance in the field of workers, managers/entrepreneurs, organisations and sport. The importance of this review derives from the fact that it is essential to understand how management literature has addressed the issue and how parallels can be drawn between workers and players, managers/entrepreneurs and coaches and organisations and teams. The results of the studies have been seen to be influenced by several factors such as performance measurement and the context in which this relationship is studied. For this reason it was seen that the literature in all three of the above mentioned cases has mixed and in some cases has been conflicting results. Obviously the B Series is an absolutely particular context and none of the contexts dealt with in the management literature can be considered similar.

In addition, the sports literature on the subject was also analyzed. Also in this case the contexts are different from those of Serie B and as far as the players are concerned, the most discussed topics are those concerning the stability of a team and what derives from it. Moreover, experience has not been considered in any case as in the present study, except for the study by Ruigrok et al. (2011) which, however, considers a very different context with very different elements from experience because it is an international context. This review draws attention to the fact that each study has its own peculiarities and each context has its own result caused by the characteristics of the environment itself. For this reason, since this context has not been treated previously with regard to the experience-performance relationship, the next chapter will focus on the description of the data, the criteria for their collection and the preliminary analysis of the dataset to

distinguish which elements of the players' experience characterize HPs from LPs in order to facilitate the selection of players by managers.

CHAPTER 3

DATA, METHODS AND ANALYSIS

3.1. INTRODUCTION

This chapter will first of all describe the collected datasets, both from a qualitative point of view with the description of the literature used for the choice of their creations, and from a quantitative point of view with descriptive statistical analysis for each of them. These preliminary analyses will be performed taking into account one variable at a time and comparing the two clusters in each season. This will be used to see if already with these analyses a difference can be seen in at least 4 out of 5 seasons between one cluster and the other. As said several times, since the main focus of the study is on the players dataset, the technique called PCA will be used to understand if there are variables that characterize more the HP and LP datasets and if these variables differ. First of all it will be explained how the PCA works and what we will use it for (reduction of variables and dimensionality of the dataset). Then it will be used on the unique player dataset per year to see if the projection of individuals on the first two PCs sees two distinct groups between HP and LP and then the clusters will be separated and analyzed year by year to see which original variables, which are repeated in at least 4 out of 5 seasons, explain most of the variance of the one and of the other cluster.

3.2. DESCRIPTION OF DATASET AND COMPILATION CRITERIA

The data taken into consideration for the construction of the dataset were taken from the www.transfermarkt.it website and entered by hand for the creation of the database. Data were collected for five Serie B seasons, i.e. the following season: 2018/19; 2017/18; 2016/17; 2015/16; 2014/15. For each of the seasons taken into consideration the dataset is divided into three parts, one part relating to the data concerning the history of the

club and therefore can be understood as organizational experience; one part relating to the history of the coaches and their past experiences that can be understood as managerial experience and finally a part relating to the experience of the players that can be understood as the experience relating to the workers. Of the 5 seasons considered, only those teams that have been directly promoted to Serie A or that have participated in the playoffs, considered as high performers (HP), and those teams that have been directly relegated or that have participated in the playouts, considered as low performers (LP), have been selected. For the choice of the teams, the official final classification was not considered, but the avulsed ranking. In fact, in each season considered there have been penalties imposed by the federation and therefore there have been influences in the score not deriving from the sports performance. For this reason we have considered the ranking for how it would have been without penalties.

Each dataset for a season contains all the data relating to the object under consideration (team, coach or player) until the start of the season under consideration. For example, if one considers the 2018/19 season, data will be recorded dating back to June 30, 2018, i.e. the date of the end of the previous season. All seasons have been taken into account for the data entry, starting from the 1929/30 season, the year in which the Serie A and the Serie B with a single group are held for the first time. In addition, the data are organized considering the statistics of the professional leagues only, so they are considered first level championships (in Italy the Serie A), second level (Serie B), third level (Serie C) and international competitions (FIFA and UEFA competitions). This criterion is applied in each of the divisions made, so for example for a player will not be considered experiences in amateur series such as the D Serie or for what concerns the trophies won by teams, will not be considered any trophy level lower than the C Serie.

For the collection of data and the choice of the related variables for each of the datasets, as a first step we decided to look for a definition of experience within the literature for players, coaches and teams. Starting from the players, that are the main focus of this study, in the literature there is no real definition of experienced player, in fact there is no list of factors that defines a player as experienced or inexperienced. Despite this, the experience of a player has been considered in several studies with different purposes than this, such as studies on the determinants of the value or salary of a player. The

study by Kuethe & Motamed (2010) on the salaries of "superstars" in MLS uses two distinct measures for experience, namely the years and number of years of experience in the league. Among other things, their results show that those who have played in the national team, all other things being equal, receive a 63% higher salary. To the variables "Age" and "Total league experience", the study by Gerrard (2001) adds the "appearance rate in previous season" to the "Player quality index". In this case average team age has a positive but insignificant effect on league performance beyond the age (and experience) effects captured in the PQI score, this suggests the need for further studies for the use of the PQI. Also with regard to the determinants of players' wages, the study conducted by Bryson; Frick and Simmons (2009) adds to the variables "Age" and "Experience" also the number of appearances in the league and in international competitions. A different definition of experience is given by Barros (2001) who considers the number of years of career as a footballer in his formula for determining wages. Lucifora & Simmons (2003) have studied the context of Italian football and the effect of superstars in determining wages and have considered as elements that influence the experience of the player several variables such as age, the number of cumulative attendance in Serie A and B up to the season analyzed, attendance in Serie A and Serie B in the season analyzed, attendance in the national major team and in under 21. In addition, for their calculation they have also considered the career points ratio of coaches in Serie A and Serie B. They also considered as different the cumulative experience and the most recent one following the model of Carmichael, Forrest and Simmons (1999). In general, the literature concerning players' salaries finds a positive relationship between the player's salary and age, the number of international appearances and a U-shaped relationship between age and salary, for obvious reasons related to a decrease in performance for reasons of "old age" (Bryson, Frick & Simmons 2012; Garcia-del-Barrio & Puyol 2007; Huebl & Swieter 2002; Lehmann 2000; Lucifora & Simmons 2003; Wicker et. al. 2013). Other examples of studies that have used experience are Gius & Johnson (2000); Hamilton (1997); Idson & Kahane (2000) Jones, Nadeau & Walsh (1999); Kahn (1993) who have considered age and presences while Montanari, Silvestri & Bof (2008) consider two different indicators as "player experience": "age" and "career seniority", taking into account only Serie A in the five seasons they considered for their study. Starting from these studies we have collected in the data set players 8 distinct variables that will be

listed later in the paragraph. As already mentioned, the above studies have been used as a theoretical basis for the construction of the players' experience macro variable. Table 3.1 summarizes these studies. Obviously, although the purpose of these studies has a different objective than the present one, they have been used as a basis for the components.

AUTHORS	TOPIC ANALYSED	VARIABLES USED
Kuethe & Motamed (2010)	Salaries of "superstars" in MLS	"Age"; "Number of years of experience in the league
Gerrard (2001)	Measure of quality for players and teams	"Age"; "Total league experience"; "Appearance rate in previous season"
Bryson et al. (2009)	Determinants for players' wage	"Age"; "Experience"; "Appearances in the league and in international competitions"
Barros (2001)	Players' wages	Experience as number of years of career as a footballer
Lucifora & Simmons (2003)	Effects of superstars in the determination of wages in the italian Serie A	"Age"; "Number of cumulative attendance in Serie A and B up to the season analysed"; "Attendance in the Serie A and B in the season"; "Attendance in major national team and U-21"
Montanari et al. (2008)	Predictors of pay levels in Serie A	Player experience intended as "Age" and "Career seniority" considering only 5 Serie A years

Table 3.1: Literature for the construction of the "Players' experience"

As far as coaches are concerned, a study carried out by Singell (1993) analysed the relationship between a coach's skills and the performance of the individual player. The author identified as "managers' skills" the years of experience as a coach and as a player. The study by Hadley, Poitras, Ruggiero and Knowles (2000) on the relationship between the performance of a team and the coach in the context of the NFL, considers as characteristics of the experience the number of years as a coach in the NFL league and in this case was found a relationship between the most experienced coaches and the best performance. The same result was found in the study conducted by Porter and Scully (1982), Clement and McCormick (1989) and Ruggiero et. al. (1996), while Horowitz (1994) did not find a significant relationship between the performance and the qualities of the coach. Finally, Kahn (1993) identified as manager experience the experience accumulated in the same league he is analyzing, namely Major League

Baseball. So here too it can be seen that the experience used in these studies can give conflicting results. As in the case of the players, these researches have been analyzed for the construction of the macro variable coach experience.

Finally, to consider the "experience of the teams" in itself, the literature related to profit maximisation and win maximization was held as a reference point. In particular, Del Barrio and Szymanski (2006) include among the causes related to the change of perspective from profit maximization to win maximization "the historical status of the club and the expectations of fans", therefore considering teams that have won in the past as more likely to spend to achieve results in the future. Based on this assumption, all the trophies won by the team in its history and the number of participations in the different competitions have been collected to see if a more "expert" team continues to win or if its past does not influence its future performances.

Speaking more generally, we started from the characteristics attributed to the experience in the various studies discussed above for the choice of variables to be used for the compilation of the dataset. As already mentioned above, we will now describe in more detail all the datasets.

In general, the three datasets consist of 9 variables for each year, as regards organizational experience, with 13 observations for the seasons 2018/19 (8 for HP cluster and 5 for LP cluster), 2017/18 (8 for HP cluster and 5 for LP cluster), 2015/16 (8 for HP cluster and 5 for LP cluster), 2014/15 (8 for HP cluster and 5 for LP cluster) and 12 observations for the season 2016/17 (due to the lack of playouts for a gap between the fourth and fifth last exceeding 5 points, 8 for HP cluster and 4 for LP cluster) with a total therefore of 64 teams.

As far as management experience is concerned, on the other hand, 8 variables were considered. All the coaches who have followed one another during the season have been considered, so for 2018/19 we have 25 observations; for 2017/18 25; for 2016/17 16; for 2015/16 23 and for 2014/15 19, for a total of 108 coaches. For ease of analysis, all registered coaches have been reduced to one per team per year, pondering their cumulative experience based on the number of benches in the season.

Finally, for the players 8 variables were decided for 449 observations in the 2018/19 season (256 for HP cluster and 193 for LP cluster); 417 for the 2017/18 season (249 for HP cluster and 168 for LP cluster); 384 for the 2016/17 season (242 for HP cluster and 142 for LP cluster); 422 for the 2015/16 season (252 for HP cluster and 170 for LP cluster) 438 observations for the 2014/15 season (275 for HP cluster and 163 for LP cluster) for a total of 2110 registered players.

3.2.1. ORGANIZATIONAL'S EXPERIENCE DATASET

As previously mentioned, the dataset related to the organizational experience of the team consists of 64 observations spread over 5 years. The variables that were chosen for the compilation of the dataset are:

1. "NSACW" → "Number of Serie A championships won"
2. "NSBCW" → "Number of Serie B championships won"
3. "NSCCW" → "Number of Serie C championships won"
4. "NNCW" → "Number of national cups won"
5. "NICW" → "Number of international cups won"
6. "NPIIC" → "Number of participations in international competitions"
7. "NPISA" → "Number of participations in Serie A"
8. "NPISB" → "Number of participations in Serie B"
9. "NPISC" → "Number of participations in Serie C"

The purpose of these variables is to try to describe as well as possible all the experience that the team has had in terms of society from season 1929/30 to season t-1.

The first five variables listed relate to the national and international professional championship prizes that a team can win and the number of trophies that a team has won in its history.

The variables from 6 to 9 concern national and international participations, always taking as initial reference point the season 1929/30. In the counting of the participations, all levels have been considered, therefore first, second, third and international

competitions. As far as international cups are concerned, only the competitions organised directly by UEFA have been taken into account, not considering the various friendly competitions or those which do not include all the UEFA member countries that have followed one another over the years.

3.2.2. MANAGERIAL'S EXPERIENCE DATASET

This dataset collects data relating to all coaches who have trained the selected teams during the 5 seasons considered. The number of coaches registered is 108. Also in this case the variables chosen try to better explain the experience that the coach has, related to his experience as a coach and as a player. As said before, every team has only a dummy coach who is the result of the weighted average, based on the number of benches in the season, of the cumulated experience of all the coaches of that season.

The variables are:

1. "NBIIC" → "Number of benches in international competitions"
2. "NBIFLC" → "Number of benches in first level championships"
3. "NBISLC" → "Number of benches on second level championships"
4. "NBITLC" → "Number of benches in third level championships"
5. "NPAPIIC" → "Number of presences as a player in international competitions"
6. "NPAPIFLC" → "Number of presences as a player in first level championships"
7. "NPAPISLC" → "Number of presences as a player in second level championships"
8. "NPAPITLC" → "Number of presences as a player in third level championships"

Variables from 1 to 4 relate to the coach's experience as such, i.e. the number of benches in all the professional competitions. Also in this case the starting point is considered the 1929/30 season.

The remaining variables relate to the experience of coaches as players, so they record the attendance of players at various levels of national competitions and international competitions. Some coaches have not played football or have never landed in professional series, so the values for these variables will all be 0.

3.2.3 PLAYERS' EXPERIENCE DATASET

This dataset is composed by all the players that have played in the HP and LP teams. The individual players are 2110 and the variables account for their accumulated experience in their careers.

The dataset has as variables:

1. "Age"
2. "NYIT" →"Number of years in the team"
3. "NPIICN" → "Number of presences in international competitions (national team)"
4. "NPIICcc" → "Number of presences in international competitions (club competitions)"
5. "NPIFLC" → "Number of presences in first level championships"
6. "NPISLC" → "Number of presences in second level championships"
7. "NPITLC" → "Number of presences in third level championships"
8. "NPITS" → "Number of presences in 2018/19 season"

The first variable records the age of each player considering the difference between the first year of the season and the year of birth of the player. So for example if it is considered the 2018/19 season and the player was born in 1994 then the age will be 24 years.

Variable number 2 is the number of years the player has been in the team.

Variables 3 to 7 strictly concern the experience that the player has accumulated in his career divided into international experience with the club and the national team, first, second and third level championships.

Finally, the variable number eight was collected to be able to weigh the contribution of each player in terms of attendance during the analysed season.

3.3 DESCRIPTIVE STATISTICS OF THE DATASET

The 3 datasets presented from a qualitative point of view in the previous paragraph will now be analysed through central tendency and measure of variability measures. In particular, the average and the median will be taken into consideration for the central tendency measures and for the measure of variability the standard deviation and the maximum and minimum ones. For each of the three datasets the High Performers will be distinguished from the Low Performers. Each variable constituting each of the datasets will be compared with it during the period under consideration, i.e. 5 seasons. This is in order to have a deeper knowledge of the data collected also from a quantitative point of view. From these preliminary analyses we will try to have a general view of the presence over the years of similarities or differences in the data.

3.3.1. PLAYERS DATASET

Considering the measures of central tendency we start from the variable "Age" (Figure 3.1). Over the years under consideration, the average of the observations assumes value 25.11 in the 2018/19 season; 25.95 in 2017/18; 25.27 in the 25.27 season; 25.18 in 2015/16 and finally 24.59 in 2014/15. The values assumed by the median in the same seasons do not differ so much from the values assumed by the average, assuming values of 25, 25, 25, 24.5 and 24 respectively. This tells us that the distribution is not very asymmetrical and this is probably due to the same nature of the variable. This is also demonstrated by the values that represent the maximums and minimums over the years. In fact, the minimum value recorded during the seasons is 17, which is present in 4 out of 5 seasons and the maximum is 40, present in two out of five seasons (the other three have a maximum value of 38). All this is also visible in the distribution represented by the boxplot which shows how in principle also the distribution between the first and third quartile is symmetrical. There are also few outliers. As far as the estimation of the variability within the population is concerned, also the variance and the standard deviation show rather low values. Taking the standard deviation into account, the values

are 4.51 in 2018/19, 4.89 in 2017/18, 4.38 in 2016/17, 4.94 in 2015/16 and finally 4.53 in 2014/15.

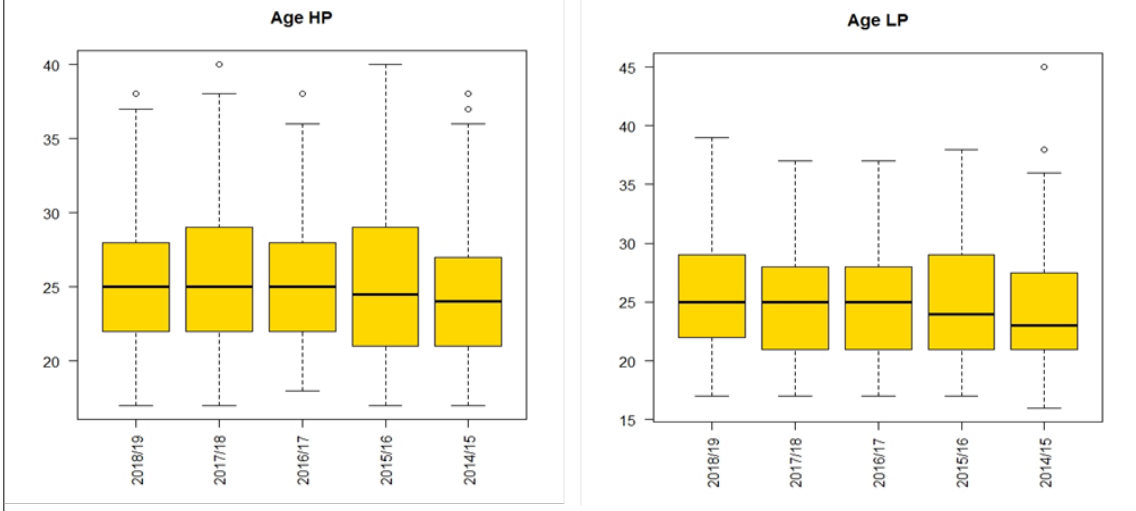

Figure 3.1: "Age" variable, HP and LP cluster. X axis: seasons; Y axis: age

If we consider the teams belonging to the LP, the averages for the same variable over five years are similar to those of HP teams. The values are 25.73, 25.02, 24.95, 24.85 and 24.59 respectively. Also in this case, the values relating to the medians do not differ so much from those of the averages except, in a relatively poor way in the 2014/15 season. This is probably due to the values that assume the maximum and minimum distribution. In fact, in that season the maximum is 45 and the minimum is 16. In the other seasons instead the minimum is 17 and the maximums are 37 in two seasons and 38 and 39. Also in this case the boxplots show that the distribution of the population is symmetrical over the years, with the exception of the 2014/15 season. Comparing it with HP data, it can be seen that there are no major differences in terms of age between the two clusters. Finally, looking at the variability indices, we can see that in this cluster the standard deviation assumes a value of 4.76 in 2018/19, 4.35 in 2017/18, 4.86 in 2016/17, 4.83 in 2015/16 and 4.83 in 2014/15, which are very similar to those assumed in the HP cluster.

Instead, the variable "NYIT" (Figure 3.2), or number of years in team, presents for HP a minimum of 0.5 years for each of the seasons and a maximum ranging from 8 in the 2014/15 season to 13 in 2018/19. Even in LPs the minimum is equal to 0.5 years while for the maximums we have 10 in the last three seasons, 6 in 2015/16 and 13 in 2014/15. As far as the median is concerned, it assumes a value of 1 in all seasons for both datasets. Instead, the average for HP is 1.95 in 2018/19, 1.8 in 2017/18, 1.97 in 2016/17, 1.61 in

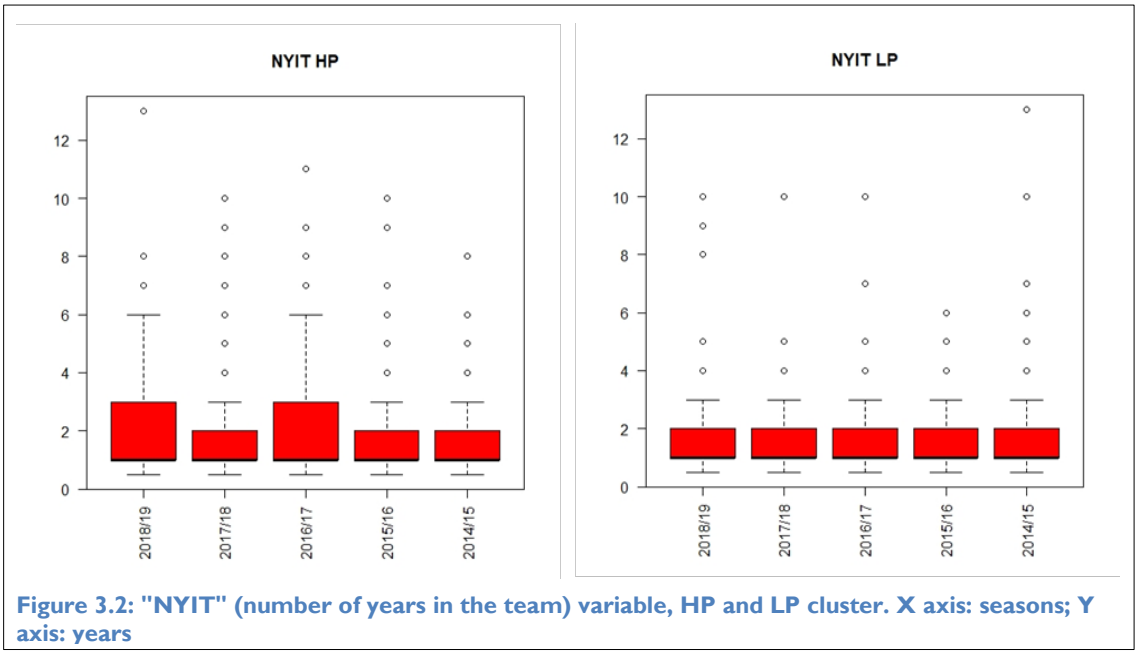

Figure 3.2: "NYIT" (number of years in the team) variable, HP and LP cluster. X axis: seasons; Y axis: years

2015/16 and 1.42 in 2014/15. For LPs instead 1.67 in 2018/19, 1.59 in 2017/18, 1.57 in 2016/17, 1.43 in 2015/16 and 1.54 in 2014/15. It can be seen that the averages for HP assume higher values over the years. This can be caused by a higher stability of teams with higher performance because they probably have a lower turnover of players than those with bad performance. However, this could be denied by the values assumed by the median, which is equal to 1 in all cases. In fact, the boxplots show a much more uniform distribution over the years for LP teams while HP teams are more asymmetrical. In both cases there is a strong distribution of values in the range between the median and the third quartile. In addition, both clusters have numerous outliers over the years.

As far as standard deviation is concerned, in general HP have higher values than LP. This is probably due, as previously mentioned, to the more asymmetric distribution of the data. Respectively, HP have a standard deviation of 1.72 for the 2018/19 season, 1.53 for the 2017/18 season, 1.67 for the 2016/17 season, 1.35 for the 2015/16 season and 1.15 for the 2014/15 season. Instead, LPs have values of 1.53 for the 2018/19 season, 1.32 for the 2017/18 season, 1.33 for the 2016/17 season, 1.02 for the 2015/16 season and 1.62 for the 2014/15 season.

The variable "NPIICN" (Figure 3.3) represents the number of players' appearances in international competitions with their national team. Over the course of the seasons, the averages for HP assume values between 8.05 in the 2016/17 season and 10.12 in the

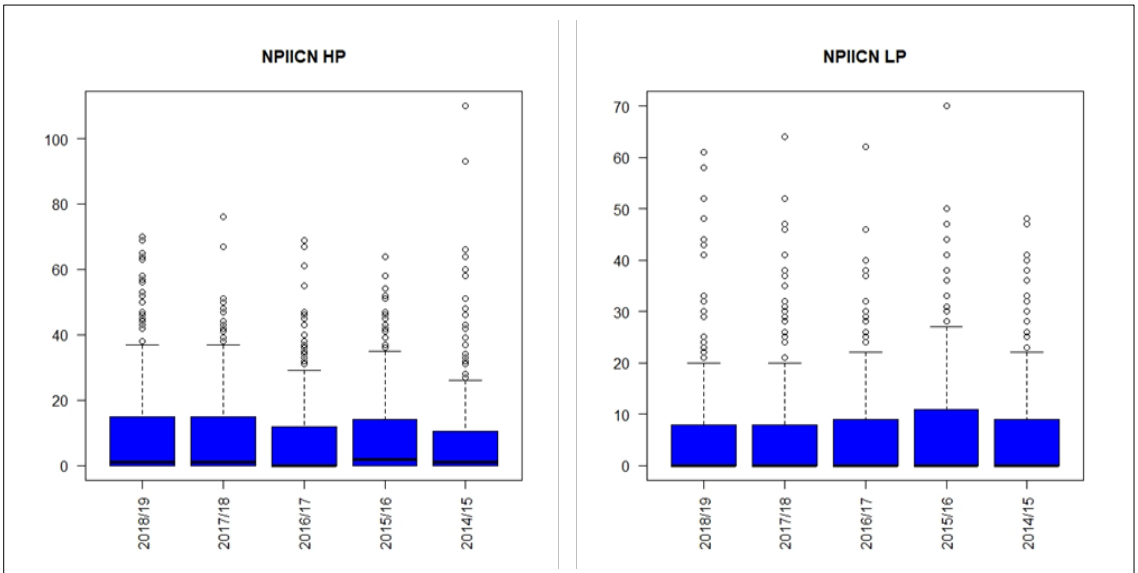

Figure 3.3: "NPIICN" (number of presences in international competitions (national competitions)) variable, HP and LP cluster. X axis: seasons; Y axis: presences

2018/19 season. For LPs instead the averages are lower with a minimum value of 6.35 in the 2016/17 season and a maximum of 7.44 in 2017/18. Always with regard to LPs, the median always assumes a value of 0 in all seasons while for HP the median is 1 in the 2018/19 season as for the 2017/18 and 2014/15 seasons, while it is 0 in 2016/17 and 2 in 2015/16. For both clusters there are significant gaps between the average and the median. This is influenced by the presence of numerous outliers in both groups and these outliers assume higher values in the HP group than in the LP group. This would explain the higher averages of the HP cluster accompanied by similar medians between the two groups. This is further confirmed by the values assumed by the maxima and minima in the two sets. In general, the minimum value common to all seasons and to both clusters is 0, while the maximums assume values of 70, 76, 69, 64 and 110 for HP while for LPs they are 61, 64, 62, 70, 48. Therefore, HP are generally distributed in a less uniform way than LPs. This can also be seen from the standard deviation which is higher in each year than the LP cluster. In fact, if in the 2018/19 season the standard deviation is 16.02 for HP, for LP it is 12.12, in the 2017/18 season it is 14.38 and 12.46, in the 2016/17 season 13.48 and 10.93, in the 2015/16 season 13.83 and 12.62 and finally in the 2014/15 season 15.04 against 11.05.

The variable "NPIICcc" (Figure 3.4) is closely related to whether a player has played in a first level league or not. In fact, international competitions with clubs organised by UEFA can only be played by clubs that have qualified by being ranked in a top league. As can

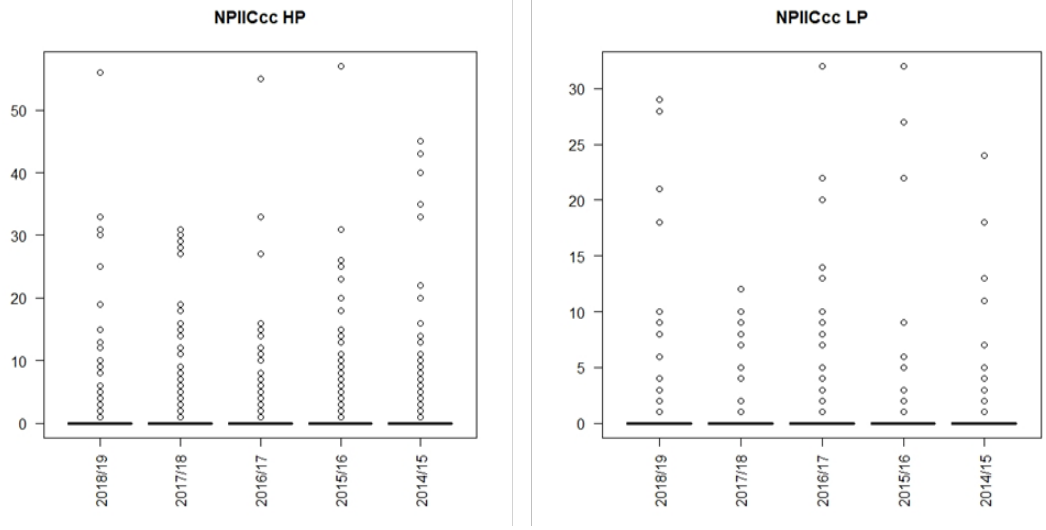

Figure 3.4: "NPIICcc" (number of presences in international competitions (club competitions)) variable, HP and LP cluster. X axis: seasons; Y axis: presences

be seen from the two boxplots the median for both clusters is always 0 and there are numerous outliers that take on higher values in the HP cluster, both in terms of number and in terms of absolute values. This is reflected in the values assumed by the average which generally has higher values in the group of HP with 1.67 in the 2018/19 season against a value of 1.06 of LP, 1.85 in the 2017/18 season for HP and 0.58 for LP, 1.15 in the 2016/17 season against a value of 1.18 (the only case in which the LP exceeds HP), 1.84 in 2015/16 and 0.74 and finally in 2014/15 a value of 1.96 against a value of 0.73. As far as data dispersion is concerned, the HP cluster has a higher standard deviation than the LPs in each of the years considered, the same for the maximum values. Instead, the minimum population value is 0 in both clusters for all 5 years.

As already mentioned, the fact that a player has had international experience with clubs is due to the fact that he has participated in first level championships and the variable "NPIFLC" represents the number of presences of a player in first level championships (Figure 3.5). In both clusters the boxplots show an asymmetric distribution is a large number of outliers. Considering HP we can see that the distribution between the median is the third quartile is larger than that of the LP in all five seasons considered. As far as the outliers are concerned, we can see that in the HP group they have on average a higher value of presences, even if in the LPs, in the 2018/19 season, they reach the absolute maximum in terms of presences in first level championships for a player with 474 presences. In general, the average for this variable assumes values considerably

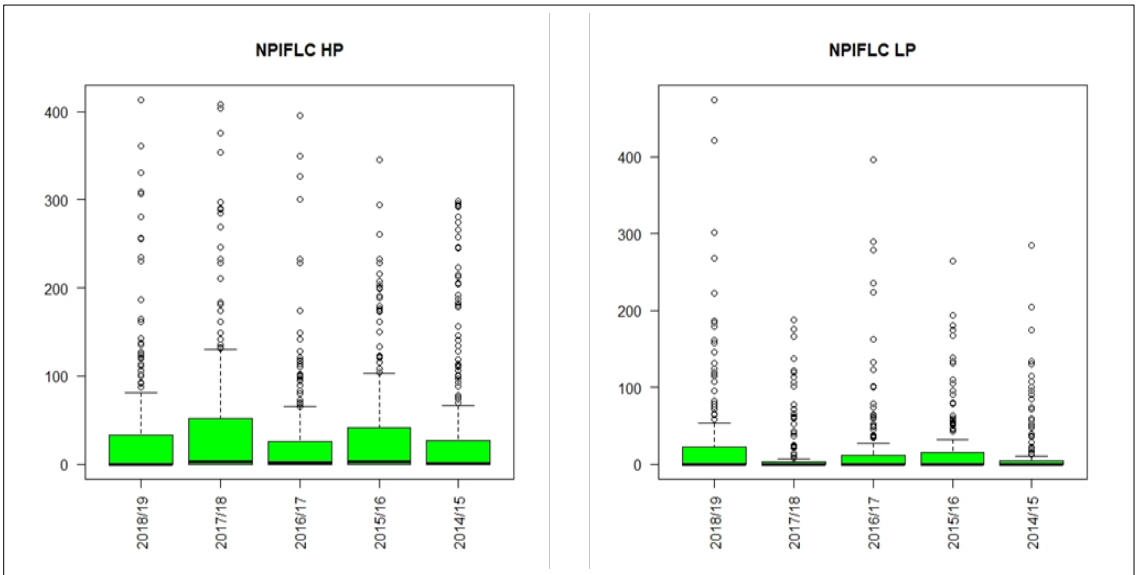

Figure 3.5: "NPIFLC" (number of presences in first level championships) variable, HP and LP cluster. X axis: seasons; Y axis: presences

higher in HP compared to LPs. Respectively the values are 31.44, 40.89, 27.42, 34.65 and 32.64 for the former and 27.19, 13.90, 22.40, 20.26 and 15.33 for the latter. The values of the median are partly in contrast with what has been said about the average. As it can be seen from the boxplots, the medians are practically all the same, with values of 0 in all seasons for LPs and with values of 0, 3, 2, 3 and 1 for HPs. This gap between the average and median values can be explained once again by taking into account the asymmetric distribution and the large number of outliers. Continuing to talk about data dispersion within datasets, HP values are on average higher but not in all seasons. In addition, for both clusters the standard deviation values also vary greatly between seasons. In particular, the values are 66.2, 75.13, 57.99, 61.27 and 66.42 for HP and 65.88, 35.66, 58.62, 44.08 and 39.25 for LP.

With regard to the experience of players in second-level leagues, two variables were collected, namely "NPITS", the number of appearances that the player had in the season and "NPISLC", the number of appearances accumulated in second-level leagues. The first will have as a minimum 0 and as a maximum the number of games played by the team in the season considered, while the second will have as a minimum 0 if the player has never participated in a second level championship and as a maximum the value of accumulated attendance.

As can be seen from the boxplots, the variable "NPITS" has no outliers points given the range of values it can assume (Figure 3.6). As for the average and the median of HP, the values are equal to 15.61, 18.69, 19.16, 18.69 and 17.1 for the first and equal to 15, 18, 18, 18 and 15 for the second. Therefore in this case there is not a big difference between the average and the median, index of the absence of outliers. As far as LPs are concerned, the situation is similar but with a greater difference between average and median than in the other cluster. Moreover, the values assumed by these two indices are for four seasons out of five lower than those of the other cluster, showing a higher turnover in the LP teams. Respectively, the average values are 13, 17.38, 16.48, 17.09 and 17.7, while the median values are 12, 17, 15.5, 15 and 17.

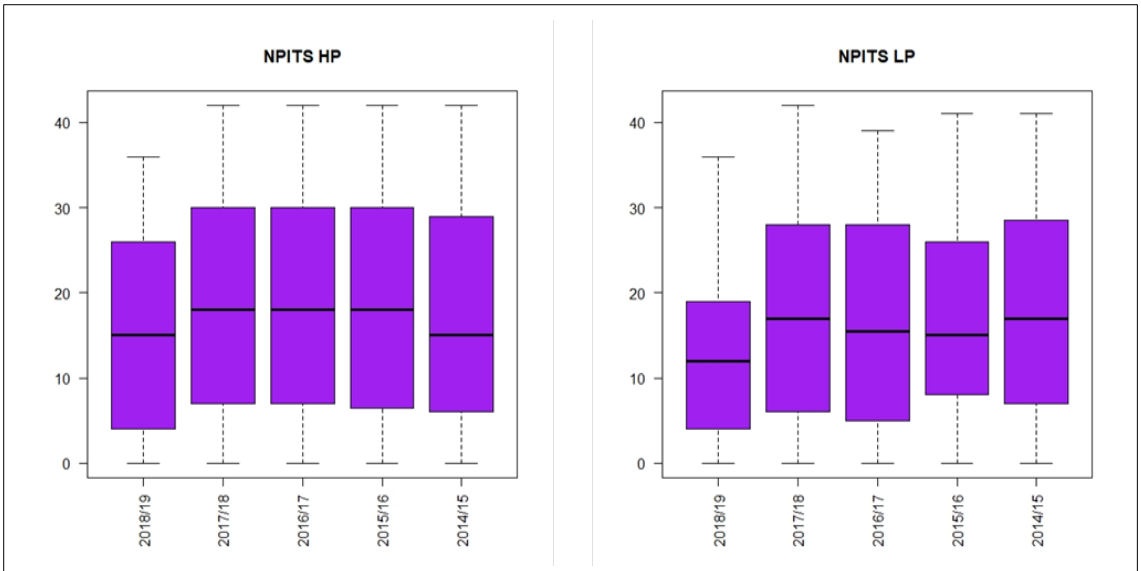

Figure 3.6: "NPITS" (number of presences in this season) variable, HP and LP cluster. X axis: seasons; Y axis: presences

On the other hand, the variable "NPISLC" (Figure 3.7) is by nature similar to the variables related to the player's experience in international competitions and first level championships and therefore does not have a specified maximum value. In particular in HP the absolute maximum value recorded in the five seasons is 301 in the 2015/16 season and for LPs the absolute maximum value is 372. In general, in LPs the maximum value in the 5 seasons collected is always higher than that of HP. In fact, the values for LPs are 343, 308, 372, 302 and 326 while for HPs are 295, 284, 268, 301 and 243. The same thing is not reflected as far as the average is concerned, in fact HP have a higher average in 3 out of 5 seasons and instead in terms of median the same thing happens. Looking at the two boxplots we can see how in both clusters the distribution of the

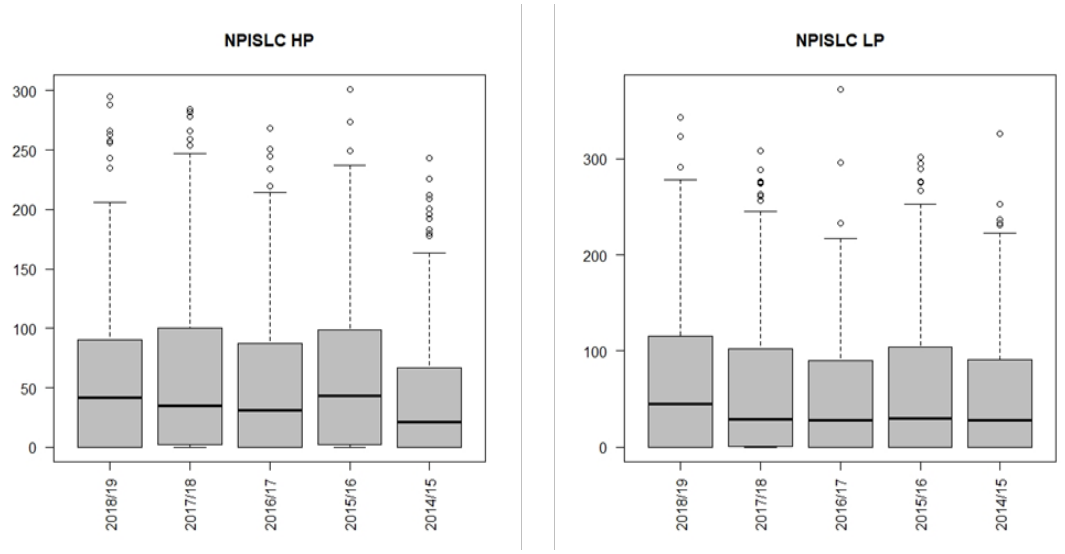

Figure 3.7: "NPISLC" (number of presences in second level championships) variable, HP and LP cluster. X axis: seasons; Y axis: presences

population is more unbalanced between the median and the third quartile even if this thing is noticed in a more prominent way in the cluster related to the LPs.

The second graph (Figure 3.7) visually presents also fewer outliers even if the distribution of data in HP seems more uniform. This is also demonstrated by the values assumed by the standard deviation. In fact, in all five seasons LPs have higher values than HPs. Respectively, the values for LPs are 77.76, 79.05, 68.94, 75.1 and 68.18 while for HPs are 66.34, 73.13, 63.88, 67.09 and 52.72.

Finally, the last variable related to players' experience concerns the number of cumulative third-level league appearances, i.e. "NPITLC" (Figure 3.8). Again, as in the previous variables except for the variables "Age" and "NPITS", there is no defined maximum. In particular, for HP the maximum values are 289, 293, 250, 324 and 208 while for LPs they are 240, 247, 258, 231 and 290. On average, the teams belonging to LPs have players with more experience in third level leagues than HP except in the 2016/17 season. As far as the median is concerned, instead, the LPs always have higher values. This seems to indicate a higher number of cumulative presences in third level championships in the teams of the bad performers cluster.

The graphs also show how in the HP cluster there are more outliers and how the distribution of data is almost uniform in the seasons 2017/18 and 2016/17 but much more asymmetrical in the rest. Instead, LPs have a uniform distribution in four out of

five seasons except for the 2014/15 season, which is particularly unbalanced between the median and the third quartile. To be more specific, the HP cluster averaged 42.39, 40.73, 39.33, 28.26, 25.97 and 23, 26, 24, 7 and 0 as medians over the five years. For LPs, on the other hand, the average and median values were 48.91, 47.67, 37.72, 34.9 and 30.58 for the first and 29, 33.5, 24.5, 23.5 and 10 for the second. The data dispersion synthesized by the standard deviation shows very similar values between the two clusters over the course of the seasons with slightly higher values for the LP cluster.

