



Uniwersytet Ekonomiczny  
we Wrocławiu

KIERUNEK STUDIÓW

Analityka gospodarcza

**Oskar Wenerowicz**

Nr albumu 173469

**PRACA DYPLOMOWA LICENCJACKA**

**Zarobki zawodników a efektywność występów w meczach  
piłkarskich – analiza statystyczno-ekonometryczna**

*Promotor:*

Dr Alicja Grześkowiak

Katedra Ekonometrii i Badań Operacyjnych

**WROCLAW 2021**

# **Player earnings and performance efficiency in football matches - a statistical and econometric analysis**

## **Abstract**

The aim of the study is to compare the contracts of top football players playing in the top European leagues with their performances in matches. Statistical and econometric analysis was used to obtain the results. The analyses were carried out on real football players statistics. The database was created for 100 players based on the 2018/2019 season. Each of the one hundred observations was described by ten variables.

The main purpose of the research is to examine if there are any relations between earnings and sports performance. Econometric modeling, composite indicator method and principal component analysis were used for this purpose. The study used software such as: R environment and RStudio, Microsoft Excel and Statistica.

## Spis treści

|  |           |
|--|-----------|
| <b>Wstęp</b> .....   | <b>4</b>  |
| <b>1. Sport w kontekście ekonomicznym</b> .....  | <b>6</b>  |
| 1.1 Znaczenie sportu w gospodarce .....  | 6         |
| 1.2 Źródła finansowania klubów sportowych .....  | 10        |
| 1.3 Wybrane aspekty finansowania klubów na przykładzie Premier League oraz Leo<br>Messiego ..... | 13        |
| <b>2. Charakterystyka podejścia badawczego</b> .....   | <b>17</b> |
| 2.1 Charakterystyka danych .....   | 17        |
| 2.2 Analiza korelacji .....  | 17        |
| 2.3 Analiza głównych składowych .....  | 19        |
| 2.4 Model ekonometryczny .....   | 22        |
| 2.5 Metoda wzorca rozwoju Hellwiga .....   | 30        |
| <b>3. Ocena zależności zarobków i efektywności występów w meczach piłkarskich</b> .....          | <b>32</b> |
| 3.1 Budowa rankingu z wykorzystaniem głównych składowych .....                                   | 32        |
| 3.2 Badanie zależności za pomocą modelu ekonometrycznego .....                                   | 43        |
| 3.3. Analiza na podstawie wzorca rozwoju .....   | 50        |
| <b>Zakończenie</b> .....   | <b>58</b> |
| <b>Bibliografia</b> .....  | <b>59</b> |
| <b>Spis tabel</b> .....  | <b>62</b> |
| <b>Spis rysunków</b> .....   | <b>63</b> |

# Wstęp

Celem pracy jest zestawienie i ocena zależności wielkości kontraktów topowych piłkarzy grających w czołowych europejskich ligach z ich występami na boisku. Jest to temat interesujący oraz ważny dla części społeczeństwa zainteresowaną ekonomią sportu, piłką nożną ale również finansami w sporcie.

Sformułowano następujące pytania badawcze:

- Czy istnieją związki pomiędzy zarobkami piłkarzy a efektywnością na boisku?
- Jakimi metodami ilościowymi można oceniać te związki?
- Czy i które narzędzia statystyczno-ekonometryczne mogą służyć do syntetycznej oceny efektywności występów w meczach piłkarskich?
- Czy wyniki otrzymane na podstawie rankingów stworzonych metodą głównych składowych oraz metodą wzorca rozwoju są zbieżne?
- Czy modelowanie ekonometryczne pozwala na wskazanie najważniejszych czynników wpływających na zarobki?

Do badania użyto danych zgromadzonych z dwóch portali fbref.com oraz capology.com. Zebrane dane przedstawiają zarobki oraz poszczególne statystyki zawodników w sezonie 2018/2019. Podczas tworzenia pracy oraz przeprowadzania badań korzystano z wielu źródeł literaturowych. Pośród nich znajdują się np.: publikacje książkowe, artykuły naukowe, raporty instytucji rządowych, raporty firm audytowych, artykuły z portali internetowych. Do przeprowadzenia badania skorzystano z kombinacji różnych metod, instrumentów i narzędzi analitycznych. W pierwszym etapie przeprowadzono analizę głównych składowych, która posłużyła do stworzenia rankingu. Drugą zastosowaną metodą jest modelowanie ekonometryczne, które pozwoliło wskazać najważniejsze czynniki determinujące zarobki piłkarzy. Natomiast trzecią metodą użytą jest metoda wzorca rozwoju Hellwiga, dzięki której powstał ranking piłkarzy zestawiony z rankingiem z metody pierwszej. Przy prowadzeniu obliczeń dla badań wspierano się oprogramowaniem wielu programów tj. GeoDa, R-Studio, QGIS, Microsoft Excel, Statistica.

Praca składa się ze wstępu, trzech rozdziałów oraz zakończenia. Na samym początku pracy zostały przedstawione: cel pracy, przedmiot badawczy, pytania badawcze, podmiot badań, wykorzystane metody naukowe oraz narzędzia użyte do przeprowadzenia analiz. W pierwszym rozdziale zostaną omówione zagadnienia teoretyczne związane ze sportem, gospodarką oraz zostaną przywołane przykłady z czołowych lig europejskich. W drugim

rozdziale zostanie przedstawiona metodologiczna część pracy, która zawiera charakterystykę podejścia badawczego zastosowanego do opracowania tematu pracy. Ostatni trzeci rozdział ma charakter rozdziału badawczego, który przedstawia zastosowanie metod zaprezentowanych w rozdziale drugim.

# 1. Sport w kontekście ekonomicznym

## 1.1 Znaczenie sportu w gospodarce

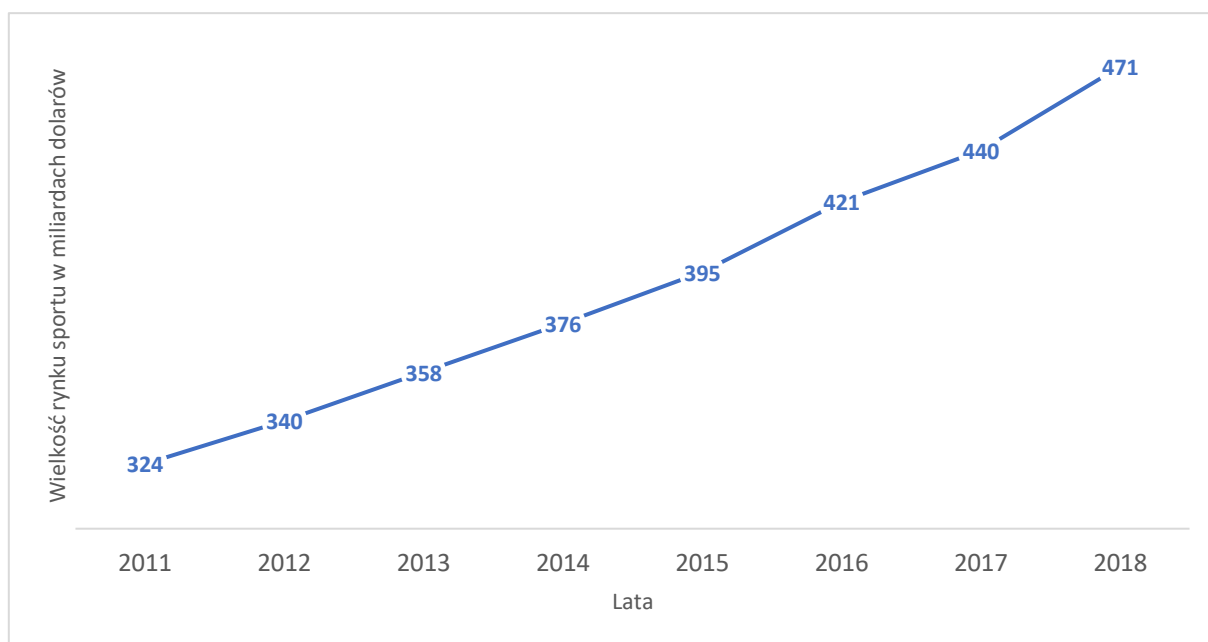
Sport od dawna, od czasów epoki starożytnej, odgrywa pewną rolę w życiu społeczeństwa. Obecnie stanowi on duży sektor gospodarki w większości krajów świata. W 2018 r. globalna wartość rynku sportu wyniosła około 488,5 mld USD. Szacuje się, że w roku 2022 globalna wartość rynku sportowego będzie wynosić 614,1 mld USD (Kutwa, Rafał, 2019). Jest to około 1% światowego PKB (Śniadek, 2012).

Ameryka Północna jest obecnie największym sportowym rynkiem. 30% światowego rynku, taki udział szacuje się dla tego kontynentu. Niemniej w niedalekiej przyszłości przewiduje się, że to Azja, Pacyfik oraz Bliski Wschód będą rynkami, które w porównaniu do reszty będą rosły najszybciej (The Business Research Company, 2019). Przykładowo, w Polsce sektor sportu kreuje 2,26% produktu krajowego brutto. W liczbach bezwzględnych wynosi to 32,1 mld zł w 2010 r. (Baran i in., 2016). Warto napomnieć, że w Polsce liczba osób zatrudnionych w sektorze sportu szacowana jest na 294 tys. Oznacza to, że 2,1% wszystkich zatrudnionych ludzi w polskiej gospodarce pracuje w tym sektorze.

Należałoby dodać, że rynek sportu wygenerował również 9,63 mld PLN rocznie wartości dodanej w 2018 r. (Kutwa, Rafał, 2019). Dla Polski wielką szansą jest energicznie rosnący globalny rynek sportu, która przy odpowiednim zagospodarowaniu pozwoli na szybszy rozwój gospodarczy kraju.

Trzeba pamiętać, że rynek sportu to nie tylko rywalizacja sportowa pomiędzy klubami lub zawodnikami. Jest to również wiele powiązanych branż, z których warto wymienić tj. rekreację, produkcję odzieży i sprzętu sportowego, pamiątek, gadżetów sportowych, rynek reklamy i marketingu, turystykę, a nawet budownictwo.

Na rysunku 1.1 przedstawiono kształtowanie się globalnego rynku sportu w latach 2011-2018.

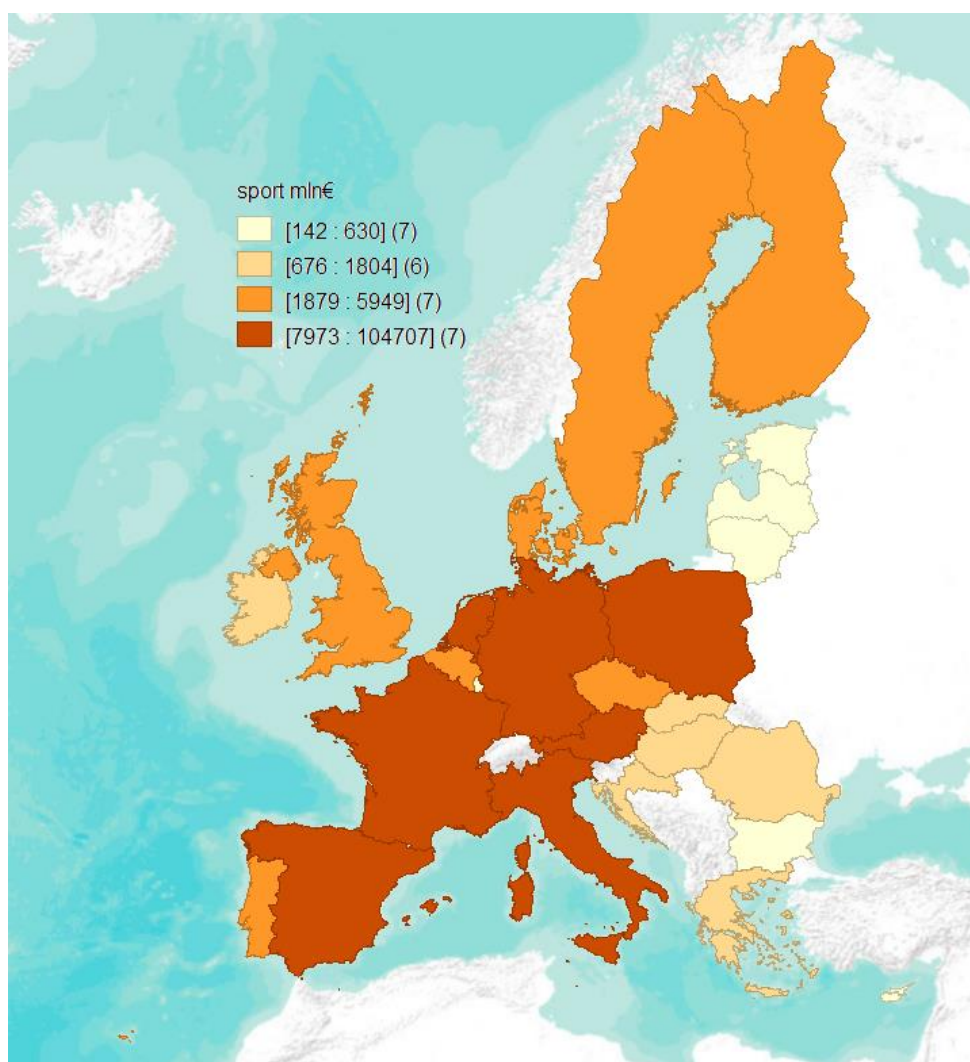


Rysunek 1.1 Wielkość rynku globalnego sportu w miliardach dolarów

Źródło: opracowanie własne na podstawie: statista.com

Rysunek 1.1 pokazuje, jak rynek sportu ciągle się rozwija i rozrasta. Oczekiwania ludzi odnośnie jakości przekazu, poziomu sportowego stale rosną, co za tym idzie wszystkie osoby związane ze sportem starają się popularyzować swoje dyscypliny, rozgrywki itd. Średnie tempo zmian dla wymienionych lat wyniosło 5,49%. Oznacza to, że w ciągu kolejnych 8 lat od roku 2011 do roku 2018 wielkość globalnego rynku sportu rosła średnio o 5,49% w stosunku do roku poprzedniego. W każdym z lat wartość rynku się zwiększała. W 2018 roku w porównaniu do 2011 roku wartość tego rynku wzrosła aż o 45,37%. Udowadnia to, że branża sportu jest szybko rozwijającą się gałęzią gospodarki. Na rysunku 1.2 został przedstawiony wskaźnik obrazujący ile sport generuje PKB w 2012 roku w każdym z krajów EU-28.