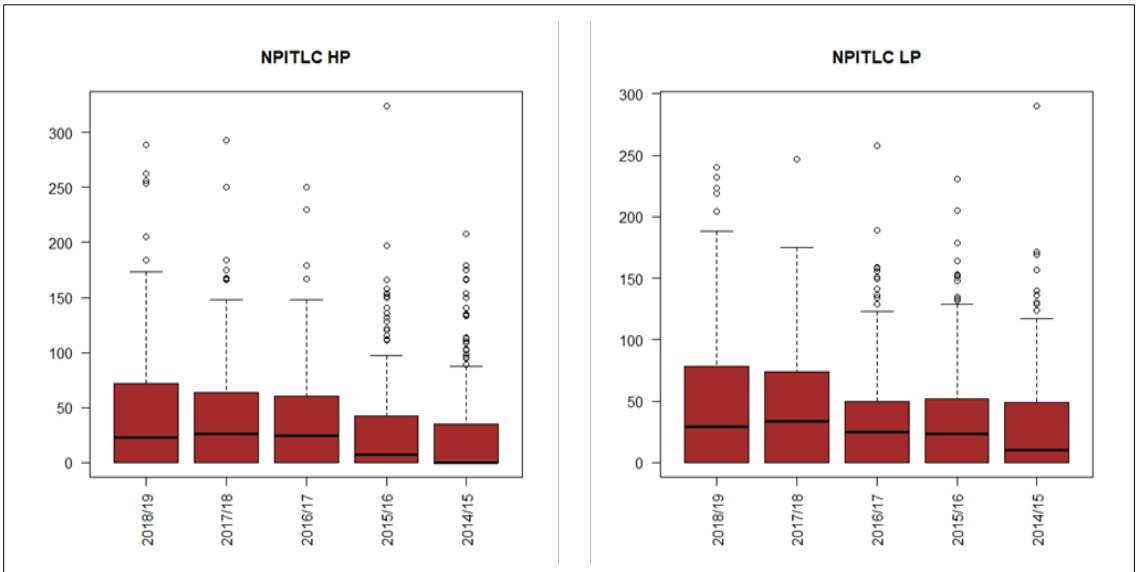

Figure 3.8: "NPITLC" (number of presences in third level championships) variable, HP and LP cluster. X axis: seasons; Y axis: presences

Summarizing the results that emerged from the descriptive statistics on the players' dataset, with regard to the variable "Age" and "NPISLC" (number of presences in second level championships) there are no significant differences in terms of average and median between the HP and LP cluster. Instead the variables "NYIT" (number of years in the team), "NPIICcc" (number of presences in international competition (club competitions)), "NPIFLC" (number of presences in first level championships) show average values that are higher for the HP cluster but the medians show very similar values between the HP and LP cluster. The variables "NPIICN" (number of presences in international competitions (national club)) and "NPITS" (number of presences in this season) show higher values for both the average and the median in the HP cluster while the variable "NPITLC" (number of presences in third level championships) shows the same thing but for the LP cluster.

3.3.2. COACH DATASET

As explained in the previous paragraph, the data collected in relation to the experience of coaches is mainly divided into two parts, namely a part relating to their cumulative experience as coaches and a part relating to their cumulative experience as players. In both parties, as for the players, their experience in the international arena and in first, second and third level leagues was analysed.

Starting from the experience as an international coach ("NBIIC"), the distribution of data over the five years for the two clusters can be seen in the two bar charts (Figure 3.9).

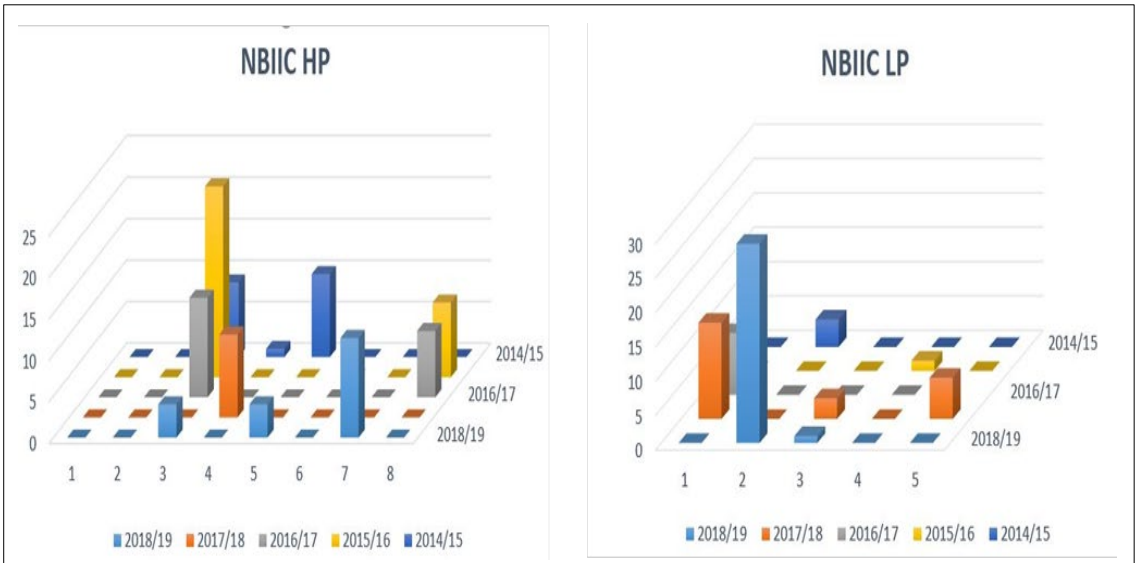

Figure 3.9: "NBIIC" (number of benches in international competitions) variable, HP and LP cluster. X axis: teams; Y axis: benches; Z axis: seasons

As far as the average of the values is concerned, it can be seen that the HP cluster has a value 3 times out of 5 higher than the LP cluster, even if in median terms the values are always 0 in both clusters, except in the 2017/18 season for the LPs, where the median value is 3. To analyze the data relating to the coaches, it must be taken into account that the datasets contain few observations and therefore even just an outlier can significantly change the summary values. In particular, to solve the problem of multiple coaches during the same season, all teams have within the dataset a single coach who has as cumulative experience the average weighted by the number of benches in the season of all coaches of the year. In the case of this variable it can be seen that in practically all seasons for both clusters there is at least one outlier. As far as maximum values are concerned, the HP cluster has more uniform values than the LP cluster. In fact, the values

for HP are 12, 10, 12, 23 and 10 while for LPs 29, 14, 9, 1.5 and 4. Finally, the data dispersion shown by the values assumed by the standard deviation is very low for LPs in the 2015/16 and 2014/15 seasons, respectively with values of 0.67 and 1.79, and assumes maximum value in 2018/19 with 12.86. Instead, HP have similar values in four out of five seasons with 4.24 in the 2018/19 season, 3.53 in the 2017/18 season, 4.75 in 2016/17, 8.3 in 2015/16 and 4.34 in 2014/15. In general, this index assumes 3 times out of 5 a higher value for HP than LP.

Continuing to talk about the experience accumulated as a coach, we will move on to analyze the experience in first level championships ("NBIFLC" - Figure 3.10). The averages are very different from each other, both during the seasons and within the two clusters, but in 3 out of 5 seasons the LP cluster has higher values. Respectively, the average values are 52.12, 6, 54, 75, 49.87 and 32.125 for HP and 67.2, 70.6, 31.75, 16.6 and 33.8 for LP.

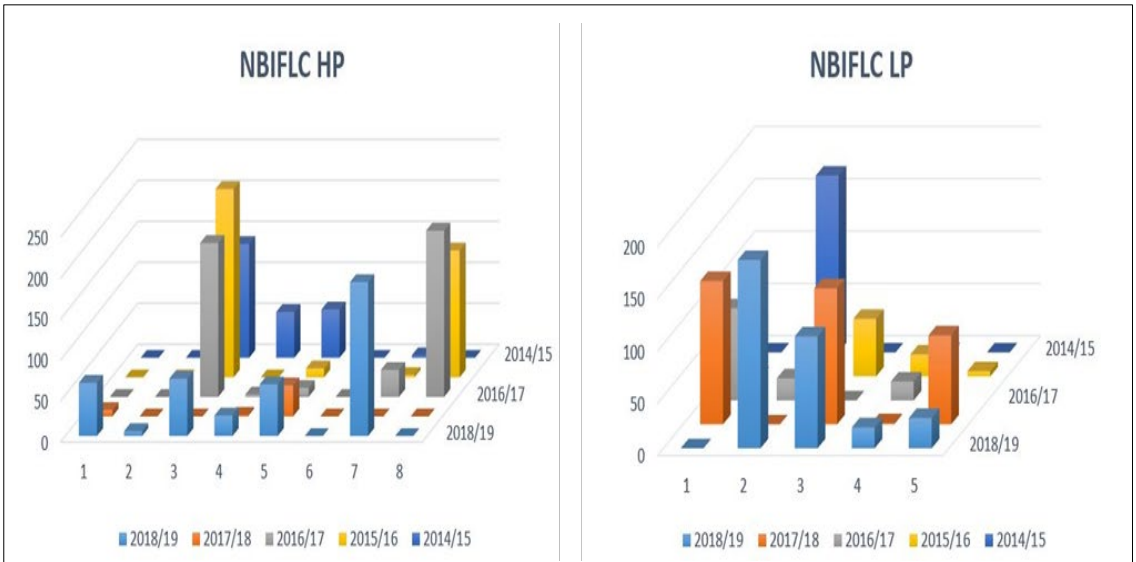

Figure 3.10: "NBIFLC" (number of benches in first level championships) variable, HP and LP cluster. X axis: teams; Y axis: benches; Z axis: seasons

Considering instead the median, also in this case we observe the same characteristics assumed by the values of the average, that is values very different from each other both for what concerns the clusters and for what concerns the seasons and the cluster of the LPs has higher values compared to the HP in 3 seasons out of 5. The maximum values also reflect this trend. If we look at the data distribution represented in the graphs we can see that in both clusters the distribution does not show a unique trend. Starting from the HP group we can clearly see that in the 2018/19 season there is an important

outlier that came seventh in that season. In the same season it can be seen how the sixth and seventh have no experience in first level championships and how the rest have experience in first level championships but with different levels between them.

The same can be said of the other seasons where there are important outliers in the season 2016/17 and 2015/16 and in the season 2017/18 we see how the level of experience in first level championships is generally low, if not zero. The situation in the LP cluster is also uneven. Finally, the values of the standard deviation assume generally high values in both clusters which for HP are 62.52, 13.22, 87.66, 90.36 and 50.27 while for LP are 75.02, 67.03, 38.63, 23 and 75.58.

For the variable that collects the experience accumulated by the coach in second-level championships ("NBISLC" - Figure 3.11) we find values for the standard deviation even higher than the previous variable, with values for the cluster of LPs higher than HP 3 times out of 5. As in the previous case, the distribution does not follow a defined trend but is very varied. In both clusters we notice the presence of outliers.

Considering the median, it can be noticed how it assumes different values over the 5 seasons for HP, while for LP it is similar in 3 out of 5 seasons. Respectively, the median for HP is 65, 31, 64, 85 and 42, while for LPs 184, 88, 77, 77 and 104. So it can be argued that over the five seasons considered, coaches who have coached a team in the LP

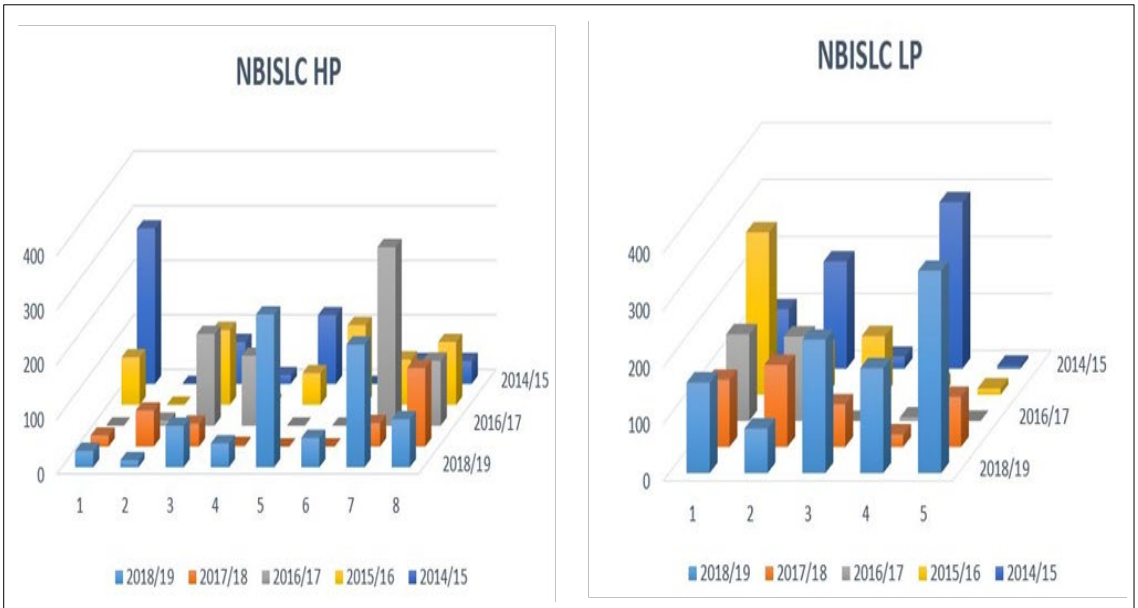

Figure 3.11: "NBISLC" (number of benches in second level championships) variable, HP and LP cluster. X axis: teams; Y axis: benches; Z axis: seasons

cluster have more accumulated experience in second-level leagues than colleagues of HP, having a value higher than the median in 4 out of 5 seasons.

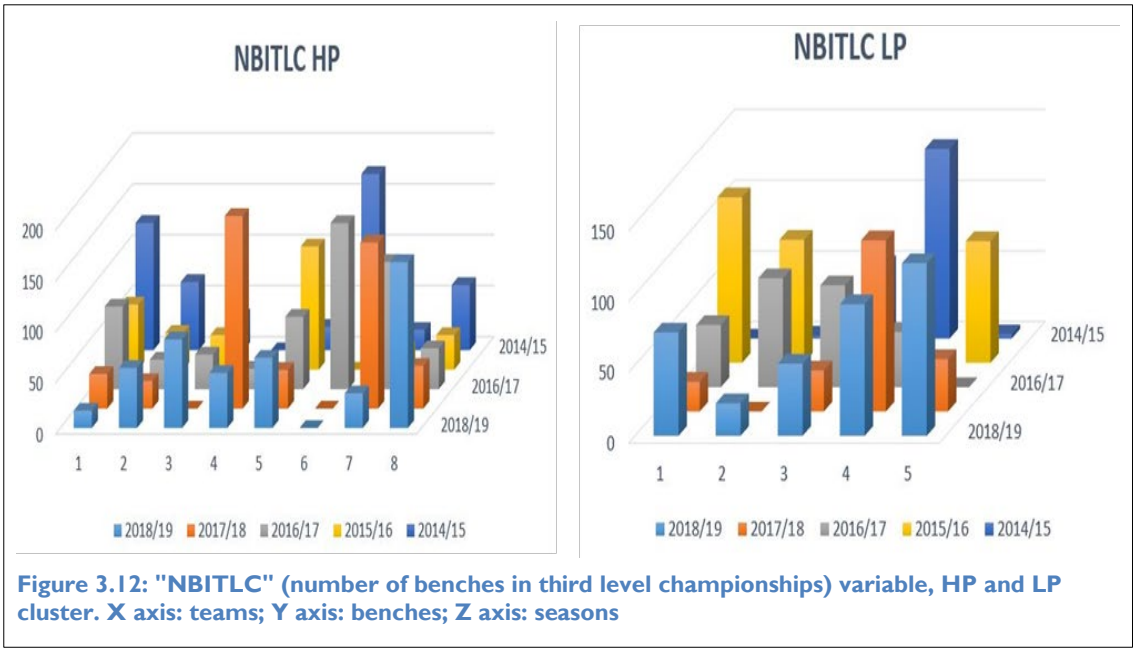

Figure 3.12: "NBITLC" (number of benches in third level championships) variable, HP and LP cluster. X axis: teams; Y axis: benches; Z axis: seasons

To conclude the descriptive analysis of the coach's experience as such, the accumulated experience in third-level championships will now be taken into account ("NBITLC"- Figure 3.12).

On average, there is no cluster that has higher values on a regular basis over the 5 years, in fact, the average for HP is 60.37, 61.62, 70.37, 38.62 and 63.12, while for LP 72.4, 41.6 58, 58.2 and 38.2. For the medians you can say the same thing even if in HP, there is more regularity of values in that cluster. Again, in both clusters and for all five seasons, the population distributions are very different. As for the other variables, we find also in this case high values for the standard deviation with 50,11, 72,7, 50,95, 39,3 and 58,89 for HP and 38,02, 46,47, 19,27, 54,13 and 57,83 for LP. It should be noted that the values for HP are in 4 seasons out of 5, higher than those of LP.

As previously mentioned, the second part of the data relating to coaches concerns their experience as players. The dataset presents the data relating to the cumulative experience in terms of attendance in all types of competitions. There are several cases of coaches who have not had a career as a professional footballer and this will affect the distribution of the population.

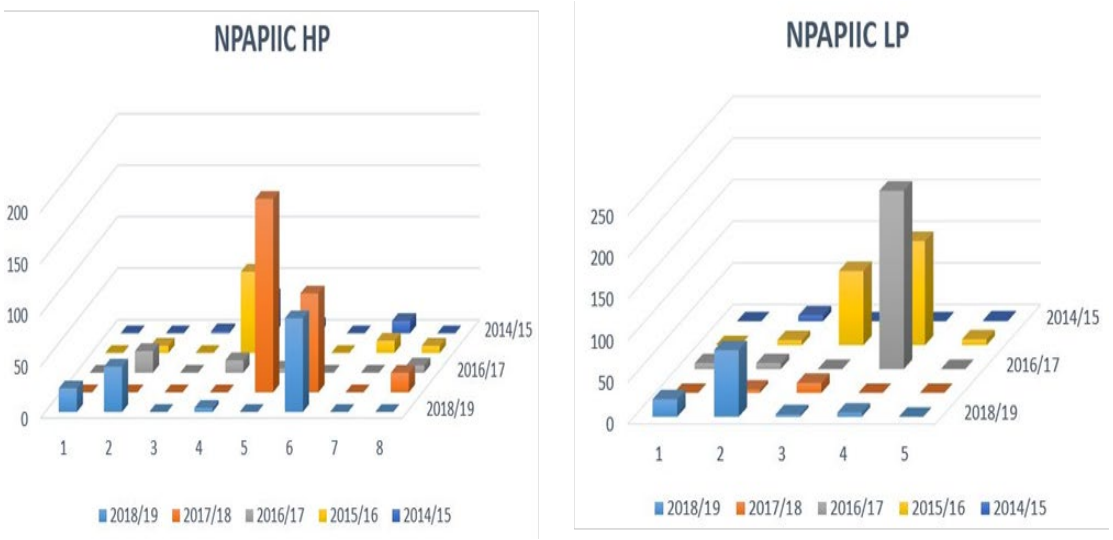

Figure 3.13: "NPAPIIC" (number of presences as a player in international competitions) variable, HP and LP cluster. X axis: teams; Y axis: presences; Z axis: seasons

The average number of players' coach's presences ("NPAPIIC" - Figure 3.13) in international competitions does not see values higher than one cluster than the other and values within groups vary greatly from season to season. As far as the median is concerned, the values are for HP 2, 0, 2, 3.5 and 1 while for LP 5, 0, 7, 7, 0. So even in this case there is no cluster with a clear superiority or inferiority in terms of values.

The standard deviation relative to the two populations assumes very different values over the course of the 5 years. Also in this case, as for the average and the median, there is not a cluster that has only ever higher or always lower values than the other over the years. If we look at the distribution represented in the bar charts we can see that there is little uniformity between seasons within the same cluster and the presence of outliers every year.

Considering the national championships, the distribution of the two clusters over the 5 years as far as the player coach's presence in first level competitions is concerned is represented in the figure 3.14 ("NPAPIFLC").

What can be seen in both clusters is a strong presence of outliers, both in terms of maximum and minimum values. The data dispersion represented by the standard deviation assumes very large and very different values over the course of the seasons for both clusters. Respectively, for HP, it assumes values of 135.11, 138.69, 119.13, 99.68 and 84.09, while for LP 79.4, 101.86, 161.82, 127.48 and 93.41. Again, it can be seen that

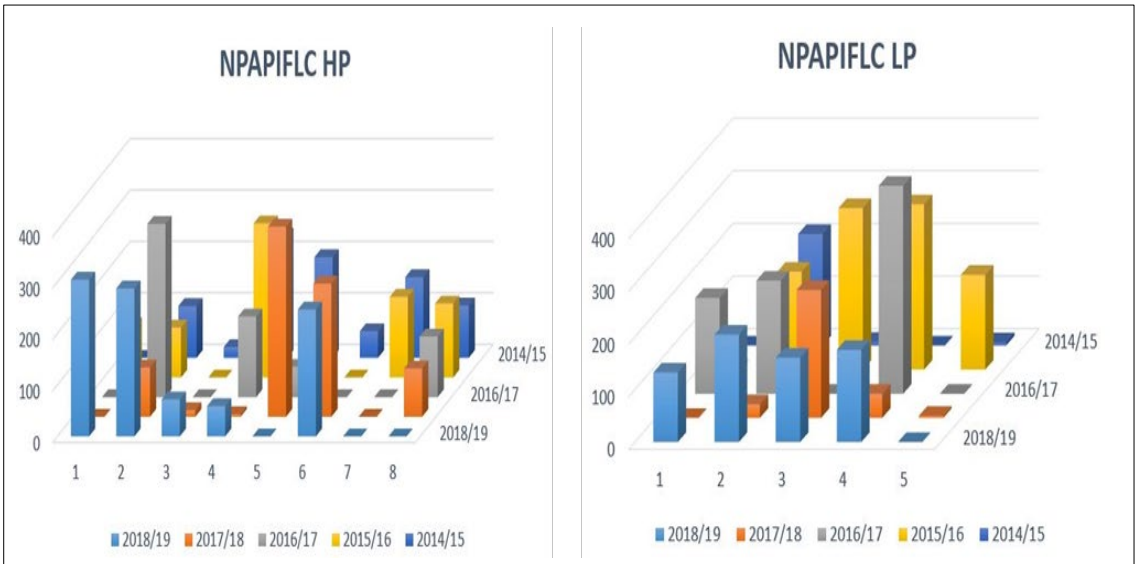

Figure 3.14: "NPAPIFLC" (number of presences as a player in first level championships) variable, HP and LP cluster. X axis: teams; Y axis: presences; Z axis: seasons

there are no constantly higher values for one cluster or another. It can also be deduced that the values assumed by the median are not uniform and that even for this value, there is not a cluster that presents in at least 4 out of 5 seasons, one value higher than the other.

The coach's player's presence in second-level leagues ("NPAPISLC") presents a four out of five seasons of higher average for the HP group than that of the LP. Respectively, the values for HP are 83.5, 87.37, 84.5, 153.25 and 160.125 while for LPs are 81.8, 94.4, 49, 74.6 and 34.8. If you look at the median instead we can see that it assumes higher values

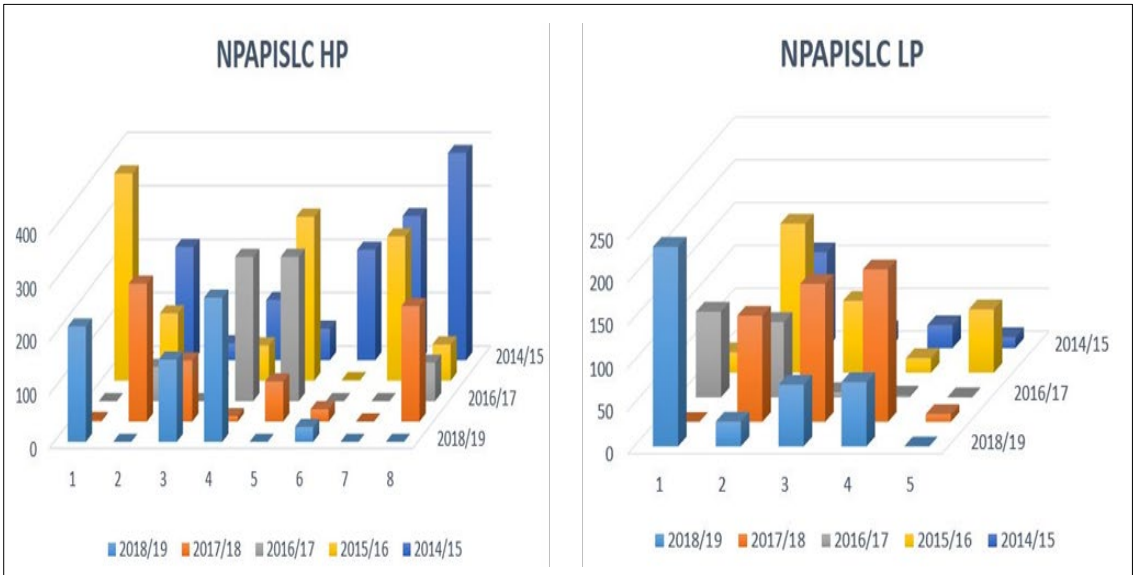

Figure 3.15: "NPAPISLC" (number of presences in second level championships) variable, HP and LP cluster. X axis: teams; Y axis: presences; Z axis: seasons

for HP in just two seasons out of five. The reason for this can be seen in the bar charts for the two populations (Figure 3.15). First of all, the first thing that can be seen is the different scale that is used for the two clusters. That of HP is a distribution that presents in each season considered an outlier with values much higher than the rest of the population.This would explain the gap between the values assumed by the average and those assumed by the median in the cluster analyzed. The population distribution for LPs is even more unbalanced. We can see the presence of important outliers in the 2018/19, 2015/16 and 2014/15 seasons. For the rest, there seems to be a rather symmetrical distribution in the 2016/17 and 2014/15 seasons, but with very different values. To give a dimension to the dispersion of data considering the standard deviation, for HP it assumes respectively 112.3, 101, 118.3, 148.64 and 132.77 values while for LP 90.11, 84.42, 52.22, 62.94 and 44.24. So in all seasons considered we see a much greater dispersion of data in HP compared to LP.

The same does not apply to the last variable taken into account for the experience of coaches, namely the experience of players in third-level leagues ("NPAPITLC"). In fact, for HP is worth 0.71, 15.62, 17.53, 29.76 and 26.52 while for LP 11.1, 18.15, 1.5, 8.14 and 25.42. Therefore, the dispersion relative to the data for this variable is very different between the years but assumes much smaller dimensions than those relative to the previous variable. In the HP diagram (Figure 3.16) we see in all five seasons the presence of outliers that influence the distribution and the average. In fact, the average for HP

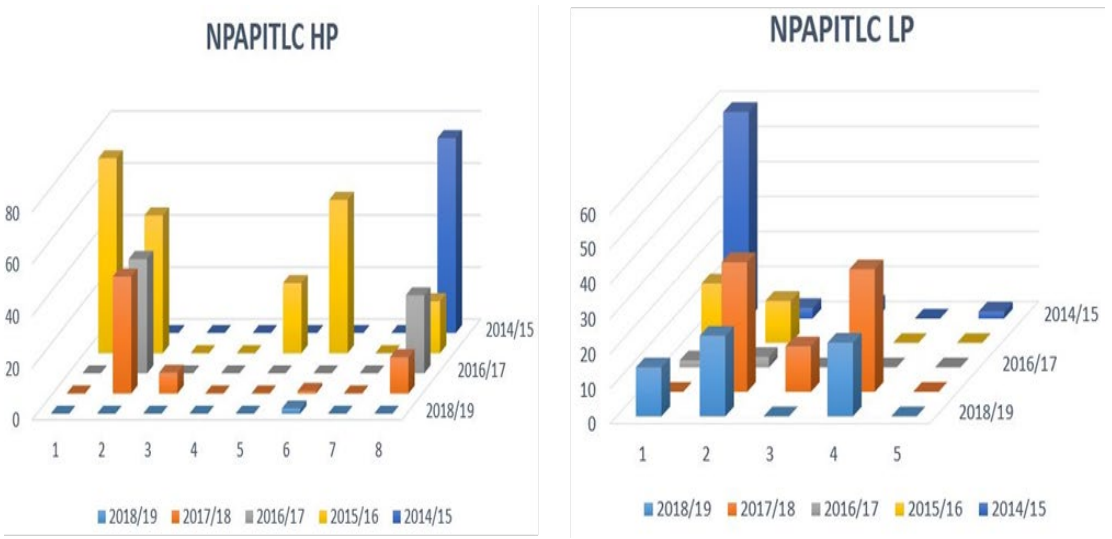

Figure 3.16: "NPAPITLC" (number of presences as a player in third level championships) variable, HP and LP cluster. X axis: teams; Y axis: presences; Z axis: seasons

assumes values very different from the median except in the case of the 2015/16 season. Respectively for HP the average assumes values of 0.25, 8.5, 9.25, 29.25 and 9.37 and the median assumes values of 0, 0.5, 0, 23.5 and 0. For LPs the situation is different. In fact, it can be seen the presence of an outlier in the 2014/15 season while for the latter, if present, have lower values. Therefore, the gap between the average and the median is less marked than in the other cluster. In fact, the values that the average assumes are 11.6, 17, 1.25, 5.8 and 13.6 while the median assumes values 14, 13, 1, 0 and 3. In general, in both clusters it can be seen how the population distributions in the various years are different from each other.

Summarizing the results obtained from the descriptive statistics of the coaches dataset, it emerged that practically no variable shows in at least 4 seasons out of 5 highest average values in either cluster. The only exceptions are the variables "NBISLC" (number of benches in second level championships) and "NPAPISLC" (number of presences as a player in second level championships). In the first case the LP cluster has a higher median in at least 4 out of 5 seasons while in the second case there are higher values for the HP cluster but not for the median. So there are no big differences between the two clusters in general.

3.3.3. TEAM DATASET

The variables relating to teams are mainly divided into two groups, one relating to the number of trophies per level of competition won by the team in its history and one relating to the number of participations at each level of competition.

The first variable collected is "NSACW" or "Number of Serie A championships won". The value of this variable in the HP cluster practically always assumes value 0 with the presence of some outliers in the course of the 5 seasons but however with very low values, given the nature of the variable itself (Figure 3.17). For example, the absolute maximum of Serie A championships won in the HP cluster is 5, recorded in the 2014/15 season. In the LP cluster instead the value assumed by the variable is always 0 (Figure

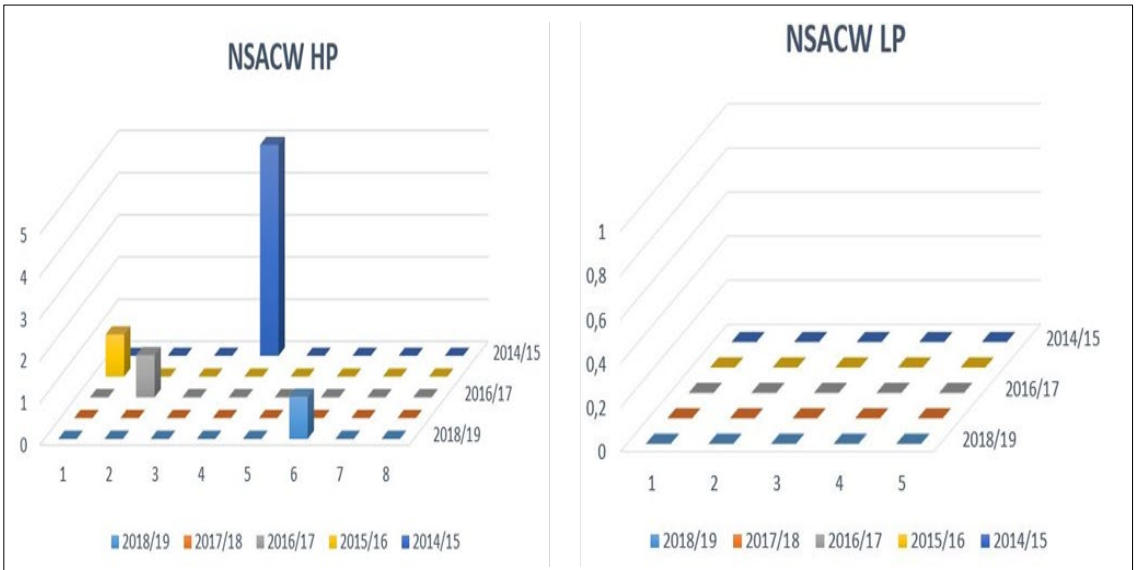

Figure 3.17: "NSACW" (number of serie A championships won) variable, HP and LP cluster. X axis: teams; Y axis: championships won; Z axis: seasons

3.17). This said above can be confirmed by reading the bar charts that in the case of HP have outliers, while in the case of LP we see that the values always assume value 0.

The variable "NSBCW", that is the one that records the number of Series B championships won by a team, presents a more diversified situation but also in this case, given the nature of the variable, the value of the data will be low. The average for the 5 seasons for the HP cluster is 1.75 in 2018/19, 1.37 in 2017/18, 0.75 in 2016/17, 0.75 in 2015/16 and 1 in 2014/15, while for the LP cluster it is 1.6 in 2018/19, 1 in 2017/18, 1.25 in 2016/17, 1.4 in 2015/16 and 1 in 2014/15. The values of the average in the two clusters

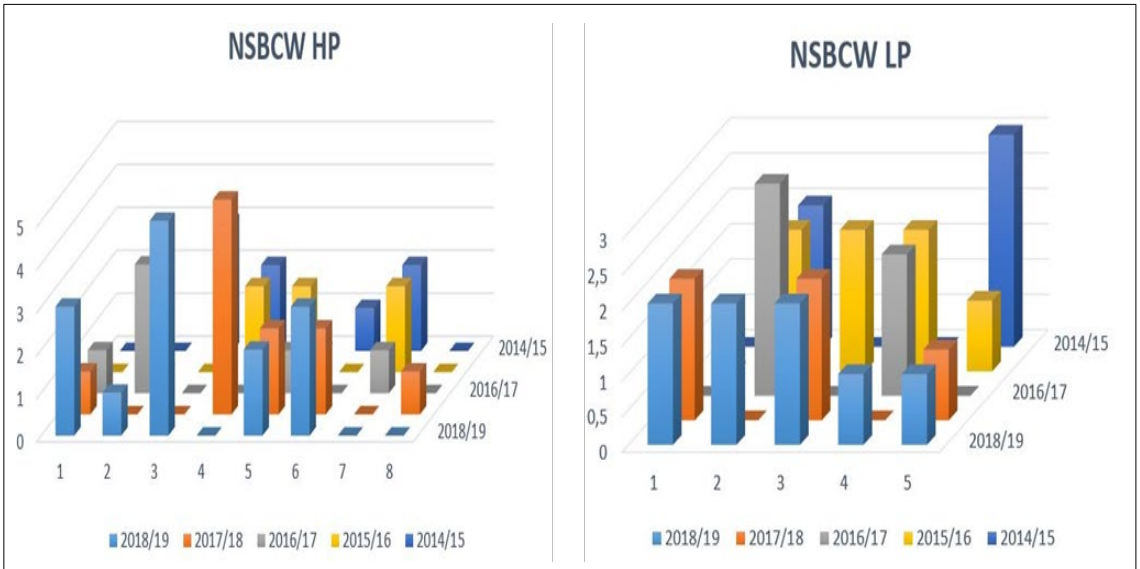

Figure 3.184:"NSBCW" (number of Serie B championships won) variable, HP and LP cluster. X axis: teams; Y axis: championships won; Z axis: seasons

do not differ much from those assumed by the median, in fact in the HP cluster assumes values 1.5; 1; 0.5; 0; 0.5 while in the second cluster has values 2, 1, 1, 2, 0. The maximum value assumed by the variable in the HP cluster is 5 which is repeated in the 2018/19 season and in the 2017/18 season, in 2016/17 and 2014/15 3 and in 2015/16 2. In the LP cluster there are values similar to those of the HP cluster but the maximum value assumed by the variable is 3 in 2016/17 and in 2014/15 while it assumes value 2 in the remaining seasons.

The data distribution represented by the diagrams shows in the HP cluster the presence of outliers in four out of five seasons (Figure 3.18). The LP cluster presents in the 2016/17 and 2014/15 seasons outliers in terms of maximum while the same thing happens in terms of minimum in the 2015/16 season. If we look at the standard deviation, in the HP cluster it assumes value 1.83 in 2018/19, 1.68 in 2017/18, 1.03 in 2016/17, 1.03 in 2015/16 and 1.19 in 2014/15 while in the LP cluster it assumes value 0.55; 1; 1.5; 0.89 and 1.41. From this it can be seen that in three out of five seasons this one assumes a higher value in the HP cluster than the LP one.

Continuing to analyze the number of trophies won by teams in their history, we moved on to analyze the amount of C series championships won by the teams analyzed, through the variable "NSCCW". The average value for this variable is higher in both clusters than in other similar variables. In fact, for the HP cluster it assumes value 2.25 in 2018/19; 3

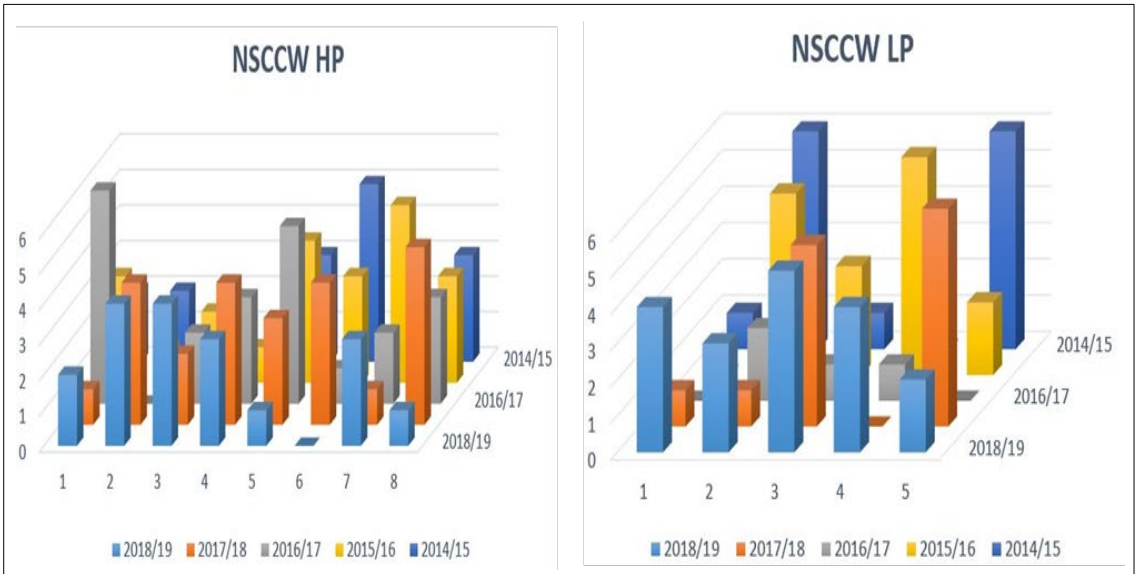

Figure 3.19: "NSCCW" (number of serie C championships won) variable, HP and LP cluster. X axis: teams; Y axis: championships won; Z axis: seasons

in 2017/18; 2.75 in 2016/17; 2.75 in 2015/16 and 2.25 in 2014/15 while for the LP cluster it is 3.6 in 2018/19, 2.6 in 2017/18, 1.5 in 2016/17, 3.2 in 2015/16 and 2.8 in 2014/15. From these values it can be seen that neither cluster has consistently higher or lower values in the five years. Considering the median instead, in the case of HP never deviates too much from the value of the average assuming value 2.5; 2.5; 2.5; 3 and 2. The same thing happens in the case of LPs but in three seasons out of five, namely in 2018/19, 2016/17 and 2015/16. Comparing the two clusters in terms of median value, the HP cluster has a higher value in 3 out of 5 seasons, in the 2015/16 season the values are equal, while in the 2018/19 season the value for the LP cluster is higher. Speaking instead of the maximum values, in the HP cluster this takes value 4, 5, 6, 5, 5, while in the LP cluster is equal to 5, 6, 2, 6, 6, so in four out of five seasons the maximum value is higher in the LP cluster than that of HP.

The distribution of the data represented by the diagrams (Figure 3.19) presents in practically all seasons outliers both in the sense of maximum and minimum. The HP cluster has a more uniform median value than the LP cluster and the same applies to population distribution. The characteristics expressed by the representations of the graphs are also found in the values assumed by the standard deviation. In fact, in the case of the HP cluster, this assumes more uniform values over the years (1.49 in 2018/19; 1.51 in 2017/18; 1.98 in 2016/17; 1.39 in 2015/16 and 1.39 in 2014/15) than the LP cluster (1.14 in 2018/19; 2.7 in 2017/18; 0.58 in 2016/17; 2.39 in 2015/16 and 2.95 in 2014/15).

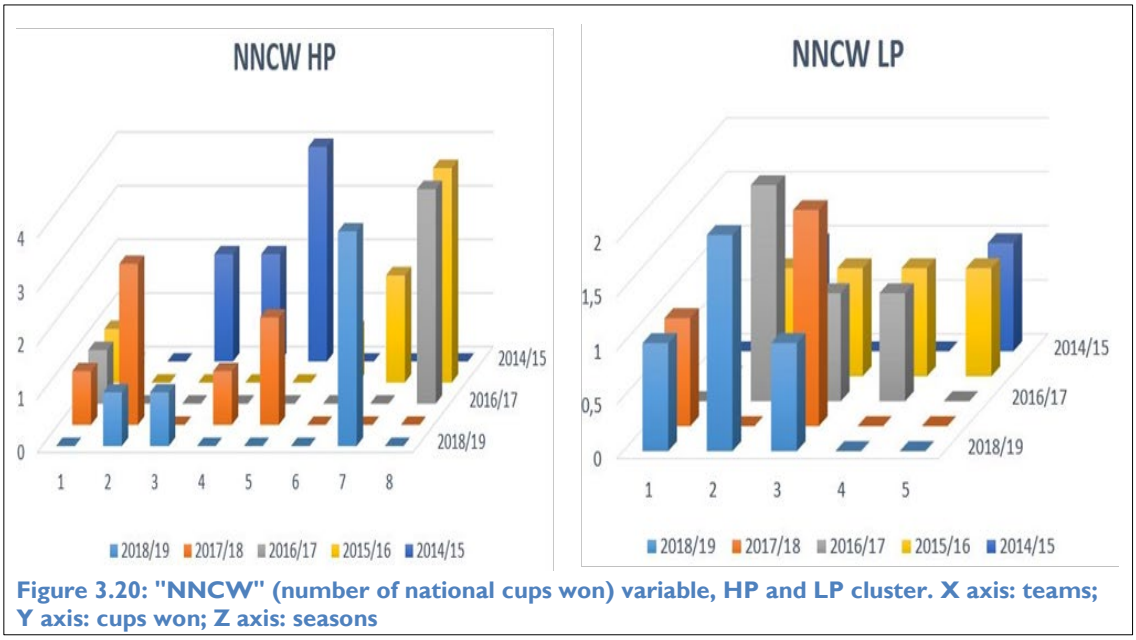

Figure 3.20: "NNCW" (number of national cups won) variable, HP and LP cluster. X axis: teams; Y axis: cups won; Z axis: seasons

Considering instead the distribution of the population relative to the variable "NNCW", i.e. the number of national cups won, the graphs show the distribution plotted in the diagrams in figure 3.20. The standard deviation for the two clusters assumes values of 1.39; 1.12; 1.41; 1.41; 1.51 for the HP cluster and for the LP cluster 0.7; 0.8; 0.67; 0.2 and 0.3. In all seasons considered, the value in the HP cluster is higher than that of the LPs, indicating a more uneven distribution in the first case. This is also testified by the different scales that the graphs assume. Moreover, this difference can be explained by looking at the maximums of the two clusters, where in the case of the HP cluster it assumes value 4 in four seasons out of five, except in the 2017/18 season where it assumes value 3, while in the LP cluster it assumes value 2 in the 2018/19, 2017/18 and 2016/17 seasons and value 1 in the remaining. On average the teams belonging to the HP cluster have won more national cups than the teams in the LP cluster, except in the 2018/19 season. Instead, the value assumed by the median seems to say the opposite. In fact, the median assumes a value of 0; 0.5; 0; 0.5; 0 for the HP cluster while for the LP cluster it is 1; 0; 1; 1 and 0. This difference between the average and the median can be explained by the afore mentioned maximum values and by the presence of important outliers in the HP cluster.

Considering the last variable that deals with the trophies won by the teams, international competitions will be considered. In this case the descriptive statistics for the variable "NICW" will be similar to those for the variable "NSACW". In fact, even in this case, the

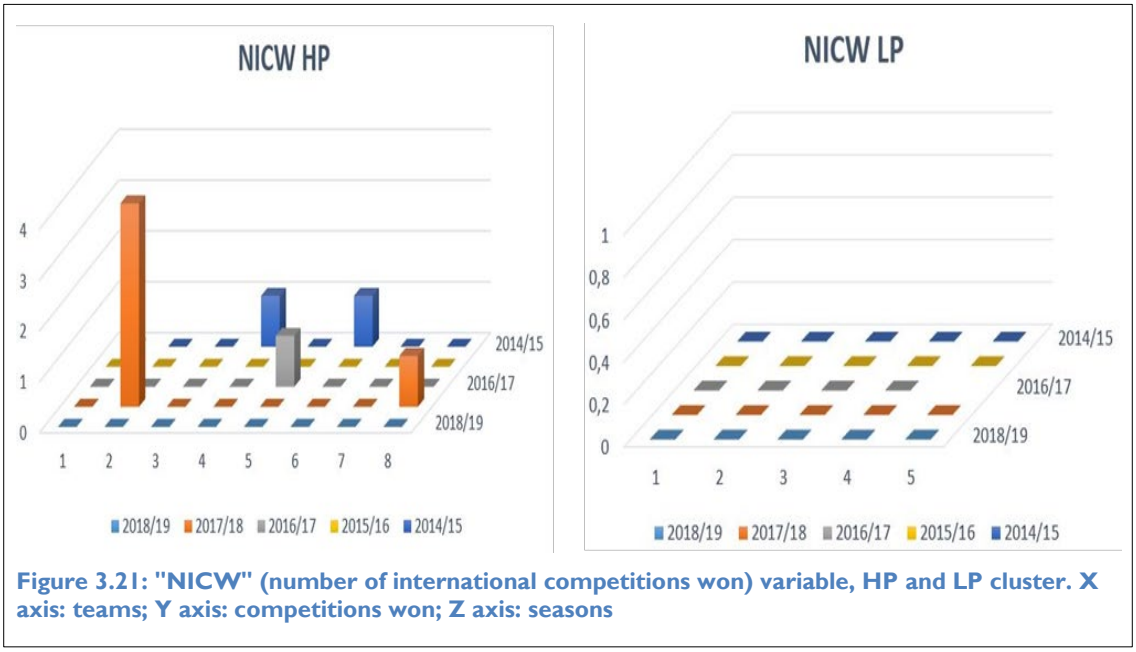

Figure 3.21: "NICW" (number of international competitions won) variable, HP and LP cluster. X axis: teams; Y axis: competitions won; Z axis: seasons

value of the variable that can be changed is practically always 0. In the LP cluster the variable is always zero while in the HP cluster it always assumes zero in the 2018/19 and 2015/16 seasons but has outliers in the remaining ones. From the diagrams (Figure 3.21) it is possible to see these descriptive traits of the population.

Moving on to the analysis of the group of variables relating to the number of participations that each team has had in a given competition, we will begin by considering the number of participations in an international competition, i.e. the variable "NPIIC" (Figure 3.22).

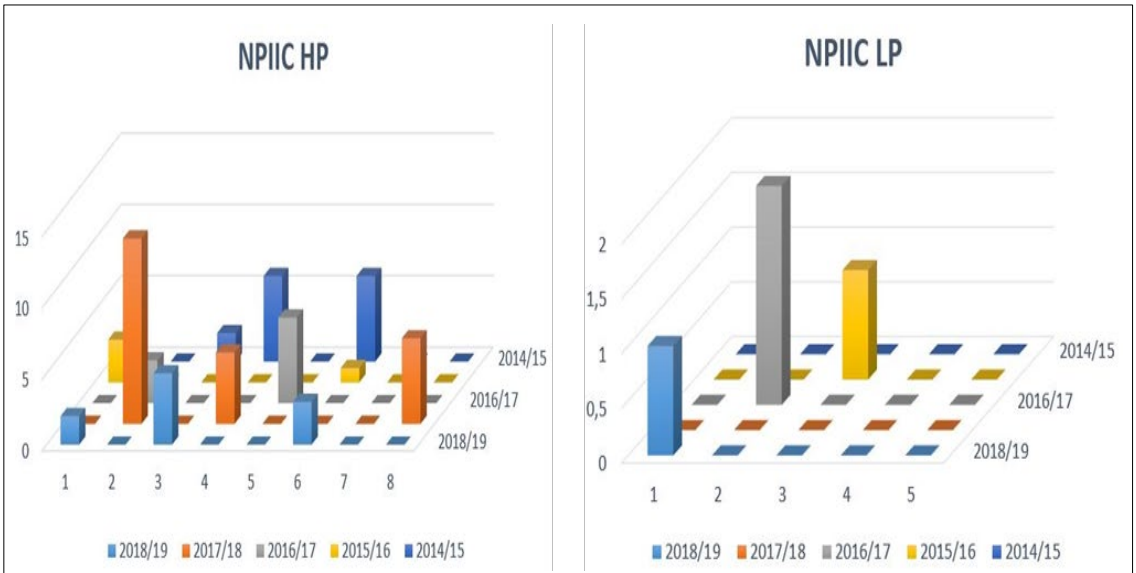

Figure 3.225: "NPIIC" (number of participations in international competitions) variable, HP and LP cluster. X axis: teams; Y axis: participations; Z axis: seasons

Starting from the consideration of the value assumed by the average, this assumes in the cluster of HP value 1.25; 3; 1.12; 0.5 and 1.75 while in the cluster of LP value 0.2; 0; 0.5; 0.2 and 0. Therefore, on average HP teams have had in their history a greater number of participation in official competitions organized by UEFA. However, this is denied by the value that the median assumes in the five seasons for both clusters, in fact the value of this is always 0. This may be due to the presence of valuable outliers in the HP cluster. In fact, the maximum value in the cluster assumes value 5; 13; 6; 3 and 6 while in the case of the LP the value is 1; 0; 2; 1 and 0. The presence of these imbalances in the population considered is also witnessed by the values assumed by the standard deviation, in fact for the HP cluster assumes value 1.91; 4.75; 2.23; 1.07 and 2.71 while in the cluster of the LP assumes value 0.45; 0; 1; 0.45 and 0. Analyzing the diagrams we can see how

first of all the scale of the two clusters is very different and how in the HP cluster there are many outliers of significant value, indicating in fact a distribution of the population much more uneven compared to the other cluster (Figure 3.22).

If we consider the variable that records the number of participations in Serie A of a team, that is the variable "NPISA", from the diagrams we can see the presence of outliers in both clusters but the distribution seems more homogeneous in the HP cluster. The LP cluster, on the other hand, presents a much more uneven situation even if the outliers have much lower values than those of HP and this is also demonstrated by the different scale of the two diagrams.

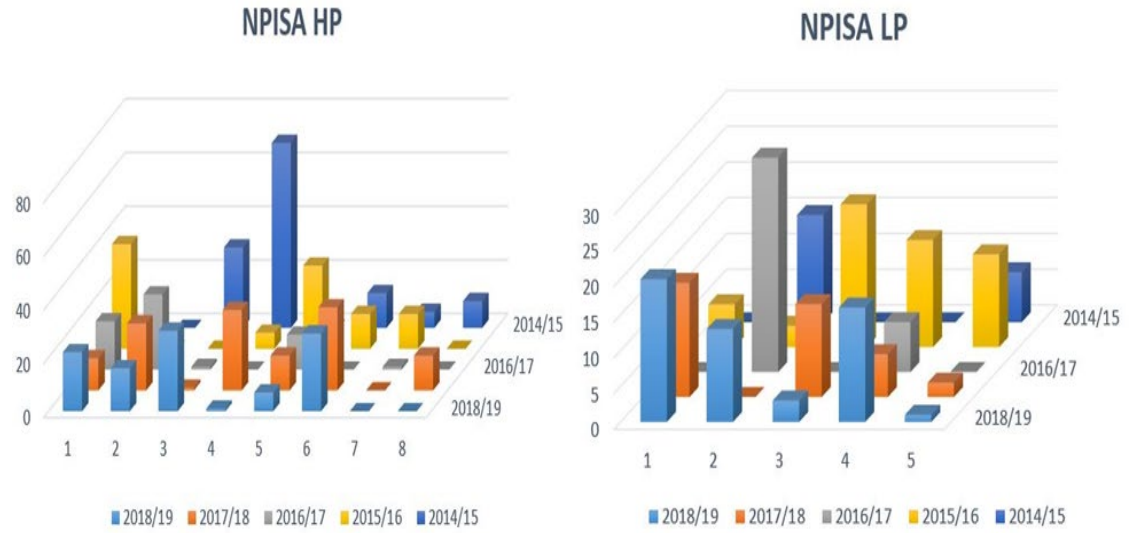

Figure 3.23: "NPISA" (number of participations in Serie A) variable, HP and LP cluster. X axis: teams; Y axis: participations; Z axis: seasons

In this regard, the standard deviation assumes values of 12.83; 12.05; 10.78; 14.91 and 23.66 in the HP cluster while for the LP cluster it is 8.26; 6.91; 14.22; 6.88 and 6.65. The consistently lower values in the LP cluster, despite the fact that from the diagrams it seems that the distribution of the population is more inhomogeneous than that of the HP cluster, may be due to the value of the outliers (Figure 3.23). For this reason it is interesting to look at the maximum and minimum values of this variable. In the HP cluster the minimum always assumes value 0 while the maximum value assumes value 30; 31; 28; 39 and 69 while in the LP cluster the minimum value is 0 in the seasons 2017/18, 2016/17 and 2014/15 while in 2018/19 it assumes value 1 and in 2015/16 value 3. If we consider instead the central trend indices, i.e. the average and the median, these assume values respectively 13.125; 15.625; 7.625; 12.75 and 16 for what concerns the average

of HP, instead for the LP is 10.6; 7.4; 9.25; 11.4 and 4.4. On the other hand, the median for HP is 11.5; 13; 1; 9.5 and 8, while for LP 13; 6; 3.5; 13 and 0. As far as the average is concerned, the HP cluster has a higher value than the LP in all seasons except the 2016/17 season. As for the median, on the other hand, it has a higher value in the HP cluster for three out of five seasons. To report the difference between the average and the median in the HP cluster in the 2014/15 season, due largely to the value of the maximum in that season, namely 69.

The maximum assumed by the variable "NPISB" (number of participations in Serie B championships) in the HP cluster is 60, while for the LP is 49. In general in the HP cluster the minimum value is 1; 7; 0; 2 and 1 while the maximum is 60; 45; 52; 43 and 35. For the LP cluster instead, the maximum value is 38; 34; 34; 49 and 48 while the minimum value is 4; 3; 3; 11; 0.

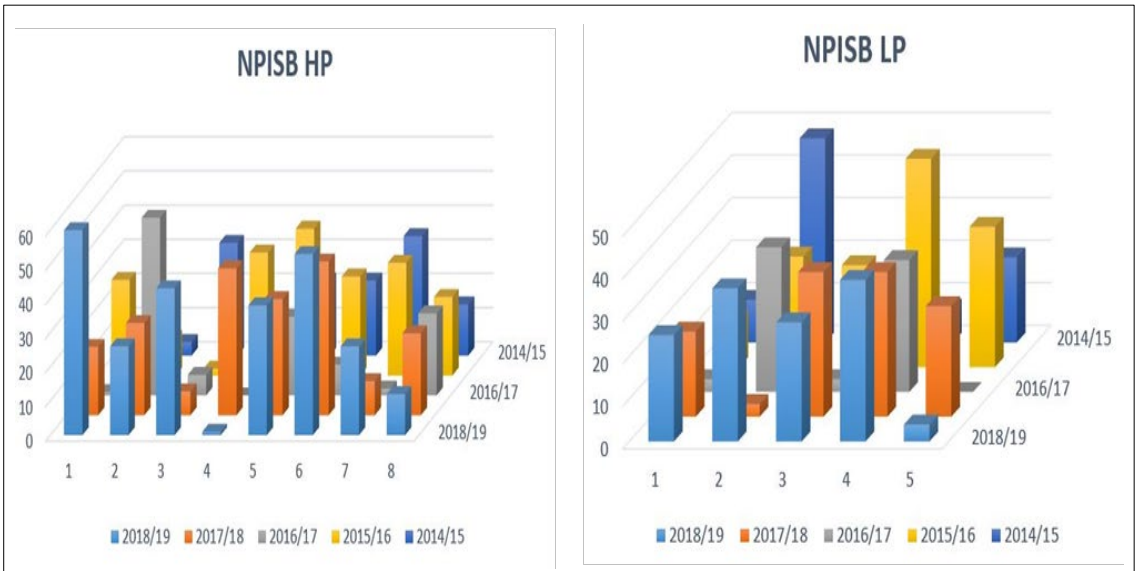

Figure 3.24: "NPISB" (number of participations in Serie B) variable, HP and LP cluster. X axis: teams; Y axis: participations; Z axis: seasons

Looking at the diagrams (figure 3.24) it is possible to see also in this case the presence of outliers in both clusters, both in terms of maximum and minimum. The standard deviation values are 20.05; 14; 17.81; 13.41 and 12.47 for the HP cluster and 13.53; 12.84; 17.07; 13.9 and 18.63 for the LPs. As we can see in three out of five seasons the value is higher for the HP cluster and this may be due to the maximum values that are higher in the HP in four out of five seasons. Considering instead the averages, in three out of five seasons the value is higher for HP while for the median the same thing

happens. Respectively the average is higher for the HP cluster in the seasons 2018/19; 2017/18 and 2014/15 while the median is the same thing.

The last variable records the number of participations in third-level championships. The variable "NPISC" averages 21.5, 21.37, 26, 23.25 and 24 for the HP cluster, while in the case of LP the value is 34.2, 23.6, 24.5, 31.8 and 23. As can be seen, in three out of five seasons the value is higher in the LP cluster. Considering instead the median as far as HP is concerned, the values do not differ much from those of the average, assuming values of 18; 18.5; 31.5; 23.5 and 27.5. The same is true for the LP cluster where the values are 31; 21; 25; 32 and 25.

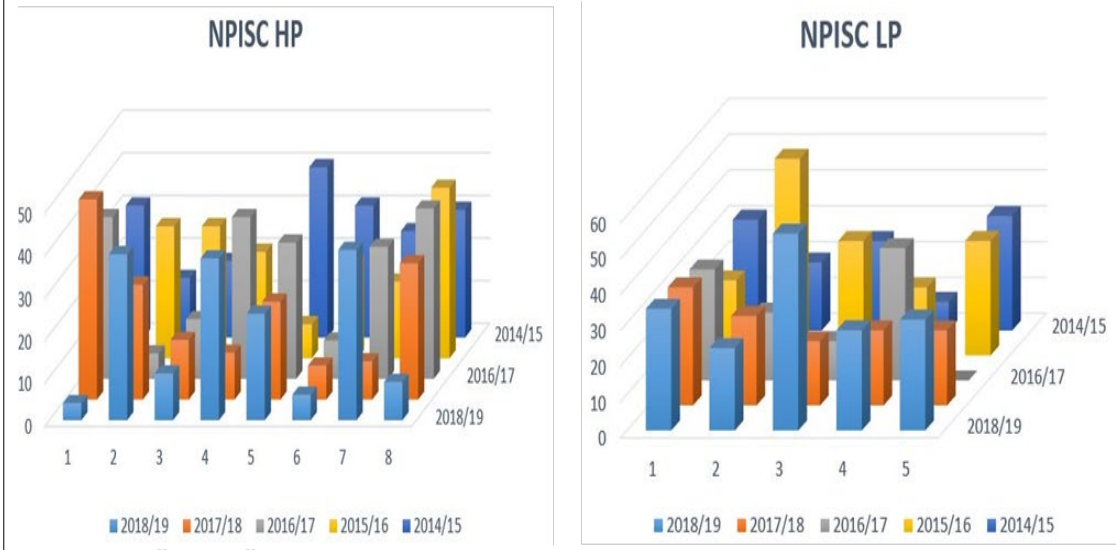

Figure 3.25: "NPISC" (number of participations in Serie C) variable, HP and LP cluster. X axis: teams; Y axis: participations; Z axis: seasons

From the diagrams (Figure 3.25) it can be seen how the two clusters have two different scales. This is due to the maximum values of the two clusters. In the HP cluster the maximum values are 40 for all seasons except for the 2017/18 season where the value is 47. The minimum values for this cluster are 4; 8; 6; 8; 3. For the LP cluster instead the values are 55; 33; 37; 55 and 32 while the minimum values are 23; 18; 11; 19 and 8. As can be seen, the minimum values in the LP cluster are always higher than those in the HP cluster. Always analyzing the diagrams we can see the presence of outliers in all the years of both clusters. The standard deviation is in four out of five seasons higher for the HP cluster than for the LP cluster. Respectively this happens in all seasons except in the 2015/16 season.

Summarizing the results of the descriptive statistics made on the teams dataset, it can be seen that in almost all the variables there is not a variable with higher average values in at least 4 out of 5 seasons except for the variables "NNCW" (number of national cups won), "NPIIC" (number of participations in international competition) and "NPISA" (number of participations in Serie A), which show higher average values for the HP cluster but not for the median values.

3.4. PCA METHODOLOGY

The main idea behind the PCA is to reduce the dimensionality of datasets consisting of a large number of interrelated variables trying to keep the variation as much as possible within it. This effect is obtained through the transformation of the original variables in PCs that are independent and are created in such a way that the first PCs explain most of the variance present in all the original variables (Jolliffe, 2002).

This methodology has been applied over the years to many fields such as genetics, geology, psychology, economics, biology, agriculture, chemistry, etc.. One of the first descriptions of the PCA was made by Pearson (1901). In his work Pearson tried to find a set of line and planes that could best describe a set of points in a p-dimensional space (Jolliffe, 2002). The second father of the PCA can be considered Hotelling (1933). To get to what the PCA is used for, you need to follow some steps (Jeffers, 1967):

- Choice of the variables to be included in the analysis;
- Construction of the basic data matrix;
- Transformation of the basic data, if required;
- Calculation of the dispersion or correlation matrix;
- Calculation of eigenvalues and eigenvectors of the dispersion or correlation matrix;
- Examination and interpretation of the eigenvalues;
- Interpretation of the eigenvectors;
- Calculation of the transformed values;
- Plotting or further analysis of transformed values.

To better understand how the methodology of the PCA works, suppose that **x** is a vector of **n** random variables that have a variance and a structure of covariance or correlation between them that is of interest. Normally we use the covariance matrix when the scale of the variables is similar while the correlation matrix is used when we have a different scale. In this case, for the players' dataset will be used the correlation matrix (CrossValidated Website, 2019).

Using the correlation matrix is equivalent to standardizing each of the variables (to mean 0 and standard deviation 1). In general, PCA with and without standardizing will give different results. Especially when the scales are different (CrossValidated Website, 2019). In this particular study we will use the correlation matrix because the variables have different scale caused by their nature. So unless the number of **n** is very small, many times it may not be very useful to look only at n variance and all **1/2*n*(n-1)** correlations or covariances (Jolliffe, 2002). To avoid this, it may be useful to look at a small number of derived variables that preserve the information in the dataset as much as possible.

First we create a linear function **a1'x** of elements of **x** with maximum variance, where **a1** is a vector with n constants **a11, a12, ..., a1n** and **" ' "** denotes transpose, so that:

$$\boldsymbol{\alpha}_1'\mathbf{x} = \alpha_{11}x_1 + \alpha_{12}x_2 + \cdots + \alpha_{1p}x_p = \sum_{j=1}^{p} \alpha_{1j}x_j.$$

After finding the first linear function we look for the second linear function **a'2x**, independent from the previous one but created following the same criteria. This happens for a number of linear functions equal to or less than the number of the original variables. Although we have a number of PCs equal to or less than the number of original variables, the aim is to find a number of variables s<<n (Jolliffe, 2002).

To better understand how the PCA works we can start by considering a small dataset that for example contains measurements of two genes in a population of 6 mice. Obviously the advantage of having n=2 is that the data can be represented exactly in two dimensions (Starmer, 2018-Table 3.2).

So if we consider the gene number 1 for the 6 observations, we can represent the data on a straight line and if we include in the representation also the second gene then we can use a canonical two-dimensional graph. Obviously, based on the number of variables n, we will have a number n of chart sizes. One of the functions of the PCA is just to represent a number of

	Mouse 1	Mouse 2	Mouse 3	Mouse 4	Mouse 5	Mouse 6
Gene 1	10	11	8	3	2	1
Gene 2	6	4	5	3	2.8	1

Table 3.2: Genes and mice data set. Source: Starmer, 2018.

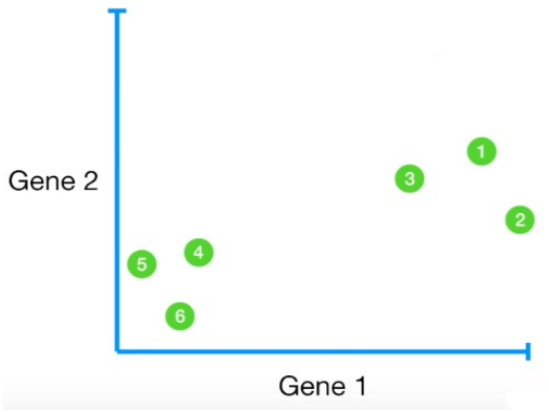

Figure 3.26: Rapresentation of observations of Genes and mice dataset. Source: Starmer, 2018

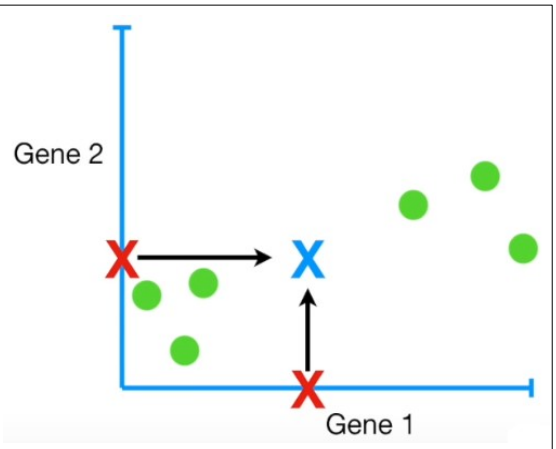

Figure 36.27: Central value of dataset. Source: Starmer, 2018

n-avriables in a two-dimensional graph (Starmer, 2018; Jolliffe, 2002). So the first thing is to represent the observations in the graph. Then we calculate the average of the measures of the distance of the observations from the axis of the x for the first gene and from the y for the second. The average of values is then used to calculate the central value of the data. Once the central value of the data has been found, it must be "moved" so that the central value corresponds to the origin of the axes (Figure 3.27). This moves the data within the graph but does not change the position of the data relative to each other. Once the move has taken place, it must be found a line that describes the data in the best possible way. The PCA methodology can calculate the distance between the points and the line and minimize it, or maximize it. These two solutions are equivalent and usually the PCA methodology maximises the distance between the points and the line. From a mathematical point of view, this can be explained using the Pythagorean theorem (Figure 3.28). Consider a point in the graph that is fixed as well as its distance from the origin, so this does not change depending on the movement of the "best line". So if we project the point

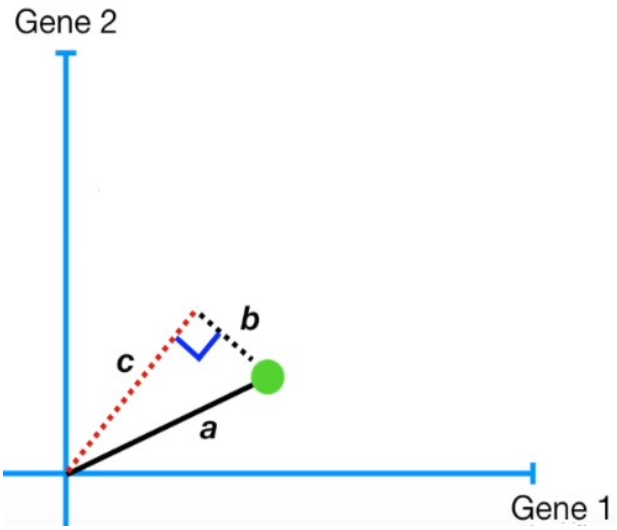

Figure 3.28: Matematical explation of PCA. Source: Starmer,2018

on this line, we will have a rectangular triangle that will have as hypotenuse "a" that is the distance between the point and the origin, as "b" the projection of the point on the "best line" and as "c" the "best line". So "b" and "c" are inversely proportional to each other. The PCA maximizes or minimizes "b" depending on the "best line". As previously mentioned, the PCA maximizes the sum of the square of the distance of the projected point from the origin. This happens for each point in the graph. Once the distances have been calculated, they are doubled so that the negative values do not compensate for the positive ones. This gives us the sum of the square of the distance ie SS. Considering this sum we find the best fitting line called Principal Component 1. Being a straight line, it will have a slope. So for example if the slope is 0.25 then for every 4 units we have along the x-axis then we have one unit on the gene 2 axis. This means that most of them are mostly distributed on the axis of gene 1. So the ratio of gene 1 relative to gene 2 tells us that gene 1 is more important as far as the description of the distribution of the data is concerned. Always taking the same example into consideration, the "recipe" for PC1 is called "linear combination" of gene 1 and 2. Knowing the values to form the linear combination, it can be used Pythagoras' theorem again to calculate the length of the "best line". In this case the length will be 4.12. If, on the other hand, the data were to be standardised, the linear combination would result in the length being 1. So in this case it is sufficient to divide each side by 4.12, in this way the values change but the ratio remains the same. This vector of length 1 consists of about 0,97 parts for the gene 1 and about 0,242 for the gene 2, it is called "eigenvector" while the proportion of each gene is called "Loading score". Considering again the SS distance for PC1, this is called "eigenvalue for PC1" and the square root of the eigenvalue is called "singular value for PC1" (Starmer, 2018; STHDA Website, 2019). Dwelling on the eigenvalue, we can say that this measures the amount of variation that explains each PC. Therefore the eigenvalues will be larger for the first PCs and smaller for the following ones. Therefore the eigenvalues are taken in consideration in order to determine the number of PCs to consider (Kaiser, 1961). Usually the total of eigenvalues values equals the number of original variables. Going down specifically, if a eigenvalue ≥ 1 means that the PC accounts for a variance equal to or greater than that of an original variable if the data are standardized (Kaiser, 1961).

Another way to choose the number of PCs to consider is to use the total accumulated variance (Kaiser, 1961).

Going back to the previous example, if using the graph, we have to find the second PC, just find the line perpendicular to the PC1. So in this case the "recipe" is -1 for gene 1 and 4 for gene 2, so standardizing the data, the eigenvector for PC2 will have value -0.242 for gene 1 and 0.97 for gene 2. So to get the final graph, once we get the 2 PCs, we have to rotate the two "best lines" until the PC1 becomes horizontal, after which we can use the points to see how the observations are distributed. Also in this case to obtain the eigenvalues it will be sufficient to project the data on the main components and then measure the distances from the origin and double them and add them together.

Once found they can be converted into variation around the origin by dividing by the sample size minus 1. In this way we find the variation for PC1 and PC2. If, for example, the variation for PC1 is equal to 15 and that for PC2 is equal to 3, then the total variation is equal to 18. This means that PC1 represents 15/18=0.83=83% of the total variation, instead PC2 will be 3/18=0.17=17% of the total variation. A "scree plot" can be used to represent these percentages. If the number of PCs is more than two PCs it can be selected the ones that represent most of the variation. For example, if we have 3 PCs and the first represents 79% of the variation, the second 15% and the third 6%, then if we consider only the first two we will have a representation of 94% of the variation. In this way we will have a good approximation and we could still represent the data on a two-dimensional graph. If the number of genes exceeded the 3 units, for example they were 4, then it would be practically impossible to draw the graph. This is not important because we can still use the mathematical operations of the PCA and look at the scree plot to choose the number of PCs to consider. In case there are many PCs describing a small part of the total variation then the choice could be more difficult.

In general, the uses that can be made of the PCA are as follows (Jeffers, 1967):

- The examination of the correlations between variables of a selected set;
- The reduction of the basic dimensions of the variability in the measured set to the smallest number of meaningful dimensions;
- The elimination of variables which contribute relatively little extra information

- The examination of the grouping of individuals in n-dimensional space;
- Determination of the objective weighting of measured variables in the construction of meaningful indices;
- The allocation of individuals to previously demarcated groups;
- The recognition of misidentified individuals;
- Orthogonalization of regression calculations.

Depending on the dataset analysed and the study, these objectives will not be of equal importance and in some cases will not be present.

For our study we will use the PCA methodology for reduce the basic dimensions of the variability in the measured set and for the elimination of variables which contribute relatively little extra information in the definition of the main characteristics of HP and LP clusters.

3.5. EMPIRICAL ANALYSIS THROUGH PCA METHODOLOGY

In this paragraph the methodology called PCA will be used through the R programme for each year taken into consideration, in the players' datasets for both clusters. First we will try to analyze the two clusters together to try to see if the projection of the individuals on the graphs of the two first PCs will show a clear distinction between the observations of the two clusters. Then we will analyze the two cluster separately. The objective is to find similarities within a cluster, in at least 4 out of 5 seasons, with respect to the composition of the main components that explain a cumulative amount of variance as close as possible to 80%. If similarities are found then this means that the factors that occur with some regularity within PCs are those that affect the HP cluster or the LP cluster. We will start to analyze the dataset for players with the HP cluster. Then we will move on to the analysis of players related to the cluster of LPs. For the other two dataset we will stop at the descriptive statistics and in the fourth chapter i twill be proposed a model for the analysis.

3.5.1. PCA ON PLAYERS' DATASET

As said in the previous paragraph, the analysis of the players dataset will be done in two ways, one using the PCA on the whole dataset to see if in the graph of the first two PCs there are two distinct clusters representing the HP cluster and the LP cluster and the other doing the PCA on the distinct datasets to see if in 4 out of 5 seasons the same factors are repeated within a cluster and if they are different between clusters.

Starting from the 2018/19 season, considering the two clusters, the first two PCs explain respectively 31.3% and 24% of the total variance, for a total of about 55.3%. The first PC is composed for a little less than 30% by the variable "NPIFLC" (number of presences in

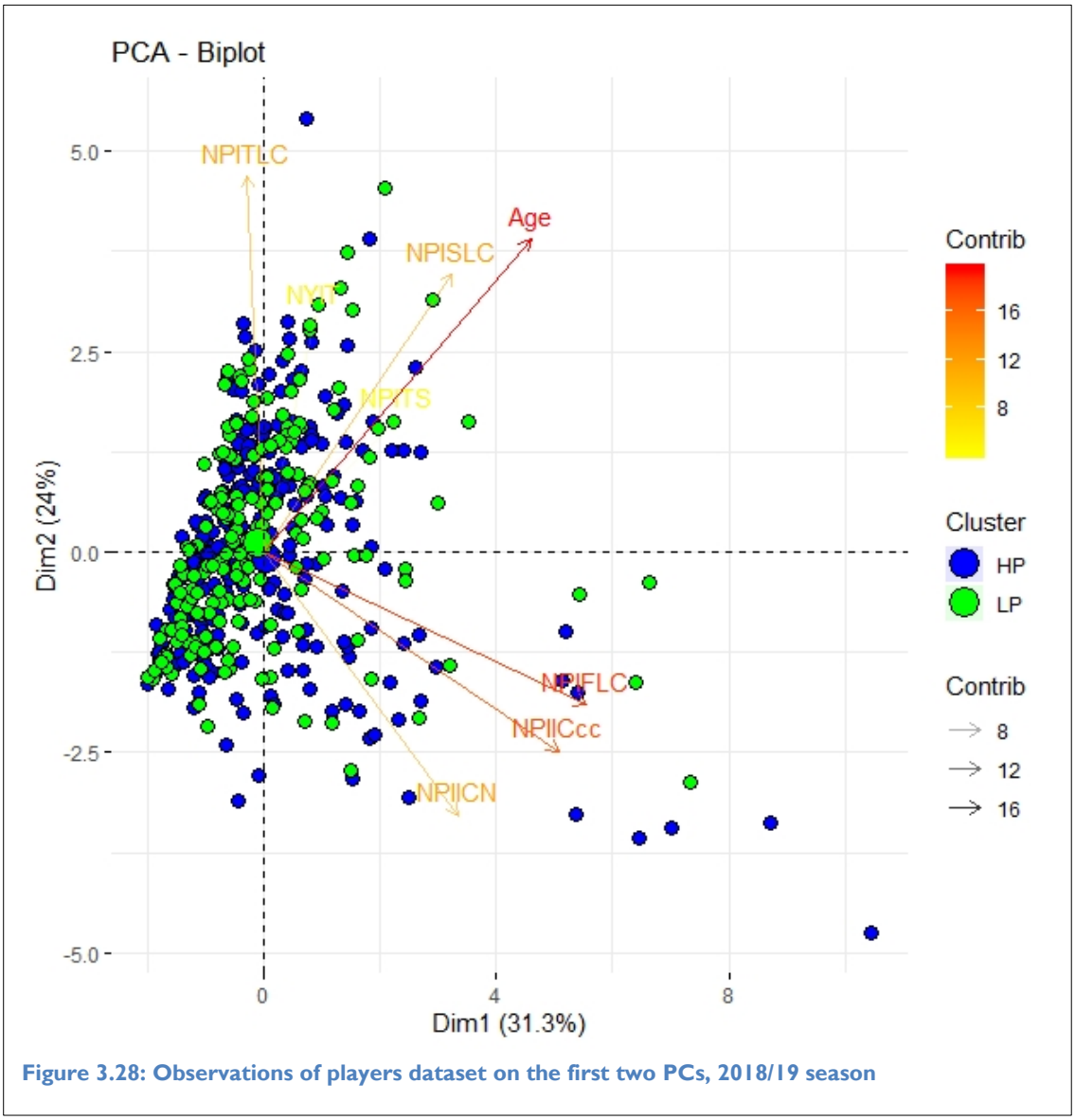

Figure 3.28: Observations of players dataset on the first two PCs, 2018/19 season

first level championships), followed by the variable "NPIICcc" (number of presences in international competitions (club competitions)) with a little less than 25% contribution and finally the variable "Age" with about 20% contribution. For this dataset the variables with a contribution of less than 12.5% will not be considered, i.e. the contribution that each variable should give if the distribution and population characteristics of each variable were equal. As far as the second PC is concerned, this is formed by about 27.5% by the variable "NPITLC" (number of presences in third level championships), a little more than 17.5% by the variable "Age" and a little less than 15% by the variable "NPISLC" (number of presences in second level championships). Finally, just above the minimum contribution threshold there is the variable "NPIICN" (number of presences in international competitions (national competitions). Considering them together we can see that the variable that contributes most to the creation of the first two PCs is "Age" with a little less than 20%, followed by the variable "NPIFLC" (number of presences in first level championships) with a little more than 17.5% contribution and finally for a little less than 17.5% by "NPIICcc" (number of presences in international competitions (club competitions)).

Looking at the projection of the individuals on the first two PCs (Figure 3.28), we do not see a clear separation of the two clusters between LP and HP. The center of the two clusters, represented by the larger points are close to the origin of the axes and the representation shows that they overlap. Most of the observations for HP are distributed between the first and the fourth quadrant but especially in the fourth quadrant we see a strong dispersion of this data. The fourth quadrant sees how the HP observations follow the direction of the arrows representing the variables "NPIFLC" (number of presences in first level championships) and "NPIICcc" (number of presences in international competitions (club competitions)). As far as the LPs are concerned, we see a strong grouping of the data that develops between the third, second and first quadrant, passing near the origin. The trend of this grouping seems to follow the axis relative to the first dimension, i.e. the one formed by "NPIFLC" (number of presences in first level championships), "NPIICcc" (number of presences in international competitions (club competitions) and "Age". Despite this, there is no a clear separation between the two clusters.

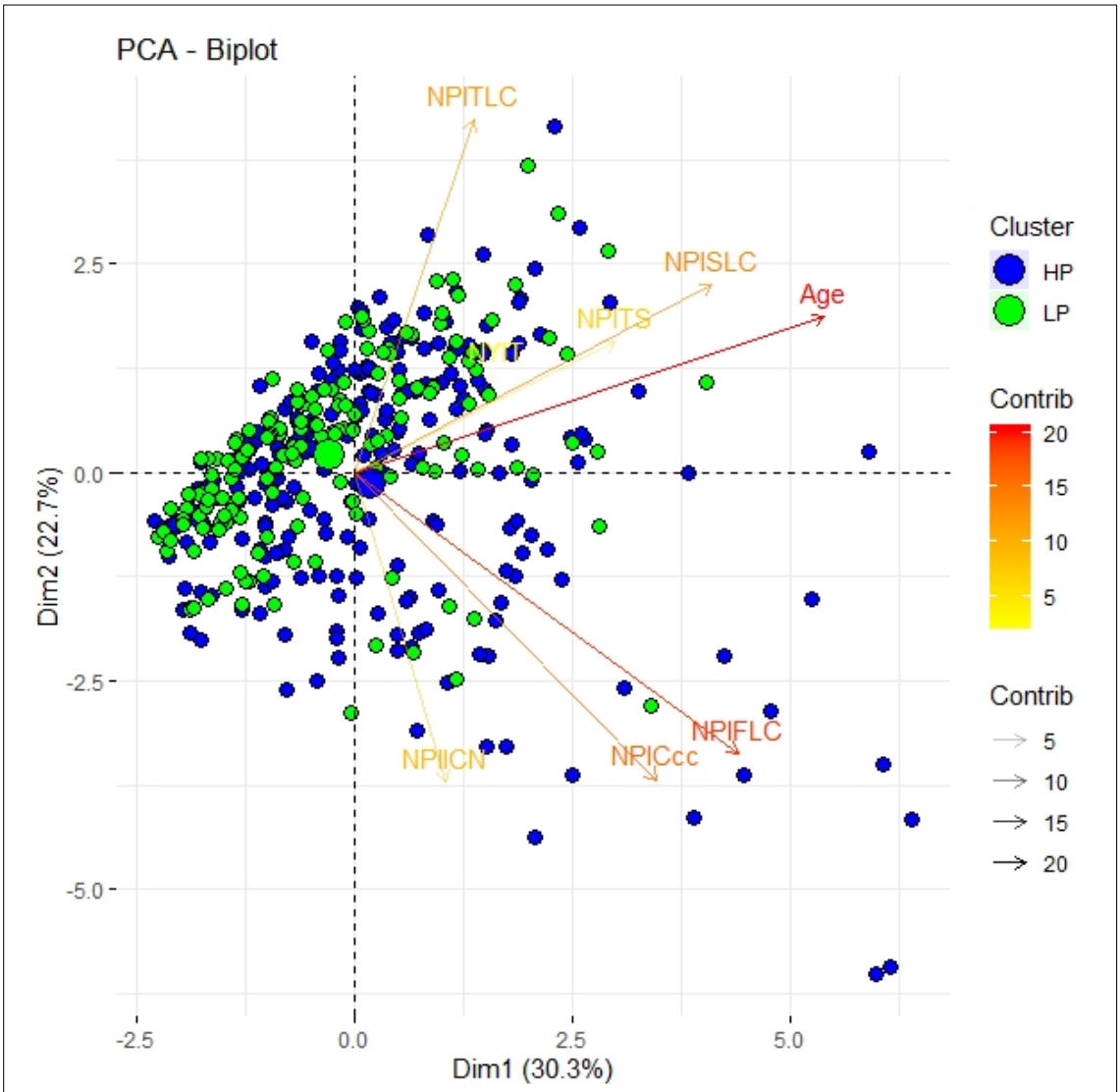

Figure 3.29: Observations of players dataset on the first two PCs, 2017/18 season

The 2017/18 season sees the first two PCs expressing about 53% of the total variance, against 55.3% in the first season (Figure 3.29). In this case the first PC is composed for more than 30% by the variable "Age", then by the variable "NPIFLC" (number of presences in first level championships) for a little more than 20%, by the variable "NPISLC" (number of presences in second level championships) for a little less than 20% and the variable "NPIICcc" (number of presences in international competitions (club competitions)) slightly exceeds the barrier threshold. The second variable is made up of more than 25% of the variable "NPITLC" (number of presences in third level championships), 20% of the variables "NPIICN" (number of presences in international competitions (national competitions)) and "NPIICcc" (number of presences in international competitions (club competitions)) and just over 15% of the variable

"NPIFLC" (number of presences in first level competitions). Considering the two PCs together, the variable that contributed most to their creation is "Age", followed by the variable "NPIFLC" (number of presences in first level championships) for more than 17.5%, "NPIICcc" (number of presences in international competitions (club competitions)) for just over 15% and finally for just over 12.5% by the variable "NPISLC" (number of presences in second level championships). The variables common to both seasons analysed so far are "Age", "NPIFLC" (number of presences in first level championships) and "NPIICcc" (number of presences in international competitions (club competitions)).

Looking at the individuals (Figure 3.29), starting from the HP, it can be seen that the data are very scattered and that in the fourth quadrant there are the data further away from the origin. These data are developed following the arrows related to the original variables "NPIFLC" (number of presences in first level championships) and "NPIICcc" (number of presences in international competitions (club competitions)) which contribute respectively more than 17.5% and 15% to the formation of the first two PCs. Looking instead at the LP data, we can see that there is an important grouping that starts in the third quadrant, passes through the second and ends in the first of the data, following the arrows related to the variables "Age", "NPISLC" (number of presences in second level championships) and "NPITLC" (number of presences in third level championships). However, within this grouping there are several observations related to the HP population. Considering the centre of the data, it can be noted that their distance is greater than that found in the previous season, which proves that in this season the distinction between the population is clearer but there is not yet the presence of the two populations in two distinct areas of the graph.