Sport oddziałuje pośrednio na wiele branż i gałęzi gospodarki. Najmocniejszy wpływ ma jednak na rynek rekreacji. Wartość tego rynku na świecie w 2018 r. wyniosła 1,44 bln USD. Eksperci szacują, że w 2022 r. wzrośnie do 1,81 bln USD. Rynek sportu natomiast miał aż 34% udziału w rynku rekreacji, nieznacznie ulegając branży rozrywkowej, która miała 35,6% udziału (Baran i in., 2016).

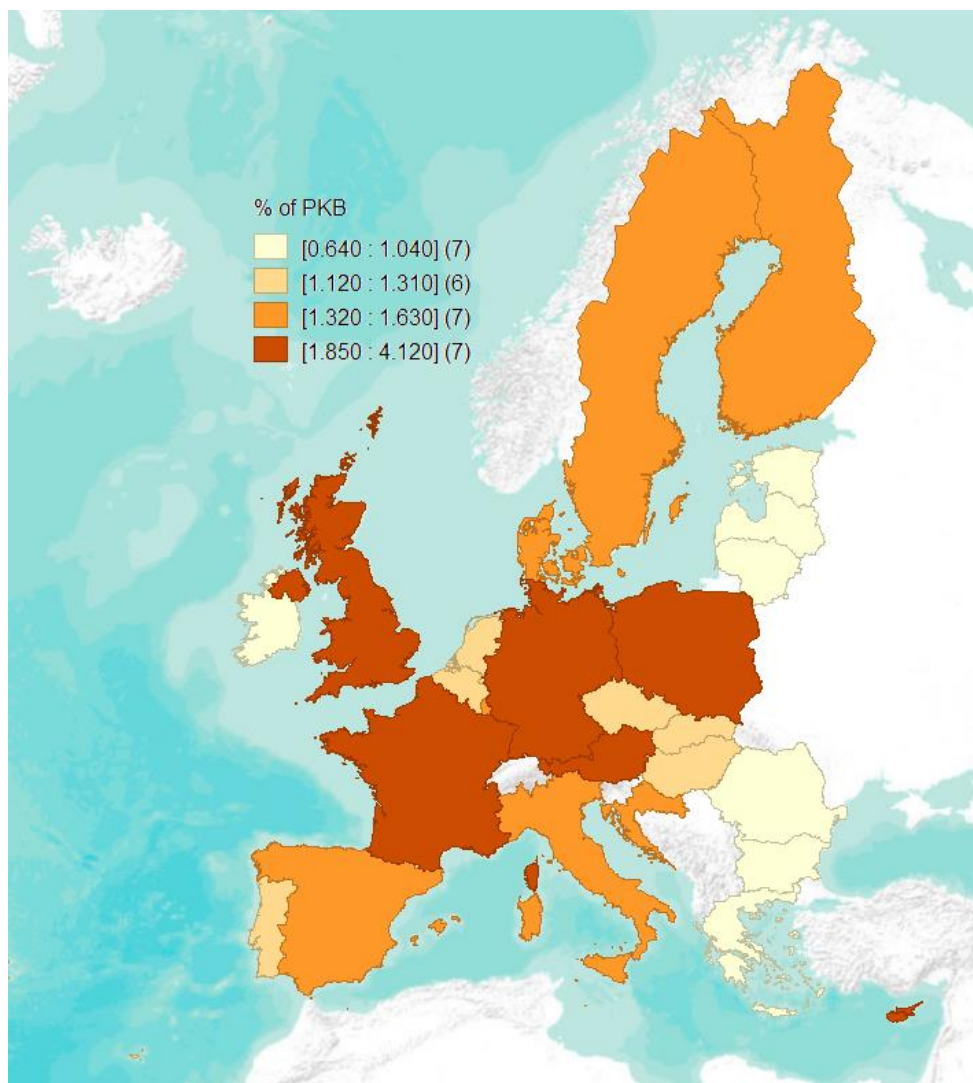


Rysunek 1.2 PKB generowane przez sport w 2012 roku w EU-28 (oprócz Słowenii – B/D)

Źródło: opracowanie własne na podstawie: Komisja Europejska (2018)

Na podstawie rysunku 1.2 można wywnioskować, że największe gospodarki Unii Europejskiej generują najwięcej PKB ze sportu w porównaniu do innych krajów. Najwięcej PKB ze sportu generują Niemcy, Hiszpania, Francja, Włochy, Austria, Polska oraz Holandia. Są to zdecydowanie jedne z najbardziej licznych państw EU-28. Warto zauważyć, że w pierwszym kwartyle znajdują się państwa małe, posiadające relatywnie mało mieszkańców. Znajdują się tam takie państwa jak np.: Cypr, Estonia, Łotwa, Litwa. Należy zwrócić uwagę na dominację Europy zachodniej w tym zestawieniu. Jednocześnie kraje Europy Wschodniej wypadają bardzo słabo w porównaniu do zachodu Europy czy nawet Skandynawii. Na rysunku 1.3 został przedstawiony procentowy udział sportu w PKB każdego z państw z EU-28.





Rysunek 1.3 procentowy udział PKB generowany przez sport w całości PKB w roku 2012 w EU-28 (oprócz Słowenii – B/D)

Źródło: opracowanie własne na podstawie: Komisja Europejska (2018)

Na podstawie rysunku 1.3 można wywnioskować, że rozkład udziału sportu w PKB wygląda podobnie jak na rysunku 1.2. Dominują największe gospodarki oraz najbardziej liczne. Ponownie najgorzej wypadają państwa ze wschodniej Europy takie jak np.: Estonia, Litwa, Łotwa, Grecja, Irlandia. Polska ponownie znajduje się w czwartym kwartylu z wielkimi europejskimi gospodarkami. Możliwe, że jest to spowodowane tym, że dane pochodzą z roku 2012, w którym w Polsce odbyły się Mistrzostwa Europy w piłkę nożną. Z pewnością tak wielka impreza wpłynęła mocno na gospodarkę kraju.

Wydarzenia sportowe na żywo często przyciągają jedne z największych widowni telewizyjnych na całym świecie. Nic w tym dziwnego, że branża jest ogromna i stale rośnie. Jednym z najpopularniejszych sportów na świecie jest piłka nożna. To ona przyciąga ogromne

grono widzów przed telewizory, napędzając tym samym koniunkturę. Dominacja piłki nożnej nad światem jest tak oczywista, że nie powinno nikogo dziwić, że piłka nożna jest tak popularna na całym świecie. Piękna Gra, jak ją nazywają, ma fanów na wschodzie, jak Japonia, i na zachodzie jak Ameryka, ale nigdzie na świecie ta gra nie jest tak popularna jak w Europie. Mistrzostwa Świata FIFA to najbardziej kultowe wydarzenie tego sportu, które jest nie tylko najbardziej oglądanym wydarzeniem drużynowym, ale także jednym z najwyższej opłacanych turniejów w całym świecie sportu.

Piłka nożna jest jednym z najbardziej dostępnych sportów na świecie. Piłka nożna zdominowała sportowe nagłówki w niektórych z głównych krajów na świecie, w tym prawie wszystkich narodów europejskich, takich jak Wielka Brytania, Niemcy, Francja, Hiszpania. W Ameryce Południowej piłka nożna to coś więcej niż tylko gra. Dla wielu dzieci w Ameryce Południowej piłka jest odskocznią od codziennych problemów i wszechobecnego ubóstwa. Wiele dzieci marzy o tym, żeby zostać piłkarzem i móc wyrwać się z trudnego środowiska. Jednocześnie wielu z nich chciałoby zapewnić odpowiedni byt dla swoich bliskich. Bardzo chcieliby przeprowadzić się do Europy, gdzie są najlepiej grające kluby piłkarskie i zarazem płacące bardzo wysokie pensje swoim zawodnikom.

Mistrzostwa Świata w piłce nożnej to zdecydowanie najchętniej oglądane zawody sportowe na świecie. Mistrzostwa Świata w 2014 roku przyciągnęły aż 3,9 mld ludzi przed telewizory. Tych, którzy przynajmniej przez moment oglądali te zawody. Natomiast sam finał tych mistrzostw w piłce nożnej w 2014 roku oglądało aż 700 milionów ludzi na całym świecie (Totalsportek, 2019).

## 1.2 Źródła finansowania klubów sportowych

Istnieje wiele sposobów na pozyskiwanie finansowania przez kluby sportowe. Głównymi z nich są (Deloitte, 2020):

- Wpływy ze sprzedaży praw do transmisji meczów w telewizji lub na platformach streamingowych,
- Sponsoring,
- Dochody z dnia meczowego,
- Reklama,
- Inne źródła finansowania.

Dawno minęły czasy, kiedy kluby piłkarskie opierały się głównie na dochodach z dnia meczu jako głównego źródła przychodów. Ten przychód jest nadal bardzo ważny dla mniejszych klubów w mniejszych ligach. Tak naprawdę patrząc na najwyższy szczebel to drużyny piłkarskie w głównej mierze opierają swoje zarobki na kontraktach komercyjnych i sprzedaży praw do transmisji. Te zarobki są powszechnie znacznie większe niż przychód z dnia meczowego. Przede wszystkim dzięki pieniądзом ze sprzedaży transmisji z czołowych lig na całym świecie. To znacznie zwiększyło ekspozycję, potencjał handlowy oraz moc przetargową samych klubów. Obecnie kluby to globalne marki, rozpoznawane na całym świecie.

Transmisje to źródło, z którego większość klubów z wyjątkiem kilku największych opiera większość swoich dochodów. Zapoczątkowane przez telewizję Sky we wczesnych latach 90-tych, stale rosnące prawa do transmisji podzielone równo pomiędzy dwadzieścia klubów najwyższej klasy rozgrywkowej w Anglii, które wynoszą obecnie 31,8 mln GBP za sezon. Do tego dochodzą dodatki w zależności od liczby transmitowanych spotkań danej drużyny. Są to kwoty jedynie dotyczące umowy z telewizją krajową. Otrzymują również po 5 mln GBP udziału w przychodach z działalności komercyjnej. Następnie 43,2 mln GBP każdy z klubów otrzymuje z zagranicznych praw telewizyjnych. Do tego trzeba doliczyć pieniądze płynące z wyników sportowych. W sezonie 2019/2020 Liverpool FC sumując wszystkie dodatki mógł liczyć na 174,6 mln GBP wpływających do budżetu (Planet Football, 2020). Kluby z całego świata opierają obecnie swoje koszty operacyjne na dochodach generowanych przez partnerstwa telewizyjne. W przypadku Premier League, kluby zdegradowane otrzymują dwuletnie (jeśli klub spadł w pierwszym sezonie po awansie, inaczej trzyletnie) płatności spadochronowe (ang. Parachute payments), czyli mimo relegacji nadal otrzymują odsetek z równo podzielonej części praw do transmisji, ponieważ bez nich większość z nich zbankrutowałaby. Z drugiej strony, finał play-off Championship o awans do Premier League jest prawdopodobnie słusznie uważany za najbogatszy pojedynczy mecz w piłce nożnej, którego zwycięzca z pewnością zarobi kilkaset milionów.

W przeszłości w innych ligach europejskich niektóre kluby wynegocjowały swoje własne umowy telewizyjne, które powodują nieproporcjonalnie duży podział dochodów. Tutaj przykładem jest hiszpańska La Liga. Liga ta nie jest nadzwyczaj konkurencyjna i wyrównana. Dominują w niej trzy drużyny tj. FC Barcelona, Real Madryt i Atletico Madryt. W sezonie 2014/2015 te trzy kluby razem zgromadziły ponad 40% pieniędzy ze sprzedaży praw telewizyjnych, gdzie reszta klubów w liczbie 17 musiała podzielić 60% „tortu” między sobą.

Na szczęście dla mniejszych klubów scentralizowany model został uchwalony w 2015 roku i został w pełni wdrożony w sezonie 2016/17. W sezonie 2015/16 zastosowano model hybrydowy jako etap wstępny do pełnego wdrożenia. Obecnie ten odsetek spadł do mniej niż 23% dla tych trzech klubów. Natomiast pieniądze z praw telewizyjnych znacznie wzrosły. W sezonie 2018/2019 FC Barcelona mogła liczyć na 166,5 mln EUR, Real Madryt na 155,3 mln EUR a Atletico Madryt na 119,2 mln EUR. Mimo, że są to kwoty zbliżone do pieniędzy zarobionych przez klub Liverpool FC w sezonie 2019/2020, liga angielska nadal mocno dominuje scenę europejską i światową pod względów wpływów pieniężnych ze sprzedaży praw (McMahon, 2019).

Kolejnym źródłem dochodów klubów jest sponsoring komercyjny. Aktualnie kluby cały czas zwiększają swoją ekspozycję na rynku globalnym, zwłaszcza te największe podmioty. Kluby piłkarskie mają lukratywne umowy sponsorskie i reklamowe. Według możliwości zasięgowych, reklama w sporcie ma najszerze grono odbiorców. Firmy z różnych nisz mają więcej instrumentów do promowania swojej marki. Kibice lepiej zapamiętują te produkty, które widzą w trakcie meczu. Wiele topowych klubów posiada koszulki sygnowane sloganem Emirates Airlines. Arsenal, Real Madryt, Paris Saint Germain, Milan, wielkie marki w czterech największych ligach świata. Wszystkie te kluby mają na środku koszulki napis "Fly Emirates" jako swojego globalnego sponsora. Firma ta przeznaczająca 35% swojego budżetu reklamowego na reklamę w piłce nożnej.

Barcelona otrzymała w 2018 roku największe kontrakty sponsorskie prawie 200 mln GBP. Lwia część tej kwoty została uzyskana tylko dzięki dwóm dużym umowom. Pierwsza z nich to umowa z firmą produkującą sprzęt sportowy Nike oraz z Rakutenem jako głównym sponsorem globalnym. Zgodnie z niedawno podpisaną umową z dostawcą zestawów, Nike będzie płać Barcelonie 100 milionów za sezon przez następne 10 sezonów (Llimós, 2016). Nie można lekceważyć potęgi sprzedawanej koszulki. Czasami, dzięki popularnemu nazwisku na plecach, klub może zbić koszty transferu drogiego zawodnika w kilka tygodni. Oczywiście ta opcja sprawdza się głównie w przypadku topowych klubów. Niedaleko w tyle, Real, Manchester United, Bayern Monachium i Liverpool znajdują się w pierwszej piątce listy pod względem najwyższych dochodów ze sponsoringu.

Największe podmioty w piłce nożnej robią fortunę tworząc globalne partnerstwa komercyjne, w większości przypadków zarabiając nawet więcej niż z niewiarygodnie dużych przychodów telewizyjnych. Mniejsze podmioty pozostawione są w walce o mniejsze firmy. W niższych ligach w całej Europie większość z nich nadal opiera swoje dochody komercyjne

na lokalnych firmach lub dotacjach rządowych, a proporcja dochodów jest znacznie mniej znacząca. Jeśli jednak osiągną wystarczająco dobre wyniki, aby awansować do najwyższej klasy rozgrywkowej w swoim kraju lub rozgrywek europejskich to bez względu na reputację zostaną zasypani ofertami.

Na tej podstawie można wywnioskować, że obecnie wpływy z transmisji telewizyjnych oraz wpływy komercyjne to największe źródła dochodów dla największych klubów. Piłka nożna stała się globalnym sportem, który został spieniężony bardzo dokładnie. Można tylko się zastanawiać do czego posuną się jeszcze kluby w celu powiększenia rocznych przychodów.

## **1.3 Wybrane aspekty finansowania klubów na przykładzie Premier League oraz Leo Messiego**

Angielska Premier League jest najchętniej oglądaną ligą na świecie. Przyciąga najlepszych graczy na świecie oraz najbogatszych ludzi i największe firmy, które inwestują w kluby ligi angielskiej. Wśród sponsorów są banki, firmy telewizyjne, firmy fast food. Premier League stała się ryzykownym rynkiem dla marek chcących zainwestować w zespoły angielskiej ligi. Marketerzy nie mogą jednak pozwolić sobie na stanie w miejscu, jeśli chcą zmaksymalizować swoje nadchodzące strategie piłkarskie. Premier League generuje wiele przychodów z różnych źródeł, jednak sponsoring jest zdecydowanie jednym z głównych źródeł przychodów. Te umowy sponsorskie wraz z dobrymi technikami marketingowymi pozwalają lidze być konkurencyjną nie tylko w Europie, ale i na świecie.

Dawno minęły czasy, kiedy marki po prostu przyklejały swoją nazwę do stroju drużyny lub banneru reklamowego przy boisku w nadziei na zarobienie pieniędzy na pięknej i widowiskowej grze. Marketing w piłce nożnej jest dziś złożonym przedsięwzięciem, rozgrywanym na wielu kanałach i przeciwko niezliczonym konkurentom. Marki, które osiągną sukces w danym sezonie i zapewnią sobie miejsce w świadomości fanów, będą tymi, które potrafią skutecznie wykorzystać dane do zwiększenia zaangażowania.

Jak wynika z raportu Marketing Week, wielu sponsorów Premier League integruje swoje działania związane z piłką nożną z szerszymi planami medialnymi w większym stopniu niż kiedykolwiek wcześniej.

Na przykład marka kanapek Subway stara się wykorzystać wnioski płynące ze sponsorowania klubu Liverpool FC, aby wykorzystać je w marketingu treści i działaniach w mediach społecznościowych przez cały rok. W międzyczasie Carlsberg rozwija platformę internetową, która będzie śledzić potencjalne treści viral'owe od fanów piłki nożnej do wykorzystania we własnych kanałach (Bacon, 2014).

Strategie te odzwierciedlają fakt, że marki nie mogą po prostu nadawać swoich własnych komunikatów do fanów i oczekiwać rezultatów. W sytuacji, gdy urządzenia mobilne oraz media społecznościowe stają się coraz ważniejszym elementem oglądania piłki nożnej, marki muszą znaleźć atrakcyjne sposoby na włączenie się w rozmowy, które kibice już prowadzą.

Marketerzy danych odgrywają tu kluczową rolę - nie tylko w przyswajaniu tego, co mówią fani i gdzie to mówią, ale także w pomocy w informowaniu marek o ich ewoluujących strategiach. Kreatywny talent jest oczywiście nadal niezbędny, o czym świadczy fakt, że coraz więcej marek tworzy wewnętrzne newsroomy, które skupiają się wyłącznie na tworzeniu treści wokół sportu i reagowaniu na wydarzenia w czasie rzeczywistym.

Marki, które poświęcają taki sam poziom uwagi swoim strategiom w zakresie danych, mogą pomóc w zapewnieniu, że inwestycja ta jest ukierunkowana, skuteczna i, co najważniejsze, przynosi rezultaty. Nie jest to otwarta bramka, ale przy odpowiednim podejściu na pewno jest możliwa do zdobycia przez marki.

Warto wspomnieć, że bardzo dużo zmieniło się w roku 1992. W 1992 więc powstał twór zwany Premier League. Liga miała być niezależna od ligi piłkarskiej i F.A. czyli angielskiej federacji, co dawało jej swobodę w organizowaniu własnych transmisji i umów sponsorskich. To okazało się być kluczowym we wzroście z powrotem na szczyt dla angielskiej piłki nożnej. Ogromny krok w marketingu ligi nastąpił dzięki telewizji Sky. Premier League podjęła zaskakujący krok, aby przenieść prawa do transmisji do tej nowej firmy, która miała zamiar pobierać opłaty od konsumentów za oglądanie meczów na żywo. Połączenie jakości sportowej oraz strategii marketingowej i oprawy telewizyjnej wartość ligi znacznie wzrosła. Liga znów zaczęła przyciągać najlepszych zagranicznych piłkarzy, co tylko zwiększyło sprzedaż biletów, oglądalność w telewizji i sprzedaż towarów. Aby spojrzeć na to z innej perspektywy, pierwotna umowa Sky z ligą podpisana w 1992 była warta 191 mln GBP w

ciągu pięciu lat, ponad 307 mln USD. Sky, BT oraz Amazon niedawno zapłaciły blisko 5,14 mld GBP za trzy lata. Od momentu powstania ligi w 1992 roku wartość praw do pokazywania meczów Premier League wzrosła z 38 mln GBP rocznie do ponad 1,665 mld GBP rocznie (BBC, 2015).

Nie tylko prawa do transmisji ogromnie zwiększyły przychody Premier League, pomogły w tym również umowy sponsorskie. Pierwszym sponsorem ligi była firma piwowarska Carling. Zapłacili za prawa do posiadania swojej nazwy obok Premier League na cztery lata, co kosztowało ich 12 mln GBP. Zawody te stały się znane jako FA Carling Premier League. Carling odnowił swój sponsoring na kolejne cztery lata, płacąc podwyżkę w wysokości trzystu procent. W 2001 roku nowy sponsor w postaci Barclaycard zapłacił 48 mln GBP za trzy lata. Następnie odnowił na kolejne trzy lata kosztem blisko 66 mln GBP. Dzięki temu, że liga generuje obecnie tak duże przychody, angielskie kluby mogą konkurować nie tylko w Europie, ale także na skalę światową. Kluby mogą teraz konkurować z każdym na świecie pod względem opłat transferowych, zarobków i stadionów. Dzięki temu Premier League przyciąga najlepszych zawodników i trenerów z całego świata, co sprawia, że jest rozpoznawalna na całym świecie. Wiele klubów rywalizujących w lidze jest obecnie postrzeganych jako globalne marki, tak jak Jankesi w Stanach Zjednoczonych. Przychody generowane przez zagraniczne transmisje telewizyjne, sponsoring i sprzedaż towarów oznaczają kolejne miliony zysku dla Premier League i jej klubów. Premier League jest obecnie prawdopodobnie najbogatszą ligą świata ze względu na generowane przez nią przychody. W sezonie 18/19 kluby Premier League wygenerowały łączny przychód w wysokości około 5,157 mld GBP (Deloitte, 2019).

Angielska Premier League przeszła długą drogę od czasu jej powstania na początku lat dziewięćdziesiątych. Anglicy zreformowali ligę i wyciągnęli ją z dna dzięki doskonałym decyzjom marketingowym i biznesowym. Obecnie Premier League stała się jednym z najbogatszych podmiotów nie tylko w piłce nożnej, ale we wszystkich dyscyplinach sportowych na całym świecie. Wynika to z globalnego marketingu ligi. Wszystkie drużyny mają otrzymać więcej pieniędzy dzięki lepszym umowom telewizyjnym.

Równie ważnym zagadnieniem jest marka osobista piłkarzy występujących w klubach piłkarskich. To oni w dużym stopniu wpływają na rozpoznawalność klubu, identyfikację kibiców z klubem oraz umowy sponsorskie. Wielu piłkarzy obecnie posiada swoje zespoły ludzi zarządzających ich mediami społecznościowymi. Najpopularniejsi piłkarze są w stanie zgromadzić setki milionów fanów na portalach, np. Instagramie. Dzięki tym mediom piłkarze

są bardziej „ludscy”. Kibic natomiast czuje większą więź z piłkarzem, którego treści trafiają bezpośrednio do niego. Co za tym idzie często kibice bardzo mocno utożsamiają się z danym piłkarzem, co częściowo wiąże się z promocją klubu. Za pomocą tych narzędzi kluby piłkarskie trafiają do jeszcze większej rzeszy fanów oraz coraz częściej do ludzi nieinteresujących się sportem. Według hiszpańskiego ekonomisty Ivana Cabeza, sam Leo Messi jest w stanie wygenerować przychody w wysokości 130-200 mln EUR dla FC Barcelony (Navarro, 2021). Argentyńczyk jest najlepszym piłkarzem na świecie oraz jest bardzo popularny w mediach społecznościowych. Większość umów handlowych lub sponsoringowych ma klauzulę, która jest ustalana według tego gracza. Również na poziomie rozpoznawalności Barcelona i Messi to pojęcia tożsame i dobra komplementarne. Nie ma Barcelony bez Messiego tak jak Messiego bez Barcelony. Według badań przeprowadzonych w 2021 roku przez nowego prezydenta klubu FC Barcelona, Joana Laporty wynika, że Leo Messi generuje około 1/3 rocznego przychodu (Wołowski, 2021). Każdy ze sponsorów domaga się posiadania Leo. Dla samej stolicy Katalonii związek tego piłkarza z klubem ma ogromny wpływ na gospodarkę regionu. Są turyści, którzy przyjeżdżają do miasta specjalnie, aby zobaczyć piłkarza.