The following season, i.e. the 2016/17 season (Figure 3.30), has the first PC explaining 30.4% of the variance and the second explaining 25.4%, with a total cumulative variance of 55.8%, the highest portion explained so far, against 55.3% of the 2018/19 season and 53% of the 2017/18 season. The original variables that contribute to the creation of the first two PCs are "Age" and "NPIFLC" (number of presences in first level championships) for a contribution higher than 17.5%, while the variable "NPIICcc" (number of presences in international competitions (club competitions)) contributes for a portion around 17%.

These variables are common to all three seasons analyzed. In detail, the first PC is composed in this case of the original variables "Age" (about 30% contribution), "NPIFLC" (number of presences in first level championships about 22.5%) and "NPIICcc" (number of presences in international competitions (club competitions) about 17%), while the second is composed of the variables "NPITLC" (number of presences in third level competitions a little more than 20% contribution), "NPIICN" (number of presences in international competitions (national competitions) about 18% contribution), "NPIICcc" (number of presences in international competitions (club competitions) about 16% contribution) and "NPIFLC" (number of presences in first level championships just above the barrier).

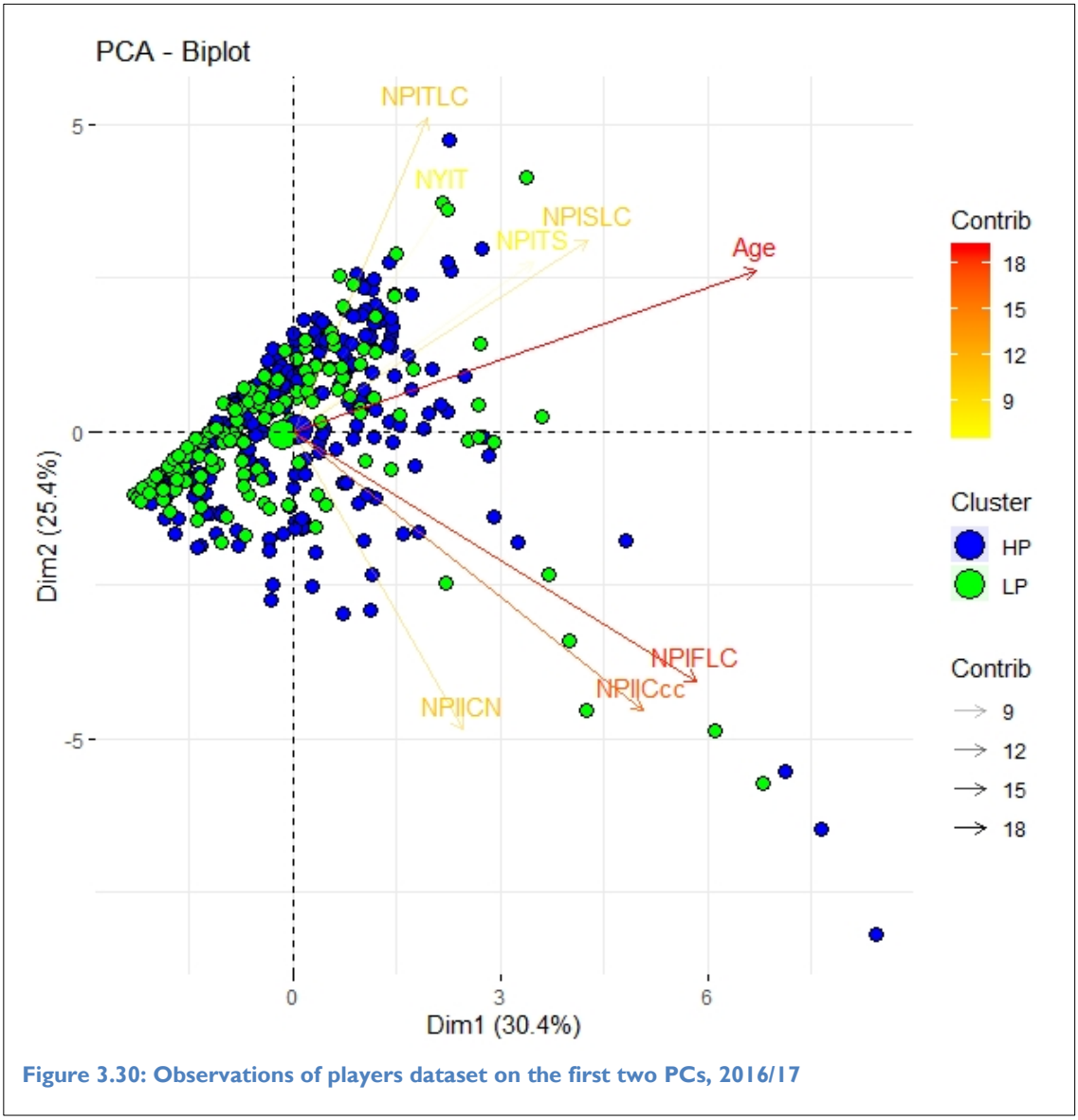

Figure 3.30: Observations of players dataset on the first two PCs, 2016/17

Moving on to the observation of the individuals on the first two PCs (Figure 3.30), in this season we notice a similar representation to that seen in the other two seasons, with an evident clustering of the LP cluster located between the third, second and first quadrant and the HP population much more dispersed. Going in order, as far as HPs are concerned, it can be noticed that the point representing the data center is practically placed on the origin, with a strong proximity to the LP cluster and this indicates that on average the data of the two clusters are similar. As it can be seen, the blue dots do not seem to follow a precise direction and certainly many are not visible because they are in the same position as LP observations. One can clearly see how the individuals are positioned disparately. As far as LPs are concerned, it can be seen how the above mentioned cluster seems to go in the direction indicated by some original variables, but they do not exceed the threshold in terms of contribution. Also as far as LPs are concerned there are some distant and scattered points with respect to the centre and the pool but less than the other cluster. Also in this case a grouping of the LP population can be distinguished but the same thing cannot be said for the HP population. Despite these divisions, also in this case the two populations do not occupy separate areas of the graph.

The penultimate season presents the first two PCs that respectively explain 30.7% and 23% of the variance, for a cumulative variance of 53.7% (Figure 3.31). Considering in particular the first PC, this one is formed for the most part by the original variables "Age" for about 35% of the variance, followed by the variable "NPISLC" (number of presences in second level championships) with just over 20% and finally the variable "NPIFLC" (number of presences in first level championships) with about 15%. Then, the second variable is composed for more than 25% by the variable "NPIICcc" (number of presences in international competitions (club competitions)), a little less than 25% by the variable "NPIFLC" (number of presences in first level championships), for about 17.5% by the variables "NPITLC" (number of presences in third level championships) and "NPIICN" (number of presences in international competitions (national club)). Considering them together, the first two PCs are formed by the variables "Age" for a contribution share of about 20%, by "NPIFLC" (number of presences in first level championships) for a share of just under 20%, by "NPIICcc" (number of presences in

international competitions (club competitions)) for a share exceeding 15% and just above the barrier by the variables "NPISLC" (number of presences in second level championships) and "NPITLC" (number of presences in third level championships). Even after analyzing this season, the variables common to all the years in terms of major contribution to the creation of the first two PCs are the variables "Age", "NPIFLC" (number of presences in first level championships) and "NPIICcc" (number of presences in international competitions (club competitions)).

Proceeding with the observation of the representation of the individuals on the first two PCs (Figure 3.31), we can see how the two points representing the center of the data

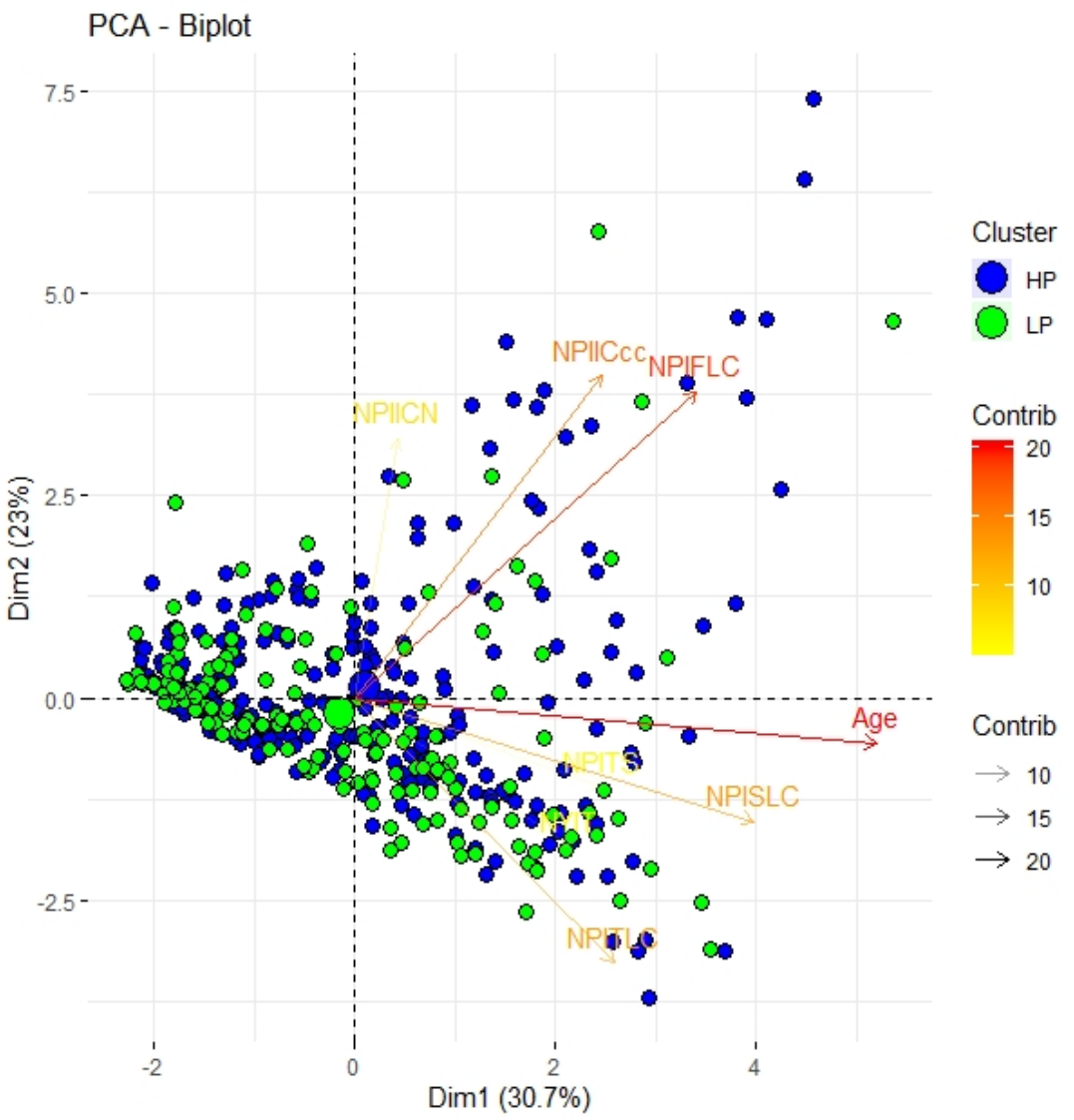

Figure 3.31: Observations of players dataset on the first two PCs, 2015/16 season

are very close to each other and respectively that of the HP is in the first quadrant while that of the LP is in the third. Considering the HPs, one can see how also in this case they have a very dispersed distribution with many observations in the first quadrant following the direction of the "NPIICcc" (number of presences in international competitions (club competitions)) and "NPIFLC" (number of presences in first level championships) arrows. Another direction followed by the data seems to be that of the variables "Age" and "NPISLC" (number of presences in second level championships) but in this case there is a strong presence of observations related to LPs. Turning to LPs, also in this case, as in previous seasons, there seems to be a more marked grouping of these, compared to the HP population. This grouping develops between the second and third quadrant and continues also in the fourth one but with a much more accentuated dispersion of data. The direction that follows this grouping is mainly that given by the variables "NPISLC" (number of presences in second level championships) and "NPITLC" (number of presences in third level championships). Although also in this case we see a grouping for LPs while for HPs it is more difficult to define, there is no clear definition of either group and even in this case the populations are not distributed in two distinct regions of the graph. Moreover, the proximity of the two central points of the data seems also in this case to testify a similarity in average terms of the data.

In the last season considered, i.e. the year 2014/15, the first PC is composed for a little more than 30% of the variable "Age", for about 23% from the variables "NPITLC" (number of presences in third level championships) and "NPISLC" (number of presences in second level championships) and just at the barrier level from the variable "NYIT" (number of years in the team), while the second is composed for a little more than 40% from the variable "NPIFLC" (number of presences in first level championships), for a little less than 35% from the variable "NPIICcc" (number of presences in international competitions (club competitions)) and for about 15% from the variable "NPIICN" (number of presences in international competition (national competitions)). The joint consideration of the first two PCs sees as the first contribution variable "NPIFLC" (number of presences in first level championships) with a little less than 20% contribution, as well as the variable "Age", while with a little more than 15% there is the variable "NPIICcc" (number of presences in international competitions (club

competitions)) and the last variable that exceeds the barrier is the variable "NPITLC" (number of presences in third level championships). Comparing the first two PCs with the other seasons analyzed, you can see that the variables common to all seasons are still "Age", "NPIFLC" (number of presences in first level championships) and "NPIICcc" (number of presences in international competitions (club competitions)). Taking into consideration the percentage of cumulated variance expressed by the first two PCs, in this case it is 55.5%, where the first PC expresses 29.3% variance while the second one expresses 26.2% (Figure 3.32). Looking at the other seasons, the cumulative percentage expressed by the first two CPs is 55.3% in 2018/19, 53% in 2017/18, 55.8% and 53.7% in 2015/16.

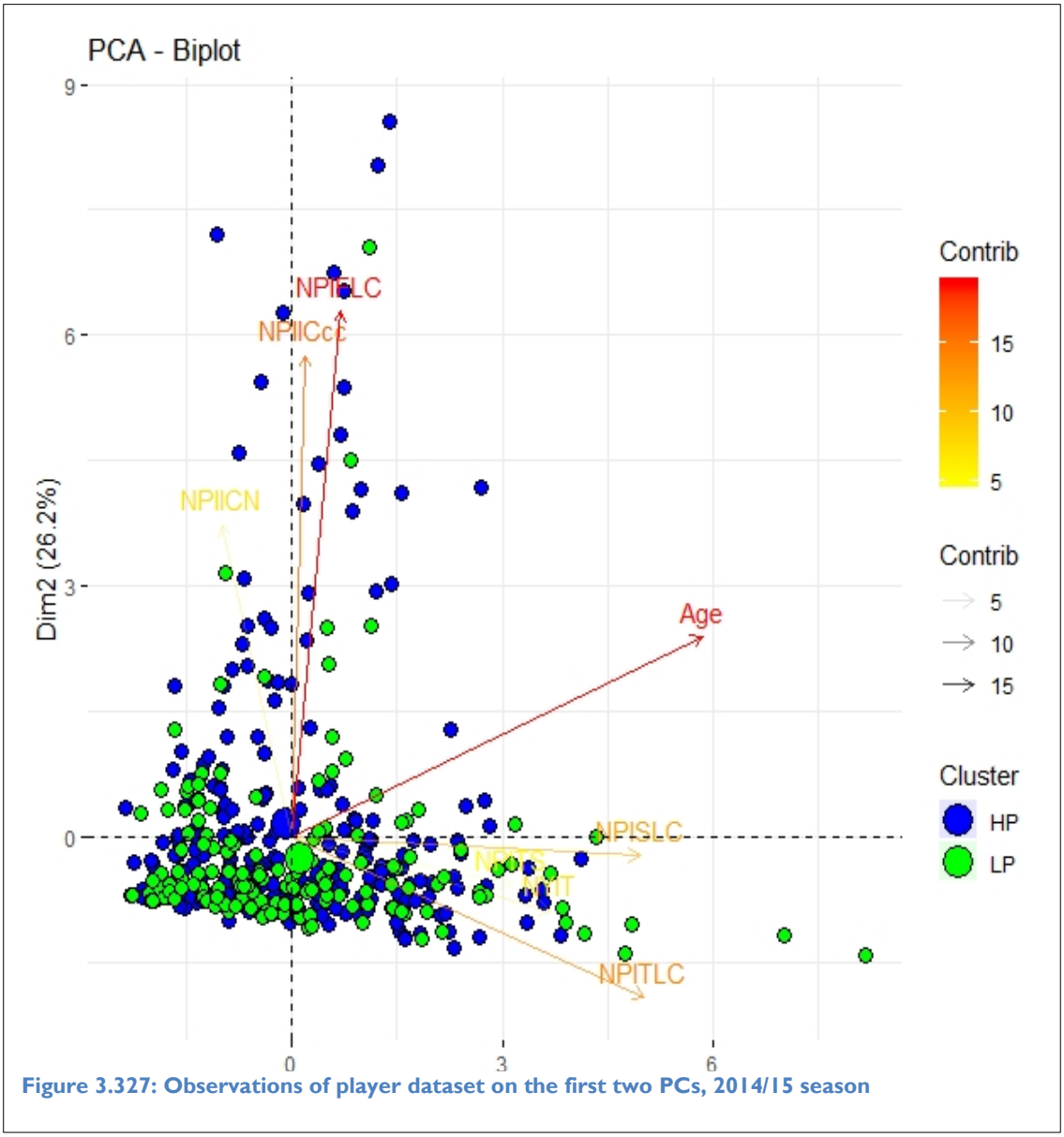

Figure 3.327: Observations of player dataset on the first two PCs, 2014/15 season

Looking at the points on the plane of the first two PCs (Figure 3.32), in this case we can see a almost unique grouping of the whole population between the third and the fourth quadrant in the direction of the "NPISLC" (number of presences in second level championships) variables, but this is not one of the variables that most constitute the first two PCs and the "NPITLC" (number of presences in third level championships) variable. Instead, these two variables are particularly important for the formation of the first PC. Going in order and starting from the HP cluster, it can be seen how the data center is positioned very close to the origin in the second quadrant. As already said, many of the observations seem to fall between the third and fourth quadrant in the formation of a large agglomerate of points that joins LP and HP. The HP cluster can be seen as having a dispersion of data along the axis of the first dimension, following the direction of the variables "NPIFLC" (number of presences in first level championships) and "NPIICcc" (number of presences in international competitions (club competitions)) which are among the most important in the formation of the first two PCs, The same thing that has been said for HP occurs in LP but with much less observations. The data center of the LPs results from the representation rather close to the origin and the central value of the HPs. It is also practically on the axis of the first dimension. The grouping that is seen developing between the third and fourth quadrant has a sense that is parallel to the axis of the second dimension. In this particular season, more than in the others, it is difficult to see a distinct grouping between the two clusters because both seem for the most part to be grouped together. Also in this case it can be argued that we do not see two distinct clusters in the plan concerning the two populations.

The analyses carried out so far have seen the two clusters analyzed together and the result of the representation of the graphs has not given in any season analyzed, a separate representation of the population in the two clusters. This has made the interpretation of the graphs very complex and this analysis is not sufficient to answer the research question asked. Moreover, the graphs see the individuals represented on the first two PCs expressing between 53% and 56% of the total variance explained. So in the next paragraph we will analyze the separate clusters and we will take into account a number of PCs that can explain a portion of variance as close as possible to 80%, so that we can consider a larger portion of variance and have more easily interpretable results. In

addition, considering the two separate clusters will give an immediate interpretation of what characterizes most of the variance in the HP and LP clusters.

3.5.2. PCA ON HP PLAYERS DATASET

	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	2.5695491	32.119364	32.11936
Dim.2	1.9598688	24.498360	56.61772
Dim.3	1.0912328	13.640410	70.25813
Dim.4	0.8111242	10.139053	80.39719
Dim.5	0.7546210	9.432762	89.82995
Dim.6	0.5024608	6.280760	96.11071
Dim.7	0.2103023	2.628778	98.73949
Dim.8	0.1008410	1.260513	100.00000

Table 3.3: Eigenvalues table, 2018/19 HP players

The eigenvalues measure the amount of variation retained by each principal component. So we consider them to determine the the number of principal components to consider.

The table 3.3 is the eigenvalue table for the players of the HP cluster for the 2018/19 season. The first column is about eigenvalues. Normally if the data are standardized, as in this case, the PCs that have a eigenvalue with a value greater than one means that account for more variance than accounted by one of the original variables. This can be used as a cutoff point as an alternative to the cumulative variance for the choice of PCs. In this case we have the first two components that explain about 56.62% of the variance and the first four are just over 80%. Instead, as far as the value of the eigenvalues is concerned, only the first 3 PCs exceed the value of 1 for a total of explained variance of about 70.26%. The scree plot gives a graphical representation of the percentage of variance explained by the various

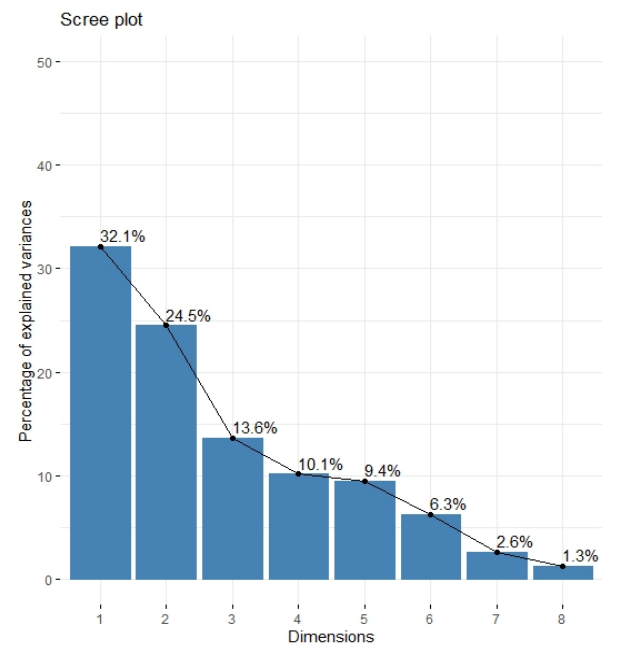

Figure 3.33: Scree plot, 2018/19 HP players

PCs. As can be seen there is a jump close to 8% between the first and the second PC, a difference that is even more pronounced if you consider the second and the third PC. Going forward with the PCs we can see how the differences in terms of explained variance are attenuated by the normal effect of the applied methodology. The contribution that the variables give to the formation of the main components can be expressed as scores, called loadings or as percentages.

Regardless of the amount of variance explained by the first two PCs, the original variables associated with these two, are the most important to explain the variability of the dataset. As can be seen from the graph relating to the contribution of the original variables to the formation of the main components for the first 5 PCs, we can see how for PC1 (ie Dim.1), the original variables that contribute to its formation, are "NPIFLC" (Number of presences in first level championships) with a contribution of 31.23%, followed by the variable "NPIICcc" (Number of presences in international competitions (club competitions)) with about 26.52%. The other variables contributing significantly are "Age" and "NPIICN" (Number of presences in international competitions (national club)), with 13.9% and 16.1% respectively (Table 3.4).

As far as the second PC is concerned, the two variables that seem to contribute most are "NPITLC" (Number of presences in third level championships) and "Age", with a contribution of 26.38% and 25.6% respectively. Other variables that contribute significantly are "NPISLC" (Number of presences in second level championships) with 17.44%, "NYIT" (Number of years in the team) with 8.26% and "NPIICN" (number of presences in international competitions (national competitions) with 9.83% (Table 3.4).

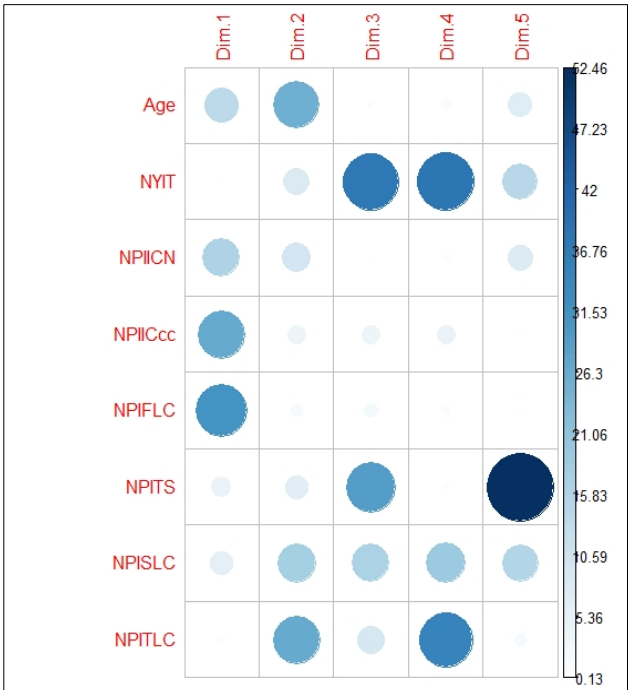

Table 3.4: Variables contribution to PCs formation, 2018/19 HP players

If we consider as crossover point the value of the eigenvalues and

therefore we consider only the PCs that equal or exceed 1 for this value, then we must consider the first 3 components.

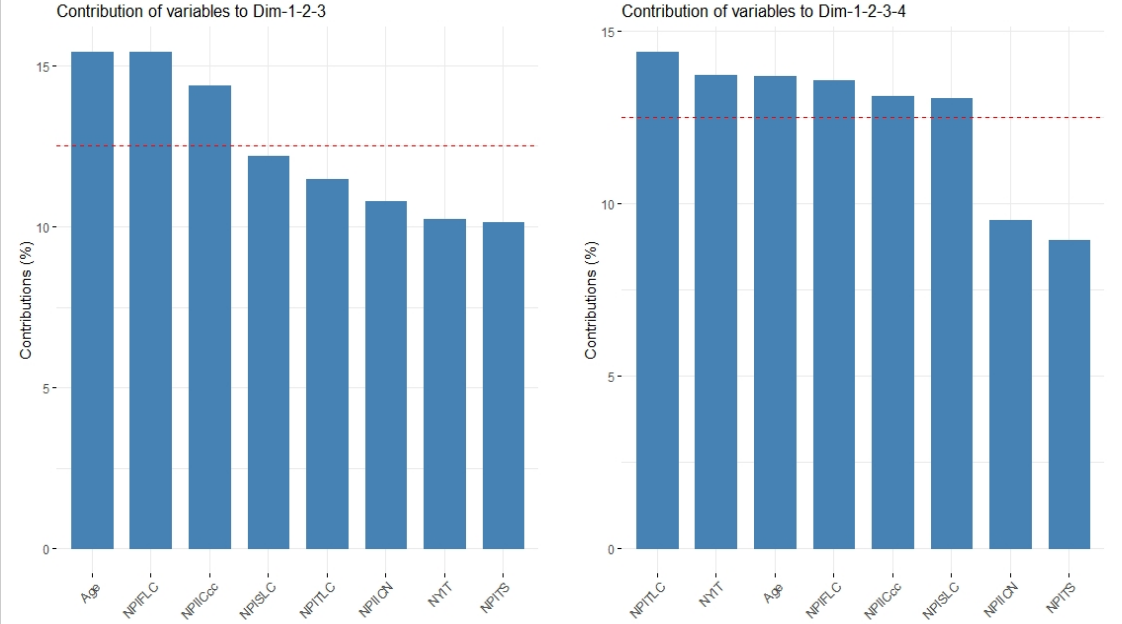

Figure 3.34: Contributions of variables according to eigenvalues and cumulative variance, 2018/19 HP players

Instead, if we consider the cumulative variance, then we must consider the first 4 (Figure 3.34). The graphs show the average contribution of the original variables to the PCs in the first and second cases. The red dotted line indicates the average contribution that each original variable could make to the composition of the PCs if all the variables had the same distribution and data. In this case having 8 variables, each original variable should contribute 12.5%. In considering the original variables that contribute most to PCs we will consider only those that exceed this threshold. So in this case, if we consider the eigenvalues, the original variables that contributed most to the formation of the first three PCs are as first the variable "Age", in second position the variable "NPIFLC" (Number of presences in first level championships) and third "NPIICcc" (Number of presences in international competitions). If instead we consider the cumulative variance explained up to 80% we have to consider the first 4 PCs. In this case the first variable for contribution in percentage terms is "NPITLC" (Number of presences in third level championships), followed by "NYIT" (Number of years in the team), "Age", "NPIFLC" (Number of presences in first level championships), "NPIICcc" (Number of presences in

international competitions (club competitions) and finally "NPISLC" (Number of presences in second level championships).

	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	2.42640689	30.330086	30.33009
Dim.2	1.84996301	23.124538	53.45462
Dim.3	1.05731693	13.216462	66.67109
Dim.4	0.93044760	11.630595	78.30168
Dim.5	0.74037107	9.254638	87.55632
Dim.6	0.62955708	7.869463	95.42578
Dim.7	0.27509081	3.438635	98.86442
Dim.8	0.09084662	1.135583	100.00000

Table 3.5: Eigenvalues table, 2017/18 HP players

For the second season, i.e. the 2017/18 season, the values for eigenvalues, explained variance and cumulative explained variance are as follows in the table 3.5. Also in this season the Pcs that exceed or equal the score 1 in terms of eigenvalues are 3 but in this case the first three main components explain 66.67% of the variance while in the previous season was about 70.26%. If you consider the accumulated variance instead, the first 4 PCs explain 78.3% of the variance while to exceed 80% you need to use 5 Pcs. With the first 5 PCs you can explain about 87.56 of the variance.

If we consider the screeplot of this season (Figure 3.35) we can see that the difference in variance explained between the first and second PC is about 7% while between the

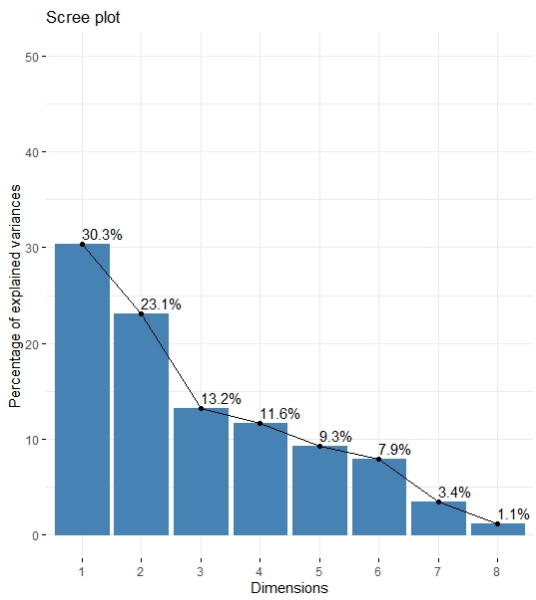

Figure 3.35: Scree plot, 2017/18 HP players

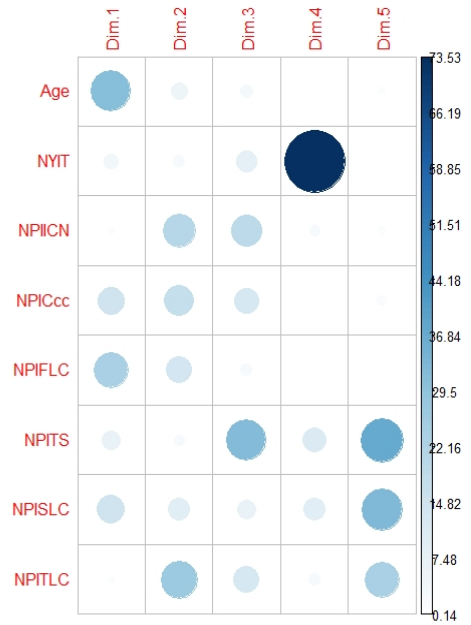

Table 3.6: Variables contribution to PCs formation, 2017/18 HP players

second and third PC you have almost 10%, so differences very similar to those seen in the first season. If, on the other hand, we consider PCs from the third to the sixth season later, we can see how there is little difference between the previous and the next one in terms of explained variance, with a more homogeneous distribution compared to that seen in the 2018/19 season.

Turning instead to considering the loadings (Table 3.6) and the relative percentages of contribution of the original variables to the formation of the PCs, we note that for the explanation of the first PC the most important variable is "Age" which contributes for about 31.43%, followed by the variable "NPIFLC" (Number of presences in first level championships) with a score of 24.28%. Moreover, above the threshold of 12.5% of contribution we find the variable "NPISLC" (Number of presences in second level championships) with a contribution of 15.27% and the variable "NPICcc" (Number of presences in international competitions (club competititons)) with a contribution of 14.99%. Moving on to the second PC, the most important original variable is the "NPITLC" (Number of presences in third level championships) with a contribution of about 26.83%, followed by the variable "NPIICN" (Number of presences in international competitions (national teams)) with the 21,16%, then the variable "NPICcc" (Number of presences in internationa competions (club competitions)) with 17.65% and finally the variable "NPIFLC" (Number of presences in first level championships) with 13.41%.

With regard to the first two components in the season 2018/19 and 2017/18 we can see that the original variables that exceed 12.5% of contribution common to both the first main components are "NPIFLC" (number of presences in first level championships) (31.23% 2018/19; 24.28% 2017/18), "NPIICcc" (number of presences in international competitions (club competitions)) (26.52% 2018/19; 14.99% 2017/18) and "Age" (13.9% 2018/19; 31.43% 2017/18). On the other hand, as far as the second PC is concerned, the only variable common to both is the "NPITLC" (number of presences in third level championships) with a contribution of 26.38% in 2018/19 and 26.83% in 2017/18.

Also in this case it is interesting to observe which are the original variables that contribute most to the composition of the PCs in relation to the crossover points considered (Figure 3.36). As previously said, also in this season only 3 PCs exceed or equal the value 1 in terms of eigenvalues, while to overcome the cumulated explained

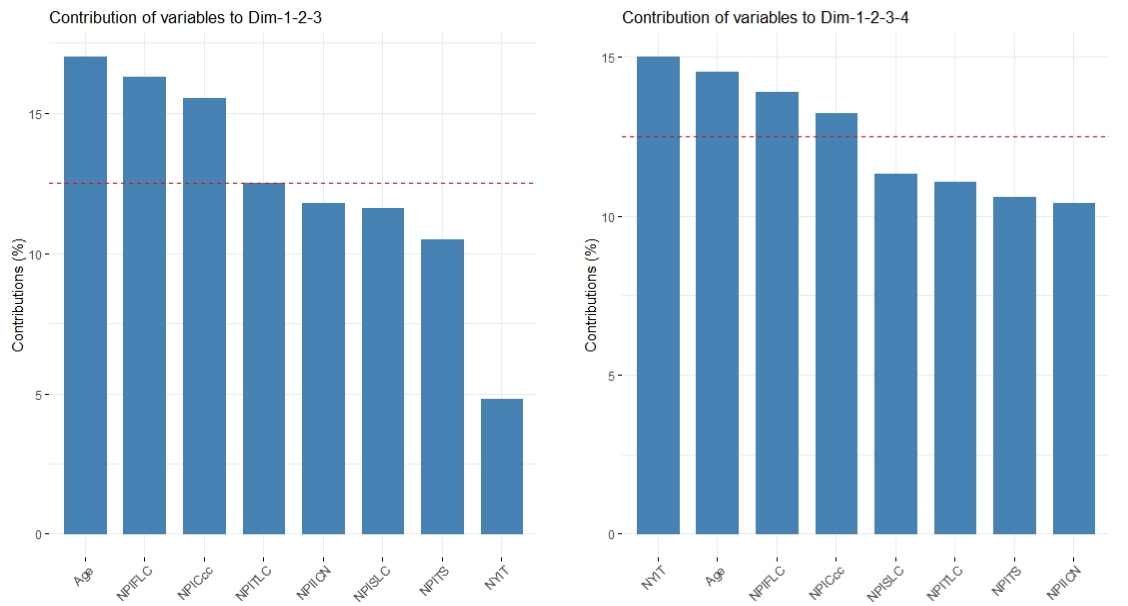

Figure 3.36: Contributions of variables according to eigenvalues and cumulative variance, 2017/18 HP players

variance higher than 80% we have to consider 5 components. Since we have taken into consideration as crossover point 80% of the variance, in this case if we took into account 5 PCs we would arrive at a cumulative variance of 87%, instead taking only 4 we would arrive at 78.3%. For reasons of greater closeness in terms of explained cumulative variance and simplicity of comparison between seasons, only 4 PCs will be considered. The two histograms explain respectively the original variables that contribute on average in a more important way to the composition of the first 3 and the first 4 PCs.

In the first case we see how the variable "Age" contributes more than 15% to the composition of the 3 PCs, followed by the variable "NPIFLC" (number of presences in first level championships), the variable "NPIICcc" (number of presences in international competitions (club competitions)) and the variable "NPITLC" (number of presences in third level championships) behind the average theoretical contribution. If we consider the case of the cumulative explained variance then we can see that in this case the most important variable is "NYIT" (number of years in the team), followed by the variable "Age", "NPIFLC" (number of presences in first level championships) and "NPIICcc" (number of presences in international competitions (club competitions)).

Between the two analysed seasons we can see how the common variables in the composition of the PCs considering the value of the eigenvalues are "Age", "NPIFLC"

(number of presences in first level championships) and "NPIICcc" (number of presences in international competitions (club competitions)) while considering the accumulated variance the common variables are "NYIT" (number of years in the team), "Age", "NPIFLC" (number of presences in first level championships) and "NPIICcc" (number of presences in international competitions (club competitions)).

	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	2.30251443	28.781430	28.78143
Dim.2	2.02117691	25.264711	54.04614
Dim.3	0.99981174	12.497647	66.54379
Dim.4	0.81923424	10.240428	76.78422
Dim.5	0.76861969	9.607746	86.39196
Dim.6	0.64038384	8.004798	94.39676
Dim.7	0.34854771	4.356846	98.75361
Dim.8	0.09971146	1.246393	100.00000

Table 3.7: Eigenvalues table, 2016/17 HP players

The 2016/17 season is instead characterized by the values relating to eigenvalues and variance recorded in table 3.7. As it can be seen, the first two PCs are the only ones to exceed the value 1 in terms of eigenvalues, with the third component that has virtually equal score to 1. The first two main components come to explain about 54% of the accumulated variance and if we consider all the PCs that have a score relative to the eigenvalue equal to or greater than 1, then we must consider the first 3 PCs for a total of explained variance of about 66.5%. We remember that in previous seasons the first two PCs explained 56.62% and 53.45% of the variance, while if we consider the PCs with eigenvalue greater than or equal to 1, in the season 2018/19 was explained 70.26% of the variance while in the season 2017/18 66.67%. In this case, to exceed 80% of the cumulative variance explained, as in the 2017/18 season it is necessary to consider 5 PCs for a total of 86.39% of the variance explained, while with 4 PCs it is possible to reach 76.78%.

From the analysis of the scree plot for this season it can be seen that the first two components explain a relatively similar portion of variance, respectively with about 28.8% and 25.3% (Figure 3.37). We notice instead that there is a strong difference between the second PC and the third PC with a difference of more than half of the variance explained by the second PC equal to 12.8%. There is instead a certain similarity in terms of the percentage of variance explained between the PCs between the third

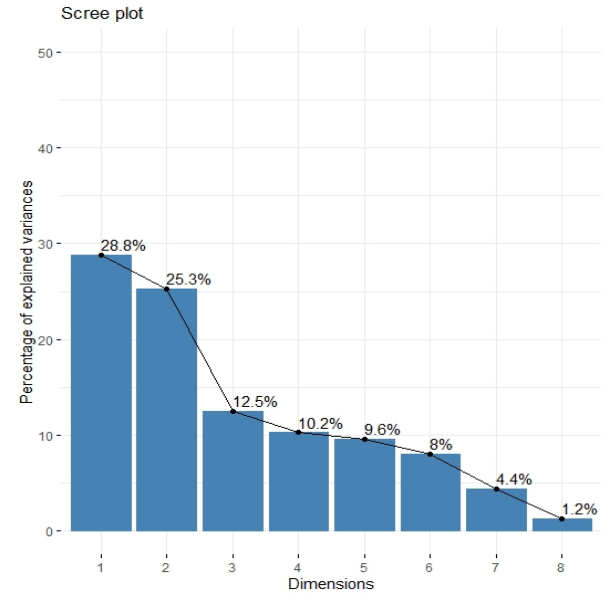

Figure 3.37: Scree plot, 2016/17 HP players

and the sixth. Considering in particular the third PC, of the three seasons analyzed so far, is the one with the lowest portion of explained variance. The same can be said for the first PC, while the second PC in this case is the one with the highest explained variance value.

Analyzing the loadings related to the latter, for the first PC, the variables that exceed the minimum threshold of contribution, ie 12.5%, we find the variable "Age" which contributes for 35, 63% about, followed by the variable "NPIFLC" (number of presences in first level championships) which contributes for 16.93% and finally the variable "NPISLC" (number of presences in second level championships) with 14.73% of contribution (Table 3.7). Considering the first PC of the two previous seasons, the variables common to all three years are the variables "Age" and "NPIFLC" (number of presences in first level championships) (respectively 13.9% in 2018/19; 31.43% in 2017/18; 35.63% in 2016/17 and 31.23% in 2018/19; 24.28% in 2017/18; 16.93 in 2016/17). On the other hand, the variable "NPISLC" (number of presences in second level championships) is common to the 2017/18 seasons with 15.27%. The second PC, on the other hand, has four variables that contribute to its composition by more than 12.5% (Table 3.7). The four variables are

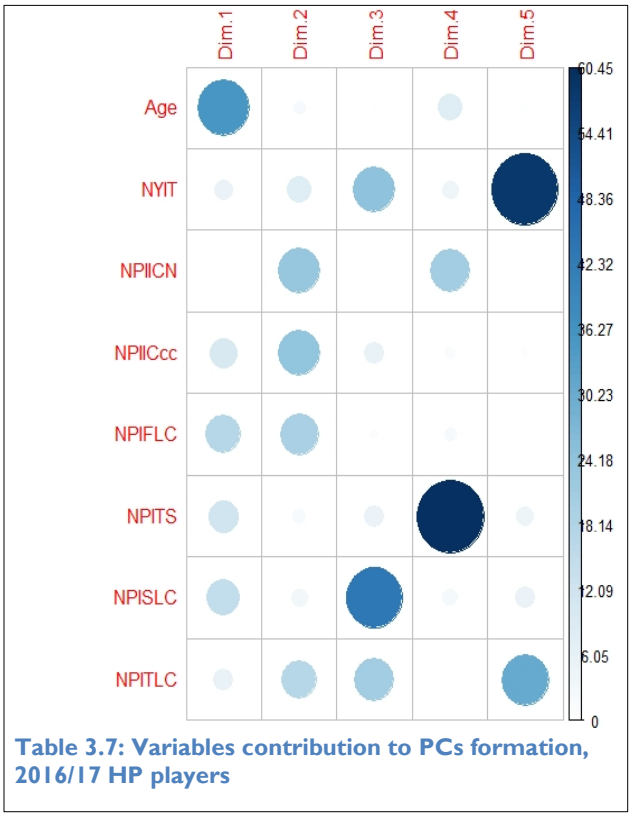

Table 3.7: Variables contribution to PCs formation, 2016/17 HP players

"NPIICcc" (number of presences in international competitions (club competitions)) with 23.95% contribution, "NPIICN" (number of presences in international competitions (national competitions)) with 23.08%, "NPIFLC" (number of presences in first level championships) with 19.81% and finally "NPITLC" (number of presences in third level championships) with 17.16%. The only variable common to all three second PCs is the variable "NPITLC" (number of presences in third level championships) which contributed 26.38% in 2018/19 and 26.83% in 2017/18. Considering always the usual crossover points related to the choice of the number of PCs to be considered, as previously said the PCs with eigenvalues equal to or greater than 1 are 3 while to exceed 80% of variance 5 PCs are needed. As for the previous season only 4 PCs will be considered because anyway the explained cumulative variance exceeds 76% and therefore for the same reasons previously listed the first 4 PCs will be considered.