Kluby również zarabiają dzięki sprzedaży pamiątek, gadżetów, w szczególności koszulek oraz ogólnie odzieży sygnowanej logo klubu. Najlepiej sprzedającym koszulką klubem na świecie w sezonie 2018/2019 był Manchester United (bestseller firmy adidas). Klub z Manchesteru sprzedał w powyższym sezonie 3,25 mln koszulek. Wcale nie dużo gorszy był inny angielski klub Liverpool, który znalazł się na 5 miejscu zestawienia z 1,67 mln sprzedanych koszulek. Interesujące będzie zobaczyć, gdzie Premier League może przenieść swoją markę w ciągu najbliższych dziesięciu lat i czy nadal będzie w stanie generować ogromne przychody (Statista, 2020).



## 2. Charakterystyka podejścia badawczego

### 2.1 Charakterystyka danych

Dane zostały pozyskane z dwóch następujących stron internetowych. Pierwsza tj. capology.com zajmuje się pozyskiwaniem danych dotyczących wynagrodzeń sportowców, obejmuje 18 profesjonalnych lig sportowych. Z tej strony pobrano dane dotyczące wynagrodzeń piłkarzy. Kolejnym źródłem danych była strona fbref.com. Zajmuje się ona gromadzeniem statystyk drużyn piłkarskich i zawodników z całego świata. FBref został stworzony przez Sports-Reference, zespół zajmujący się popularnymi stronami statystycznymi tj. Basketball-Reference. Zebrane dane dotyczą występów w sezonie ligowym 2018/2019. W bazie umieszczono 100 piłkarzy występujących w sezonie 2018/2019 w ligach z tak zwanego Top 5 czyli z najwyższych klas rozgrywkowych w pięciu krajach Europy. Z angielskiej Premier League, z hiszpańskiej La Liga, z włoskiej Serie A, z francuskiej Ligue 1 oraz z niemieckiej Bundesligi. Wartości kontraktów piłkarzy w tych ligach są najwyższe na całym świecie. Z każdej powyższych lig wybrano około 20 najlepiej zarabiających piłkarzy, tylko ofensywnych. Do każdego z nich przyporządkowano wartość wynagrodzenia w danym sezonie, liczbę występów, liczbę minut spędzonych na boisku, liczbę strzelonych bramek, liczbę zaliczonych asyst, liczbę udanych podań celnych, liczbę podań ogółem, liczbę strzałów ogółem, liczbę strzałów celnych, liczbę popełnionych fauli oraz liczbę pozycji spalonych.

### 2.2 Analiza korelacji

Zależność korelacyjna charakteryzuje się tym, że dla każdej wartości jednej zmiennej przyporządkowuje się dokładnie określone średnie wartości drugiej zmiennej. Siłę i kierunek tej zależności liniowej bada się za pomocą współczynnika korelacji  $r_{xy}$ . Występują dwa typy korelacji, poprzez, które można określić kierunek korelacji. Pierwszym jest korelacja dodatnia, która występuje wtedy, gdy wraz ze wzrostem wartości jednej cechy średnie wartości drugiej cechy również rosną. Natomiast korelacja ujemna wiąże się z tym, że wraz ze wzrostem jednej cechy wartości średnie drugiej cechy spadają. Aby obliczyć współczynnik korelacji liniowej Pearsona należy korzystać z poniższego wzoru (Ostasiewicz, 2006):

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{C(X, Y)}{s_y s_x} \quad (2.1)$$

gdzie:

$\bar{x}, \bar{y}$  – średnie arytmetyczne cech

Współczynnik przyjmuje wartości z przedziału  $[-1; 1]$ . Na podstawie znaku współczynnika można określić kierunek korelacji. Natomiast wartość bezwzględna współczynnika definiuje siłę zależności. Gdy współczynnik wynosi 1 lub -1 pomiędzy cechami występuje zależność liniowa. Kiedy współczynnik korelacji liniowej Pearsona wynosi 0 to wtedy zmienne są ze sobą nieskorelowane (Ostasiewicz, 2006). Przeważnie przyjmuje się następującą interpretację współczynnika (Ostasiewicz, 2006):

- $0 \geq r_{xy} < 0,2$  – praktycznie brak związku liniowego między zmiennymi
- $0,2 < r_{xy} < 0,4$  – zależność niska
- $0,4 < r_{xy} < 0,7$  – zależność umiarkowana
- $0,7 < r_{xy} < 0,9$  – zależność znacząca
- $0,9 < r_{xy} \leq 1$  – bardzo silna zależność pomiędzy zmiennymi

Najczęściej używaną i najbardziej popularną nieparametryczną miarą korelacji pomiędzy dwoma cechami jest współczynnik korelacji rang Spearmana. Jeśli badane jest  $n$  obiektów opisanych za pomocą dwóch cech należy przyporządkować rangi wszystkim obserwacjom pierwszej zmiennej. Należy je porządkować od najmniejszej do największej. Tak samo trzeba postąpić z drugą zmienną. Powinno się zdefiniować rangi dla każdej obserwacji, od najmniejszej do największej. Jeżeli po przyporządkowaniu żadna z wartości którejkolwiek zmiennej nie posiada tej samej rangi można zastosować poniższy wzór (Aczel, 2000):

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (2.2)$$

gdzie:

$d_i$  – różnice pomiędzy rangami zgodnymi sobie wartościami cechy  $x_i$  i cechy  $y_i$  ( $i = 1, 2, \dots, n$ ).

Współczynnik korelacji rang Spearmana tak jak współczynnik korelacji liniowej Pearsona przyjmuje wartości z przedziału  $[-1; 1]$ . Im wartość bezwzględna współczynnika wyższa tym większa siła badanej zależności.

## 2.3 Analiza głównych składowych

Podstawowym celem analizy głównych składowych jest utworzenie nowej przestrzeni wektorów o mniejszej liczbie wymiarów, na które mogą być rzutowane oryginalne punkty, w taki sposób, aby można było wykryć rozważaną strukturę. W wyniku analizy głównych składowych otrzymujemy:

- nowy zbiór osi zwanych osiami czynnikowymi w przestrzeni o mniejszej liczbie wymiarów, na które można rzutować wyjściową przestrzeń
- nowy zbiór zmiennych, które stanowią kombinacje liniowe oryginalnych zmiennych, a ponadto są nieskorelowane.

Pierwsza główna składowa jest wyodrębniona w taki sposób, że w możliwie największym stopniu wyjaśnia wariancję danych pierwotnych. Podstawowe równanie analizy głównych składowych ma postać:

$$z_j = b_{j1}S_1 + b_{j2}S_2 + b_{j3}S_3 + \dots + b_{jn}S_n \quad (2.3)$$

gdzie:

$z_j$  –  $j$ -ta zmienna ( $j = 1, 2, \dots, n$ ),

$S_1 \dots S_n$  – główne składowe,

$b_{j1} \dots b_{jn}$  – współczynniki głównych składowych

Główne składowe są wydzielone w taki sposób, że pierwsza główna składowa wyjaśnia największą część wariancji zmiennych wyjściowych. Następna składowa wyjaśnia największą część łącznej wariancji zmiennych, która nie została wyjaśniona przez pierwszą składową, aż zostanie wyjaśniona ona w całości. Liczba głównych składowych odpowiada liczbie zmiennych. Aczkolwiek racjonalnym rozwiązaniem jest branie pod uwagę jedynie kilku pierwszych składowych, które wyjaśniają większość całkowitej wariancji zmiennych. Do zaakceptowania jest rezultat, w którym główne składowe wyjaśniają przynajmniej 60% całkowitej wariancji (Kaczmarek, 2016). Pierwszym krokiem analizy jest obliczenie macierzy

korelacji. Następnie należy dokonać wstępnego oglądu macierzy korelacji oraz podjąć decyzję o usunięciu z analizy zmiennych, które nie są w ogóle skorelowane za innymi zmiennymi. Również na podstawie macierzy należy obliczyć wartości własne i wektory własne. Wartości własne macierzy korelacji pozwolą w późniejszych krokach na redukcję wymiarów oraz dobór odpowiedniej liczby głównych składowych. Wektory własne zawierają współczynniki kolejnych składowych głównych. Kolejnym krokiem w procedurze analizy głównych składowych jest dobór liczby głównych składowych. Problemem stojącym przed badaczem jest to ile głównych składowych wybrać. Do zredukowania liczby składowych proponuje się kilka następujących kryteriów.

- Kryterium wartości własnej Kaisera – założenie jest takie, że wartość własna każdej głównej składowej powinna być większa od 1, czyli wariancji pojedynczej zmiennej (Kaczmarek, 2016);
- Wykres osypiska (Cattela)– wartości własne należy ponumerować malejąco. Na osi x należy przedstawić kolejne numery wartości własnych natomiast na osi y wielkość wartości własnych. Następnie połączyć punkty odcinkami. Należy wybrać taką liczbę składowych, które tworzą „stok” natomiast zredukować liczbę składowych o te, które na wykresie tworzą osypisko (Górniak, 1998).
- Kryterium procentu wariancji – należy wybrać taką liczbę głównych składowych aby wyjaśniały pewien ustalony procent ogólnej wariancji. Składowe poniżej wybranego przedziału należy pominąć (Krzyśko, 2009). Wysokość odsetka, który zostanie przyjęty zależy od typu przeprowadzanego badania. W przypadku badań technicznych zalecane jest około 90% aczkolwiek w badaniach społecznych lub psychologicznych wystarczającym uznaje się 60% wyjaśnianej wariancji. Natomiast badania z dziedziny ekonomii mieszczą się w przedziale obu powyższych typów badań.

Jeżeli badacz zamierza uporządkować obiekty według ich relatywnego znaczenia lub rangi powinien wybrać tylko pierwszą główną składową, która będzie pełnić rolę wskaźnika, w tym przypadku rankingu. Potrzebna do tego jest znajomość metod porządkowania liniowego. W przypadku porządkowania obiektów wielocechowych ważne jest zdefiniowanie jaki charakter mają zmienne. Zmienne mogą przyjmować charakter stymulant, destymulant, nominant (Dziechciarz, 2003).

Definicje tych rodzajów zmiennych są następujące:

- Stymulantą nazywa się cechę, której wzrost wartości świadczy o wzroście poziomu zjawiska złożonego
- Destymulantą nazywa się taką cechę w przypadku, której wzrost wartości doprowadza do spadku poziomu zjawiska złożonego
- Nominantą ustala się cechę, która ma określoną najkorzystniejszą wartość zwaną wartością nominalną. Wzrost wartości tej cechy do poziomu wartości nominalnej powoduje wzrost poziomu zjawiska złożonego. Natomiast wzrost cechy ponad wartość nominalną powoduje spadek poziomu zjawiska złożonego. Oznacza to, że do pewnego poziomu ma charakter stymulandy, zwanego wartością nominalną, a powyżej wartości nominalnej cecha przyjmuje charakter destymulandy.

Zamiana charakteru cechy:

- Zamiana destymulandy na stymulantę bazuje na pomnożeniu wartości destymulandy przez -1
- Zamiana nominandy na stymulantę: w pierwszej kolejności badacz musi ustalić wartość nominalną dla cechy (czyli wartość, która jest najkorzystniejsza dla badanego zjawiska):  $N_j$ . Następnie nominantę na stymulantę przekształca się w następujący sposób:

$$z_{ij} = \begin{cases} 1 & \text{dla } x_{ij} = N_j, \\ \frac{-1}{x_{ij} - N_j - 1} & \text{dla } x_{ij} < N_j, \\ \frac{1}{x_{ij} - N_j + 1} & \text{dla } x_{ij} > N_j, \end{cases} \quad (i = 1, 2, \dots, n) \quad (2.4)$$

gdzie:

$N_j$  – to wartość nominalna j-tej zmiennej,

$x_{ij}$  – i-ty obiekt j-tej zmiennej – nominandy,

$z_{ij}$  – i-ty obiekt j-tej zmiennej po zmienieniu na stymulantę.

Po powyższym przekształceniu stymulandy będą przyjmowały wartości z przedziału od 0 do 1 (Dziechciarz, 2003).

## 2.4 Model ekonometryczny

Model ekonometryczny jest próbą odwzorowania rzeczywistości za pomocą liczb. Model ekonometryczny za Z. Pawłowskim (1978) jest to konstrukcja formalna, która za pomocą pewnego równania, lub układu równań przedstawia zasadnicze powiązania występujące pomiędzy rozpatrywanymi zjawiskami ekonomicznymi. To równanie bądź układ równań nazywamy również równaniem regresji. Pomaga ono zrozumieć zmiany powstające w zjawiskach badanych. Równanie regresji ma postać matematyczną, w której zmienna, nazywana zmienną objaśnianą, zdefiniowana jest jako funkcja deterministyczna (przeważnie liniowa) najbardziej istotnych zmiennych, które określamy jako zmienne objaśniające. W takim równaniu występuje także zaburzenie losowe zwane również błędem losowym lub składnikiem losowym, którego podstawowym celem jest przedstawienie sumarycznego oddziaływania na zmienną objaśnianą wszystkich innych czynników, pominiętych w równaniu, ze względu na ich drugorzędne znaczenie dla opisu badanego związku (Górecki, 2010).

Podstawowe równanie regresji wygląda następująco:

$$y = f(X_1, X_2, \dots, X_k, \varepsilon) \quad (2.5)$$

gdzie:

$y$  – zmienna zależna (objaśniana),

$X_1, X_2, \dots$  – zmienne niezależne (objaśniające),

$\varepsilon$  – błąd losowy modelu.

Specyfikacja modelu to zdeterminowanie zmiennej objaśnianej przez model oraz zmiennych objaśniających, podjęcie decyzji obejmującej charakter zachodzących w modelu zależności. Ważnym czynnikiem również jest zdefiniowanie problemu ekonomicznego, z którym badacz się zmierzy budując model.

Pierwszym krokiem przy budowie modelu ekonometrycznego jest sprecyzowanie zmiennych objaśniających. Do kryterium decyzji należy merytoryczna znajomość badanego zjawiska. Powinno wybierać się takie czynniki (zmienne objaśniające), które znacznie oddziałują na kształt badanego zjawiska (zmienną objaśnianą). Takie zmienne nazywane są

zbiorem potencjalnych zmiennych objaśniających. Pierwszym warunkiem, aby zmienna (np.  $X_j$ ) mogła zostać wybrana do modelu jako objaśniająca, jest jej zróżnicowanie. Zmienną objaśniającą nie może być zmienna, której zbiór obserwacji nie jest zróżnicowany lub jest bardzo mało zróżnicowany. Do wyznaczenia poziomu zróżnicowania używa się współczynnika zmienności:

$$V_j = \frac{s_j}{\bar{x}_j}, \quad (2.6)$$

gdzie:

$s_j$  – odchylenie standardowe zmiennej  $X_j$ ,

$\bar{x}_j$  – średnia arytmetyczna zmiennej  $X_j$ .

Przeważnie wartością krytyczną współczynnika zmienności  $V^*$  jest 0,1. Zmienne spełniające nierówność  $V_j < V^*$  odrzuca się z modelu, ponieważ są za mało zróżnicowane. Trzeba je wyeliminować. Podstawowym narzędziem wyboru zmiennych do modelu jest jednak analiza korelacji (Dziechciarz, 2003).

Dobór zmiennych objaśniających do modelu został wykonany metodą analizy współczynników korelacji. Pierwszym krokiem przy doborze zmiennych było ustalenie wartości krytycznej współczynnika korelacji  $r^*$  ze wzoru:

$$r^* = \sqrt{\frac{(t^*)^2}{n - 2 + (t^*)^2}} \quad (2.7)$$

gdzie:

- $t^*$  – wartość statystyki t-Studenta
- $n$  – liczba obserwacji

Metoda doboru zmiennych jest następująca:

- Ze zbioru potencjalnych zmiennych objaśniających usuwa się każdą zmienną, dla której zachodzi nierówność:

$$|r_j| \leq r^*, \quad (2.8)$$

te zmienne są nieistotnie skorelowane ze zmienną objaśnianą.

- Następnie spośród niewykorzystanych zmiennych wybiera się zmienną  $X_h$ , dla której odpowiednio:

$$|r_{hi}| = \max_j |r_j|, \quad (2.9)$$

ponieważ zmienna  $X_h$  jest nośnikiem największego zasobu informacji o zmiennej zależnej (Dziechciarz, 2003).

- Kolejno ze zbioru potencjalnych zmiennych objaśniających, które nie zostały jeszcze w żaden sposób wykorzystane eliminuje się każdą zmienną, dla której:

$$|r_h| > r^*, \quad (2.10)$$

są one za mocno skorelowane ze zmienną niezależną  $X_h$ . Co za tym idzie odtwarzają w dużym stopniu informacje przekazane w zmiennej  $X_h$ .

Jeżeli zostały jakieś zmienne, należy przejść do punktu 2, w którym zostanie sprawdzone, czy potencjalna zmienna objaśniająca jest skorelowana wystarczająco ze zmienną objaśnianą. Postępowanie trzeba powtarzać aż do momentu wykorzystania wszystkich potencjalnych zmiennych objaśniających ze zbioru.

Następnym krokiem w modelowaniu ekonometrycznym jest estymacja parametrów strukturalnych modelu. Zakładając, że będzie miał on charakter liniowy można zapisać równanie regresji w następujący sposób:

$$Y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_m x_m + \varepsilon \quad (2.11)$$

Parametry modelu:  $\alpha_0, \alpha_1, \alpha_2, \alpha_m$  należy wyestymować klasyczną metodą najmniejszych kwadratów. Zastosowanie tej metody wiąże się z przyjęciem następujących założeń (Dziechciarz, 2003):

- w gronie zmiennych egzogenicznych nie występuje dokładna zależność liniowa, są one niezależne
- zmienne niezależne są wielkościami nielosowymi
- $r(X) = m + 1 \leq n$
- równanie regresji jest liniowe względem parametrów (lub można sprowadzić je do postaci liniowej)
- $E(\varepsilon) = 0$ , błąd losowy dla wszystkich obserwacji posiada wartość średnią równą zero



- $E(\varepsilon\varepsilon^T) = \sigma^2 I$  składnik losowy dla każdej obserwacji ma skończoną wariancję równą  $\sigma^2$ , jednakże kowariancje między różnymi składnikami losowymi muszą być równe zeru
- zaburzenie stochastyczne  $\varepsilon$  nie może być skorelowane ze zmiennymi objaśniającymi

Gdy zostaną spełnione wszystkie założenia, można przejść do estymacji parametrów. Przy doborze współczynników  $\alpha_j$  w Klasycznej Metodzie Najmniejszych Kwadratów suma kwadratów odchyleń estymowanych wartości zmiennej endogenicznej  $\hat{y}$  od wartości empirycznych  $y$  musi być minimalna.

$$S = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \alpha_0 - \alpha_1 x_{i1} - \alpha_2 x_{i2} - \dots - \alpha_m x_{im})^2 = \min \quad (2.12)$$

Kolejne obliczenia prowadzono w zapisie macierzowym. Model przyjmuje następującą postać:

$$y = X\alpha + \varepsilon, \quad (2.13)$$

gdzie:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{bmatrix} - \text{wektor obserwacji zmiennej objaśnianej},$$

$$X = \begin{bmatrix} 1 x_{11} & x_{12} & \dots & x_{1j} & \dots & x_{1m} \\ 1 x_{21} & x_{22} & \dots & x_{2j} & \dots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 x_{i1} & x_{i2} & \dots & x_{ij} & \dots & x_{im} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 x_{n1} & x_{n2} & \dots & x_{nj} & \dots & x_{nm} \end{bmatrix} - \text{macierz obserwacji zmiennych objaśniających},$$

$$\alpha = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_j \\ \vdots \\ \alpha_m \end{bmatrix} - \text{wektor parametrów strukturalnych modelu},$$

$$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_i \\ \vdots \\ \varepsilon_n \end{bmatrix} - \text{wektor składników losowych.}$$

Funkcja przyjmuje swoje minimum w następującym punkcie:

$$a = (X^T X)^{-1} X^T y \quad (2.14)$$

Tylko i wyłącznie jeżeli  $\det(X^T X) \neq 0$ .

Wektor  $a$  jest wektorem ocen parametrów strukturalnych  $\alpha$ . Po wcześniejszych obliczeniach można zapisać równanie w postaci:

$$\hat{y} = a_0 + a_1 x_1 + \dots + a_j x_j + \dots + a_m x_m \quad (2.15)$$

gdzie:

$\hat{y}$  – oznacza wartości teoretyczne zmiennej zależnej  $Y$ .

Wartości parametrów strukturalnych  $\alpha_j$  ( $j = 1, 2, \dots, m$ ) oznajmniają o ile jednostek zmieni się zmienna endogeniczna  $Y$ , jeśli zmienna  $X_j$  zmieni się o jedną jednostkę przy pozostałych czynnikach niezmiennych. Podczas szacowania powstały również reszty modelu. Są one różnicą pomiędzy wartościami rzeczywistymi a wartościami teoretycznymi zmiennej objaśnianej. Otrzymuje się je w następujący sposób:

$$e_i = y_i - \hat{y}_i (i = 1, 2, \dots, n) \quad (2.16)$$

Ocena wariancji zaburzenia stochastycznego ilustruje poniższy wzór :

$$s_e^2 = \frac{\sum_{i=1}^n e_i^2}{n - m - 1} = \frac{1}{n - m - 1} (y^T y - a^T X^T y) \quad (2.17)$$

Natomiast aby oszacować macierz wariancji i kowariancji ocen parametrów należy użyć

$$D^2(a) = s_e^2 (X^T X)^{-1} \quad (2.18)$$

W niniejszej macierzy na głównej przekątnej znajdują się wariancje ocen parametrów  $V(a_j)$ . Natomiast  $S(a_j)$  mówi o tym, o ile jednostek wartość oceny  $a_j$  odróżnia się od rzeczywistych wartości parametru  $a_j$ .

$$S(a_j) = \sqrt{V(a_j)} \quad (2.19)$$

Gdy model został już wstępnie wyestymowany można przejść do jego weryfikacji. Pierwszym krokiem w weryfikacji modelu ekonometrycznego jest współczynnik determinacji. Współczynnik determinacji  $R^2$  informuje, jaką część całkowitej zmienności zmiennej zależnej stanowi zmienność wyjaśniona przez model (Borkowski, 2004).

$$R^2 = \frac{\sum_{t=1}^n (\hat{y}_t - \bar{y})^2}{\sum_{t=1}^n (y_t - \bar{y})^2} \quad (2.20)$$

Powyższy współczynnik przyjmuje wartości z przedziału  $[0, 1]$ . Gdy jest on bliski 1 dopasowanie modelu do danych empirycznych jest bardzo dobre, jednakże gdy zbliża się do 0 to model jest słabo dopasowany do danych. Następnym narzędziem do weryfikacji modelu jest współczynnik zbieżności. Oznacza się go wzorem:

$$\varphi^2 = \frac{\sum_{t=1}^n e_t^2}{\sum_{t=1}^n (y_t - \bar{y})^2} \quad (2.21)$$

Współczynnik indeterminacji bo tak też jest nazywany określa jaka część całkowitej zmienności zmiennej endogenicznej nie została wyjaśniona przez model. Przyjmuje również wartości z przedziału  $[0, 1]$  jak współczynnik determinacji. Aczkolwiek jego interpretacja jest odwrotna. Im wyższa wartość miary tym model jest gorzej dopasowany do danych. Zatem im niższa miara współczynnika zbieżności tym model jest lepiej dopasowany do danych (Borkowski, 2004).

Kolejnym z etapów weryfikacji modelu ekonometrycznego jest testowanie istotności parametrów strukturalnych na podstawie rozkładu t-Studenta. Ma on na celu sprawdzenie czy parametry strukturalne zostały oszacowane z dostateczną precyzją oraz czy zmienne objaśniające stojące przy tych parametrach, istotnie oddziałują na zmienną objaśnianą. Stawia się następujące hipotezy badawcze:

$$H_0: a_j = 0$$

$$H_1: a_j \neq 0$$

Następnie oblicza się statystykę testową:

$$t_i = \frac{a_i}{S(a_i)} \quad (2.22)$$

Z rozkładem t-Studenta o  $n - m - 1$  stopniach swobody. Następnie dla danego poziomu istotności (przeważnie 0,05) oraz odpowiedniej liczbie stopni swobody z tablic rozkładu t-

Studenta odczytuje się wartość krytyczną  $t^*$ . W kolejnym kroku należy porównać statystykę  $t$  oraz wartość krytyczną  $t^*$ . Wnioskowanie jest następujące: jeśli  $|t_i| > t^*$  to należy odrzucić hipotezę zerową  $H_0$  na korzyść hipotezy alternatywnej  $H_1$ . Oznacza to, że zmienna objaśniająca  $X_i$  (przy której stoi parametr  $\alpha_i$ ) ma statystycznie istotny wpływ na zmienną objaśnianą  $Y$  (Borkowski, 2004). Następnym testem weryfikującym model może być test symetrii reszt t-Studenta. Sprawdza on, czy reszty modelu są symetryczne. Hipotezy badawcze w teście prezentują się następująco:

$H_0$ : reszty modelu są symetryczne

$H_1$ : reszty modelu nie są symetryczne

Statystykę testową oblicza się ze wzoru:

$$t = \frac{\frac{m}{n} - \frac{1}{2}}{\sqrt{\frac{\frac{m}{n}(1 - \frac{m}{n})}{n}}} \quad (2.23)$$

gdzie:

$m$  – liczba reszt dodatnich,

$n$  – liczba obserwacji.

Wartość krytyczną odczytuje się z tablic rozkładu t-Studenta dla  $m - n - 1$  stopni swobody. Kolejno porównuje się wartość statystyki testowej i wartości krytycznej  $t^*$ . Jeżeli  $|t| < t^*$  to nie ma podstaw do odrzucenia hipotezy zerowej, reszty w takim przypadku są symetryczne. Aczkolwiek w przeciwnym przypadku reszty nie są symetryczne. Dalszym etapem weryfikowania modelu jest test White'a, który sprawdza występowanie zjawiska heteroskedastyczności w modelu. Na podstawie testu White'a można określić czy reszty mają stałą wariancję.

$H_0$ : homoskedastyczność wariancji reszt - stałość wariancji

$H_1$ : heteroskedastyczność wariancji reszt - zmienność wariancji

Statystykę testową oblicza się ze wzoru:

$$\chi_N^2(White) = nR^2 \quad (2.24)$$

gdzie:

$R^2$  – współczynnik determinacji z danego modelu,

$n$  – liczba obserwacji.