Analyzing PCs with eigenvalues greater than or equal to 1, it can be seen that the most important variable in terms of contribution is the variable "Age", followed by "NPISLC" (number of presences in second level championships), "NPIFLC" (number of presences in first level championships), "NPIICcc" (number of presences in international competitions (club competitions)) and "NPITLC" (number of presences in third level championships) (Figure 3.38). Compared to the seasons already analysed, in this case the common variables in the three years are the variable "Age", the variable "NPIFLC"

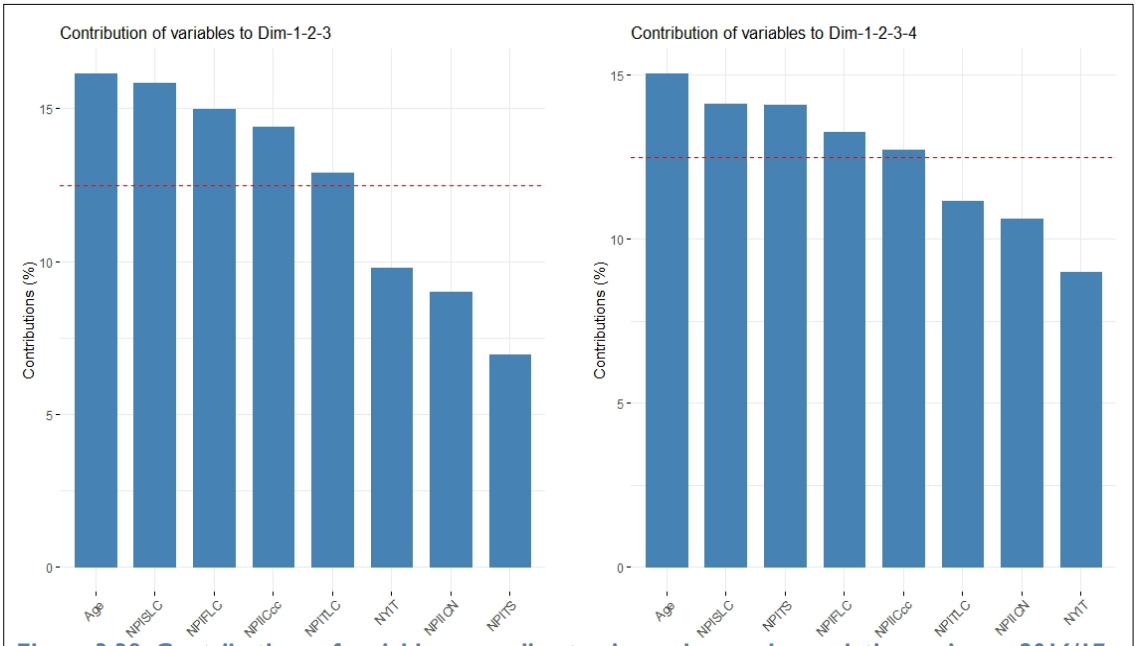

Figure 3.38: Contributions of variables according to eigenvalues and cumulative variance, 2016/17 HP players

(number of presences in first level championships) and the variable "NPIICcc" (number of presences in international competitions (club competitions)). On the other hand, the variable "NPITLC" (number of presences in third level championships) is common only to the analysed season and to the 2017/18 season. Finally, as regards the crossover point of the cumulative variance (Figure 3.38), we still see how the variable "Age" is the most important in terms of average percentage contribution with about 15%, followed by "NPISLC" (number of presences in second level championships) and "NPITS" (number of presences this season) which have about the same percentage of contribution and finally "NPIFLC" (number of presences in first level championships) and "NPIICcc" (number of presences in international competitions (club competitions)). Also considering this crossover point, the variables common to the 3 seasons analyzed are "Age", "NPIFLC" (number of presences in first level championships) and "NPIICcc" (number of presences in international competitions (club competitions)). Instead, the variable "NPISLC" (number of presences in second level championships) is common only with the 2018/19 season.

	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	2.4022411	30.028014	30.02801
Dim.2	1.7949889	22.437361	52.46538
Dim.3	1.0558269	13.197837	65.66321
Dim.4	0.9143555	11.429443	77.09266
Dim.5	0.7785576	9.731969	86.82462
Dim.6	0.6032046	7.540057	94.36468
Dim.7	0.3288543	4.110678	98.47536
Dim.8	0.1219712	1.524640	100.00000

Table 3.8: Eigenvalues table, 2015/16 HP players

The penultimate season considered is that of 2015/16. The latter has a number of PCs that exceed or equal the value 1 in terms of eigenvalues of 3 for a total of explained variance of 65.66% about and to exceed 80% of variance must be used 5 PCs with a value of cumulative variance of 86.82% about. The first two PCs account for about 52.46% and the first four for 77.09% (Table 3.8).

Comparing the seasons already analysed, it can be seen that the first two components in this case explain a lower cumulative variance with 52.46% compared to 56.62% in the 2018/19 season, 53.45% in 2017/18 and 54% in 2016/17. The first 4 PCs also explain a lower cumulative variance than the analyses already made. In fact, in this case they

explain 77.09% of the cumulative variance while in 2018/19 it explained 80.40%, in 2017/18 78.3% and in 2016/17 75.78%. Also in terms of the values assumed by the eigenvalues, this season represents the lowest point with a value of the eigenvalue of the first highest PC only of the 2016/17 season and with a value relative to the second lowest component of the four years analyzed.

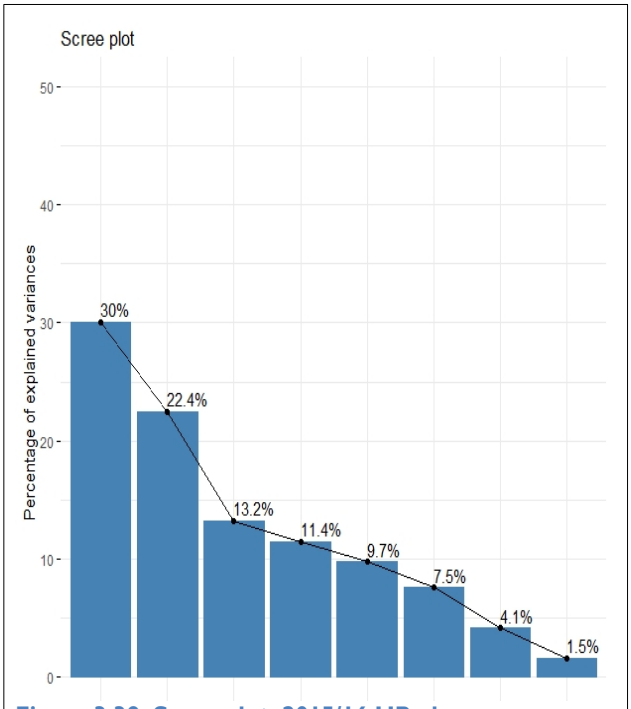

Figure 3.39: Scree plot, 2015/16 HP players

From the analysis of the screeplot it can be seen graphically what the table of eigenvalues expresses (Figure 3.39). The trend of the graph is very similar to that of the 2017/18 season. As can be seen, there is a difference between the first and the second PCs of just over 7% and between the second and the third PC there is an even more marked jump of 9.2%. From the third to the sixth PC instead there are no big differences in terms of cumulative variance explained.

Considering instead the loadings that constitute the PCs (Table 3.9), we start considering the first two PCs. The first main component has as main contributor the original variable "Age" with a percentage contribution of about 34.87%. The following is the variable "NPISLC" (number of presences in second level championships) with a contribution practically equal to the variable "NPIFLC" (number of presences in first level championships) (18.54% and 18.5% respectively).

The first main component has as main contributor the original variable "Age" with a percentage contribution of about 34.87%. The following is the variable "NPISLC" (number of presences in second level championships) with a contribution practically equal to the variable "NPIFLC" (18.54% and 18.5% respectively). Under the 12.5% contribution, the contribution of the variable "NPICcc" (number of presences in international competitions (club competitions)) should be noted with 10.05%. If we

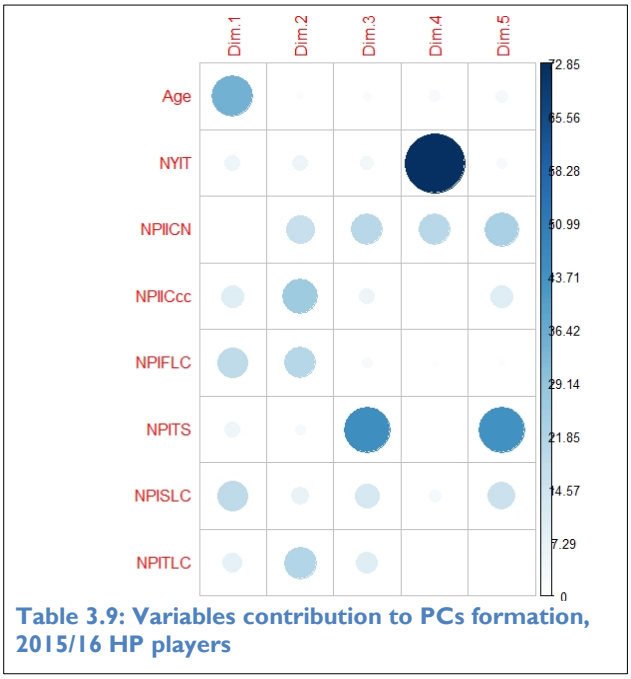

Table 3.9: Variables contribution to PCs formation, 2015/16 HP players

consider only those variables that exceed the contribution of 12.5% and consider the first main components of each season, the variables common to all years are "Age", which in 2018/19 contributed to the formation of the first PC for 31.23%, in 2017/18 for 31.43 and in 2016/17 for 35.63 and "NPIFLC" (number of presences in first level championships) with a contribution of 31.23% in 2018/19, 24.28% in 2017/18 and 16.93% in 2016/17. Instead, the variable "NPISLC" (number of presences in second level championships) is common to the season 2017/18 (15.27%) and 2016/17 (14.73%) but not to the season 2018/19. Considering instead the second main component, the first original variable for contribution is "NPIICcc" (number of presences in international competitions (club competitions) with about 25.85% contribution, followed by "NPITLC" (number of presences in third level championships) with 21.69%, "NPIFLC" (number of presences in first level championships) with 20.55% and "NPIICN" (number of presences in international competitions (national competition)) with 16.28%. In this case, the common characters that contribute to the creation of PCs in all seasons analyzed are represented only by the variable "NPITLC" (number of presences in third level championships) that in 2018/19 contributed 26.38%, in 2017/18 for 26.83% and in 2016/17 for 17.16%. Instead, the variable "NPIICcc" (number of presences in international competitions (club competition)) is common to all years except 2018/19, as well as to the variables "NPIFLC" (number of presences in first level championships) and "NPIICN" (number of presences in international competitions (national competition)).

Considering, as usual, as crossover point the value of the eigenvalues equal to or greater than 1 and the cumulated variance equal to 80%, the first three PCs will be considered for the first case while for the second the first 4 (Figure 3.40). The first 4 PCs express a

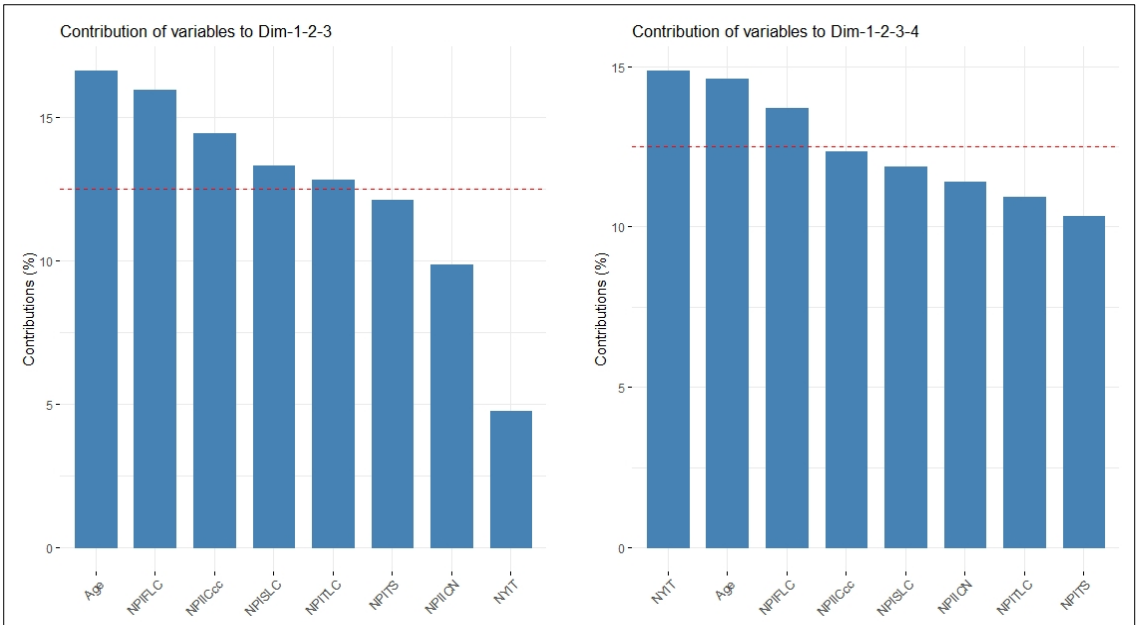

Figure 3.40: Contributions of variables according to eigenvalues and cumulative variance, 2015/16 HP players

cumulative variance equal to 77,09%, therefore a percentage lower than 80% but, as for the other seasons that have presented the same problem, only the first 4 will be considered for a question of greater proximity to the variance considered as crossover point compared to the one that would be expressed with 5 PCs.

Starting from the crossover point represented by the eigenvalues (Figure 3.40), it can be seen how the most important variable in terms of contribution is "Age", with an average contribution higher than 15%. Same thing goes for the variabile “NPIFLC” (number of presences in first level championships). Below 15% but still above 12.5% are the variables "NPIICcc" (number of presences in international competitions (club competitions), "NPISLC" (number of presences in second level championships) and "NPITLC" (number of presences in third level championships). Comparing this case with those of the other analysed seasons, it results that the variables common to all four years are the variable "Age", the variable "NPIFLC" (number of presences in first level championships) and the variable "NPIICcc" (number of presences in international competitions (club competitions)). Instead, the common three-season variable out of four is "NPITLC" (number of presences in third level championships), common to 2017/18, 2016/17 and 2015/16, while the variable "NPISLC" (number of presences in second level championships) is common only to this season and to that of 2016/17.

Turning instead to the consideration of the cumulative variance (Figure 3.40), in the 2015/16 season the variables that have contributed most on average to the creation of the first 4 PCs are the variable "NYIT" (number of years in team) with about 15% of average contribution, followed by the variable "Age" and "NPIFLC" (number of presences in first level championships). Just below the threshold of 12.5% we find the variable "NPIICcc" (number of presences in international competition (club competitions)). Probably the presence of the variable "NYIT" (number of years in the team) in the second case as a variable with the highest average contribution, is due to the fact that the fourth PC has a contribution of more than 70% by the latter. The variables common to all seasons are the variable "Age" and the variable "NPIFLC" (number of presences in first level championships). The variable "NYIT" (number of years in the team), on the other hand, is present in three out of four of the seasons analyzed, i.e. in the season 2018/19, 2017/18 and the one being analyzed.

	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	2.1790654	27.238318	27.23832
Dim.2	2.0338910	25.423637	52.66195
Dim.3	1.0519065	13.148832	65.81079
Dim.4	0.9621512	12.026890	77.83768
Dim.5	0.7508964	9.386206	87.22388
Dim.6	0.5775895	7.219869	94.44375
Dim.7	0.3317015	4.146269	98.59002
Dim.8	0.1127985	1.409981	100.00000

Table 3.10: Eigenvalues table, 2014/15 HP players

The last season that will be analyzed is the most distant in terms of time of the dataset. Also in this case only 3 Pcs have a eigenvalue greater than or equal to 1, with the fourth component approaching the limit value (Table 3.10). The cumulative variance explained by these three components is 65.8%, while even in this case, to exceed 80% of cumulative variance explained, it takes 5 PCs (87.22% of variance explained). The first two components explain 52.66% of the cumulative variance, a value slightly higher than that expressed by the first two PCs of the previous season but still lower than the 2018/19, 2017/18 and 2016/17 seasons. . Instead, with regard to the cumulative variance explained by the first four PCs, this season expresses 77.84% of the cumulative variance, a value higher than the 2015/16 (77.09%) and 2016/17 (76.78%) seasons but lower than the 2017/18 (78.3%) and 2018/19 (80.40%) seasons.

This season's screeplot is very similar to that of the 2016/17 season (Figure 3.41). The first two PCs have a very low difference in variance explained but between the third and the second PC there is a difference of more than 12%. Considering the first PC and comparing it with that of the other seasons, we notice that in this case it is the one with the lowest percentage of explained variance. In fact in 2018/19 the explained variance was 32.11%, in 2017/18 30.33%, in 2016/17 28.78% and in 2015/16 30.03%. In addition, the gap between the first and second PC is the lowest of the 5 seasons.

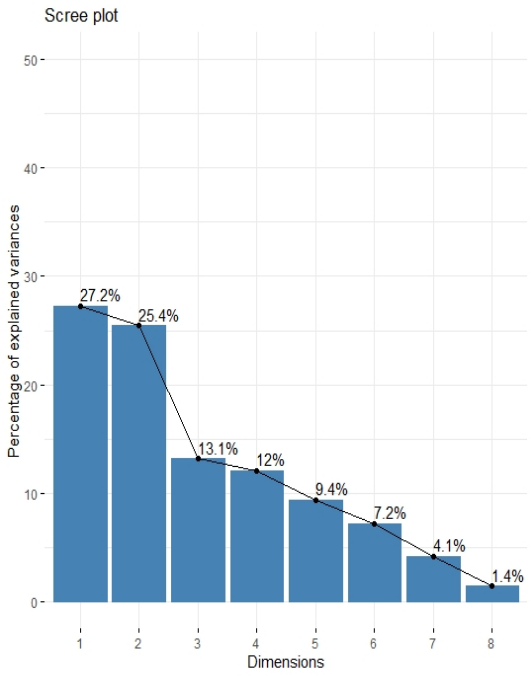

Figure 3.41: Scree plot, 2014/15 HP players

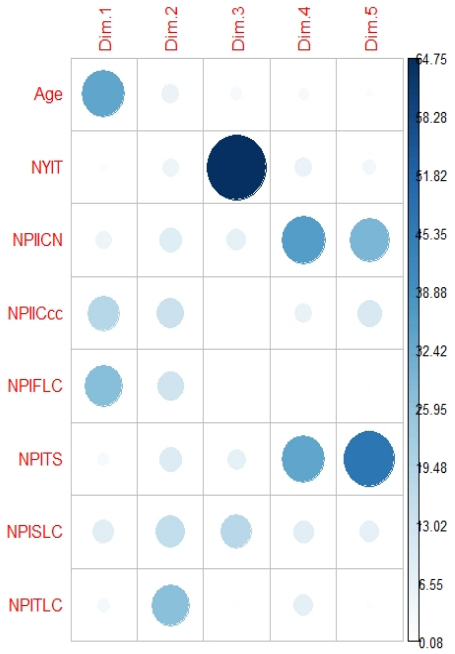

Table 3.11: Variables contribution to PCs formation, 2014/15 HP players

Also for this season we start considering the loadings of the first two main components (Table 3.11). Starting from the first PC, the variable "Age" is the most important with a contribution of about 33.73%, followed by the variable "NPIFLC" (number of presences in first level championships) with 27.41% and finally by the variable "NPIICcc" (number of presences in international competitions (club competitions))for 18.68%. The other variables that contribute to the creation of the first PC are all significantly below the threshold of 12.5%. The presence of the variable "Age" as a variable with an important contribution to the formation of the first PC is common to all seasons analyzed, as is the variable "NPIFLC" (number of presences in first level championships). On the other hand, the variable "NPIICcc" (number of presences in international competitions (club

competitions)) is common only to 3 out of 5 seasons, i.e. the 2018/19, 2017/18 and analysed seasons. Looking at the second PC, the most important variable is "NPITLC" (number of presences in third level competitions) with 27.22% of variance explained, followed by the variable "NPISLC" (number of presences in second level championships) with 15.93%, "NPIICcc" (number of presences in international competitions (club competitions) with 14.28% and "NPIFLC" (number of presences in first level championships with 13.32%. The other contributors are below the threshold limit. One element common to all the second main components is the variable "NPITLC" (number of presences in third level championships). On the other hand, one element common to four out of five seasons, with the exception of the 2018/19 season, is the variable "NPIICcc" (number of presences in international competitions (club competitions)).

Taking into account the eigenvalues it can be seen that the two variables "Age" and "NPIFLC" (number of presences in first level championships) contribute in a similar way on average with a contribution higher than 15%. After that the variable "NYIT" (number of years in the team) is found with a contribution of about 15%. Always above the minimum threshold we find the variable "NPIICcc" (number of presences in international competitions (club competitions)) and "NPISLC" (number of presences in second level championships). Just below the threshold is the variable "NPITLC" (number of presences in third level championships). If we look for common traits for all seasons, we can see

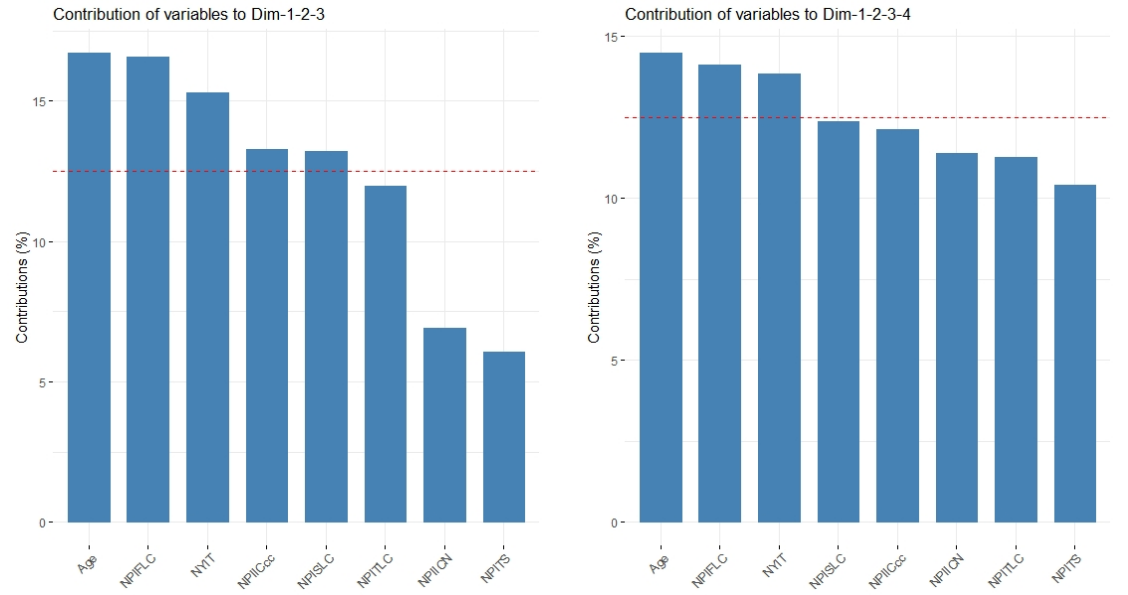

Figure 3.42: Contributions of variables to eigenvalues and cumulative variance, 2014/15 HP players

that considering the value of eigenvalues, the variables "Age", "NPIFLC" (number of presences in first level championships) and "NPIICcc" (number of presences in international competitions (club competitions)) are always present in an important way in terms of average contribution to the formation of PCs. A common feature of three out of five seasons is the variable "NPISLC" (number of presences in second level championships), respectively present in the seasons 2016/17, 2015/16 and 2014/15.

Moving on to analyze the cumulated variance, we can see how in this season again, the most important variable is "Age" followed by "NPIFLC" (number of presences in first level championships) and "NYIT" (number of years in the team). Just below the 12.5% contribution threshold we find the variables "NPISLC" (number of presences in second level championships) and "NPIICcc" (number of presences in international competitions (club competitions)). Again, the variables common to all five seasons are the variable "Age" and "NPIFLC" (number of presences in first level championships), while the variable "NYIT" (number of years in the team) is common to four out of five seasons, with the exception of the 2016/17 season.

3.5.3. PCA ON LP PLAYERS DATASET

In the previous paragraph we have applied the PCA to players belonging to the HP cluster for all five seasons taken into account. From the analysis it emerged that, considering a cumulative explained variance of an amount as close as possible to 80%, the variables that contributed to the creation of PCs for a value greater than 12.5% for each year analyzed are "Age" and "NPIFLC" (number of presences in first level championships) while the varibiable "NYIT" (number of years in the team) is present in 4 seasons out of five. Now we will check if in the LP cluster there are important variables in terms of contribution in at least 4 out of 5 seasons and if these variables differ from those of the other cluster.

	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	2.4981522	31.226903	31.22690
Dim.2	1.8038864	22.548580	53.77548
Dim.3	1.0748371	13.435464	67.21095
Dim.4	0.8209931	10.262413	77.47336
Dim.5	0.7639859	9.549824	87.02318
Dim.6	0.7447378	9.309223	96.33241
Dim.7	0.2017262	2.521578	98.85398
Dim.8	0.0916812	1.146015	100.00000

Table 3.12: Eigenvalues table, 2018/19 LP players

In the 2018/19 season, the PCs that match or exceed the value 1 in terms of eigenvalue are the first three, which explain a cumulative variance of about 67.21%. To exceed 80% of the variance, more precisely to express about 87.02% of it, it is necessary to use 5 PCs. In this case, the cumulative variance of the first two PCs explains 53.77% of the variance, with about 31.23% expressed by the first and 22.55% by the second. The first 4 PCs, on the other hand, express about 77.47% of the variance (Table 3.12).

In this case, the screeplot (Figure 3.43) shows a difference in terms of explained variance between the first and the second PC of about 8.7%, while the difference between the second and the third PC is about 9.1%. The difference between the third and fourth PCs, on the other hand, is smaller, assuming a value of 3.1%. The fourth, fifth and sixth PC explain about the same amount of variance, around 10%.

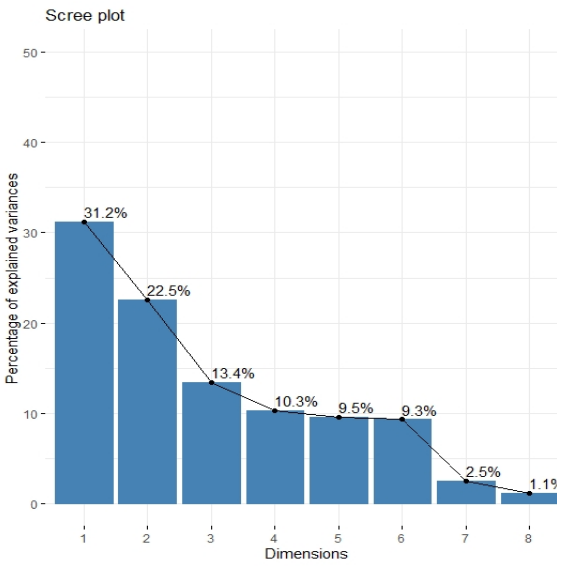

Figure 3.43: Scree plot, 2018/19 LP players

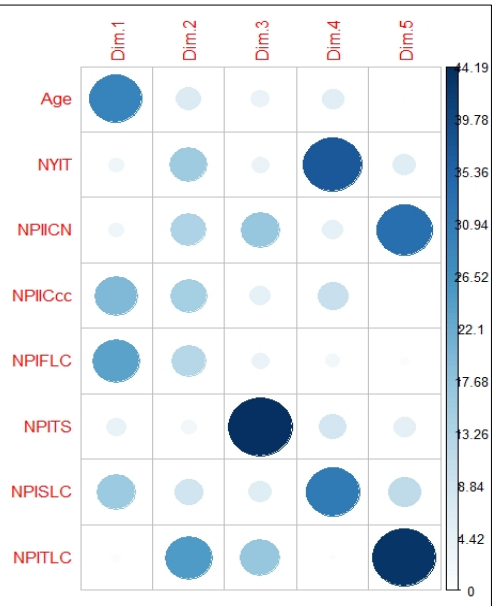

Table 3.13: Variables contribution to PCs formation, 2018/19 LP players

Considering the loadings (Table 3.13) and the relative percentages of contribution of the original variables to the PCs, the first PCs has as variable with greater contribution "Age" with 29.56% of contribution, followed by the variable "NPIFLC" (number of presences in first level championships) with 23.77%, "NPIICcc" (number of presences in international competitions (club competitions)) with 19.73% and "NPISLC" (number of presences in second level championships) with 16.15%. As far as the second PC is concerned, the most important variable is "NPITLC" (number of presences in third level championships) with about 24.77% of contribution.

The other variables that contribute to its formation, compete for about the same percentage. In fact, the variable "NYIT" (number of years in the team) competes for about 15.88%, "NPIICcc" (number of presences in international competitions (club competitions)) for about 14.99%, "NPIICN" (number of presences in international competition (national competitions)) for about 13.53% and just above the threshold, with about 12.77%, the variable "NPIFLC" (number of presences in first level championships). Special mention should be made of the third PC for its composition in terms of original variables. In fact, in this case the variable "NPITS" (number of presences in this season) contributes 44.19%, while the other two variables that exceed the threshold of 12.5%, namely "NPIICN" (number of presences in international

competitions (national competitions)) and "NPITLC" (number of presences in third level championships), contribute less than 17%.

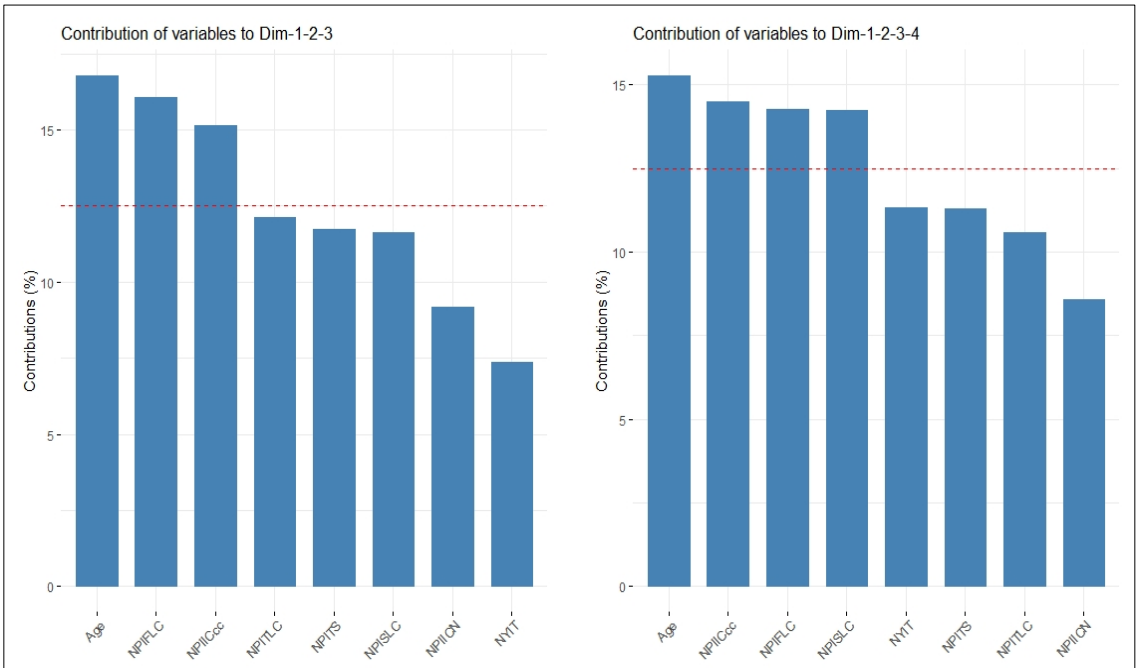

Figure 3.44: Contributions of variables according to eigenvalues and cumulative variance, 2018/19 LP players

Considering the same crossover points selected for the HP cluster, if we look at the eigenvalues, the PCs to consider are 3. In this case the original variables that on average exceed 12.5% of contribution for the first 3 PCs are "Age" with an average contribution above 15%, followed by "NPIFLC" (number of presences in first level championships) that also contributes to an average percentage above 15% and finally the variable "NPIICcc" (number of presences in international competitions (club competitions)) that is around 15% of average contribution. Analyzing instead the portion of variance explained, from the table of eigenvalues we can see how to exceed 80% of variance explained you need 5 PCs, coming to explain 87.02%. Instead, taking only the first 4 PCs, it explains 77.47% of the variance. As for the HP cluster, we will choose the PCs based on the portion of the variance explained closest to 80%, so even in this case the first 4. The original variables that on average contribute most to the formation of PCs are also in this case "Age", "NPIICcc" (number of presences in nternational competitions (club competitions)) and "NPIFLC" (number of presences in first level championships) with the addition of the variable "NPISLC" (number of presences in second level championships).

The table 3.14 represents the eigenvalues for the season 2017/18. As for the previous season, also in this case only 3 PCs have a value equal to or greater than 1 with a total explained variance of 66.79%.

	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	2.4350909	30.438636	30.43864
Dim.2	1.6690050	20.862562	51.30120
Dim.3	1.2390390	15.487988	66.78919
Dim.4	0.9011229	11.264036	78.05322
Dim.5	0.6431525	8.039406	86.09263
Dim.6	0.5866738	7.333422	93.42605
Dim.7	0.4177649	5.222061	98.64811
Dim.8	0.1081511	1.351888	100.00000

Table 3.14: Eigenvalues table, 2017/18 LP players

The first PC in this case explains about 30.44% of the variance, a lower percentage than the same in the 2018/19 season. The same applies to the second, which in this case explains 20.86% of the variance, compared to about 22.55% in 2018/19. Obviously the first two PCs in this case explain a variance of about 51.3%, which is a lower percentage than that seen in the 2018/19 season (Table 3.12-3.14). Compared to the 2018/19 season, however, in this case the first 4 PCs explain a greater portion of variance than in the 2018/19 season, with 78.05% about against 77.47%. This is due to the portion of variance explained by the third and fourth PCs that exceed the values expressed by them in the previous season.

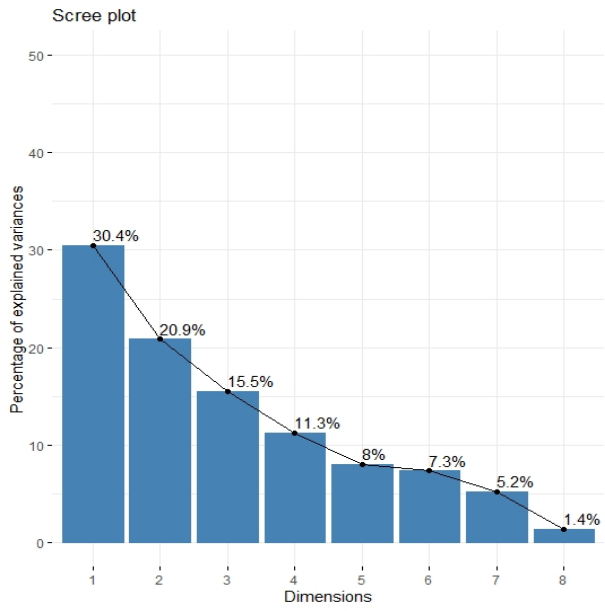

Figure 3.45: Scree plot, 2017/18 LP players

The screeplot (Figure 3.45) shows us that the difference in terms of explained variance between the first and the second PC is around 10%, a value higher than the difference found in the season 2018/19. In this season, however, the difference in variance between the second and third PCs is lower, with a value of about 5.4% compared to about 9.1% in the previous season analyzed. As far as the remaining PCs are concerned,

there are increasingly less marked differences, as is natural in the use of this methodology.

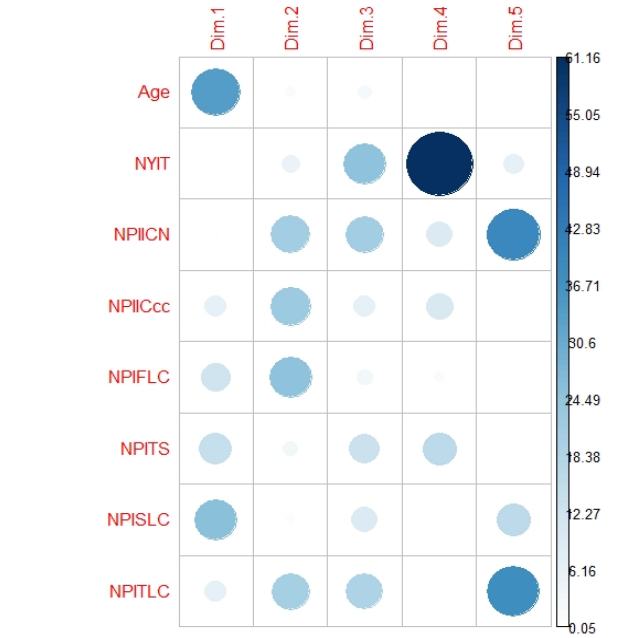

Table 3.15: Variables contribution to PCs formation, 2017/18 LP players

In this case, the percentages of contribution of the original variables to the formation of PCs (Tabe 3.15) see as the largest contributor in the creation of the first main component the variable "Age" with about 33.83%, followed by the variable "NPISLC" (number of presences in second level championships) with about 25.47% and the variable "NPITS" (number of presences in this season) with about 14.58%. Immediately below the minimum threshold, with about 12.34% of contribution, there is the variable "NPIFLC" (number of presences in first level championships). The variables common to the two seasons that contribute to the formation of the first two PCs, with an average contribution higher than 12.5%, are the variables "Age" and "NPISLC" (number of presences in second level championships). As

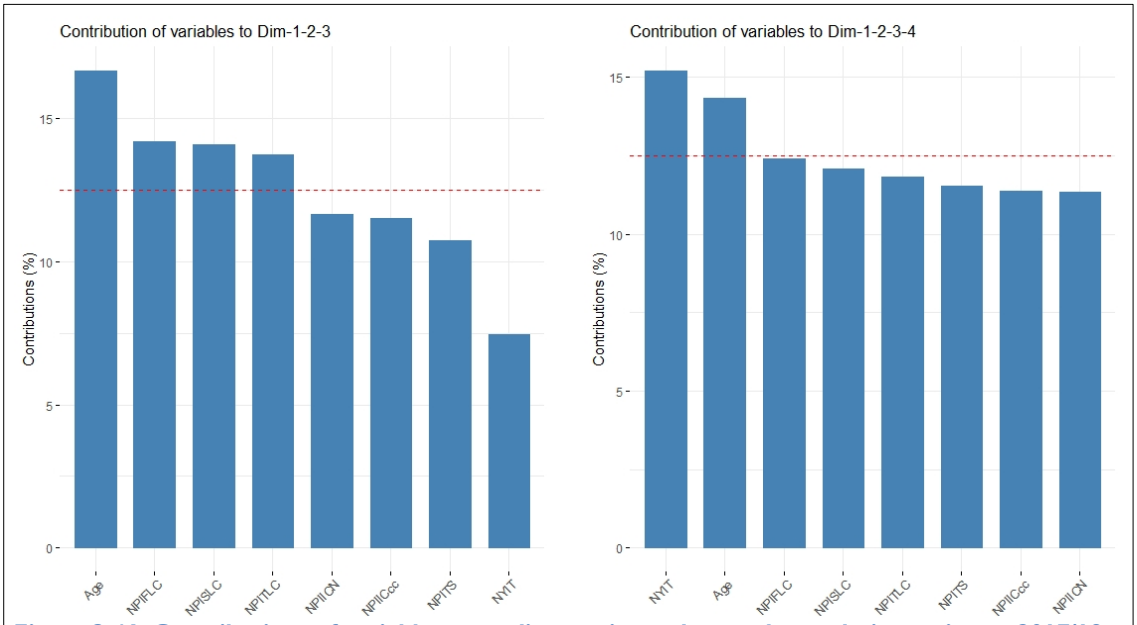

Figure 3.46: Contributions of variables according to eigenvalues and cumulative variance, 2017/18 LP players

for the variables common to the formation of the two second PCs in the two years, these are "NPIFLC" (number of presences in first level championships), "NPIICcc" (number of presences in international competitions (club competitions)), "NPIICN" (number of presences in international competiton (national competitions)) and "NPITLC" (number of presences in third level championships). These are also the only variables that exceed the contribution threshold of 12.5% in the formation of the second PC for the 2017/18 season. Respectively, they contribute about 24.75%, about 22.52%, about 21.05% and about 20.33%.