Statystyka White'a ma rozkład  $\chi^2$  z  $N$  stopniami swobody, gdzie  $N$  oznacza liczbę parametrów modelu pomocniczego (z pominięciem wyrazu wolnego). Wartość krytyczną odczytuje się z tablic rozkładu  $\chi^2$ . Następnie należy porównać statystykę testową z wartością krytyczną. Gdy  $\chi_N^2(White) \leq \chi_N^2(\alpha)$  nie ma podstaw do odrzucenia  $H_0$ . W modelu występuje stałość wariancji składników losowych (homoskedastyczność). W przeciwnym przypadku w modelu występuje zmienność wariancji składników losowych (heteroskedastyczność) (White, 1980). Ostatnim testem weryfikującym model jest test Kołmogorowa-Smirnowa. Jest to test badający zgodność rozkładu empirycznego z rozkładami teoretycznymi (w szczególności z rozkładem normalnym). Stosowany jest dla dużych prób  $n \geq 100$ . Hipotezy badawcze prezentują się następująco:

$H_0$ : rozkład reszt jest rozkładem normalnym

$H_1$ : rozkład reszt jest różny od rozkładu normalnego

Aby obliczyć statystykę testową na początku trzeba stworzyć statystykę będącą dystrybuantą empiryczną:

$$F_n(x) = \begin{cases} 0 & \text{dla } x < x_1 \\ \frac{m}{n} & \text{dla } x_m \leq x < x_{m+1}, 1 \leq m \leq n-1 \\ 1 & \text{dla } x \geq x_n \end{cases} \quad (2.25)$$

Statystyka testowa mierzy największą różnicę dystrybuant:

$$D_n = \max_x |F_n(x) - F(x)| \quad (2.26)$$

Dla dużych prób korzysta się z granicznego rozkładu Kołmogorowa, wartość krytyczną odczytuje się następująco:

$$P(\sqrt{n}D_n > \lambda_\alpha) = \alpha \quad (2.27)$$

Jeśli  $D_n$  przekracza wartość krytyczną, wtedy należy przyjąć  $H_0$ , rozkład reszt jest wówczas rozkładem normalnym. Natomiast gdy  $D_n$  jest mniejsze niż wartość krytyczna należy odrzucić  $H_0$  na rzecz  $H_1$ . Wtedy rozkład reszt nie jest rozkładem normalnym.

## 2.5 Metoda wzorcu rozwoju Hellwiga

Kolejną metodą zastosowaną w pracy jest metoda wzorcu rozwoju Hellwiga. Jest to metoda porządkowania liniowego, na podstawie którego można dokonać podziału obiektów według określonych zasad w celu wyłonienia klas. Taksonomia jest dziedziną analizy wielowymiarowej, która zajmuje zasadami i regułami klasyfikacji obiektów wielocechowych (Namyślak, 2015). W toku postępowania porządkowania liniowego charakteryzuje się etapy procesu jak: zdefiniowanie charakteru zmiennych (stymulanty, destymulanty, nominanty), normalizacja zmiennych, wyznaczenie współrzędnych wzorca w przypadku agregacji wzorcowej, agregacja wzorcowa obiektów, klasyfikacja porangowanych obiektów i rozpoznanie typów rozwojowych (Bąk, 2016). Konstrukcję miary rozwoju Hellwiga rozpoczyna się od ujednolicenia charakteru zmiennych. Wszystkie zmienne muszą być stymulantami lub destymulantami. Następnym krokiem jest normalizacja danych a dokładniej standaryzacja, która przebiega następująco (Bąk, 2016):

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}, \quad (2.28)$$

gdzie:

$x_{ij}$  – obserwacja j-tej zmiennej obiektu i,

$\bar{x}_j$  – średnia arytmetyczna z próbki dla danej zmiennej,

$s_j$  – odchylenie standardowe z próbki dla danej zmiennej.

Kolejnym krokiem w porządkowaniu jest ustalenie współrzędnych dla wzorca. Proces wygląda następująco:

$$z_{0j} = \begin{cases} \max_i \{z_{ij}\} & \text{dla zmiennych stymulant} \\ \min_i \{z_{ij}\} & \text{dla zmiennych destymulant} \end{cases} \quad (2.29)$$

Po ustaleniu idealnego obiektu, który nazywa się wzorcem rozwoju, ponieważ ma „najlepsze” wartości dla każdej badanej cechy, dalszym krokiem w postępowaniu jest zbadanie odległości obiektów od wzorca rozwoju. Wykonuje się działania według następującego wzoru:

$$d_{i0} = \sqrt{\sum_{j=1}^m (z_{ij} - z_{0j})^2}. \quad (2.30)$$

Jest to odległość euklidesowa wektorów reprezentujących konkretny obiekt od taksonomicznego wzorca rozwoju w przestrzeni wielowymiarowej. Im niższa wartość  $d_{i0}$  tym dana jednostka jest w wyższym stopniu rozwinięta. Oznacza to, że jest korzystniejsza pod względem badanego zjawiska. Następnie dla każdego obiektu należy obliczyć wartości miernika syntetycznego, czyli tzw. miarę rozwoju (Namyślak, 2015).

$$m_i = 1 - \frac{d_{i0}}{d_0} \quad (2.31)$$

gdzie:

- $m_i$  – miara rozwoju dla i-tego obiektu,
- $d_0$  – odległość między wzorcem rozwoju i antywzorcem.

Na ogół miara rozwoju przyjmuje wartości z przedziału  $[0, 1]$ . Im wyższa wartość miary rozwoju, tym wyższy poziom rozwoju jednostki. W ostatnim etapie badania można dokonać podziału obiektów na klasy. Dokonuje się tego za pomocą kryteriów statystycznych z użyciem średniej arytmetycznej ( $\bar{m}$ ) oraz odchylenia standardowego ( $S_{mi}$ ) wartości miernika syntetycznego ( $m_i$ ). Przykładowo można dokonać podziału na trzy klasy (Namyślak, 2015).

- Klasa 1:  $m_i \geq m_s$ ,
- Klasa 2:  $m_r < m_i < m_s$ ,
- Klasa 3:  $m_i \leq m_r$ .

gdzie:

$$m_s = \bar{m} + S_{mi}, \quad (2.32)$$

$$m_r = \bar{m}_l - S_{mi}. \quad (2.33)$$

### **3. Ocena zależności zarobków i efektywności występów w meczach piłkarskich**

#### **3.1 Budowa rankingu z wykorzystaniem głównych składowych**

Na podstawie danych przedstawionych w tabeli załączniku przeprowadzono analizę głównych składowych. Jej celem było zredukowanie liczby wymiarów oraz stworzenie miernika do oceny czy zarobki mają wpływ na efektywność występów w meczach piłkarskich. Kilka pierwszych głównych składowych zawiera zdecydowaną większość informacji o badanym zjawisku, dostarczanych przez zmienne wejściowe, co pozwala na redukcję liczby głównych składowych przy możliwie małej stracie informacji wejściowych. Kolejne główne składowe są tak wyznaczane, aby wariancje kolejnych składowych były coraz mniejsze.

Pierwszym krokiem było zbudowanie modelu. Następnym sporządzenie rankingu oraz porównanie go z wynagrodzeniami. Do wyboru liczby głównych składowych posłużyły trzy kryteria:

- procent wyjaśnianej wariancji,
- liczba wartości własnych składowych przekraczająca wartość 1,
- wykres osypiska

W tabeli 3.1 przedstawione zostały wartości własne otrzymanych składowych, które mogą stanowić podstawę wyboru liczby głównych składowych. Na podstawie tabeli 3.1 zostanie dobrana liczba składowych za pomocą dwóch pierwszych kryteriów.



Tabela 3.1 Wartości własne badanych składowych

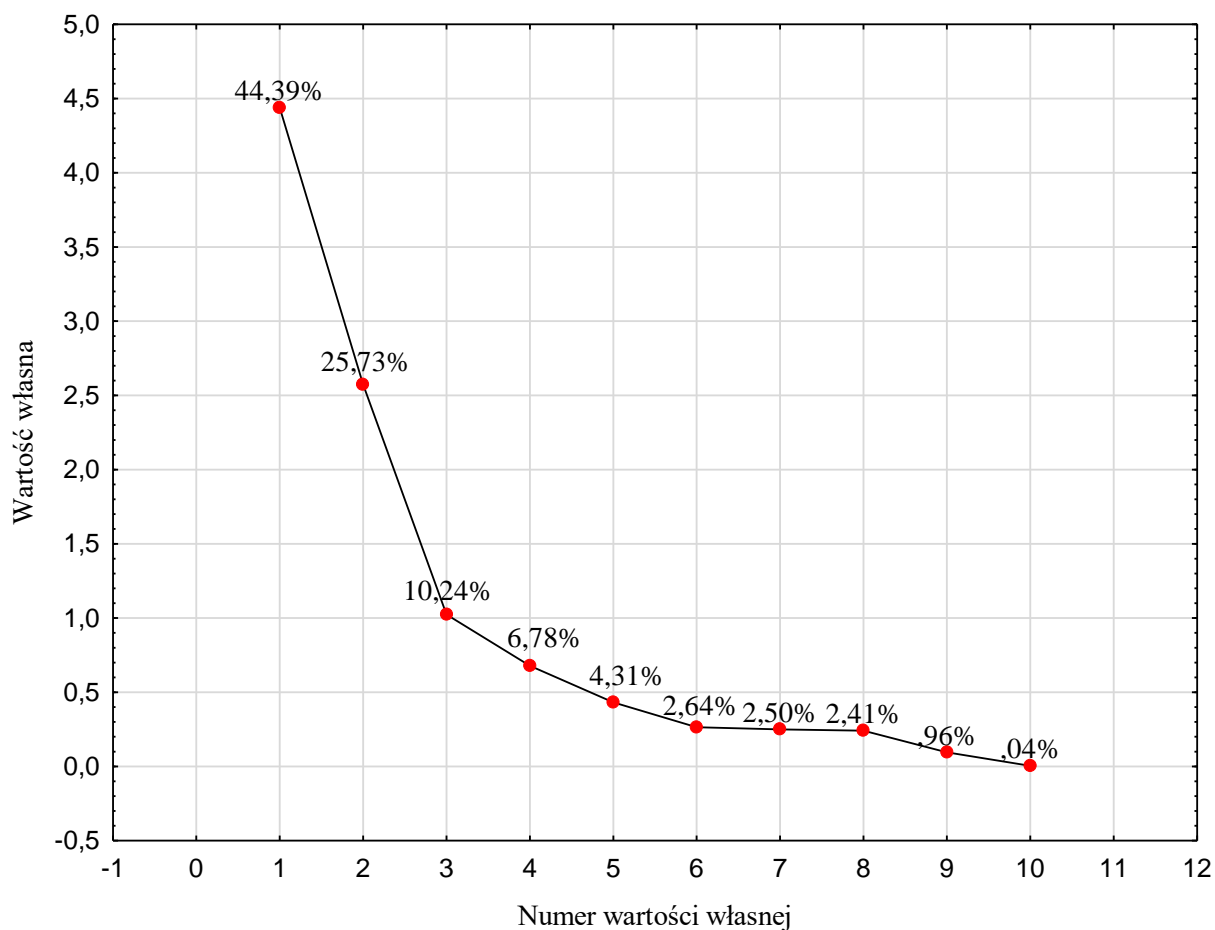
| Nr składowej | Wartość własna  | % wyjaśnienia wariancji | Skumulowane wartości własne | Skumulowany % wyjaśnienia wariancji |
|--------------|-----------------|-------------------------|-----------------------------|-------------------------------------|
| 1            | <b>4,438902</b> | 44,38902                | 4,43890                     | <b>44,3890</b>                      |
| 2            | <b>2,572914</b> | 25,72914                | 7,01182                     | <b>70,1182</b>                      |
| 3            | <b>1,023920</b> | 10,23920                | 8,03574                     | <b>80,3574</b>                      |
| 4            | 0,678469        | 6,78469                 | 8,71421                     | 87,1421                             |
| 5            | 0,431345        | 4,31345                 | 9,14555                     | 91,4555                             |
| 6            | 0,264078        | 2,64078                 | 9,40963                     | 94,0963                             |
| 7            | 0,249724        | 2,49724                 | 9,65935                     | 96,5935                             |
| 8            | 0,240916        | 2,40916                 | 9,90027                     | 99,0027                             |
| 9            | 0,095559        | 0,95559                 | 9,99583                     | 99,9583                             |
| 10           | 0,004172        | 0,04172                 | 10,00000                    | 100,0000                            |

Źródło: opracowanie własne.

W tabeli 3.1 na czerwono zaznaczone zostały wartości własne przekraczające 1, również zaznaczony został % wyjaśnienia wariancji. Pomogły one w wyborze liczby składowych na podstawie pierwszego oraz drugiego kryterium, które przedstawiają się następująco:

- kryterium subiektywne: wybrano 3 składowe, ponieważ w sumie wyjaśniają one zjawisko w 80,36%.
- kryterium obiektywne: wybrane zostały 3 składowe, dla których wartości własne są większe niż 1.

W celu doboru odpowiedniej liczby składowych posłużono się również wykresem osypiska, który przedstawiono na rysunku 3.1.



Rysunek 3.1 Wykres osypiska

Źródło: opracowanie własne.

Na podstawie wykresu osypiska z rysunku 3.1, analiza została przeprowadzona na podstawie 3 głównych składowych. Od punktu 3 wykres zaczyna się wyraźnie spłaszczać w porównaniu do pierwszych składowych, gdzie występuje stromy spadek.

Najbardziej odpowiednią liczbą głównych składowych w modelu są trzy główne składowe. Pierwsza składowa wyjaśnia badane zjawisko w 44,39%, druga w 25,73%, a trzecia w 10,24%. W tabeli 3.2 zostały przedstawione wektory własne macierzy korelacji składowych z pierwotnymi zmiennymi, które posłużą do stworzenia równań głównych składowych.

Tabela 3.2 Wektory własne macierzy korelacji

| Zmienna          | Składowa 1 | Składowa 2 | Składowa 3 |
|------------------|------------|------------|------------|
| Występy          | -0,397654  | -0,144682  | -0,035829  |
| Rozegrane minuty | -0,004220  | 0,044234   | -0,978569  |
| Bramki           | -0,367119  | 0,323139   | 0,101407   |
| Asysty           | -0,383904  | -0,080115  | -0,080640  |
| Podania celne    | -0,227943  | -0,519747  | 0,060909   |
| Podania          | -0,254174  | -0,504307  | 0,038886   |
| Strzały          | -0,392189  | 0,205512   | -0,036048  |
| Strzały celne    | -0,372943  | 0,249716   | 0,067382   |
| Popelnione faule | -0,255682  | -0,284545  | -0,113163  |
| Spalone          | -0,291309  | 0,395287   | 0,021562   |

Źródło: opracowanie własne.

Na tym etapie analizy zmienne zostały oznaczone następująco w celu ich uporządkowania oraz zaprezentowania wyników analizy głównych składowych:

- występy –  $X_1$
- rozegrane minuty –  $X_2$
- bramki –  $X_3$
- asysty –  $X_4$
- podania celne –  $X_5$
- podania –  $X_6$
- strzały –  $X_7$
- strzały celne –  $X_8$

– popełnione faule –  $X_9$

– spalone –  $X_{10}$

Następnie zostały wyznaczone trzy główne składowe. Pierwsza z nich posłuży bezpośrednio do stworzenia rankingu. Składowe prezentują się następująco.

Pierwsza główna składowa:  $Y_1 = -0,398 * X_1 - 0,0042 * X_2 - 0,367 * X_3 - 0,384 * X_4 - 0,228 * X_5 - 0,254 * X_6 - 0,392 * X_7 - 0,373 * X_8 - 0,256 * X_9 - 0,291 * X_{10}$

Druga główna składowa:  $Y_2 = -0,145 * X_1 + 0,044 * X_2 + 0,323 * X_3 - 0,080 * X_4 - 0,520 * X_5 - 0,504 * X_6 + 0,201 * X_7 + 0,250 * X_8 - 0,285 * X_9 + 0,340 * X_{10}$

Trzecia główna składowa:  $Y_3 = -0,036 * X_1 - 0,979 * X_2 + 0,101 * X_3 - 0,081 * X_4 + 0,061 * X_5 + 0,040 * X_6 - 0,036 * X_7 + 0,067 * X_8 - 0,113 * X_9 + 0,022 * X_{10}$

Tabela 3.3 przedstawia zasoby zmienności wspólnej na podstawie korelacji.

Tabela 3.3 Zasoby zmienności wspólnej korelacji

| Zmienna          | Składowa 1 | Składowa 2 | Składowa 3 |
|------------------|------------|------------|------------|
| Występy          | 0,701916   | 0,755775   | 0,757089   |
| Rozegrane minuty | 0,000079   | 0,005113   | 0,985616   |
| Bramki           | 0,598260   | 0,866920   | 0,877450   |
| Asysty           | 0,654215   | 0,670729   | 0,677387   |
| Podania celne    | 0,230636   | 0,925674   | 0,929473   |
| Podania          | 0,286773   | 0,941132   | 0,942680   |
| Strzały          | 0,682756   | 0,791424   | 0,792755   |
| Strzały celne    | 0,617391   | 0,777833   | 0,782481   |
| Popołnione faule | 0,290186   | 0,498504   | 0,511616   |
| Spalone          | 0,376689   | 0,778712   | 0,779188   |

Źródło: opracowanie własne.

Szczególną uwagę warto zwrócić na zmienne  $X_2$ ,  $X_5$  oraz zmienną  $X_6$ . Są one wyjaśniane przez model odpowiednio w najwyższym stopniu. Wartymi uwagi zmiennymi są również zmienna  $X_9$  oraz zmienna  $X_4$ . Są one najslabiej wyjaśniane przez model. Jako jedyne ze zmiennych wyjaśniane są poniżej 70%.

Następnym krokiem w analizie było zbudowanie rankingu. Pierwsza główna składowa wyjaśnia tylko 44% wariancji, co oznacza, że mało zmienności zostało wyjaśnionej za jej pomocą. Jest to wynik nie pozwalający bezgranicznie wierzyć we wnioski płynące z wyników rankingu. Mimo tego, został stworzony ranking, który będzie bardzo pomocny w późniejszym etapie w celach porównawczych z wynikami otrzymanymi metodą wzorca rozwoju. Pozwoli on na zweryfikowanie wstępnych wniosków płynących z analizy.

W tabeli 3.4 zostały zestawione zmienne wybrane do przeprowadzenia rankingu oraz przyporządkowany im charakter (stymulant bądź destymulant).

Tabela 3.4 Zmienne wybrane do przeprowadzenia rankingu

| Zmienna          | Charakter    |
|------------------|--------------|
| Występy          | stymulanta   |
| Rozegrane minuty | stymulanta   |
| Bramki           | stymulanta   |
| Asysty           | stymulanta   |
| Podania celne    | stymulanta   |
| Podania          | stymulanta   |
| Strzały          | stymulanta   |
| Strzały celne    | stymulanta   |
| Spalone          | destymulanta |

Źródło: opracowanie własne.

W rozpatrywanym modelu 8 zmiennych to stymulanty, jedna to destymulanta. Zmienne te zostały wybrane, ponieważ są one najbardziej istotnymi statystykami w grze danego zawodnika. Zależy od nich w dużym stopniu wynik meczu piłkarskiego.

W tabeli 3.5 zaprezentowano ładunki czynnikowe zmiennych. Są one niezbędne do przeprowadzenia procesu, którego zwieńczeniem jest ranking.

Tabela 3.5 Współrzędne czynnikowe zmiennych

| Zmienna          | Ładunek<br>czynnikowy |
|------------------|-----------------------|
| Występy          | -0,837804             |
| Rozegrane minuty | -0,008891             |
| Bramki           | -0,773473             |
| Asysty           | -0,808836             |
| Podania celne    | -0,480246             |
| Podania          | -0,535512             |
| Strzały          | -0,826291             |
| Strzały celne    | -0,785742             |
| Spalone          | -0,613750             |

Źródło: opracowanie własne.

Na potrzebę rankingu ładunki czynnikowe z tabeli 3.5 będące stymulantami w przypadku ujemnych wartości głównych składowych zmieniono znak na „+”. Dla zmiennej „Spalone” będącej jedyną destymulantą znaku nie zmieniono.

W tabeli 3.6 przedstawiono skorygowane wartości ładunków czynnikowych zmiennych, służące bezpośrednio do stworzenia rankingu na podstawie kombinacji liniowej.

Tabela 3.6 Skorygowane ładunki czynnikowe zmiennych

| Zmienna          | Skorygowane<br>ładunki<br>czynnikiowe |
|------------------|---------------------------------------|
| Występy          | 0,837804                              |
| Rozegrane minuty | 0,008891                              |
| Bramki           | 0,773473                              |
| Asysty           | 0,808836                              |
| Podania celne    | 0,480246                              |
| Podania          | 0,535512                              |
| Strzały          | 0,826291                              |
| Strzały celne    | 0,785742                              |
| Spalone          | -0,613750                             |

Źródło: opracowanie własne.

Na podstawie poprawionych wartości z tabeli 3.6 oraz zestandaryzowanych danych z załącznika za pomocą kombinacji liniowej wyznaczono wartości umożliwiające wykonanie rankingu. Pozwolił on na uszeregowanie zawodników malejąco względem jakości gry. Należy podkreślić, że w kolejnym kroku na podstawie rankingu, dokonano oceny wpływu zarobków na efektywność występów w meczach piłkarskich.

Tabela 3.7 przedstawia 10 najlepszych zawodników według przeprowadzonego rankingu oraz ich zarobki.

Tabela 3.7 Zestawienie 10 najlepszych zawodników według rankingu

| Miejsce w rankingu | Miejsce pod względem zarobków | Zarobki roczne | Zawodnik          |
|--------------------|-------------------------------|----------------|-------------------|
| 1                  | 1                             | \$80 023 053   | Lionel Messi      |
| 2                  | 23                            | \$15 396 581   | Eden Hazard       |
| 3                  | 15                            | \$19 895 496   | Paul Pogba        |
| 4                  | 30                            | \$13 721 032   | Mohamed Salah     |
| 5                  | 2                             | \$64 923 873   | Cristiano Ronaldo |
| 6                  | 3                             | \$51 845 682   | Antoine Griezmann |
| 7                  | 76                            | \$6 283 503    | Suso              |
| 8                  | 86                            | \$5 544 999    | Florian Thauvin   |
| 9                  | 94                            | \$4 734 115    | Memphis Depay     |
| 10                 | 11                            | \$21 084 114   | Kylian Mbappé     |

Źródło: opracowanie własne.

Szczególną uwagę warto zwrócić na pierwsze miejsce w rankingu w tabeli 3.7. Okazało się, że najlepiej zarabiający piłkarz Lionel Messi jest również pierwszym w rankingu. Warto również wskazać pozycje takich piłkarzy jak Cristiano Ronaldo oraz Antoine Griezmann, którzy będąc w pierwszej trójce według zarobków znajdują się w pierwszej szóstce rankingu pod względem efektywności.

Interesującym wynikiem jest wystąpienie w dziesiątce najlepszych takich piłkarzy jak Suso, Florian Thauvin czy Memphis Depay. Znajdują się oni relatywnie nisko w rankingu zarobków. Dzięki efektywnej grze udało im się dostać do czołówki.

Tabela 3.8 przedstawia 10 najgorszych zawodników według przeprowadzonego rankingu oraz ich zarobki.



Tabela 3.8 Zestawienie 10 najgorszych zawodników według rankingu

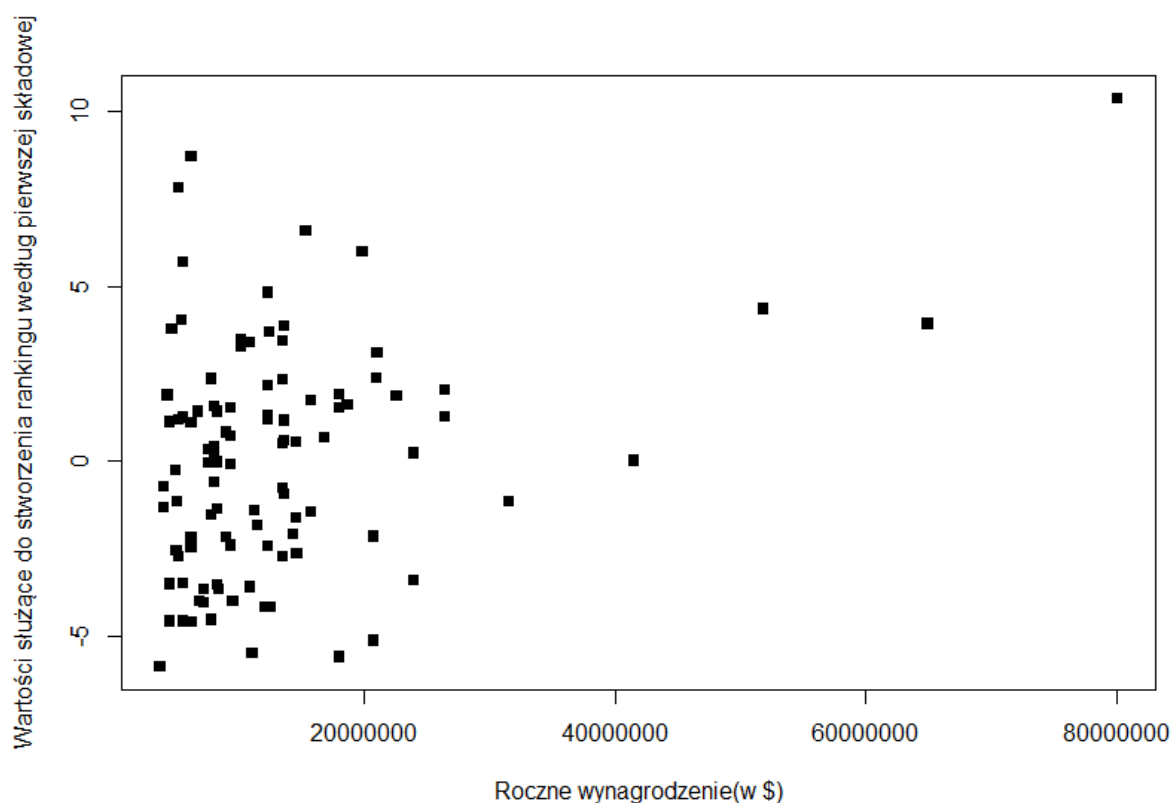
| Miejsce w rankingu | Miejsce pod względem zarobków | Zarobki roczne | Zawodnik           |
|--------------------|-------------------------------|----------------|--------------------|
| 91                 | 14                            | \$20 815 069   | Gonzalo Higuaín    |
| 92                 | 38                            | \$12 565 874   | Douglas Costa      |
| 93                 | 71                            | \$7 329 622    | Javier Pastore     |
| 94                 | 45                            | \$12 172 307   | Vitolo             |
| 95                 | 48                            | \$11 136 366   | Álvaro Morata      |
| 96                 | 77                            | \$6 283 503    | Antonio Candreva   |
| 97                 | 17                            | \$18 083 731   | Diego Costa        |
| 98                 | 69                            | \$7 889 437    | Arjen Robben       |
| 99                 | 84                            | \$5 635 474    | Yevhen Konoplyanka |
| 100                | 100                           | \$3 787 518    | Stevan Jovetic     |

Źródło: opracowanie własne.