Turning now to consider the two crossover points (Figure 3.46), as previously mentioned, the PCs that equal or exceed the value 1 in this season are 3, with a total explained variance of 66.79%. The greatest average contribution, also in this case, is given by the variable "Age" which exceeds 15% of average contribution, while the variables "NPIFLC" (number of presences in first level championships), "NPISLC" (number of presences in second level championships) and "NPITLC" (number of presences in third level championships) are at lower levels but still above the threshold limit. The variables common to both seasons in terms of contribution are "Age" and "NPISLC" (number of presences in second level championships). If we look at the cumulative variance instead, also for this season the first 4 PCs will be taken into consideration, for a cumulative value of 78.05%, for the same criterion used in each previous case. In this season there are two variables that exceed the threshold of 12.5%, namely "NYIT" (number of years in the team) with about 15% and "Age", with a contribution slightly below 15%. Instead, just below the barrier threshold there is the variable "NPIFLC" (number of presences in first level championships) with just under 12.5%. The only variable common to both seasons is the variable "Age".

The 2016/17 season presents the portion of variance explained by the first highest PC of the three seasons analyzed so far, with about 34.32% of variance explained (Table 3.16). If we consider the first two PCs, the cumulative explained variance reached exceeds 60% compared to 53.77% in 2018/19 and 51.3% in 2017/18. Even taking the first 4 PCs, the cumulative variance explained is the highest of the three seasons with a percentage of 83.20%. In terms of eigenvalues, again only 3 PCs exceed or equal the value of 1.

	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	2.74547188	34.3183986	34.31840
Dim.2	2.06273235	25.7841544	60.10255
Dim.3	1.03592763	12.9490954	73.05165
Dim.4	0.81156606	10.1445757	83.19622
Dim.5	0.66109821	8.2637276	91.45995
Dim.6	0.47233690	5.9042112	97.36416
Dim.7	0.13500068	1.6875084	99.05167
Dim.8	0.07586629	0.9483287	100.00000

Table 3.16: Eigenvalues table, 2016/17 LP players

As already mentioned, the first PC explains a higher variance compared to the same in the other two seasons with 34.3% against 31.2% about 2018/19 and 30.4% in 2017/18.

The difference in variance explained between the first and the second PC is 8.5%, the lowest difference so far found in the LP cluster, with a value of 8.7% in 2018/19 and 9.5% in 2017/18. Instead, the difference between the second and the third PC is 12.9%, the largest found in the 3 seasons analyzed. After that, between the third and the sixth PC, for each PC there is a decrease of explained variance of about 2% (Figure 3.47).

From the analysis of the contribution of the original variables to the formation of the PCs (Table 3.17), the first PC is composed for about 25.84% of the variable "NPIFLC" (number of presences in first level championships), for about 24.46% of the variable "Age" and for about 23.36% of the variable "NPIICcc" (number of presences in

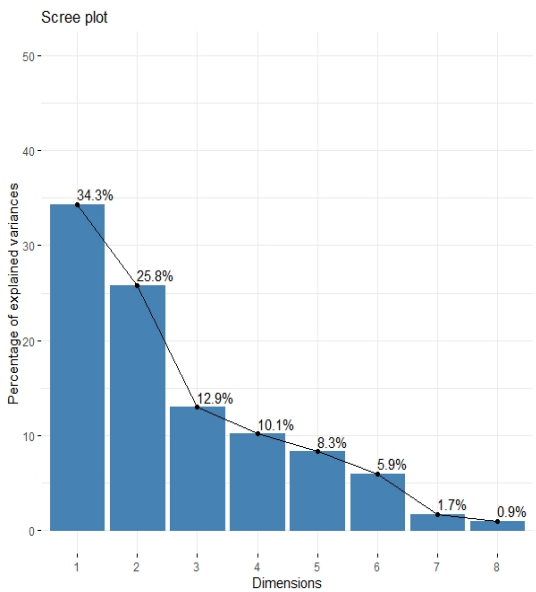

Figure 3.47: Scree plot, 2016/17 LP players

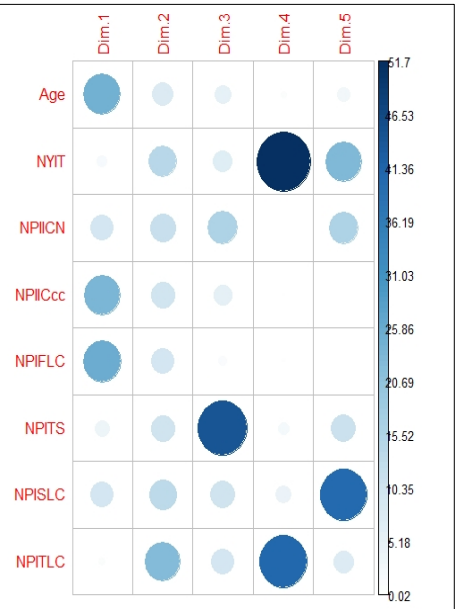

Table 3.17: Variables contribution to PCs formation, 2016/17 LP players

international competitions (club competitions)). The other contributions are all below the threshold of 12.5%. The only variable common to all three seasons is the variable "Age" while the variables "NPIFLC" (number of presences in first level championships) and "NPIICcc" (number of presences in international competitions (club competitions)) are common to this and the 2018/19 season. Considering instead the second main component, the variables that contribute most to its formation are the variable "NPITLC" (number of presences in third level championships) with about 22.29%, the variable "NYIT" (number of years in the team) with a contribution of about 14.43% and the variable "NPISLC" (number of presences in second level championships) with a contribution of 13.66%. All other variables contribute less than 12.5%. For example, just below the threshold limit, we find the variable "NPIICN" (number of presences in international competitions (national competitions)) with a contribution of 11.87% and with 10.55% the variable "NPITS" (number of presences in this season). In this case, the only variable that contributes in a meaningful way in each of the three seasons to the formation of the second main component is the variable "NPITLC" (number of presences in third level championships). Instead, the variable "NYIT" (number of years in the team) is common only to this season and to the season 2018/19. The variable "NPISLC" (number of presences in second level championships) is only present in this season for the formation of the second PC.

Considering the PCs that have a eigenvalue greater than or equal to one (Figure 3.48), as previously mentioned, we must consider the first 3 PCs, with a total explained variance of 73.05%. The first variable for average contribution of the first 3 PCs is the variable "NPIFLC" (number of presences in first level championships), which contributes more than 15%, followed by the variable "NPIICcc" (number of presences in international competitions (club competitions)), which also contributes more than 15%, as well as the variable "Age". Finally, the last variable that contributes more than 12.5% on average is the variable "NPITS" (number of presences this season). The variables common to all seasons analysed are "Age" and "NPIFLC" (number of presences in first level championships) respectively, while the variable "NPIICcc" (number of presences in international competitions (club competitions) is common only to the 2018/19 and 2016/17 seasons.

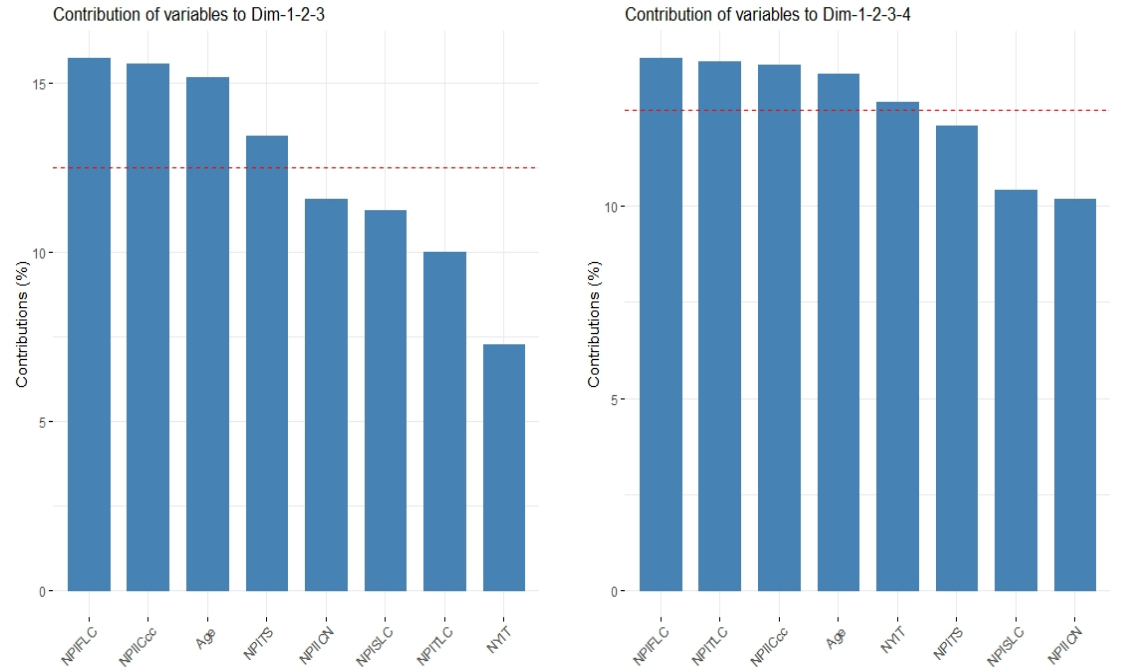

Figure 3.48: Contributions of variables according to eigenvalues and cumulative variance, 2016/17 LP players

Analyzing instead the crossover point relative to the cumulated variance (Figure 3.48), the first 4 main components express a cumulated variance of about 83,20%. In this case none of the contributory variables exceeds the average 15% of contribution. The variables to be considered are "NPIFLC" (number of presences in first level championships), "NPITLC" (number of presences in third level championships), "NPIICcc" (number of presences in international competitions (club competitions)), "Age" and "NYIT" (number of years in the team). Immediately below the threshold of dam we find the variable "NPITS" (number of presences in this season). In this case the only common variable for the contribution is the variable "Age", while the variable "NYIT" (number of years in the team) is common to the season 2017/18 and 2016/17 and the variables "NPIFLC" (number of presences in first level championships) and NPIICcc" (number of presences in international competitions (club competitions)) are common to the season 2016/17 and 2018/19.

The penultimate season of the LP cluster presents only two PCs with a eigenvalue equal to or greater than 1, i.e. the first and second PCs (Table 3.18). These two combined represent 56.37% of the total variance. To exceed 80% of the explained cumulative variance, 5 PCs must be used for a total of 86.55% of the explained cumulative variance.

	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	2.59352332	32.419042	32.41904
Dim.2	1.91646920	23.955865	56.37491
Dim.3	0.94060530	11.757566	68.13247
Dim.4	0.75761111	9.470139	77.60261
Dim.5	0.71557103	8.944638	86.54725
Dim.6	0.67231215	8.403902	94.95115
Dim.7	0.32349187	4.043648	98.99480
Dim.8	0.08041603	1.005200	100.00000

Table 3.18: Eigenvalues table, 2015/16 LP players

With 4 PCs, on the other hand, 77.6% of the cumulative variance can be explained. The first PC explains about 32.42% of the variance, a value lower only to that of the first PC in the 2016/17 season (34.32%). The same goes for the second PC, in fact in 2018/19 it counted for about 22.55%, while in the 2017/18 season it counted for about 20.86%. As a result, the cumulative variance explained by the first two PCs is also the second highest after that of the 2016/17 season.

This season's screeplot (Figure 3.49) shows a percentage difference between the variance explained by the first and second PCs of about 8.4%, an even lower difference than in the previous season. Respectively, the percentage difference between the first and second PC in 2018/19 was 8.7%, in 2017/18 9.5% and in 2016/17 8.5%. Instead, the difference between the second and the third PC is quite marked, with a percentage

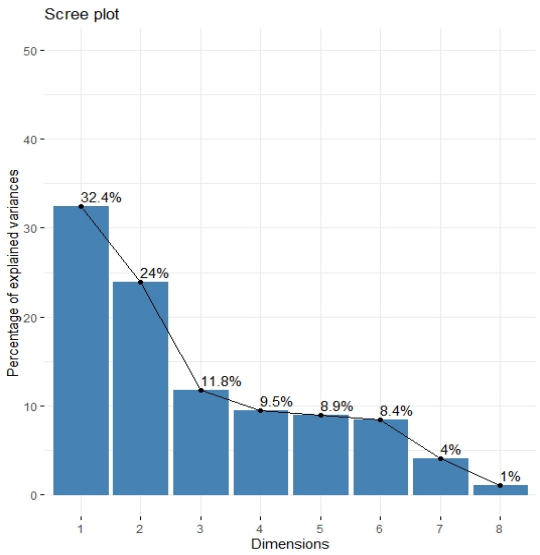

Figure 3.49: Scree plot, 2015/16 LP players

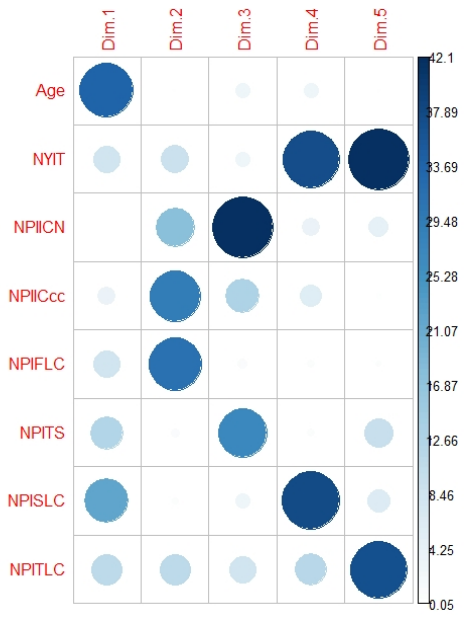

Table 3.19: Variables contribution to PCs formation, 2015/16 LP players

difference of about 12.2%, second only to the 2016/17 season which had a difference of 12.9%. The PCs from the fourth to the sixth instead do not present strong differences between them in terms of percentages of variance explained.

Turning to loadings (Table 3.19), the only two variables that exceed the threshold of 12.5% contribution are the variable "Age" and "NPISLC" (number of presences in second level championships), respectively with a contribution of 33.72% and 22.18% approximately. The variables "NPITS" (number of presences in this season) and "NPITLC" (number of presences in third level championships) are slightly below the 12.5% contribution with 12.43% and 11.23% approximately. The only variable common to all seasons for contribution to the formation of the first PC, is the variable "Age". Instead, the variable "NPISLC" (number of presences in second level championships) is common to 3 out of 4 seasons analyzed so far, namely the season 2018/19, 2017/18 and 2015/16. Turning to the analysis of the second PC (Table 3.19), there are three variables that exceed the contribution threshold, namely the variable "NPIFLC" (number of presences in first level championships) with 31.55% contribution, the variable "NPIICcc" (number of presences in international competitions (club competitions)) with 29.31% contribution and "NPIICN" (number of presences in international competitions (national competitions)) with 17.69%. The fourth highest contribution variable is "NPITLC" (number of presences in third level championships) but contributes about 11%. In this case, there is no variable that is common to all seasons taken into account for the contribution to the formation of the second PC, but the variables "NPIFLC" (number of presences in first level championships), "NPIICcc" (number of presences in international competitions (club competitions)) and "NPIICN" (number of presences in international comeptitions (national competitions)) are common to the season 2018/19, 2017/18 and 2015/16.

As already mentioned, in this case only two PCs equal or exceed the value 1 in terms of eigenvalue (Figure 3.50). Analyzing how for the other seasons the variables that on average contribute more than 12.5% to the formation of PCs can be seen from the graph as the variable "Age" contributes for almost 20%, followed by the variable "NPIFLC" (number of presences in first level championships) that far exceeds the average contribution of 15%. Below this threshold we find the variable "NPIICcc" (number of

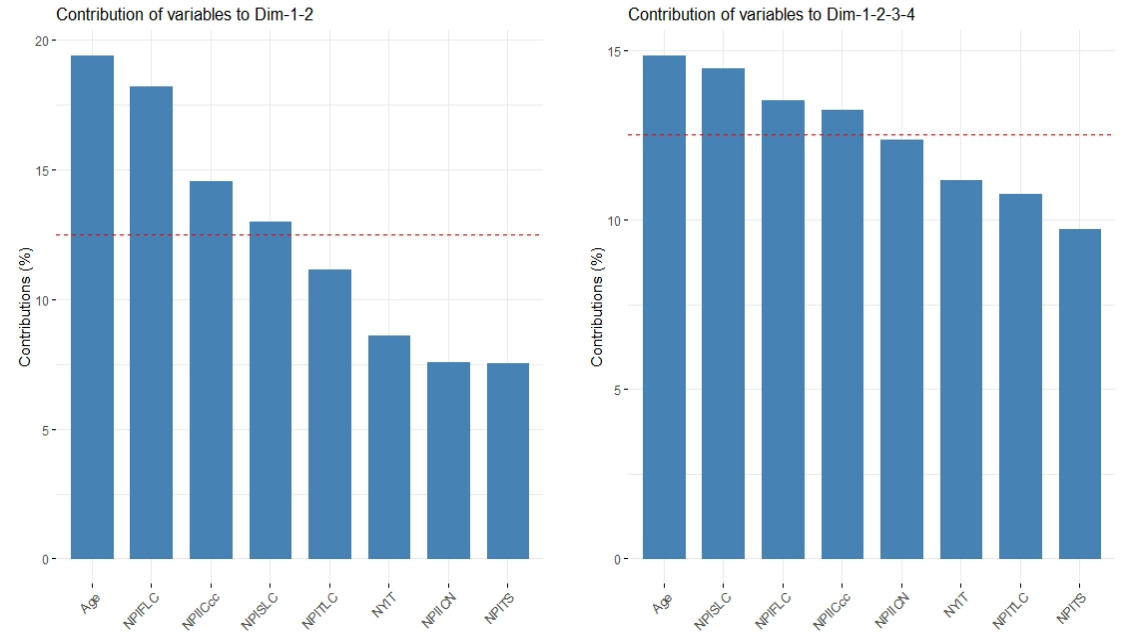

Figure 3.50: Contributions of variables according to eigenvalues and cumulative variance, 2015/16 LP players

presences in international competitions (club competitions)) followed by "NPISLC" (number of presences in second level championships). Speaking of common traits, the variable "Age" is a common contributor to the formation of PCs with eigenvalues greater than or equal to 1, for all seasons analyzed. The same can be said of the variable "NPIFLC" (number of presences in first level championships). The variable "NPIICcc" (number of presences in international competitions (club competitions)) instead is common to three seasons out of four, being present in the season 2018/19, 2016/17 and 2015/16. Finally, the variable "NPISLC" (number of presences in second level championships) is common only to this season and to the 2017/18 season.

Turning to the portion of cumulative variance explained, we must consider 4 PCs, with a cumulative variance value of about 77.6%. In this case no variable exceeds 15% of average contribution and as first variable we have the variable "Age", followed by the variable "NPISLC" (number of presences in second level championships), "NPIFLC" (number of presences in first level championships) and "NPIICcc" (number of presences in international competitions (club competitions)). Immediately below the threshold of barrage we find the variable "NPIICN" (number of presences in international competitions (national competitions)). The only variable common to all seasons is once again the variable "Age", while the variables "NPIICcc" (number of presences in

international competitions (club competitions)) and "NPIFLC" (number of presences in first level championships) are common to the seasons 2018/19, 2016/17 and 2015/16. Finally, the variable "NPISLC" (number of presences in second level championships) is common only to this season and to the 2018/19 season.

	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	2.70599914	33.824989	33.82499
Dim.2	2.13503057	26.687882	60.51287
Dim.3	0.94845197	11.855650	72.36852
Dim.4	0.79265666	9.908208	82.27673
Dim.5	0.63935074	7.991884	90.26861
Dim.6	0.46434656	5.804332	96.07295
Dim.7	0.22095591	2.761949	98.83489
Dim.8	0.09320847	1.165106	100.00000

Table 3.20: Eigenvalues table, 2014/15 LP players

The last season analysed is 2014/15 and in this case, to exceed 80% of the cumulative variance explained, 4 PCs are needed, with a percentage value of about 82.28% (Table 3.20). The first PC explains a percentage variance of about 33.82% against 32.42% explained by the first PC in the 2015/16 season, 34.32% in 2016/17, 30.44% in 2017/18 and 31.23% in 2018/19. As for the second PC, the one of the season analyzed is the one with the highest percentage of explained variance of the 5 seasons taken into account, with 26.69% of explained variance. Respectively in 2018/19 the second PC explained 22.55% of variance, in 2017/18 20.86%, in 2016/17 25.78% and in 2015/16 23.95%. Given the values of the first two PCs, the cumulative variance explained by the first and second PCs is the highest recorded with 60.51%, surpassing even that of the 2016/17 season which reached about 60.10%. Also in this season, as in the previous one, the PCs that equal or exceed the value 1 are only the first 2.

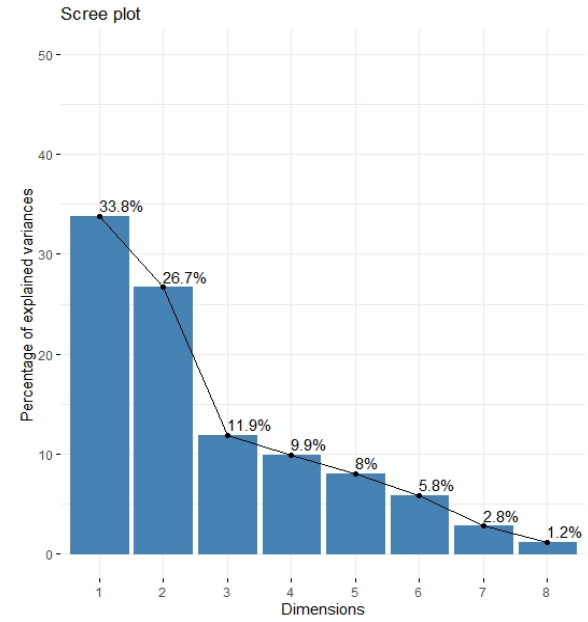

Figure 3.51: Scree plot, 2014/15 LP players

From the analysis of the screeplot (Figure 3.51) it can be seen that the percentage difference of variance

explained between the first and the second PC is equal to about 7.1%, the lowest difference recorded in the cluster of LPs. Instead, the difference between the second and the third PC in this season is the highest recorded with a percentage of about 14.8%.

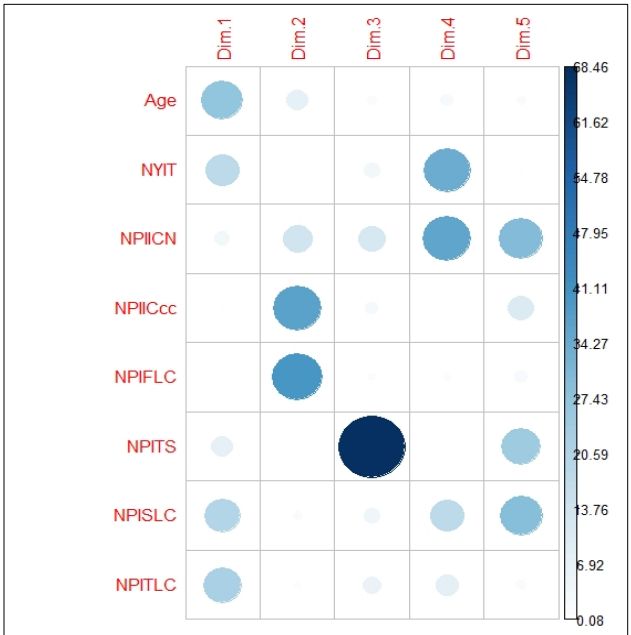

Table 3.21: Variables contribution to PCs formation, 2014/15 LP players

The variables that have contributed most to the formation of the first PC in this case are "Age" with a contribution of 27.02%, "NPITLC" (number of presences in third level championships) with 22.6% contribution, "NPISLC" (number of presences in second level championships) with 20.53% and "NYIT" (number of years in the team) with about 18.07%. The only variable common to all the first main components is once again the variable "Age". Instead, the variable "NPISLC" (number of presences in second level championships) is common to all seasons except 2016/17. Turning instead to the analysis of the second main component, in the current year the most important original variable for contribution is "NPISLC" (number of presences in second level championships) with 39.81% contribution, followed with 36.72% about by the variable "NPIICcc" (number of presences in international competitions (club competitions)) and the variable "NPIICN" (number of presences in international competitions (national competitions)) with 13.23%. Moving on to consider the traits common to all the second main components of the 5 seasons considered, there are no variables that contribute to its constitution for a percentage value greater than 12.5%. Instead, there are 3 variables, namely "NPIFLC" (number of presences in first level championships), "NPIICcc" (number of presences in international competitions (club competitions)) and "NPIICN" (number of presences in international competitions (national competitions)), which are common to four out of five seasons, with the sole exception of the 2016/17 season.

In this season, the first 4 PCs explain 82.28% of the cumulative variance and the variables that on average contribute most to their formation are all below 15% of contribution (Figure 3.52). The most important variable is still "Age", followed by "NPIFLC" (number of presences in first level championships), "NPITS" (number of presences in this season) and "NPIICcc" (number of presences in international competitions (club competitions)). The only variable common to all seasons is the variable "Age", while the variables "NPIFLC" (number of presences in first level championships) and "NPIICcc" (number of presences in international competitions (club competitions)) are common to the season 2018/19, 2016/17, 2015/16 and 2014/15.

If once again, instead of considering the cumulative variance, we look at the value of the eigenvalues, then from the graph we see how, of the two main components that have a value equal to or greater than 1, the varibaile that contributes most to the composition of the two is "Age" with more than 17.5% contribution, followed by "NPIFLC" (number of presences in first level championships) that contributes about 15% and "NPIICcc" (number of presences in international competitions (club competitions)) and just above the threshold of barrier "NPITLC" (number of presences in third level championships). Considering this crossover point there are two variables that are common to all five seasons, namely "Age" and " NPIFLC" (number of presences in first level championships)

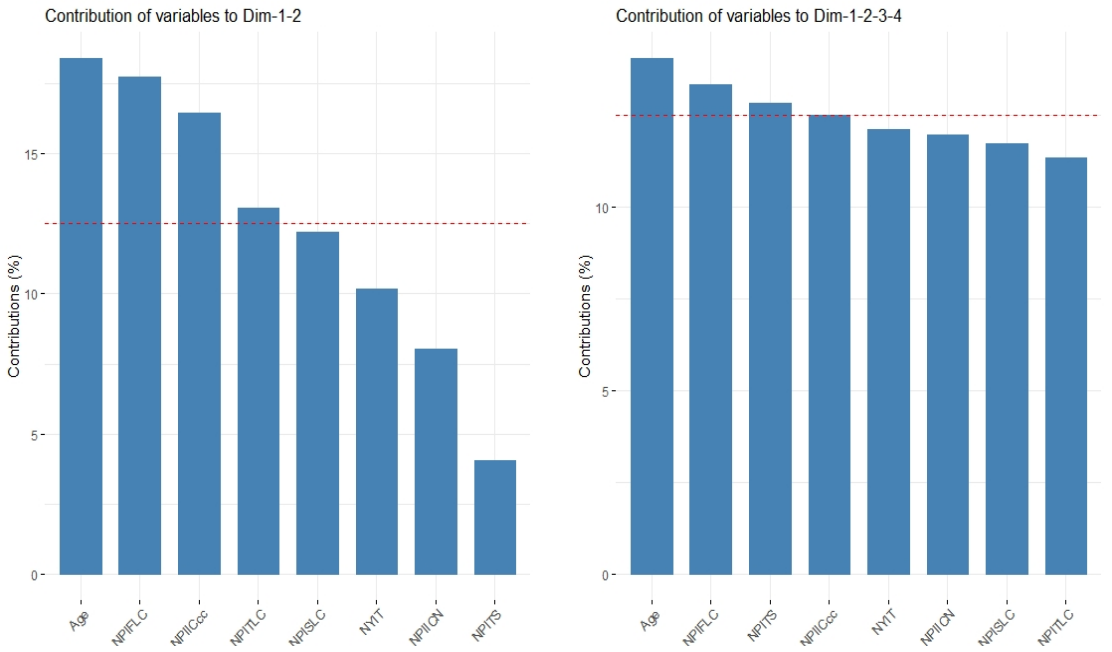

Figure 3.52: Contributions of variables according to eigenvalues and cumulative variance, 2014/15 LP players

while the variable "NPIICcc" (number of presences in international comeptitions (club competitions)) is common to all seasons except that of 2017/18.

3.6. NORMALITY TEST

To make sure that the methodology applied so far was usable and gave reliable results, we performed the Shapiro-Wilk test to test the normality of the data distribution. The "W" value of the test can take values between 0 and 1 and if the value is too small, the test rejects the null hypothesis that the sample values are distributed as a normal random variable. In our case the players dataset in the 2018/19 season assumes a value W=0.93941, in 2017/18 assumes a value W=0.92662, in 2016/17 W=0.97956, in 2015/16 W=0.84531 and in the last season W=0.87457. So in all seasons the test assumes a value high enough to accept the null hypothesis, i.e. that the data are distributed normally.

3.7. RESULTS RESUME AND CONCLUSIONS

As previously mentioned, the PCA analysis for the two joint clusters did not give definite and clear results in terms of characteristics describing one or the other cluster. Moreover, the first two PCAs that expressed a cumulative variance between 53-56% have as major contributors to their formation "Age", "NPIFLC" (number of presences in first level championships) and "NPIICcc" (number of presences in international comeptitions (club competitions)). Considering instead the separate datasets and putting as crossover point for the choice of the number of PCs to use, an eigenvalue value equal or higher than 1 and a cumulated variance explained as close as possible to 80%, in the first case we have the variables "Age", "NPIFLC" (number of presences in first level championships) and "NPIICcc" (number of presences in international competitions (club competitions)) for both clusters in four out of five seasons. So even with separated clusters, the result is practically the same as the one obtained through the joint cluster analysis. This tells us that the two datasets do not have strong differences between them but the consideration of this crossover point is relatively not solid, in fact if we examine

these values, the explained cumulative variance also varies a lot from season to season. This effect is much more evident in the LP cluster than in the HP cluster. For example in the HP cluster in 2018/19 the PCs considered explained about 70.26% of the variance, in 2017/18 66.67%, in 2016/17 66.54%, in 2015/16 65.66% and in 2014/15 60.51%. In the LP cluster, on the other hand, PCs expressed 67.21% in 2018/19, 66.79% in 2017/18, 73.05% in 2016/17, 56.37% in 2015/16 and 60.51% in 2014/15.

If instead as crossover point the accumulated variance is taken into account as close as possible to 80%, then the original variables with the highest average contribution in the HP cluster are "Age", "NPIFLC" (number of presences in first level championships) and "NYIT" (number of years in the team). The variable "NPIICcc" (number of presences in international competitions (club competitions)) did not appear in this list because in two seasons it was just below the minimum threshold of 12.5%. As for the LP cluster, the most present variables are "Age", "NPIFLC" (number of presences in first level championships) and "NPIICcc" (number of presences in international competitions (club competitions)). So the two clusters have two variables out of three in common with the difference of the presence of the variable "NYIT" (number of years in the team) for HP and the variable "NPIICcc" (number of presences in international competitions (club competitions)) for LPs. This crossover point can be considered as more solid and more explanatory than the dataset. In fact, as far as the HP cluster is concerned, in 2018/19 the explained variance is about 80.40%, in 2017/18 78.30%, in 2016/17 76.78%, in 2015/16 77.09% and in 2014/15 77.84%. Instead, for LPs in the 2018/19 season the explained variance is 77.47%, in 2017/18 78.05%, in 2016/17 83.20%, in 2015/16 77.60% and in 2014/15 82.28%. As it can be seen, the percentage of variance varies less from season to season within the two clusters than the portion of variance explained if we take into account the values of eigenvalues.

The studiesconducted in this chapter, both descriptive and PCA methodology analyses, did not show a strong differentiation between the HP and LP clusters. It would seem from these preliminary analyses that the characteristics of the experience that most distinguish the two clusters are the same with the only difference underlined in Table 3.22. Starting from these results, in the next chapter we will analyze the variables found through the PCA in depth. We will use the multiple logistic regression for each variable

with a single dataset containing the HP and LP players for each season. This will be done to understand if the variables actually affect the probability that a player belongs to the HP or LP cluster and to what extent.

MAIN VARIABLES FOR THE PCs CREATION (80% VARIANCE)	ORIGINAL VARIABLE	ORIGINAL VARIABLE	ORIGINAL VARIABLE
HP	"Age"	"NPIFLC"	"NYIT"
LP	"Age"	"NPIFLC"	"NPIICcc"

Table 3.22: Original variables that contributed most to the creation of the PCs (80% variance) in at least 4 out of 5 seasons in the HP and LP clusters

CHAPTER 4

MODELS OF ANALYSIS AND MANAGERIAL IMPLICATIONS

4.1. INTRODUCTION

After having analyzed in detail the context within which the data were collected, the literature on the relationship to be investigated in the three areas related to the three the collected datasets and the preliminary statistical analysis on the datasets, with particular focus on the players dataset, in this chapter the nature of the results of the analysis of the previous chapter will be investigated. Starting from the players dataset and the results of the PCA over the 5 years, an examination of multiple logistic regression over the 5 years will be made to see how the variables found interact with the probability that a player belongs to the HP cluster or not. For example, although the PCA has found practically the same variables for HP and LP players, it will be studied whether a variable common to both has a positive, insignificant or negative influence on the probability that the player belongs to the HP cluster. In addition to these detailed analyses on the players dataset (focus of the study) a very simplistic scoring system will be proposed for the examination of the other two datasets, given the low reliability and difficult interpretation of the same analysis made for players on datasets that have only few observations for each year. Finally the limitations of the following study and the possibility of new insights will be listed. In this chapter also the managerial implications of the results obtained will be presented.

4.2. LOGISTIC REGRESSION: A BRIEF EXPLANATION

Logistic regression is used to analyse the relationship between two quantitative variables X and Y, when Y is a categorical variable and has a binomial distribution (Samuels et al., 2015). So when the response variable is dichotomous, logistic regression can be used to

interpret this relationship that will develop through the constitution of a curve that will always assume values between 0 and 1. With a curve that can only assume values between 0 and 1, logistic regression can be used to estimate the probability that Y=1 (i.e. that the player belongs to the HP cluster) given a certain value of X. With the variable Y assuming a value of 1/0, it would still be possible to apply a linear regression model from a mathematical point of view but this would no longer be appropriate. This is due to the formula describing the linear model i.e. Y= a + bx which therefore implies that the values of Y can go from less infinite to more infinite. So if the linear regression model is used with a dichotomous variable it is possible to obtain values of Y greater than 1 when X increases and values less than 0 when X decreases. To overcome this problem the relationship between X and Y should be dictated by a logistic curve. The use of a non-linear relationship does not allow the OLS method to be applied unless the necessary transformations are made that can make the relationship linear. In order to better understand the methodology, considerations must be made regarding the nature of the dependent variable. In the logistic regression the purpose of the Y is to define the belonging to one or the other cluster. The assigned values are attributed in an arbitrary way and therefore what interests is not the expected value but the probability that a given subject belongs or not to one of the two groups.

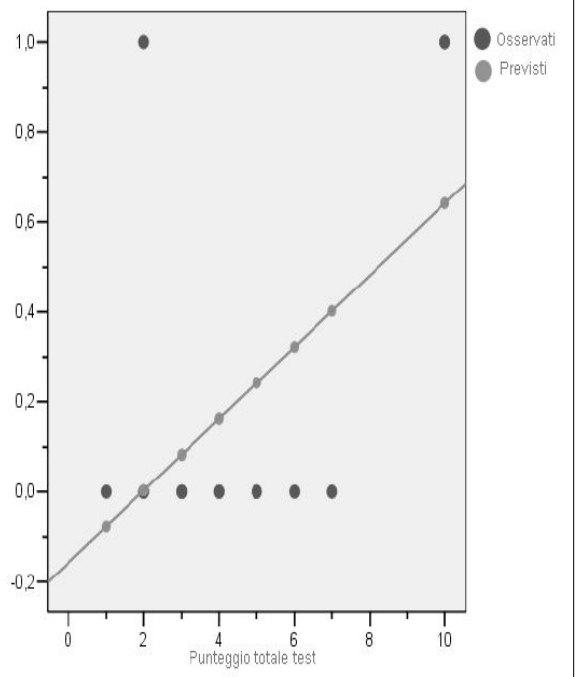

Figure 4.1: Simple linear regression model with a dichotomous dependent variable. Source: Senese, 2014

Even if the assignment of the values to the Y is arbitrary, they can influence the results of the analysis and a solution is to replace the probability with the odds. This is the way to express a probability through a relationship that is obtained by comparing the frequencies observed on the one level with the frequencies observed on the other, then the relationship between two categories is expressed. For example if in a sample we

have 30 men and 12 women, using the odds we will get 2.5, so for each woman there are 2.5 men. If instead to express the relationship between two categories as a function of another variable, it is possible to use the odds ratio or ratio between the odds obtained by making a ratio between the odds of a given variable obtained for each level of the second variable. If the odds ratio has values other than 1 then there is an association between the variables. When getting the odds, it is possible to calculate the natural logarithm of this by getting the logit. All statistichs discussed so far provide the same information but with different mathematical values. Going down in detail, if the two categories represented by the Y are equivalent, then the relative frequencies are nozzles at 0.5 for both with odds equal to 1 and logits equal to 0. If, on the other hand, the number of successes is greater than the number of failures, then the relative frequencies are higher than 0.5 for category Y=1 and lower for category Y=0. For what concerns odds, they assume values higher than 1 while logits values higher than 0. If, on the other hand, the number of failures exceeds the number of successes, the relative frequencies are lower than 0.5 for Y=1 and higher for Y=0. The odds assume values lower than 1 and the logit values negative. To summarize, the relative frequencies have a range of variability from 0 to 1, the odds a range from 0 to more infinite and the logits can go from less infinite to more infinite. Speaking of the model parameters, as in the case of linear regression, it is necessary to evaluate the model parameters to better understand the relationship between X and Y. As previously said, with logistic regression, the OLS method cannot be applied but it uses the maximum-likelihood (ML) algorithm that estimates the model parameters so that we can maximise the log-likelihooc function that indicates the probability of obtaining a certain value of Y given the values of the independent variables. To understand the adequacy of the model for the representation of the data, indicators similar to those used in linear regression are used. In the case of logistic regression the value of the log-likelihood is used multiplied by -2 which assumes the notation -2LL. If this assumes large and positive values, then there is a low ability of the model to predict data. If a model including intercept only is used, this indicator is the analogue of the deviance in linear regression. If the model contains both the intercept and the variables, then the indicator represents the part of the data variability that is not explained by the model or error deviance. The difference between the two deviances, the one related to the model with the intercept only or the

one related to the model with all variables, indicates the part of variability explained by the independent variables. This difference is also called Chi-square of the model and tells us the amount of error reduction due to the model. To see in percentage how much the model reduces the error and explains a higher percentage of variance we use McFadden's formula.

Finally, if the contribution of the individual predictors is to be considered, to evaluate the contribution of each independent variable on the dependent variable it can be tested the significance of this one. This is done when the model including the variables is better in predicting the dependent variable than the model including only the intercept. Once this has been verified, to check the contribution of each variable, regression coefficients that can be standardized or non-standardized are considered. The standardised coefficient is a coefficient that has been calculated using the standard deviation as the unit of measurement. A parameter that can be used is the odds ratio, which expresses the variation of the dependent variable as a function of changes in the independent variable. If this assumes a value greater than 1 it means that as the independent variable increases the probability of Y=1 and if it is less than 1 it means that as the independent variable increases the probability of Y=1 decreases.

4.3. LOGISTIC REGRESSION ON PLAYERS' DATASET

In this paragraph the multiple logistic regression will be used to further process the results obtained by the PCA. In the previous chapter it resulted how the HP cluster explain most of the variance with the variables "Age", "NPIFLC" (number of presences in first level championships) and "NYIT" (number of years in the team) with the variable "NPIICcc" (number of presences in international competitions (club competitions)) slightly out of consideration for the fact that for two out of 5 seasons it was just under the 12.5% barrier. Instead, the LP cluster saw as the most important variables for the explanation of the variance, the variables "Age", "NPIFLC" (number of presences in first level championships) and "NPIICcc" (number of presences in international competitions (club competitions)). From this analysis it would seem that the only variable that

differentiates the two clusters is the "NYIT" (number of years in the team) variable for HP. The PCA tells us which original variables contribute to the formation of the PCs that explain most of the variance, so through this methodology we can go back to those that are the most important variables to describe the dataset but it doesn't tell anything about the fact that the increase of the value of the variable "NYIT" (number of years in the team) corresponds to a higher ranking position and therefore a belonging to the HP cluster or vice versa. In this case, since it does not seem to be variables that distinguish one or the other cluster in an important way, the relationship between the variables found and the belonging to one or the other cluster will be analyzed by using the multiple logistic regression. In this way it will be easier to understand if, despite the great similarity of the most important variables, an increase or decrease in the value of these correspond to the belonging of the HP cluster or vice versa. The analysis will be done considering all the variables together for each year and with the two clusters of players united.

In order to carry out the study with the logistic regression method, as first thing is to create through R, a matrix of the data to analyze. This will start with the 2018/19 season. Once the matrix has been created, it will be searched if there is a relationship between the dependent variable (TN or team name, in this case the name are 1 if the player belongs to a HP team and 0 if he belongs to a LP team) and the independent variables ("Age", "NYIT", "NPIFLC" and "NPIICcc"). In order to check the hypothesis, the residual deviance of the null hypothesis model, i.e. a model that does not foresee an influence of the independent variables, and the residual deviance of the model including the alternative hypothesis, i.e. the independent variables influence the dependent variable, are calculated. To make the comparison between the two the Chi-square theoretical distribution will be used. In strictly parametric terms, the null hypothesis assumes that the parameter representing the slope of the independent variables is equal to 0 and that therefore model 2 does not succeed in significantly reducing the residual deviance of model 1. Obviously, the null hypothesis assumes that the parameter representing the slope of the independent variables is different from 0. To verify this, through R, it is possible to use a generalized linear model with a binomial function.

```
Analysis of Deviance Table

Model 1: TN ~ 1
Model 2: TN ~ Age + NYIT + NPIFLC + NPIICcc
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       448     613.58
2       444     602.96  4   10.615  0.03124 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Table 4.1: Deviance table, 2018/19 season

The analysis provides the output of table 4.1. The probability value referring to the statistical test is p=0.0312, therefore lower than the value of 0.05. This result leads to deny the null hypothesis that the variance of model 2 was random. In this way the alternative hypothesis must be accepted, i.e. the variance explained by model 2 is not random and is greater than the one explained by model 1. To understand better in quantitative terms what it is being talked about, it is necessary to apply McFadden's formula to see the percentage of variance explained by Model 2 compared to Model 1. The result is 0.01730081, so a result of about 2% better, so not a large portion more. This means that the addition of the independent variables explains only about 2% more variance than a model that only included the intercept. From the Model 2 results it can be seen that the function that explains the relationship between HP cluster membership and independent variables includes the intercept and the "Age" and "NYIT" (number of years in the team) variables that are statistically significant (Table 4.2). This means that if the level of the variable "Age" and "NYIT" (number of years in the team) is 0, the logit of the variable "TN" is about 1.4183. Probably this result is due to the higher presence of individuals belonging to HP teams in the dataset. Analyzing in detail the variables "Age" and "NYIT" (number of years in the team) it can be seen that in the case of the former the relationship is negative, so as it increases the probability of an individual to be in an HP team decreases, while for the latter it is exactly the opposite (Table 4.2).

Looking at the absolute value of these coefficients one can see that they are extremely close to 0 and therefore, although the relationship is statistically significant, the influence is really weak.

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.4182644  0.5861701   2.420   0.0155 *
Age         -0.0577952  0.0243621  -2.372   0.0177 *
NYIT         0.1449616  0.0645104   2.247   0.0246 *
NPIFLC       0.0009554  0.0025816   0.370   0.7113
NPIICcc      0.0328774  0.0331403   0.992   0.3212
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Table 4.2: Multiple logistic regression coefficients, 2018/19 season

Trying to build another model that contained only the statistically significant variables of the previous model, it was performed, as in the previous case, the residual deviance of this model in comparison to the model including all the variables.

In the comparison between the model containing only the statistically significant variables of model 1 and the model itself, the value of p=0.1245, i.e. greater than p=0.05, can be seen, so this leads to not rejecting the null hypothesis (Table 4.3). Verifying this statement through McFadden's formula, this brings a result equal to -0.006911566. So this model explains less variance than model 1 but in a really small portion. So the variables excluded, that is "NPIFLC" (number of presences in first level championships) and "NPIICcc" (number of presences in international competitions (club competitions)) together contributed to an explanation of the variance very close to zero.

```
Analysis of Deviance Table

Model 1: TN ~ Age + NYIT + NPIFLC + NPIICcc
Model 2: TN ~ Age + NYIT
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       444     602.96
2       446     607.13 -2  -4.1674   0.1245
```

Table 4.3.: Deviance table B, 2018/19 season

Therefore, analysing the 2018/19 season it is possible to conclude that the variables "NPIFLC" (number of presences in first level championships) and "NPIICcc" (number of presences in international competitions (club competitions)) contribute to the increase of the individual's probability of belonging to the HP cluster in a non-statistically

significant way and that the variable "Age" instead presents a negative relation and the variable "NYIT" a positive relation. In both cases the contribution found is very low.

If the 2017/18 season is considered, the comparison of the variance between model 1, i.e. the one including only the intercept and model 2, i.e. the one with all variables, gives the results recorded in table 4.4.

```
Analysis of Deviance Table

Model 1: TN ~ 1
Model 2: TN ~ Age + NYIT + NPIFLC + NPICcc
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1       416     562.25
2       412     537.85  4   24.396 6.651e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Table 4.4.: Deviance table, 2017/18 table

As in the case of the previous season, the null hypothesis must be denied, since the value of p is extremely significant in statistical terms. Therefore the model including the variables is more useful in explaining the investigated relationship. There is an influence between the probability that an individual belongs to the HP cluster and the variables included in the model. The application of McFadden's formula results in 0.04339081, so the variance explained by model 2 is about 4% higher than model 1. In the previous season the variance explained more in percentage terms was about 1.7%.

Looking at the coefficients of model 2 (Table 4.5), it can be seen that the intercept, the variable "Age", "NYIT" (number of years in the team) and "NPIICcc" (number of

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.266837   0.617669   0.432  0.66574
Age         -0.011317   0.025226  -0.449  0.65370
NYIT         0.096770   0.078153   1.238  0.21564
NPIFLC       0.010101   0.003304   3.057  0.00224 **
NPICcc       0.011340   0.042961   0.264  0.79181
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Table 4.5: Multiple logistic regression coefficients, 2017/18 season

presences in international competitions (club competitions)) have a value of p greater than 0.05 and therefore confirm the null hypothesis that they do not influence the probability that an individual belongs to the HP cluster. The only variable that shows a statistical significance is the variable "NPIFLC" (number of presences in first level championships) with p=0.00224. Considering the coefficient, a positive relationship can be found which implies that as the value of the variable increases, the probability of being part of the HP cluster increases. In the previous season, the two statistically significant variables "Age" and "NYIT" (number of years in the team) had assumed values of -0.0577952 and 0.1449616 respectively, therefore higher in absolute value than the one found in this season for the variable "NPIFLC" (number of presences in first level championships).

Trying to analyze the data with a second model that takes into account only the statistically significant variables of the model including all the variables obtained through the PCA, a value of p is obtained that is higher than 0.05, therefore statistically not significant (Table 4.6). For precision the value is p=0.6414. Also in this case to verify that the explained variance of the new model is lower than that of model 1, the McFadden formula is applied, which returns a value of -0.003123098. As in the previous season, the amount of variance explained less by the new model is really very low. So the variables excluded from model number 3, i.e. "Age", "NYIT" (number of years in the team) and "NPIICcc" (number of presences in international competitions (club competitions)) contributed in model number 2 to explain a variance that was practically zero. What can be deduced from this season is that, all other factors being equal, which are not statistically significant and contribute to explain a tiny portion of variance, when the value of the variable "NPIFLC" (number of presences in first level championships) increases,

Analysis of Deviance Table

Model 1: TN ~ Age + NYIT + NPIFLC + NPICcc
Model 2: TN ~ NPIFLC

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	412	537.85			
2	415	539.53	-3	-1.6798	0.6414

Table 4.6: Deviance table B, 2017/18 table

the probability of the individual to belong to the HP cluster increases but by a value very close to zero.

Moving on to the 2016/17 season to verify the hypothesis that model 2, i.e. the one including all variables, explains a greater portion of variance than the model including only the intercept, as in the other seasons the residual deviance of the null hypothesis model and the residual deviance of the model including the alternative hypothesis are calculated. In strictly parametric terms, the null hypothesis assumes that the parameter representing the slope of the independent variables is equal to 0 and that therefore model 2 does not succeed in significantly reducing the residual deviance of model 1.

```
Analysis of Deviance Table

Model 1: TN ~ 1
Model 2: TN ~ Age + NYIT + NPIFLC + NPIICcc
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       383     505.99
2       379     498.09  4   7.9056   0.0951 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Table 4.7.: Deviance table, 2016/17 season

In this season the value of p=0.0951 which is therefore higher than 0.05, the barrier threshold for statistical significance, so the alternative hypothesis that model 2 would explain more variance than the model including only the intercept can be rejected (Table 4.7). But McFadden's formula gives us a result equal to 0.0156239, so the use of variables leads to an explanation of greater variance and moreover the deviance results to have a positive value of 7.9056.

Analyzing specifically the coefficients (Table 4.8) relative to this model, only the one relative to the variable "NYIT" (number of years in the team) results to be below the threshold of 0.05, while all the others result to be higher. Therefore, the only statistically significant parameter is the one expressed by the variable "NYIT" (number of years in the team) The coefficient has a positive relation with a value equal to 0.196275, the highest absolute value found in the seasons analyzed so far.

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.568190   0.663831   0.856   0.3920
Age         -0.017732   0.028520  -0.622   0.5341
NYIT         0.196275   0.083639   2.347   0.0189 *
NPIFLC       0.004125   0.003070   1.344   0.1790
NPIICcc     -0.030204   0.033831  -0.893   0.3720
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Table 4.8: Multiple logistic regression coefficients, 2016/17 season

Trying again to build a model that contains only the statistically significant variables of the previous model (Table 4.9), as for the previous model, the value of p is higher than the 0.05 barrier threshold. Moreover, the deviance assumes a negative value. This shows that the alternative hypothesis that the model containing only the variable "NYIT" (number of years in the team) is better than the model containing all variables should not be accepted. This is also demonstrated by the value assumed by applying McFadden's formula which is -0.003841681.

```
Analysis of Deviance Table

Model 1: TN ~ Age + NYIT + NPIFLC + NPIICcc
Model 2: TN ~ NYIT
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       379     498.09
2       382     500.00 -3  -1.9135   0.5906
```

Table 4.9: Deviance table B, 2016/17 season

So the analysis of the 2016/17 season showed how, with other variables being equal, the variable "NYIT" (number of years in the team) positively influences the probability of belonging to the HP cluster. The same result was found, but with a slightly lower coefficient value and a negative effect given by the variable "Age", in the 2018/19 season. In spite of these considerations, however, this model cannot be considered since the results shown in Table 4.7 have seen the model not statistically significant.

Moving on to consider also for the 2015/16 season the deviance analysis for the two models, in this case the value of p=0.01941, therefore statistically significant and can be rejected the null hypothesis and accept the one that claims that the variables influence

```
Analysis of Deviance Table

Model 1: TN ~ 1
Model 2: TN ~ Age + NYIT + NPIFLC + NPIICcc
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       421     568.98
2       417     557.24  4   11.738  0.01941 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Table 4.10: Deviance table, 2015/16 season

the relationship between them and the probability of belonging to the HP cluster (Table 4.10). The deviance in this case is equal to 11,738, reconfirming the rejection of the model including only the intercept. The further confirmation is obtained from the result of McFadden's formula which is 0.02062898, so the model explains a 2% higher portion of variance than the model without the other variables.

From the table of coefficients for this season it can be seen that no value of p is less than 0.05, so no variable has a statistically significant influence on the relationship with probability (Table 4.11). The only two variables that have a level of significance below 0.1 are "NYIT" (number of years in the team) and "NPIFLC" (number of presences in first level championships).

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.758996   0.578342   1.312   0.1894
Age         -0.031633   0.025116  -1.259   0.2079
NYIT         0.164397   0.094268   1.744   0.0812 .
NPIFLC       0.005679   0.003035   1.872   0.0613 .
NPIICcc      0.021805   0.033135   0.658   0.5105
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Table 4.11: Multiple logistic regression coefficients, 2015/16 season

Also in this case a model containing only the statistically significant variables of the model including all the variables, even if in this season, as previously said, there is no variable with a p value lower than 0.05.

From the deviance table (Table 4.12) it can be seen that the value of p is greater than 0.05 and therefore the null hypothesis cannot be rejected. This is also confirmed by the deviance value which is negative and McFadden's formula gives a result equal to -

0.003926348, so the new model explains a lower portion of variance, even if the absolute value is practically null.

```
Analysis of Deviance Table

Model 1: TN ~ Age + NYIT + NPIFLC + NPIICcc
Model 2: TN ~ NYIT + NPIFLC
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       417     557.24
2       419     559.43 -2  -2.1879   0.3349
```

Table 4.12: Deviance table, 2015/16 season

From the analysis of this season it can be argued that the variables "NYIT" (number of years in the team) and "NPIFLC" (number of presences in first level championships), with a p value greater than 0.05 but less than 0.1, positively influence the probability of an individual to be part of HP. The variable "NYIT" (number of years in the team) had shown this same relationship also in the 2018/19 season (the variable "Age" was also part of the significant variables) and 2016/17, while the variable "NPIFLC" (number of presences in first level championships) had shown this relationship in the 2017/18 season. The coefficients shown by the variable "NYIT" (number of years in the team) are 0.1449616 in 2018/19, 0.196275 in 2016/17 and 0.164397 in 2015/16. The variable "NPIFLC" (number of presences in first level championships) instead shows much lower coefficients with 0.010101 in 2017/18 and 0.005679 in 2015/16. Obviously the relationships found in this year are to be considered with caution, given the values assumed by p.

Examining the last season it is possible to reject the null hypothesis that the model that explains most of the variance is the one in which there is only the intercept. In fact the

```
Analysis of Deviance Table

Model 1: TN ~ 1
Model 2: TN ~ Age + NYIT + NPIFLC + NPIICcc
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       437     578.24
2       433     565.63  4   12.608  0.01336 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Table 4.13: Deviance table, 2014/15 season

value of p=0.01336, therefore below the threshold of 0.05 (Table 4.13). So the model including the variables explains a higher variance than the one referred to by the null hypothesis. This is also indicated by the value of the deviance which is positive. To get the definitive confirmation, the result of McFadden's formula returns as a value 0.02180381, so the model number two explains a variance about 2% higher than model 1.

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.098532   0.574700   1.911   0.0559 .
Age         -0.028684   0.024984  -1.148   0.2509
NYIT        -0.024407   0.078699  -0.310   0.7565
NPIFLC       0.006177   0.003118   1.981   0.0476 *
NPIICcc      0.021233   0.038736   0.548   0.5836
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Table 4.14: Multiple logistic regression coefficients, 2014/15 season

Moving on to consider the coefficients of the second model (Table 4.14), it can be noted that the only variable with the value of p lower than 0.05 is the variable "NPIFLC" (number of presences in first level championships) which shows a positive relationship between the increase in its value and the increase in the probability for an individual to belong to the HP cluster. The other variables have a value of p higher than 0.05 and the intercept is slightly higher.

Also in this case it will be considered a model that considers only the variable "NPIFLC" (number of presences in first level championships) to see if is necessary to accept the null hypothesis, i.e. that the model containing all the variables is the one that explains the larger portion of variance or if is possible to accept the alternative hypothesis that

```
Analysis of Deviance Table

Model 1: TN ~ Age + NYIT + NPIFLC + NPIICcc
Model 2: TN ~ NPIFLC
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       433     565.63
2       436     567.97 -3  -2.3426   0.5044
```

Table 4.15: Deviance table B, 2014/15 season

says that the model containing only the variable "NPIFLC" (number of presences in first level championships) explains more variance than the previous one.

The value of p signals that the null hypothesis must be accepted, since it is higher than 0.05 (Table 4.15). So the model including all variables is better to explain the variance. This is also demonstrated by the value of the deviance which is negative and the value of the McFadden formula result which is -0.004141625. As in the other seasons, the value of variance explained in less is practically 0, so the contribution of the excluded variables is really low.

As a last analysis using logistic regression, all the datasets for the five seasons have been unified to create a single macro group of players. This was done to look for a further interpretation given the lack of a recurring variable in at least 4 out of 5 seasons, but also because the models including all variables, when statistically significant, explained a higher portion of variance, compared to the model including only the intercept, which was very low. Also in this study, first of all, the deviance table was created to compare the amount of variance explained by the model that included all the variables on the dataset with all the players with the variance explained by the model including only the intercept.

```
Analysis of Deviance Table

Model 1: TN ~ 1
Model 2: TN ~ Age + NYIT + NPIFLC + NPIICcc
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      2109     2833.5
2      2105     2791.3  4   42.222  1.5e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Table 4.16: Deviance table, all players

Table 4.16 reveal the results for the deviance table and these show an extremely statistically significant p-value, so model 2 should be taken into account. In addition to this, the deviance assumes a positive value and therefore model 2 explains more variance than model 1. To understand in percentage how much more variance is explained by model 2, the McFadden formula is used, which gives a result of 0.01490109, so about 1.5%. So also in this analysis, as in those divided by year, the model including all the

variables is true that it explains more variance than the model including only the intercept, but the percentage of variance explained more is really low.

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.875225   0.266238   3.287 0.001011 **
Age         -0.030958   0.011283  -2.744 0.006074 **
NYIT         0.112033   0.034029   3.292 0.000994 ***
NPIFLC       0.004944   0.001287   3.841 0.000123 ***
NPIICcc      0.009326   0.015258   0.611 0.541037
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Table 4.17: Multiple logistic regression coefficients, all players

Moving on to consider the coefficients related to this model, in Table 4.17, it can be seen that the intercept, the variable "Age", "NYIT" (number of years in the team) and "NPIFLC" (number of presences in first level championships) have a value of p below the value of 0.05 and are therefore statistically significant. The only variable that is not statistically significant is "NPIICcc" (number of presences in international competitions (club competitions)).

The variable 'Age' is the only statistically significant variable with a negative coefficient of -0.030958. Therefore, in this model, as the value of the "Age" variable increases, the probability of the player belonging to an HP team decreases. The variables "NYIT" (number of years in the team) and "NPIFLC" (number of years in first level championships) instead have a positive relationship but with extremely different coefficients. In the first case the coefficient assumes a value of 0.112033 and in the second 0.004944, so the "NYIT" (number of years in the team) variable has a much stronger relationship than the other variable. However, in all three cases, the coefficients do not have extremely high values.

```
Analysis of Deviance Table

Model 1: TN ~ Age + NYIT + NPIFLC + NPIICcc
Model 2: TN ~ Age + NYIT + NPIFLC
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      2105     2791.3
2      2106     2791.7 -1 -0.38304    0.536
```

Table 4.18: Deviance table B, all players

If in this case too, a model containing only the statistically significant variables is considered and the deviance table is calculated, the results are those reported in Table 4.18. Unlike the previous case, on this occasion the value of p is greater than 0.05 and therefore not statistically significant. Moreover also the value of the deviance is negative even if with a value equal to -0.38304. To have further confirmation that the model with all the variables is better than the one containing all the significant variables of the model itself, we use McFadden's formula which gives as value -0.0001372282. So it is true that the previous model explains a higher percentage of variance but the model that excludes the variable "NPIICcc" (number of presences in international competitions (club competitions)) explains a lower portion extremely close to 0.

To recapitulate the results of the logistic regression applied to the players dataset, in the seasons 2018/19, 2017/18, 2015/16 and 2014/15, the analysis of the deviance table gave values of p such that one could reject the null hypothesis, i.e. that most of the variance is explained by a logistic regression model that includes only the intercept, and accept the alternative hypothesis that the model that explains most of the variance is the one that includes all the variables. In the above mentioned seasons the values related to the application of McFadden's formula give as results 0.01730081, 0.04339081, 0.02062898 and 0.02180381, so the models that take into account all the variables explain a percentage of variance slightly higher than the model that does not consider them. For what concerns the 2016/17 season instead, the value of p returned by the deviance table is not statistically significant, so the null hypothesis must be accepted. Despite this, McFadden's formula gives as a result 0.0156239, the lowest value of all the seasons considered. The analysis of this year returned as the most contributory variables to the relationship between them and the probability of being in the HP cluster the variable "NYIT" (number of years in the team). Having to reject the alternative hypothesis for this season, for the reasons previously expressed, in the other years the variables that were significant in the influence of the relationship between them and the probability of being in the HP cluster are "Age" with a -0.0577952 coefficient and "NYIT" (number of years in the team) with a coefficient 0.1449616 in the 2018/19 season, "NPIFLC" (number of presences in first level championships) with coefficient 0.010101 in 2017/18, "NPIFLC" (number of presences in first level championships) with coefficient 0.005679 and "NYIT"

(number of years in the team) with coefficient 0.164397 in 2015/16 but these two variables have a p value higher than 0.05 and therefore not statistically significant and finally in the 2014/15 season the variable "NPIFLC" (number of presences in first level championships) with coefficient 0.006177. From the results obtained it can be seen that there is not a variable that has a significant contribution in at least 4 out of 5 seasons and therefore the experience of the players is not a discriminating factor in determining whether a team will achieve a high performance or not.

If instead the unified dataset is considered, the variables that influence the probability for a player to belong to the HP cluster are "Age", "NYIT" (number of years in the team) and "NPIFLC" (number of presences in first level championships), with coefficients of -0.030958, 0.112033 and 0.004944 respectively. The limit in the use of the information obtained from this precise analysis results in the fact that different actors have been put together because of the difference in each year considered. All the results obtained from the logistic regression on the players dataset are collected in table 4.19.

	2018/19	2017/18	2016/17	2015/16	2014/15	All players
Variable	"Age"	"NPIFLC"	n.s.s.	n.s.s.	"NPIFLC"	"Age"
Coefficient	-0.0577952	0.010101	n.s.s.	n.s.s.	0.006177	-0.030958
Variable	"NYIT"	/	n.s.s.	n.s.s.	/	"NYIT"
Coefficient	0.1449616	/	n.s.s.	n.s.s.	/	0.112033
Variable	/	/	n.s.s.	n.s.s.	/	"NPIFLC"
Coefficient	/	/	n.s.s.	n.s.s.	/	0.004944
McFadden formula	0.01730081	0.04339081	n.s.s.	0.02062898	0.02180381	0.01490109

Table 4.19: Resume of multiple logistic regression

4.4. COACH AND TEAM DATASET ANALYSIS: A PROPOSAL

As previously mentioned, the study contained in this thesis focuses on the players' dataset and the relationship between their experience and the success of the team of they are part. Having also the data related to the experience of the organization itself

and the coaches of the respective team, this chapter proposes a very simplistic type of analysis to see if the greater experience of the subjects in these two datasets corresponds to a better or worse performance. This will be done through a simple model of scores that for each observation, in a given year, will multiply by a certain coefficient the amount of experience recorded and, adding together the results, will give a score for each observation. This model will only apply to the teams and coaches dataset as the two datasets for each team have only one observation. This thing in the players' dataset could have given wrong interpretations as the number of players for each team is different and this could significantly affect the score.

When choosing the coefficients to use for multiplication, arbitrary values will be given according to the importance of the competition. One of the advantages of the sport is that it is easy to understand if one competition is more important than another for the organization of tournaments. So the most important type of tournament would be the international one, followed by first, second and third level championships. In the datasets in question there are also data referring to national cups. This data will be treated with a lower coefficient than the other one because of the lower number of matches and the transversality of this one, not referable to a precise level.

Therefore, starting from international competitions, they will be given a coefficient equal to 4, first level competitions a coefficient equal to 3, second level competitions a coefficient equal to 2, third level competitions a coefficient equal to 1 and finally a coefficient of 0.5 for national cups.

Once this is obtained a fake ranking will be created based on the scores obtained and will be compared with the real ranking of that year. The result of this will be an index that will be given by the sum of all the differences between the real and fake ranking positions. This index will have value 0 if all teams in the fake ranking are placed in the same position as the real ranking and will have value 84 if each team is in the position in the fake ranking farthest away from the real ranking. The higher this value will be, the lower the ability of the experience to match an HP team.

REAL RANKING 18/19	COACH'S RANKING 18/19	TEAM'S RANKING 18/19
Brescia	Venezia= 1848 pts.	Palermo= 221,5 pts.
Lecce	Brescia= 1711 pts.	Hellas Verona= 220 pts.
Palermo	Salernitana= 1480 pts.	Brescia= 206 pts.
Benevento	Livorno= 1351 pts.	Padova= 158 pts.
Pescara	Hellas Verona= 1271 pts.	Livorno= 156,5 pts.
Hellas Verona	Padova= 1237 pts.	Lecce= 145,5 pts.
Spezia	Lecce= 1143 pts.	Venezia= 142 pts.
Cittadella	Spezia= 1096 pts.	Salernitana= 129,5 pts.
Livorno	Palermo= 989 pts.	Pescara= 127 pts.
Venezia	Benevento= 950 pts.	Spezia= 97 pts.
Salernitana	Carpi= 919 pts.	Benevento= 46 pts.
Padova	Pescara= 834 pts.	Carpi= 46 pts.
Carpi	Cittadella= 339 pts.	Cittadella= 34 pts.

Table 4.20: Real ranking vs. Coach's and Team's ranking, 2018/19 season

As it can be seen, the points related to the coaches ranking have higher scores than the team (Table 4.20). This is due to the nature of the variables within the dataset but despite this the two are still comparable taking into account the index that sums up all the differences between the real and the fictitious ranking position that can take a value from 0 to 84. This value is equal to 62 for the coaches ranking, a value that exceeds half of the maximum value that the index can assume. Considering the first three teams in this ranking, Venezia gets the first place because the coach of the team has higher values than his colleagues in terms of both the number of benches and player appearances, in first level championships and international competitions. Brescia, on the other hand, obtains such a high score thanks to the number of benches in first level championships and presence as a player in the same for its coach. Finally, the coach of Salernitana has high values both in number of benches and presence as a player in first and second level championships. In the first three of the ranking of the coaches there are 2 teams that actualy have played the playout or that are relegated. This combined with the fact that the difference index assumes 62, gives an indication that the coach's experience is not a good way to explain the performance of a team. The same index assumes 50 for the team ranking, so a lower score than the one mentioned above. In the first three positions of the ranking there are three HP teams even if in the fourth position there is the Padua

team that arrived penultimate. Analyzing in detail the variables, that most influence the score of the first three teams, are their participations in many first and second level championships.

REAL RANKING 17/18	COACH'S RANKING 17/18	TEAM'S RANKING 17/18
Empoli	Venezia= 2164 pts.	Parma= 229,5 pts.
Parma	Novara= 1693 pts.	Palermo= 221,5 pts.
Frosinone	Bari= 1208 pts.	Bari= 199 pts.
Palermo	Perugia= 1134 pts.	Perugia= 154 pts.
Venezia	Parma= 1008 pts.	Venezia= 138 pts.
Bari	Ascoli= 724 pts.	Novara= 135 pts.
Cittadella	Pro Vercelli= 697 pts.	Empoli= 126,5 pts.
Perugia	Virtus Entella= 667 pts.	Ascoli= 126,5 pts.
Ascoli	Ternana= 522 pts.	Pro Vercelli= 107 pts.
Virtus Entella	Frosinone= 359 pts.	Ternana= 87 pts.
Novara	Palermo= 276 pts.	Frosinone= 33 pts.
Pro Vercelli	Cittadella= 249 pts.	Virtus Entella= 32 pts.
Ternana	Empoli= 104 pts.	Cittadella= 30 pts.

Table 4.21: Real ranking vs. Coach's and Team's ranking, 2017/18 season

In the 2017/18 season as far as the coaches' ranking is concerned, there are 2 out of 3 teams belonging to HP while the teams' ranking sees 3 out of 3 teams in the HP cluster (Table 4.21). The difference index for the ranking derived from the coaches dataset assumes value 68 while in the previous season it had assumed value 62. So in the 2017/18 season, the coaches' experience is not a good way to verify the performance even more markedly than in 2018/19. Considering the elements that most influence the first three classified, it results that the coach of Venezia heavily influences the score with very high values of the presence as a player in first level championships and international competitions, but the values related to the number of benches is low especially when compared to his colleague from Novara who instead presents high values both in terms of benches and presence as a player in first and second level championships. The same thing that was seen for the coach of Bari also applies to Bari. Moving on to consider the ranking of teams, the value of the difference index is 44, while in 2018/19 it was 49. So in the 2017/18 season the experience of the teams explains better the position in the ranking of the teams compared to the previous season and the coaches' rankings.

The value assumed by the first in the standings, i.e. Parma is characterized by high values in participation in international competitions compared to the other teams, as well as the international trophies won. Also the participations on first and second level championships assume high values. The points of Palermo and Bari, on the other hand, are very much influenced by participations on first and second level championships (Table 4.21).

REAL RANKING 16/17	COACH'S RANKING 16/17	TEAM'S RANKING 16/17
SPAL	Pisa= 2150 pts.	Hellas Verona= 215 pts.
Hellas Verona	Spezia= 1475 pts.	Vicenza= 193 pts.
Frosinone	Trapani= 1424 pts.	Perugia= 152 pts.
Benevento	Benevento= 1344 pts.	Pisa= 126,5 pts.
Perugia	Hellas Verona= 1314 pts.	SPAL= 102,5 pts.
Cittadella	Vicenza= 1288 pts.	Spezia= 93 pts.
Carpi	Frosinone= 982 pts.	Carpi= 42 pts.
Spezia	Carpi= 878 pts.	Benevento= 41 pts.
Trapani	Perugia= 837 pts.	Trapani= 39 pts.
Vicenza	Cittadella= 163 pts.	Frosinone= 31 pts.
Latina	Latina= 84 pts.	Cittadella= 28 pts.
Pisa	SPAL= 81 pts.	Latina= 18,5 pts.

Table 4.22: Real ranking vs. Coach's and Team's ranking, 2016/17 season

The 2016/17 season has one team less than the other seasons so the maximum value that the difference index can assume is 72 (Table 4.22). Starting from the coaches' ranking, this index assumes a value of 54, with 2 teams in the first 3 belonging to the real LP group. Also in this case the value exceeds, in a huge way, half of the maximum value that can be reached and therefore also in this season the coaches' experience may not be considered as something that explains the team's performance. In particular in this ranking SPAL is in the last position while, in the real ranking it is first. The opposite happens for Pisa. The factors that most influence the score of the coach of Pisa are the appearances as a player in both international competitions and first level championships. Another important value is given by the score of the number of benches in first level championships. Moving on to his colleague from La Spezia, in this case the highest score is the number of benches in first level championships. The second score by value is the

number of appearances as a player in first level leagues and the third score is the number of benches in second level leagues. Finally, the Trapani is distinguished by very high values of the coach for score related to the number of benches and attendance as a player in first and second level championships. Moving on to consider the value of the difference index for the ranking of the teams' experience, this assumes value 42, therefore a value slightly above half of the maximum value for this season. The score reported by Hellas Verona is influenced in particular by the number of Serie A championships won and the number of participations in first and second level championships. The second team is Vicenza, third last in the real championship, with a score influenced by the number of participations in first level championships, with the highest value recorded this season. Another important element is the number of participations in second level championships. Finally, the elements that most influence the score of the third in the ranking, i.e. Perugia, are the participations in international competitions, first, second and third level championships.

In the 2015/16 season the difference indices give a value of 62 for the first ranking and 60 for the second (Table 4.23). In the coaches' ranking the first three classified are Livorno, Modena and Cagliari, so 2 out of 3 teams are part of the LP. The scores that most influence the first two classified are the number of presences as a player of the coach in first level championships and international competitions, very important for the Livorno team, are also the scores related to the benches in first and second level championships. Cagliari, on the other hand, presents as the most important value the number of presences as a player of the coach in second level competitions, followed by the score related to the number of presences as a player in first level championships.

Also in this season the coach's experience is not an explanation of the team's success. Turning to the consideration of the ranking of the teams' experience, the difference index is the highest found so far, making the experience of the organization less suitable to explain the success compared to its counterparts in previous seasons. The first three are Cagliari, Bari and Modena, then 2 out of 3 belonging to the HP cluster. Cagliari's score is more influenced by the number of participations in international competitions

and first and second level championships. The score of Bari is particularly influenced by the number of participations in first and second level championships as well as Modena.

REAL RANKING 15/16	COACH'S RANKING 15/16	TEAM'S RANKING 15/16
Cagliari	Livorno= 1817 pts.	Cagliari= 202,5 pts.
Crotone	Modena= 1585 pts.	Bari= 195 pts.
Trapani	Cagliari= 1395 pts.	Modena= 172,5 pts.
Pescara	Spezia= 1381 pts.	Livorno= 151,5 pts.
Bari	Pescara= 1360 pts.	Como= 141,5 pts.
Cesena	Novara= 1254 pts.	Novara= 133 pts.
Novara	Salernitana= 1189 pts.	Cesena= 126,5 pts.
Spezia	Trapani= 1090 pts.	Salernitana= 125,5 pts.
Pro Vercelli	Bari= 1010 pts.	Pescara= 120 pts.
Salernitana	Como= 839 pts.	Spezia= 91 pts.
Livorno	Pro Vercelli= 752 pts.	Pro Vercelli= 61 pts.
Modena	Crotone= 660 pts.	Crotone= 54 pts.
Como	Cesena= 353 pts.	Trapani=37

Table 4.23: Real ranking vs. Coach's and Team's ranking, 2015/16 season

In this season, if we look at the fourth and fifth position of this ranking, there are two LP teams, so in the top 5 there are 3 LP and HP teams. Although also in this season the organizational experience explains the performance better than that of the coaches but in this case it doesn't explain the relationship as well as in the other seasons.

The last season examined sees in the first three positions in the standings for the coaches, Modena, Avellino and Bologna, then 2 out of 3 teams belonging to HP (Table 4.24). The difference index is the one with the lowest value found so far for the coaches' experience with a value equal to 46, therefore a good indicator, very near to the half of the maximum value that the index can assume. The score of Modena is more influenced by the values of the experience as a coach and as a player in first level championships, while the score of Avellino is more influenced by the number of presences as a player in first and second level championships, finally Bologna is more influenced by the presence of the coach as a player in first and second level championships and the number of benches in first level championships.

The ranking related to the experience of the teams instead sees an index of difference with value 46, so a good indicator being very close to the half of the maximum value that the index could assume. The first three classified are Bologna, Vicenza and Modena, then 2 teams out of 3 belonging to the HP cluster. The factors that most influence the score of these teams are for Bologna the number of participations in Serie A and international competitions, followed by the number of participations in the Serie B championship. Vicenza, on the other hand, is most influenced by the number of participations in first and second level championships. The same is true for Modena, with particular importance given to the number of participations in second level championships.

REAL RANKING 14/15	COACH'S RANKING 14/15	TEAM'S RANKING 14/15
Carpi	Modena= 1792 pts.	Bologna= 281 pts.
Frosinone	Avellino= 1305 pts.	Vicenza= 190 pts.
Vicenza	Bologna= 1288 pts.	Modena= 170,5 pts.
Bologna	Spezia= 1230 pts.	Perugia= 149 pts.
Spezia	Pescara= 1172 pts.	Pescara= 141,5 pts.
Perugia	Frosinone= 794 pts.	Varese= 105,5 pts.
Pescara	Vicenza= 775 pts.	Avellino= 93 pts.
Avellino	Cittadella= 772 pts.	Spezia= 89 pts.
Crotone	Perugia= 743 pts.	Crotone= 52 pts.
Modena	Carpi= 695 pts.	Carpi= 35 pts.
Virtus Entella	Crotone= 269 pts.	Virtus Entella= 26 pts.
Cittadella	Virtus Entella= 161 pts.	Frosinone= 24 pts.
Varese	Varese= 63 pts.	Cittadella= 24 pts.

Table 4.24: Real ranking vs. Coach's and Team's ranking, 2014/15 season

Summarizing the results related to the ranking obtained through the coaches' experience scores, it can be seen that the difference index is in 4 out of 5 seasons, much higher than half of the maximum value that can be reached by the same (Table 4.25). This trend can be understood as a poor correspondence between the greater experience of a coach and the success of the team. Moreover, in 4 out of 5 seasons, this index has higher values than the index related to teams' experience.

Obviously the model presented is very simplistic and does not have the validity and strength of the models applied for the players dataset, but from this it is possible to have

a certain direction. Going into detail it can be seen that virtually in each of the first three teams in each year, the variable that contributes most to the final score is "NPAPIFLC" or number of presences as a player in first level championships.

COACH'S RANKING	Difference Index	Max value Diff. Index	Min value Diff. Index	First team score contributors	Second team score contributors	Third team score contributors
2018/19	62	84	0	"NPAPIFLC", "NBIFLC", "NPAPIIC"	"NPAPIFLC", "NPAPISLC", "NBIFLC"	"NPAPIFLC", "NBISLC", "NBIFLC"
2017/18	68	84	0	"NPAPIFLC", "NPAPIIC", "NPAPISLC"	"NPAPIFLC", "NBIFLC", "NPAPISLC"	"NPAPIFLC", "NPAPIIC", "NPAPISLC"
2016/17	54	72	0	"NPAPIFLC", "NPAPIIC", "NBIFLC"	"NBIFLC", "NPAPIFLC", "NBISLC"	"NPAPIFLC", "NBISLC", "NBIFLC"
2015/16	62	84	0	"NPAPIFLC", "NPAPIIC", "NBISLC"	"NPAPIFLC", "NPAPIIC", "NBIFLC"	"NPAPISLC", "NPAPIFLC", "NBISLC"
2014/15	46	84	0	"NPAPIFLC", "NBIFLC", "NBISLC"	"NPAPISLC", "NPAPIFLC", "NBISLC"	"NPAPIFLC", "NPAPISLC", "NBIFLC"

Table 4.25: Results of the fake coach's ranking

Another very present variable is "NPAPISLC" which indicates the number of presences as a player in second level championships. Combining these two variables with the values related to the difference indexes in the interpretation, it can be argued that the coach's player experience is not decisive for the success or lack of success of the team. It is repeated that the model is very simplistic and therefore its interpretation should be taken in a very superficial way. Finally, another very present variable is "NBIFLC" or number of benches in first level championships. The presence of this, combined with the interpretation of the data already made, gives further indication of how the coach's experience is not so influential in achieving success for a team.

As far as the ranking created by the team's experience is concerned, the difference index assumes, in 4 seasons, a lower value than the one assumed by coaches' experience(Table 4.26). These results say that in principle, the experience of the organization explains better than the coach's experience the success of a team. As for the coaches, the model is extremely simplistic but still provides some indications. In most cases, the variable that

TEAM'S RANKING	Difference Index	Max value Diff. Index	Min value Diff. Index	First team score contributors	Second team score contributors	Third team score contributors
2018/19	50	84	0	"NPISA", "NPISB", "NPIIC"	"NPISA", "NPISB", "NPIIC"	"NPISA", "NPISB", "NPIIC"
2017/18	44	84	0	"NPISA", "NPISB", "NPIIC"	"NPISA", "NPISB", "NPIIC"	"NPISA", "NPISB", "NPISC"
2016/17	42	72	0	"NPISB", "NPISA", "NPIIC"	"NPISA", "NPISB", "NPISC"	"NPISB", "NPISA", "NPISC"
2015/16	60	84	0	"NPISA", "NPISB", "NPIIC"	"NPISA", "NPISB", "NPISC"	"NPISA", "NPISB", "NPISC"
2014/15	46	84	0	"NPISA", "NPIIC", "NPISB"	"NPISA", "NPISB", "NPISC"	"NPISB", "NPISA", "NPISC"

Table 4.26: Results of the fake team's ranking

best explains the results obtained by the top three over the course of the seasons is the "NPISA" variable, i.e. the number of appearances in Serie A championships. The second most important variable is "NPISB" which counts the number of participations in the Serie B league. This indicates that the most influential factors relating to organisational experience in determining a team's success are the number of Serie A and Serie B league appearances. Other variables that are influential but less influential are participation in international competitions (NPIIC) and Serie C (NPISC).

4.5. MANAGERIAL IMPLICATIONS

As explained in depth in the first chapter, football has increasingly changed from a simply sporting phenomenon to an economic phenomenon over the years and the great availability of data has allowed its literature to study the economic-financial dimension of these subjects. The typical activity of a football club is that of entertainment and the main sources of revenue for a football team are matchday ticket revenues, revenues from the granting of television rights, merchandising sales, revenues from sponsorships and the purchase and sale of players. All these components are of course influenced by

the sports results that the individual team achieves. In addition, each team receives a prize depending on the placement it achieves in each competition. So the financial performance of a football team is closely linked to the sports result in terms of revenues. A distinction must be made between large clubs and small clubs. In fact, large clubs can count on a greater number of fans not only located in the city where the club belongs, which is not found in small clubs, i.e. all those belonging to the context studied in this thesis, i.e. the Serie B teams. So there is a strong split in management and strategy between large and small clubs. In fact, the clubs that can be considered as large have tried to focus on increasing revenues from the sources mentioned above to put in place a virtuous circle between sporting and economic results that aims to win trophies of international and national level that leads to an increase in the audience base in order to earn more and more. For small clubs like those in Serie B, however, the objectives are different. In fact, their sporting goals are to be promoted to Serie A or stay in Serie B. Although the objectives are different, the principle of small and large clubs is the same, i.e. to use their financial resources to select players through which to obtain certain sporting results that can bring more revenue in terms of ticket sales, revenue from television rights, merchandising, sponsorship and capital gains from the sale of players (Figure 4.2). These higher revenues can lead to possible profitability, if the costs are managed well, and additional financial resources to start the cycle again. Obviously the form of revenue that can vary more over the years for small clubs is the one related to

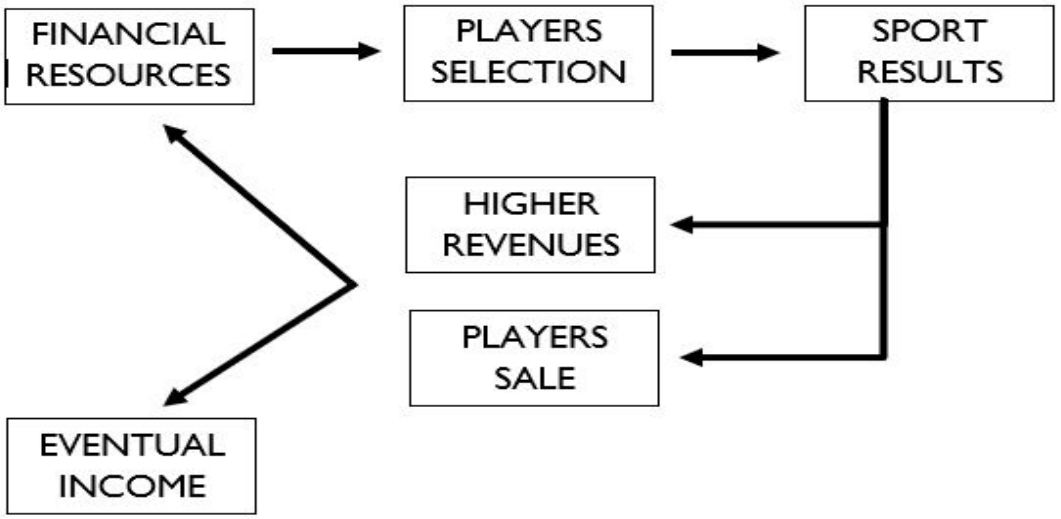

Figure 4.2: Virtuous economic cycle of professional football teams. Source: personal rielaboration from Lago et al., 2004

the purchase and sale of players because the other components, always talking about teams with a little more than local attention, do not fluctuate so much but still affect an important part of the financial resources of the team.

The present study aims, with the focus on players, to give guidance to the team selection managers on how to build a successful team, i.e. an HP team. As can be seen in figure 4.2 the selection of the right players would give as a result the achievement of an important sporting result that would then make the cycle continue. So the accumulated experience of the players has been taken into account to see whether or not its explains the success or failure of a team. In particular, it has been tried to understand what elements of the players' experience differentiate the HP cluster from the LP cluster so that a manager can better choose the team members to achieve direct promotion to Serie A or the playoffs.

However, the results obtained through multiple logistic regression models have not given homologous results in at least 4 out of 5 seasons. Proceeding neatly in the 2018/19 season, a negative relationship was found between increasing age and a player's probability of belonging to an HP team. The value of the coefficient, however, has been found to be extremely close to 0 but despite that this report confirms that of the study by Poli et al. (2018) according to which there is a statistically negative relationship between age and position in the UEFA ranking and where the leagues and teams that are more competitive are those that are neither extremely old nor extremely young, finding an ideal average age of the team around 26.5 years. The second relationship found for this year has been the positive one between increasing the value of the variable "NYIT" (number of years in the team) and the probability of belonging to the HP cluster. In this case the coefficient is significantly higher than that of the variable "Age" with a value of about 0.1449 against -0.0578. This report is in line with another study by Poli et al. (2018) where it was found that the best performing clubs were the most stable ones in terms of changed players. This report is also in line, in part with the study by Huckman et al. (2009) which saw in the team familiarity a U-shaped inverted relationship that probably does not arise in this context, given the very high turnover that players have. On the other hand, the study by Ruigrok et al (2011) found a weak relationship between team longevity and high team performance. Apart from these implications for that year

McFadden's formula explains how the percentage of extra variance explained by the model with all the variables derived from the PCA compared to the one containing only the intercept, explains about 1.7% of the extra variance. This result, combined with the fact that both significant variables have very low coefficients, shows how, with regard to this year, the players' experience and the factors that make it up, do not explain, or rather explain in a very weak way, the fact that a player is part of an HP team.

Proceeding with the 2017/18 season, a positive correlation was found between the increase in the value of the variable "NPIFLC" (number of presences in first level championships) and the probability that a player belonged to an HP team. The coefficient value in this case is even lower than those found for the other variables in the previous season, with a value of 0.010101. Moreover, the only statistically significant variable of the model that explains about 4% more than the variance of the model including only the intercept, is different from those found in the previous season. The fact that the cumulated experience in first level championships can positively influence the probability of belonging to an HP team is logical, since it is a higher championship experience. Despite this, for the same reasons expressed for the 2018/19 season, it has been seen that there is no strong explanation from one of the elements that constitutes the experience of a player that differentiates an HP team from an LP.

This is made even more evident in the 2016/17 season, where even the model including all variables is not statistically significant. The same thing happens, however, as far as the coefficients of the variables are concerned, in the 2015/16 season where McFadden's formula showed that the model including all the variables explained about 2.1% more variance than the model with only the intercept.

In the last season instead, that of 2014/15, the only variable that is statistically significant is "NPIFLC" (number of presences in first level championships) with a coefficient even lower than the one found in the 2017/18 season, that is about 0.006. In this case the model with all the variables explained about 2.1% more variance than the model including only the intercept.

Finally, a model was also created that did not take into account the differences of the seasons and considered all players of all years in a unified way. This model returned

three significant variables. These are "Age", "NYIT" (number of years in the team) and "NPIFLC" (number of presences in first level championships). In the first case a negative correlation has been found with a -0.031 coefficient and in the other two a positive correlation with a 0.112 and 0.005 coefficient respectively. Even with the unified model it can be seen that the value of the coefficients is extremely small and indicates a very weak relationship. Finally, McFadden's formula returned a value of the explained variance about 1.5% higher than the model containing only the intercept.

In the light of the results expressed by McFadden's formula, of the coefficients assumed by the statistically significant variables and of the lack of a variable that is repeated in at least 4 out of 5 seasons, it can be said that experience and in particular none of its elements differentiate the HP cluster from the LP cluster. In essence the elements collected that constituted the experience of a player, do not explain the success or not of a team. From this statement, the managerial implication that is obtained concerns the fact that in choosing a player to make his team successful, manager should not consider the experience as a discriminatory element but should take into account other elements different from those analyzed. This result is opposite to the one found by Ruigrok et al. (2011) who argue that a team manager should consider the experience of the player to make a selection and that overlapping experiential background is another very important factor in the selection. This study has already been analyzed in previous chapters and the context analyzed was that of the 2006 FIFA World Cup, so an international very short competition. The characteristics of the B Series context are very different and it is precisely the context analyzed that could have brought this opposite result. First of all, the context analyzed foresees an "exercise" much longer than that of a world championship and moreover it is not talking about an international competition but a second level national championship. Just the level of the championship could justify the fact that a greater experience is not more successful or that there are no distinctive elements between HP and LP because the players with the highest cumulative experience may no longer have any incentive to compete or have reached a level of age that, despite experience, shows descending performances. Still analyzing the context, in paragraph 1.7.1. it emerged as if analyzing the three Italian professional championships, it results that the one who wins the Serie A championship is also the team that is usually the first

in terms of production value (Table 4.28). The same is true for the total amount of salary, even in a more pronounced way (Table 4.27). The one who has the highest fee for players is usually the one who wins the championship. As far as Serie B is concerned, however, the situation is much less clear and this can be seen once again from the 4.27 and 4.28 table. The winner of the Serie B in most cases is not the first in terms of production value and total amount of salary. The Serie C is even more varied. So from this data it can be said that it is not true that whoever spends the most then wins. Obviously this table concerns only the winners of the league and does not consider the other teams arrived in the playoff area but still gives an indication of the trend characterizing the Serie B. In addition to this the Serie B championship is very irregular and with strong turnover every year. In fact, regularly the teams that change are 7, three promoted in Serie A and 4 relegated to Serie C but in addition to these there can be a turnover given by forced relegation of teams with strong irregularities. It should be

	Serie A	Serie B	Serie C
2007-2008	1	3	4
2008-2009	1	3	23
2009-2010	1	2	3
2010-2011	1	9	25
2011-2012	3	5	3
2012-2013	1	4	14
2013-2014	1	1	7
2014-2015	1	16	1
2015-2016	1	1	12
2016-2017	1	19	8
2017-2018	1	2	7
Posizionamento medio	1,2	5,9	9,7

Table 4.27: Comparison of the ranking of the salaries of the winners of Serie A, Serie B and Serie C. Source: Report Calcio Pwc, 2019

	Serie A	Serie B	Serie C
2007-2008	2	3	1
2008-2009	2	16	28
2009-2010	1	9	1
2010-2011	1	5	19
2011-2012	3	10	13
2012-2013	1	10	18
2013-2014	1	1	9
2014-2015	1	11	1
2015-2016	1	1	15
2016-2017	1	19	11
2017-2018	1	1	9
Posizionamento medio	1,4	7,8	11,4

Table 4.28: Comparison of ranking positioning in the value of production of the winners of Serie A, Serie B and Serie C. Source: Report Calcio Pwc, 2019

considered in every season at least one Serie B team receives penalty points just to understand the magnitude of the phenomenon. This high turnover could explain the lack of one or more variables that are repeated in at least 4 seasons in the analyzed time window. Another explanation that may justify the lack of a clear effect between the elements of experience and success may be given by the fact that according to a study by Poli et al. (2018), the Serie B championship is the one with the highest competitive balance in the world and with a large structural and technical gap compared to Serie A which shows the highest number of teams relegated to Serie B after one year of the 5 major European leagues (Football benchmark KPMG, 2019). This could be linked to the results found. In fact in Serie B players' experience is not a discriminatory factor for achieving success but maybe in Serie A this is not true. In such a fluid and competitive context, the analyses carried out showed that the sports manager in charge of player selection does not have to pay attention to the experience and the elements that make it up for the formation of a winning team.

Having learned that there is a lack of distinction of the elements that make up the experience in the two clusters, it can be said that practically the teams belonging to HP have the same amount of experience as those LPs and are characterized by the same elements. So probably at the basis of success or failure there is the inability to integrate, in a functional way, the various experiences of the subjects. For this reason it can be said that probably in the seasons considered, one of the elements that caused the failure of the LPs was a bad integration of the various players' experiences. In fact also in the study by Ruigrok et al. (2011) it turns out that the different experiences must be carefully managed in order to gain a competitive advantage, obtain benefits and mitigate costs. The same study underlines the importance of the coach in facilitating this integration.

In this chapter a system has also been proposed to assess whether the coach's or team's accumulated experience was linked to the success of a team and what elements of this experience were the most important for the achievement of the teams' scores. It must be repeated that the model is very simplistic but it can give general indications and gives managerial implications. The proposed model shows that the score obtained through the cumulative experience of the coaches is not a proxy for the explanation of performance. In fact the difference index has always reached values much higher than

half of the maximum value, this is evidence of a theoretical ranking rather different from the real one. As seen in paragraph 4.4 the most recurrent variables that contribute most to the score are "NPAPIFLC" and "NPAPISLC" (number of presences as a player in first level championships; number of presences as a player in second level championships). These results, combined with the interpretation given by the difference index, ensure that the broad indications given by this model are in line with the Hall & Pedace (2016) and Muehlheusser et al. (2016) studies. The study by Muehlheusser et al. (2016) even found a negative relationship between the manager's previous experience as a player and the performance as a coach, while Hall & Pedace (2016) found an insignificant relationship between the manager's previous experience as a player and the team's performance. Another important result of their study was the fact that they found an influence by the manager on the team's winning percentage but that this influence derives from non-measurable traits. In our case the measurable trait is experience. This conclusion can be considered in line with the results found by Ruigrok et al. according to which in order to gain a competitive advantage in a team, the project manager or the coach must be able to manage the different backgrounds of the players in an excellent way. This ability can certainly be traced back to an unmeasurable trait of the coach's characteristics. This draws attention to the fact that in the present study the HP cluster does not differ significantly from the LP cluster in terms of cumulative experience, so it is further confirmed that probably the teams that performed well did so because their coach was able to integrate the various experiences and the coaches who coached LP teams did not. This is just a hypothesis as to why the experience is not related to success. So it could be said that it is not so much the experience that leads to success but how it is managed and integrated, so in itself there is further confirmation that the experience and the elements that make it up do not affect performance in this particular context.

Also in paragraph 4.4 a ranking has been created following the accumulated score from the experience of the teams. The difference index in this case was much better in explaining the relationship between accumulated experience and performance. In particular, the most influential variables were "NPISA" (number of participations in Serie A) and "NPISB" (number of participations in Serie B). Given the greater adherence between the ranking obtained through the cumulated experience scores and the real

ranking, the participations in Serie A and in the same championship analyzed are the most important to explain the success or not of a team. Considering the fact that it is talking about Serie B teams and that therefore the maximum objective is to be able to participate in Serie A, this result is in line with the study by Del Barrio & Szymanski (2006) already analyzed in chapter 3. In fact, the authors in their discussion about profit maximization and win maximization, explained as a cause of change from the first to the second, the historical state of the club and the expectations of the fans. This could be linked with the sedimentation of a precise team culture, a "winning mindset". So the more times a team has participated in Serie A or has reached high positions in Serie B, the more it will have a corporate strategy that allows it to continue on that path. In addition, a corporate culture of success and victory could be another one of those non-measurable characteristics that can help better integrate the players' experience and increase incentives.

This study may also have implications for the league itself. In fact, both the fact that whoever spends or earns the most is not the one who wins as explained above and that this study shows that experience is not a factor that explains the success or failure of a team, the competitive balance of the league can be kept intact. So the data about the comparison between the team winner and its position in the salaries ranking are scientifically proven. The competitive balance can be traced back to the theory of uncertainty of the result proposed by Rottemberg (1956) according to which greater uncertainty of the result causes interest in the competition and the league to grow. The nature of the uncertainty can be related to one match, one season or the long term as proposed by Sloane (1971). Long-term uncertainty relates to the long-term dominance of a single team over several seasons which is impossible to see in this league due to organizational issues. Uncertainty about one season, on the other hand, concerns a narrower time mirror but is nevertheless important for the attractiveness of the championship. Gerrard (1999) argues instead that the dominance by a team can undermine the collective welfare because a league or a championship that becomes predictable makes the demand fall and so it creates a damage also financial. So the results found in this study are very positive for the League in terms of entertainment.The lack of elements that undermine the uncertainty of the result in the analysed context can

create two distinct implications. The former concerns the fact that, as far as television rights and sponsorship are concerned, a high degree of uncertainty increases the appeal of the league and therefore at league level this element should be developed as a strong bargaining power with match broadcasting platforms and sponsors. In order to further increase the bargaining power, it is also necessary to put in place obligations for clubs concerning the 'sale of the product' with guidelines aimed at improving the infrastructure in order to be more attractive. The latter implication, on the other hand, concerns the need for the Lega Serie B not to implement measures aimed at maintaining balance.

To conclude, the results of this study have shown that players' and coaches' experience is not a determinant of the success or failure of a team. Instead the teams' experience is a better predictor for the success or failure of a team. The managerial implications are both for the managers of the single team and for the managers of the League. For the latter the interpretation of the results it's a good news in terms of subsistence of the show because the success of a team don't seem influence in a determinant way by the team's budget and the players' and coaches' experience.

4.6. LIMITATIONS AND FURTHER DEVELOPMENTS

The limitations of this study are also the basis for further developments in this area. The first limitation concerns the fact that in all datasets the cumulative experience has been considered without caring when it has been achieved and therefore without taking into account a possible loss of knowledge given by it. Therefore a development of the study could be to consider a variable that can insert a weighting given by time. A second limit always concerns the data, in fact the datasets include only the teams that have reached the playoffs or directly promoted to Serie A and the teams participating in the playout or directly relegated to Serie B. A further development of the study could be to consider all the teams of each year and lengthen the time window of reference. A third limitation of the study concerns methodology. In fact, a step forward on the subject could be given by the use of a single analysis technique for all types of datasets but also a methodology that can relate the three elements, i.e. players, coaches and teams.

4.7. CONCLUSIONS

In the course of this chapter a brief presentation of the analysis method that would be used to analyze the players dataset after using the PCA as a preliminary analysis has been made. Then the multiple logistic regression on the dataset has been used and a very simplistic model has been proposed to analyze the relationship between the experience of the coaches and the team itself with the success of the team at the end of the championship. The results obtained through the multiple logistic regression have showed how the cumulative experience of the players and its constituent elements do not clearly distinguish players belonging to HP teams from those belonging to LP teams, demonstrating the insignificant relationship between players' experience and success. So it could be stated that in the selection of players, the manager should not take into account particular experiential elements but rather other measurable and non-measurable elements. It has also been pointed out that the result obtained, in contrast to other studies both in sport and business, may derive from the nature of the Serie B championship context, the lack of incentives from more experienced players to compete in a second level championship and the inability to integrate all the different experiences. In addition to these results, the model proposed for coaches and teams shows that in principle the cumulative experience of a coach, both in terms of player and coach experience, is not significant in terms of success, while the cumulative experience of the team, understood as organisation, is more influential in determining success and failure.

CONCLUSIONS

The results of this study show that the players' experience and the elements that make it up do not have a significant influence on whether or not a team is likely to achieve success. This leads to the argument that when selecting players, the manager should consider other factors that may or may not be measurable. In addition to the analysis of players, the main focus of the study, the analysis conducted on the experience of coaches and teams has showed that the experience of the former does not lead to success while the latter has a greater ability to identify teams that have performed well. It is repeated once again that the analysis on these datasets are based on a very simplistic model proposal and therefore the results should be considered only as an indication in principle.

In general it has been assumed that the nature of the results derive deeply from the nature of the competition context, the lack of ability to integrate the human capital of each player within the group or the lack of incentives from more experienced players to play in a second level league.

The results obtained have positive managerial implication for the League, in fact it was demonstrated that there is not something that gives a precise and definite competitive advantage for a team. For this reason the competitive balance is maintained and the interest in the competition can remain intact.

The following thesis does not claim to be exhaustive in investigating the experience/performance relationship in this context but can be a starting point for further investigation, perhaps using data from all teams participating in the league, lengthening the time window and using a weighting for each player's cumulative experience based on when it occurred. In addition, a single methodology for the analysis of all datasets and an analysis that includes all three datasets simultaneously could also be the subject of interesting developments.

The articulation of this thesis has also allowed to understand the importance of the sport business and how it can grow and develop in the future, attracting more and more attention from scholars and managers. In fact, the study of relationships already widely

analyzed in other areas, such as the experience-performance relationship, in sport could lead to increasingly better managerial implications for the management of professional sports clubs.

This work, despite its limitations and possible developments, represents one of the few managerial studies on the context of Serie B and the only one that investigates the experience-performance relationship in the latter.

APPENDIX A

R codes used for this thesis:

#Construction of boxplot chapter 3.3.1.#

```
boxplot(name_dataset,col='colour')
```

#Projection of individuals on the first two PCs chapter 3.5.1.#

```
library(sjmisc)

library("FactoMineR")

library("factoextra")

library(corrplot)
```

Transform column 1 into factor variable

```
name_dataset[,1] <- to_factor(name_dataset$`TN`)

M <- name_dataset

Res.pca <- PCA(M[,-1], graph = FALSE)

x11()

fviz_pca_biplot(Res.pca, # Individuals  geom.ind = "point", fill.ind = M$TN, col.ind = "black", pointshape = 21, pointsize = 3, palette = c ("blue", "green"), addEllipses = TRUE, ellipse.type= "convex" , center.pch =TRUE, # Variables alpha.var ="contrib", col.var = "contrib", gradient.cols = c ("yellow", "red"), legend.title = list(fill = "Cluster", color = "Contrib", alpha = "Contrib"))
```

```
#PCA on players' dataset chapter 3.5.2. and 3.5.3.#

library("FactoMineR")

library("factoextra")

library(corrplot)

#data import and exploration

head(name_dataset)

n <- dim(name_dataset)[1]

p <- dim(name_dataset)[2]

#PCA

PCA(name_dataset, scale.unit = TRUE, ncp = 5, graph = TRUE)

A <- PCA(name_dataset)

#eigenvalues

eig.val <- get_eigenvalue(A)

eig.val

x11()

fviz_eig(A, addlabels = TRUE, ylim = c(0, 50))

#graph_of_variables

var <- get_pca_var(A)

var

var$contrib
```

```
head(var$contrib)

#Contributions of variables on PCs

head(var$contrib, 8)

x11()

corrplot(var$contrib, is.corr=FALSE)

#Contributions of variables to PCx

x11()

fviz_contrib(A, choice = "var", axes = 1, top = 10)

#Contributions of variable PCx-PCy

x11()

fviz_contrib(A, choice = "var", axes = 1:4, top = 10)
```

#Multiple logistic regression chapter 4.3.#

```
logisticRegressionModel =glm(TN~ Age + NYIT + NPIFLC + NPIICcc, data= name_dataset,family=binomial)

#Sintax for the matrix of data

A <-data.frame (name_dataset)

A

M0 <- (glm(TN ~ 1, data= A, family=binomial))

M1 <- (glm(TN ~ Age + NYIT + NPIFLC + NPIICcc, data= A, family=binomial))

anova (M0,M1, test= "Chisq")

(M0$deviance-M1$deviance)/M0$deviance
```

summary(M1)

Bibliography

Anand B, Khanna T. (2000). *Do firms learn to create value? The case of alliances.* Strategic Management Journal **21**(3): 295–315.

Anand J., Mulotte L. & Ren C. R. (2016). *Does experience imply learning?* Strategic Management Journal, 37(7), 1395-1412.

Argote L. (1999). *Organizational learning: Creating, retaining and transferring knowledge.* Norwell, MA: Kluwer.

Argote L. (2011). *Organizational learning research: Past, present and future.* Management learning, 42(4), 439-446.

Aversa P., Furnari S. & Haefliger S. (2015). *Business model configurations and performance: A qualitative comparative analysis in Formula One racing, 2005–2013.* Industrial and Corporate Change, 24(3), 655-676.

Avolio B. J., Waldman D. A. & McDaniel M. A. (1990). *Age and work performance in nonmanagerial jobs: The effects of experience and occupational type.* Academy of Management Journal, 33(2), 407-422.

Baglioni S. (2011). *Social Capital and sport governance in Europe*. New York-Londra, Routledge Taylor & Francis Group, pp 146-162

Barbieri L., (2005). *Il settore del calcio professionistico: spunti per un'analisi.* Analisi Giuridica dell'Economia, 2, pp. 267-300.

Barros C.P. (2001). *Economic return on schooling for soccer players.* Journal of sport economics, vol. 2 No. 4, pp. 307-327

Benoldi A. e Sottoriva C. (2011). *La disciplina della redazione del bilancio d'esercizio delle società di calcio. confronto con l'esperienza internazionale ed impatto del c.d. «financial fair play»*. Rivista di Diritto ed Economia dello sport. VII (1), pp. 147-191.

Berman S. L., Down J. & Hill C. W. (2002). *Tacit knowledge as a source of competitive advantage in the National Basketball Association.* Academy of management Journal, 45(1), 13-31.

Bohn RE (1995). *Noise and learning in semiconductor manufacturing*. Management Science 41(1): 31–42.

Bonacini A. (2012/13). *Vincere o guadagnare? Organizzazione e strategie dei settori giovanili delle società di calcio di Serie A*. Tesi magistrale in Economia e direzione aziendale, Università degli studi di Padova, pp. 20-22

Bruderl J. & Preisendorfer P. (2000). *Fast growing businesses: empirical evidence from a German study*. International Journal of Sociology, Vol. 30, pp. 45-70.

Bryson, A., Frick, B., & Simmons, R. (2012). *The returns to scarce talent: Footedness and player remuneration in European soccer*. Journal of Sports Economics, 13, 1-23.

Carta M., (2005). *La disciplina normativa delle società di calcio: un quadro di sintesi*. Analisi giuridica dell'economia, vol. 2, pp: 301-311.

Carmichael, F., Forrest, D. & Simmons, R. (1999). *The labour market in association football: Who gets transferred and for how much?* Bulletin of Economic Research, 51(2), 125-150.

Ciampaglia G.M. (2013). *La creazione del valore nelle società di calcio*. Accademia Italiana di Economia Aziendale AIDEA

Clement RC, McCormick RE. (1989). *Coaching team production*. Economic Inquiry 27(2): 287–304.

Cohen W. & Levinthal D. (1990). *Absorptive capacity: a new perspective on learning and innovation*. Administrative Science Quarterly, Vol. 35, pp. 128-152.

Colombo M. & Grilli L. (2005). *Founder's human capital and the growth of new technology-based firms: a competence-based view.* Research Policy, Vol. 34, pp. 795-816.

Cooper A. C. & Gimeno-Gascon F. J. (1992). *Entrepreneurs, Processes of Founding, and New-Firm Performance.* In D. L. Sexton & J. D. Kasarda (Eds.), The State of the Art of Entrepreneurship (pp. 301-340). Boston: PWS-Kent Publishing Company.

Dahlin K. B., Chuang Y.-T. & Roulet T. J. (2018). *Opportunity, Motivation, and Ability to Learn from Failures and Errors: Review, Synthesis, and Ways to Move Forward*. Academy of Management Annals, 12(1): 252-277.

Ehrenberg R. G. & Smith R. S. (2000). *Modern labor economics: Theory and public policy.* Reading, MA: Addison Wesley.

Ericsson K. A. (2006). *The influence of experience and deliberate practice on the development of superior expert performance.* The Cambridge handbook of expertise and expert performance, *38*, 685-705.

Ericsson K. A., Krampe R. T., & Tesch-Römer C. (1993). *The role of deliberate practice in the acquisition of expert performance.* Psychological review, 100(3), 363.

Ericsson K. A. & Lehmann A. C. (1996). *Expert and exceptional performance: Evidence on maximal adaptations on task constraints.* Annual Review of Psychology, 47, 273–305.

Ferrara M., (2003). *L'organizzazione dello sport.* Torino, Giappichelli

Fiol CM & Lyles MA (1985). *Organizational learning*. Academy of Management Review 10(4): 803–813.

Fitts P. & Posner M. I. (1967). *Human performance*. Belmont, CA: Brooks/Cole.

Garcia-Del-Barrio, P., & Pujol, F. (2007). *Hidden monopsony rents in winner-take-all markets—Sport and economic contribution of Spanish soccer players*. Managerial and Decision Economics, 28, 57-70.

Gerrard, B. (1999). *Team sports as a free-market commodity***.** New political economy, 4(2), 273-278.

Gerrard B. (2001). *A new approach to measuring player and team quality in professional team sports.* European Sport Management Quarterly, 1:3, 219-234.

Gimeno J., Folta T., Cooper A. & Woo C. (1997). *Survival of the fittest? Entrepreneurial human capital and the persistence of underperforming firms.* Administrative Science Quarterly, Vol. 42, pp. 750-783.

Gius M. & Johnson D. (2000). *Race and compensation in professional football.* Applied Economics Letters, 7(2), 73-75.

Hadley L., Poitras M., Ruggiero J. & Knowles S. (2000). *Performance evaluation of national football league teams.* Managerial and decision economics, 21, pp. 63-70

Haleblian J. & Finkelstein S. (1999). *The influence of organizational acquisition experience on acquisition performance: A behavioral learning perspective.* Administrative Science Quarterly, 44(1), 29-56.

Hall C. M. & Pedace R. (2016). *Do managers matter? Manager effects on organization performance.* Managerial and Decision Economics, 37(8), 541-551.

Hamilton B. (1997). *Racial discrimination and professional basketball salaries.* Applied Economics, 29(3), 287-296.

Haunschild PR. & Sullivan BN. (2002). *Learning from complexity: Effects of prior accidents and incidents on airlines' learning*. Administrative Science Quarterly 47(4): 609–643.

Hayward M. (2002). *When do firms learn from their acquisition experience? Evidence from 1990 to 1995.* Strategic Management Journal **23**(1): 21–39.

Holzer H. J. (1990). *The determinants of employee productivity and earnings*. Industrial Relations: A Journal of Economy and Society, 29(3), 403-422.

Horowitz I. 1994a. Pythagoras, Tommy Lasorda, and Me*: on evaluating baseball managers*. Social Science Quarterly 75: 187–194.

Huckman R. S., Staats B. R. & Upton D. M. (2009). *Team familiarity, role experience, and performance: Evidence from Indian software services.* Management science, 55(1), 85-100.

Huebl L., & Swieter D. (2002). *Der Spielermarkt in der Fußball-Bundesliga*. Zeitschrift fur Betriebswirtschaft, Erg ¨ anzungsheft 4, Sport ¨ okonomie, 72, 105-125.

Huizinga J. (1938). *Homo Ludens*, Amsterdam.

Idson T. & Kahane L. (2000). *Team effects on compensation: An application to salary determination in the National Hockey League.* Economic Inquiry, 38(2), 345-357.

Jeffers J.N.R (1967). *Two case studies in the application of principal component analysis.* Journal of the Royal Statistical Society, Series C (Applied Statistics), vol. 16, n°3, pp 225-236

Jolliffe I.T. (2002). *Principal component analysis*. 2 ed., Springer, pp. 1-9

Jones J.C.H., Nadeau S. & Walsh W. (1999). *Ethnicity, productivity and salary: Player compensation and discrimination in the National Hockey League*. Applied Economics, 31(5), 593-608.

Kahn L. (1993). *Free agency, long-term contracts and compensation in Major League Baseball: Estimates from panel data*. Review of Economics and Statistics, 75(1), 157-164.

Kahn L.M. (1993). *Managerial Quality, Team Success, and Individual Player Performance in Major League Baseball*. ILR Review, Vol. 46, No. 3 (Apr., 1993), pp. 531-547.

Kaiser H. (1961). *A Note on Guttman's Lower Bound for the Number of Common Factors*. British Journal of Statistical Psychology 14, pp 1–2.

Kessene S. (2006). *The Bosman case and European football*. In Andreff W., Szymanski S. (eds), Handbook of Economics of sport, Edward Elgar, Cheltenham, UK • Northampton, MA, USA, pp. 636-642

Kotur B. R. & Anbazhagan S. (2014). *Education and work-experience-influence on the performance.* Journal of Business and Management, 16(5), 104-110.

Krampe, R. T., & Charness, N. (2006). *Aging and expertise.*

Kuethe T.H. & Motamed M. (2010). *Returns to Stardom: evidence from U.S. Major League Soccer*. Journal of Sport Economics, 11(5), pp. 567-579

Lago U. (2006). *The state of italian football industry*. In Andreff W., Szymanski S. (eds), Handbook of Economics of sport, Edward Elgar, Cheltenham, UK • Northampton, MA, USA, pp. 463-473

Lago U., Baroncelli A., Szymanky S., (2004). *Il business del calcio: successi sportivi e rovescifinanziari,* Egea, Milano, pp. 149-190

Lampel J., Shamsie J. and Shapira Z. (2009). *Special Issue on Learning from Rare Events*. Organization Science 20(5): 835–845.

Lehmann E. (2000). *Verdienen Fußballspieler was sie verdienen?* In H.-M. Schellhaaß (Ed.) Sportveranstaltungen zwischen Liga- und Medieninteresse (pp. 97-121). Schorndorf, Germany: Karl Hofmann.

Levinthal DA and March JG (1981). *A model of adaptive organizational search*. Journal of Economic Behavior and Organization 2(4): 307–333.

Levitt B and March JG (1988). *Organizational learning*. Annual Review of Sociology 14: 319–340.

Lucifora C., Simmons R. (2003). *Superstar effects in sport: evidence from italian soccer*. JOURNAL OF SPORTS ECONOMICS, Vol. 4 No. 1, 35–55.

Madsen P. M, & Desai V. (2010). *Failing to learn? The effects of failure and success on organizational learning in the global orbital launch vehicle industry*. Academy of management journal, 53(3), 451-476.

Maguire, J. and Stead, D. (1998). *Border crossing, soccer labour migration and the European Union*. International Journal for the Sociology of Sport, 33 (1), pp. 146–68.

Mancin M., (2009). *Il bilancio delle società sportive professionistiche*. Padova, Cedam.

Martin S. (2011). *SPORT ITALIA: the italian love affair with sport*. Londra, Bloomsburg Publishing

McCarthy P. (2000). *Sport and society in Italy today*. Journal of Modern Italian Studies, 5:3, 322-326.

McDaniel M. A., Schmidt F. L. & Hunter J. E. (1988). *Job experience correlates of job performance*. Journal of Applied Psychology, 73: 327–330.

McDonald, M., Milne. G. (1997). *Conceptual framework for evaluating marketing relationship in professional sport franchises*. Sport Marketing Quarterly, 6(2), 27-.14.

McEnrue M. P. (1988). *Length of experience and the performance of managers in the establishment phase of their careers*. Academy of Management Journal, 31(1), 175-185.

McNamara P.; Peck S.I., and Sasson A. (2013). *Competing Business Models, Value Creation and Appropriation in English Football*. Long Range Planning 46 (6): 475–487.

Michael S. C. & Palandjian T. P. (2004). *Organizational learning and new product introductions*. Journal of product innovation management, 21(4), 268-276.

Mion G. & Opromolla L. D. (2014). *Managers' mobility, trade performance, and wages*. Journal of International Economics, 94(1), 85-101.

Montanari F., Silvestri G. & Bof F. (2008). *Performance and Individual Characteristics as Predictors of Pay Levels: The Case of the Italian 'SerieA'*. European Sport Management Quarterly, 8:1, 27-44.

Mullin, B., Hardy, S., & Sutton, W. (1993). *Sport marketing*. Champaign,IL: Human Kinetics, first edition.

Mulotte L. (2014). *Do experience effects vary across governance modes?* Organization Science **25**(3): 757–775.

Musaji S., Schulze W. S. & De Castro J. O. (2019). *How Long Does It Take to "Get to" the Learning Curve?* Academy of Management Journal, (ja)

Nash, R. (2000). *Contestation in modern English football*. Internallonal Review for the Sociology of Sport, J5(4), 419- 452.

Nerkar A, Roberts P. (2004). *Technological and product-market experience and the success of new product introductions in the pharmaceutical industry*. Strategic Management Journal **25**(8/9):779–799.

Peake W. O. & Marshall M. I. (2009). *Does experience determine performance? A meta-analysis on the experience-performance relationship* (No. 319-2016-9685).

Peres-Neto, R. Pedro, A. J. Donald, and K.M. Somers (2005). *How Many Principal Components? Stopping Rules for Determining the Number of Non-Trivial Axes Revisited.* British Journal of Statistical Psychology 49, pp 974–97.

Perkins S. E. (2014). *When does prior experience pay? Institutional experience and the multinational corporation.* Administrative Science Quarterly, 59(1), 145-181.

Pfeffer S. (1998). *The human equation: building profits by putting people first.* Harward Business school press, Harvard.

Pitts, B.G. (1988). *Sport produci markets*. Unpublished manuscript, University of Louisville.

Pitts, B. G., Fielding, L. W., & Miller, L.K. (1991, June). *Sport product-markets: A conceptual model.* Paper presented at the conference of the North American Socìety for Sport Management,Ottawa, Canada.

Pitts, B. G., Fielding, L. W., & Miller, L. K. (1994). *Industry segmentation theory and the sport industry: Developing a sport industry segment model.* Sport Marketing Quarterly, 3 (1), 15-24.

Pitts J. D. (2016). *Determinants of Success in the National Football League's Postseason: How Important Is Previous Playoff Experience?* Journal of Sports Economics, 17(1), 86-111.

Ployhart R. E. & Hakel M. D. (1998). *The substantive nature of performance variability: Predicting interindividual differences in intraindividual performance.* Personnel Psychology, 51: 859–901.

Poli R., Ravenel L. & Besson R. (2018). *Is there an optimum squad age to win in football?* CIES Football Observatory Monthly Report Issue 32 - February 2018

Poli R., Ravenel L. & Besson R. (2018). *The importance of squad stability: evidence from European football.* CIES Football Observatory Monthly Report Issue 34 - April 2018

Poli R., Ravenel L. and Besson R. (2018). Monthly Report 40 - *Competitive balance: a spatio-temporal comparison*, CIES.

Porter, M.E. (1985). *Competitive advantage: Creating and sustaining superior performance.* New York: The Free Press.

Porter PK, Scully G. (1982). *Measuring managerial efficiency: the case of baseball.* Southern Economic Journal 48: 642–650.

Quiñones M. A., Ford J. K., & Teachout M. S. (1995). *The relationship between work experience and job performance: A conceptual and meta-analytic review.* Personnel Psychology, 48: 887–910.

Ranieri M., (2005). *Calcio e diritti televisivi: il ruolo della Tv nella competizione sportiva ed economica.* Analisi Giuridica dell'Economia, 2, pp. 503-513.

Reagans R., L. Argote D. Brooks (2005). *Individual experience and experience working together: Predicting learning rates from knowing who knows what and knowing how to work together*. Management Sci. 51(6) 869–881.

Reuber A. R. & Fischer E. M. (1994). *Entrepreneurs' Experience, Expertise, and the Performance of Technology-based Firms*. IEEE Transactions on Engineering Management, 41(4), 365-374.

Reuber A. R. & Fischer E. (1999). *Understanding The Consequences of Founders' Experience*. Journal of Small Business Management, 37(2), 30-45.

Rico R., Sánchez-Manzanares M., Gil F. & Gibson C. (2008). *Team implicit coordination processes: A team knowledge–based approach.* Academy of Management Review, 33(1), 163-184.

Rohm, A. J., G. R. Milne, and M. McDonald (2006). *A Mixed-Method Approach for Developing Market Segmentation Typologies in the Sports Industry*. Sport Marketing Quarterly, 15 (1), 29-39.

Rottemberg S. (1956). *The baseball players' labor market.* Journal of Polit. Econ, vol 64, n.3.

Rubin I. M. & Seelig W. (1967). *Experience as a factor in the selection and performance of project managers.* IEEE Transactions on Engineering Management, (3), 131-135.

Ruggiero J, Hadley L, Gustafson E. (1996). *Technical efficiency in Major League Baseball.* In Baseball Economics: Current Research, Fizel J, Gustafson E, Hadley L (eds). Praeger: Westport, CT; 191–200.

Ruigrok W., Greve P. & Engeler M. (2011). *International experiential diversity and performance at project organizations: The case of national football teams.* Sport, Business and Management: An International Journal, 1(3), 267-283.

Sampson R. (2005). *Experience effects and collaborative returns in R&D alliances.* Strategic Management Journal **26**(11): 1009–1031.

Sandulli P. (2005). *La giurisdizione "esclusiva" in materia di diritto sportivo*. "Analisi Giuridica dell'Economia, Studi e discussioni sul diritto dell'impresa" 2/2005, pp. 395-404

Sanino M. (2005). *La nuova collocazione delle leghe nell'organizzazione dello sport*. "Analisi Giuridica dell'Economia, Studi e discussioni sul diritto dell'impresa" 2/2005, pp. 440-448

Schmidt F. L., Hunter J. E. & Outerbridge A. N. (1986). *The impact of job experience and ability on job knowledge work sample performance, and supervisory rating of performance*. Journal of Applied Psychology, 71: 432–439.

Schulz R. & Curnow C. (1988). *Peak performance and age among superathletes: Track and field, swimming, baseball, tennis, and golf.* Journal of Gerontology: Psychological Sciences, 43, 113–120.

Sener I.; Karapolatgil A.A. (2015). *Rules of the game: strategy in football industry*. Procedia-Social and Behavioral science 207, Cankaya University of Ankara, Turkey, pp 10-19

Serna Rodríguez M., Ramírez Hassan A. & Coad A. (2018). *Uncovering Value Drivers of High Performance Soccer Players.* Journal of Sports Economics, 1527002518808344.

Shaver M, Mitchell W & Yeung B. (1997). *The effect of own-firm and other-firm experience on foreign direct investment survival in the United States (1987–1992).* Strategic Management Journal **18**(10): 811–824.

Simon H. A. & Chase W. G. (1973). *Skill in chess.* American Scientist, 61, 394–403.

Simpson M., Tuck N. & Bellamy S. (2004). *Small business success factors: the role of education and training.* Education+ Training, 46(8/9), 481-491.

Singell L. (1993). *Managers, specific human capital, and firm productivity in Major League Baseball.* Atlantic Economic Journal 21: 47–59.

Sitkin SB (1996). *Learning through failure: The strategy of small losses*. In Cohen MD and Sproull LS (eds) Organizational Learning. Thousand Oaks, CA: Sage, 541–578.

Sloane P.J. (1971). *The Economics of professional football: The Football club as a utility maximize*, Journ. Of Pol. Econ. vol 18, n.2.

Sorci C. (2002). *Lezioni di economia aziendale.* Giuffrè, Milano, pag 451

Stewart, B., Smith, A. and Nicholson, M. (2003). *Sport consumer typologies: A critical review*. Sport Marketing Quarterly, vol. 12, no. 4, pp. 206-216.

Stuart R. W. & Abetti P. A. (1990). *Impact of entrepreneurial and management experience on early performance.* Journal of business venturing, 5(3), 151-162.

Sturman M. C. (2003). *Searching for the inverted U-shaped relationship between time and performance: Meta-analyses of the experience/performance, tenure/performance, and age/performance relationships.* Journal of management, 29(5), 609-640.

Tarlow J. (2012). *Experience and winning in the national basketball association*. Paper presented at the MIT Sloan Analytics Conference, March 23, 2012, Boston, MA.

Teddlie, C., & Tashakkori, A. (2003). *Major issues and controversies in the use of mixed methods in the social and behavioral sciences.* In A. Tashakkori & C. Teddlie (Eds.), Handbook of mixed methods in social & behavioral research. Thousand Oaks, CA: Sage Publications.

Teodor D. (2015). *The Business Model of European Football Club Competitions.* Procedia Economics and Finance 23 (October 2014), pp. 1245–52.

Tesluk P. E. & Jacobs R. R. (1998). *Toward an integrated model of work experience.* Personnel Psychology, 51:321–355.

Tirrito G.D. (2005). *Il modello inglese: il calcio business.* Analisi Giuridica dell'Economia, 2, pp. 427-440.

Toft-Kehler R., Wennberg K. & Kim P. H. (2014). *Practice makes perfect: Entrepreneurial-experience curves and venture performance.* Journal of Business Venturing, 29(4), 453-470.

Trąpczyński P. & Banalieva E. R. (2016). *Institutional difference, organizational experience, and foreign affiliate performance: Evidence from Polish firms.* Journal of World Business, 51(5), 826-842.

Whitler K. A., Krause R. & Lehmann D. R. (2018). *When and How Board Members with Marketing Experience Facilitate Firm Growth.* Journal of Marketing, 82(5), 86-105.

Wicker P., Prinz J.,Weimar D., Deutscher C., & Upmann T. (2013). *No pain, no gain? Effort and productivity in professional soccer.* International Journal of Sport Finance, 8, 124-140.

Wise S. & Valliere D. (2014). *The impact on management experience on the performance of start-ups within accelerators.* The Journal of Private Equity, 18(1), 9-19.

Wright J. D. & Hamilto, R. F. (1978). *Work satisfaction and age: Some evidence for the job change hypothesis*. Social Forces, 56: 1140–1158.

Writh P., Smart D. & McMahan G. (1995). *Matches between human resources and strategy among NCAA basketball teams*. Academy Management Journal, Vol. 39, pp. 441-463.

Wu W. Y. & Lin C. Y. (2010). *Experience, environment, and subsidiary performance in high-tech MNEs.* Journal of Business Research, 63(12), 1301-1309.

Sitography

http://www.treccani.it/enciclopedia/l-organizzazione-dello-sport-nel-mondo-e-in-italia-nella-societa-moderna_%28Enciclopedia-dello-Sport%29/

https://it.wikipedia.org/wiki/Calcio_(sport)#Organi_di_governo

http://www.figc.it/

https://www.olympic.org/about-ioc-olympic-movement

http://www.coni.it/it/coni/leggi-e-statuto.html

https://it.wikipedia.org/wiki/Calcio_nel_mondo

https://www2.deloitte.com/it/it/pages/strategy/articles/_l_ecosistema-dello-sport-per-rispondere-al-cambiamento-delle-nu.html

https://europeanleagues.com/

https://www.statista.com/

https://www.legab.it/

https://www.legab.it/lega-nazionale-professionisti-serie-b/storia-della-lega/

https://it.wikipedia.org/wiki/Serie_B

https://www.footballbenchmark.com/library/promoted_clubs_who_are_the_new_kids_in_the_big_five_leagues

https://www.youtube.com/watch?v=FgakZw6K1QQ

http://www.sthda.com/english/articles/31-principal-component-methods-in-r-practical-guide/112-pca-principal-component-analysis-essentials/#dimension-description

https://stats.stackexchange.com/questions/53/pca-on-correlation-or-covariance/78#78

https://www.R-project.org/

https://www.transfermarkt.it/

Other Sources

Annual Review of Football Finance, **Deloitte**, 2019

Aspetti della vita quotidiana, **ISTAT**, 2017

Baroncelli A. (2019). *Slides from the Summer School in Football Stadia Management*, Udine.

Bellia O.D. (2019). *Slides from the Summer School in Football Stadia Management*, Udine.

Cambridge dictionary

Centro Studi CONI Servizi, da Monitoraggio **CONI-FSN-DSA** 2017

Deloitte Football Money League, **Deloitte**, 2019

Deloitte's sports industry starting lineup 2019

Il libro bianco dello sport italiano, **CONI**, 2012

Mancin M. (2018). *Slides from the course "Economia delle imprese sportive"*, University of Venice, Ca' Foscari.

Oxford dictionary

PwC's Sports Survey 2019

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Report Calcio 2019, **Arel and Pwc**, 2019

Statuto Federale **FIGC,** 2019