Ranking w tabeli 3.8 jasno ukazuje piłkarzy, którzy są mocno przepełceni i nie spisują się na zielonej murawie odpowiednio dobrze w porównaniu do zarabianych pieniędzy. W najgorszej dziesiątce znaleźli się tacy piłkarze jak Gonzalo Higuaín czy Diego Costa, którzy znajdują się w czołowej dwudziestce pod względem zarobków.

W najgorszej dziesiątce znajduje się aż pięciu piłkarzy, którzy są w pięćdziesiątce najlepiej zarabiających. Pozwala to wysunąć wniosek, że niekoniecznie pieniądze mają wpływ na ich występy na boisku. Szczególnie widać to po zestawieniu pierwszej dziesiątki, w której znaleźli się piłkarze z końca listy płac, będący jednak mocnymi punktami swoich zespołów pod względem statystyk.

Kolejną wykonaną czynnością był wykres korelacyjny pomiędzy wartościami służącymi do stworzenia rankingu według pierwszej składowej a wynagrodzeniem, który znajduje się na rysunku 3.2.



Rysunek 3.2 Wykres korelacyjny pomiędzy wynagrodzeniem a wartościami służącymi do stworzenia rankingu według pierwszej składowej

Źródło: opracowanie własne.

Na podstawie rysunku 3.2 można zauważyć, że korelacja pomiędzy wartościami rankingu a wynagrodzeniem jest dosyć niska. Punkty są rozproszone. Następnym krokiem było obliczenie współczynnika korelacji Pearsona w celu potwierdzenia wcześniejszych obserwacji. Korelacja między wartościami rankingu a wynagrodzeniem jest przeciętna, wynosi zaledwie 0,3423. Wartości z rankingu są przeciętnie skorelowane z rocznym wynagrodzeniem piłkarzy. Następnym krokiem będzie zweryfikowanie czy korelacja jest istotna.

$H_0$  : korelacja pomiędzy dwoma cechami jest równa zero

$H_1$  : korelacja pomiędzy dwoma cechami jest różna od zera

Tabela 3.9 Wartość statystyki testowej oraz wartość krytyczna t odczytana z tablic rozkładu t-Studenta

| Element testu       | Wartość |
|---------------------|---------|
| Statystyka testowa  | 3,6068  |
| Wartość krytyczna t | 1,984   |

Źródło: opracowanie własne.

Na podstawie tabeli 3.9 statystyka testowa znajduje się w obszarze krytycznym. Co za tym idzie odrzucono hipotezę zerową na rzecz hipotezy alternatywnej. Korelacja jest istotna, choć słaba. Wnioski wysnute podczas analizy wydają się słuszne.

Warto wspomnieć, że do rankingu posłużyły wartości dotyczące pierwszej składowej, która nie wyjaśniła zbyt dużo ogólnej wariancji. Aczkolwiek ranking nadal był użyteczny i posłużył w celach porównawczych.

## 3.2 Badanie zależności za pomocą modelu ekonometrycznego

Jako następna metoda badania wpływu zarobków na efektywność występów piłkarzy posłużył model ekonometryczny. Celem tych działań było sprawdzenie wielkości wpływu konkretnych czynników na zarobki piłkarzy w sezonie 2018/2019.

Pierwszym krokiem podjętym w analizie było wytypowanie zmiennej objaśnianej oraz zmiennych objaśniających. Jako zmienną objaśnianą wybrano pensję roczną w sezonie 2018/2019. Zmienną objaśnianą oznaczono  $Y$ .

Jako potencjalne zmienne objaśniające zmienne przyjęto następujące zmienne:

- występy –  $X_1$
- rozegrane minuty –  $X_2$
- bramki –  $X_3$
- asysty –  $X_4$

- podania celne –  $X_5$
- podania –  $X_6$
- strzały –  $X_7$
- strzały celne –  $X_8$
- popełnione faule –  $X_9$
- spalone –  $X_{10}$

Pierwszym krokiem przy doborze zmiennych do modelu metodą analizy współczynników korelacji było ustalenie wartości krytycznej współczynnika korelacji  $r^*$  ze wzoru (2.7).

W tabeli 3.10 zostały umieszczone wartości statystyki t-Studenta oraz współczynnika korelacji  $r^*$ , które zostały obliczone w programie Excel oraz posłużyły do wstępnego doboru zmiennych.

Tabela 3.10 wartość t-Studenta oraz współczynnika korelacji  $r^*$

| Element testu                                   | Wartość  |
|---|----------|
| Statystyka t-Studenta                           | 1,984467 |
| Wartość krytyczna współczynnika korelacji $r^*$ | 0,19655  |

Źródło: opracowanie własne.

W tabeli 3.11 zaprezentowano wektor korelacji pomiędzy zmiennymi objaśniającymi a zmienną objaśnianą. Na jego podstawie zostały odrzucone zmienne objaśniające, których współczynnik korelacji jest niższy niż obliczone wcześniej  $r^*$ , również została wybrana pierwsza zmienna do modelu, która jest najbardziej skorelowana ze zmienną objaśnianą. Następnie tabela 3.11 posłużyła w doborze kolejnych zmiennych objaśniających.

Tabela 3.11 Wektor współczynników korelacji  $R_{0i}$  pomiędzy zmienną objaśnianą a potencjalnymi zmiennymi objaśniającymi

| Zmienna          | 2018-19 Pensja roczna |
|------------------|-----------------------|
| Występy          | 0,116                 |
| Rozegrane minuty | -0,113                |
| Bramki           | 0,451                 |
| Asysty           | 0,335                 |
| Podania celne    | 0,200                 |
| Podania          | 0,213                 |
| Strzały          | 0,435                 |
| Strzały celne    | 0,370                 |
| Popelnione faule | -0,072                |
| Spalone          | 0,231                 |

Źródło: opracowanie własne.

Wnioskując na podstawie wartości w tabeli bezzwłocznie z modelu odrzucono zmienne „Występy”, „Rozegrane minuty” oraz „Popelnione faule”, ponieważ wartości bezwzględne współczynnika korelacji ze zmienną objaśnianą są mniejsze niż  $r^*$ . Kolejno przyjęto do modelu zmienną „Bramki”, która charakteryzuje się najwyższym współczynnikiem korelacji wobec zmiennej objaśnianej.

Następnie w celu podjęcia decyzji dotyczącej kolejnych zmiennych potrzebna była tabela 3.12 z której odczytane zostały wartości współczynników korelacji pozostałych zmiennych objaśniających ze zmienną  $X_5$  czyli „Bramki”.

Tabela 3.12 Macierz współczynników korelacji zmiennych objaśniających

| Zmienna         | X <sub>1</sub> | X <sub>2</sub> | X <sub>3</sub> | X <sub>4</sub> | X <sub>5</sub> | X <sub>6</sub> | X <sub>7</sub> | X <sub>8</sub> | X <sub>9</sub> | X <sub>10</sub> |
|-----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|-----------------|
| X <sub>1</sub>  | 1,00           | -0,04          | 0,45           | 0,60           | 0,52           | 0,57           | 0,61           | 0,51           | 0,52           | 0,39            |
| X <sub>2</sub>  | -0,04          | 1,00           | -0,06          | 0,03           | -0,10          | -0,08          | 0,02           | -0,04          | -0,01          | -0,01           |
| X <sub>3</sub>  | 0,45           | -0,06          | 1,00           | 0,54           | -0,04          | 0,01           | 0,76           | 0,84           | 0,16           | 0,71            |
| X <sub>4</sub>  | 0,60           | 0,03           | <b>0,54</b>    | 1,00           | 0,46           | 0,52           | 0,58           | 0,58           | 0,25           | 0,25            |
| X <sub>5</sub>  | 0,52           | -0,10          | -0,04          | 0,46           | 1,00           | 0,99           | 0,12           | 0,03           | 0,41           | -0,23           |
| X <sub>6</sub>  | 0,57           | -0,08          | 0,01           | 0,52           | 0,99           | 1,00           | 0,18           | 0,10           | 0,42           | -0,19           |
| X <sub>7</sub>  | 0,61           | 0,02           | <b>0,76</b>    | 0,58           | 0,12           | 0,18           | 1,00           | 0,71           | 0,25           | 0,62            |
| X <sub>8</sub>  | 0,51           | -0,04          | <b>0,84</b>    | 0,58           | 0,03           | 0,10           | 0,71           | 1,00           | 0,22           | 0,56            |
| X <sub>9</sub>  | 0,52           | -0,01          | 0,16           | 0,25           | 0,41           | 0,42           | 0,25           | 0,22           | 1,00           | 0,06            |
| X <sub>10</sub> | 0,39           | -0,01          | <b>0,71</b>    | 0,25           | -0,23          | -0,19          | 0,62           | 0,56           | 0,06           | 1,00            |

Źródło: opracowanie własne.

Na podstawie tabeli 3.12 odrzucone zostały zmienne „Asysty”, „Strzały”, „Strzały celne” oraz „Spalone”, ponieważ ich korelacja ze zmienną „Bramki” wybraną już do modelu jest większa niż przyjęte  $r^*$ . Zostały one w tabeli pogrubione. Co za tym idzie, pozostały dwie zmienne, które należało wybrać lub odrzucić. Dlatego powrócono do tabeli 3.11 sprawdzając, która z pozostałych zmiennych „Podania” oraz „Podania celne” jest bardziej skorelowana ze zmienną objaśnianą. Porównując wartości współczynników korelacji powyższych zmiennych ze sobą ostatecznie do modelu wybrano zmienną „Podania”, ponieważ zmienne były ze sobą zbyt mocno skorelowane.

Dobór zmiennych został ukończony, co pozwoliło przejść do estymacji parametrów modelu. Za pomocą funkcji „Regresja” w module „Analiza danych” w programie Excel parametry modelu zostały wyestymowane. Postać modelu prezentuje się następująco:

$$Y = 3078524,65 + 698263,18X_3 + 4197,13X_6$$

Po oszacowaniu parametrów przystąpiono do weryfikacji modelu. W pierwszym etapie oceniono dopasowanie modelu do danych. Użyto do tego współczynnika  $R^2$  oraz współczynnika zbieżności  $\phi^2$ . Na ich podstawie zbadano procent wyjaśniania danych przez model. Zaprezentowane są one w tabeli 3.13

Tabela 3.13 Wartości miar oceny dopasowania modelu do danych

| Miara dopasowania | Wartość |
|-------------------|---------|
| $R^2$             | 0,2482  |
| $\phi^2$          | 0,7518  |

Źródło: opracowanie własne.

Skorygowany współczynnik  $R^2$  wyniósł 0,2482 oznacza to, że w 24,82% zmienności jest wyjaśniane przez model. Natomiast współczynnik zbieżności  $\phi^2$  przyjmuje wartość 0,7518, co świadczy o 75,18% zmienności niewyjaśnianej przez model. Wynik ten jest mocno niezadowolający. Oznacza, że model nie jest dobrze dopasowany. Mimo tego, można z niego wyciągnąć pewne wnioski.

Drugim krokiem w weryfikacji modelu było wykonanie testu t-Studenta przeprowadzanego w celu sprawdzenia istotności parametrów strukturalnych. Test został przeprowadzony dla obu parametrów przy zmiennych objaśniających. W obu przypadkach hipoteza zerowa zakładała, że parametr nie różni się istotnie od zera (czyli że badana zmienna objaśniająca nieistotnie wpływa na zmienną objaśnianą), natomiast hipoteza alternatywna wskazywała, że parametr różni się istotnie od zera (czyli że badana zmienna objaśniająca wpływa istotnie na zmienną objaśnianą).

$H_0$  : zmienna przy parametrze  $\alpha_i$  nie wpływa istotnie na zmienną objaśnianą

$H_1$  : zmienna przy parametrze  $\alpha_i$  wpływa istotnie na zmienną objaśnianą

W tabeli 3.14 zaprezentowano elementy związane z testowaniem istotności parametrów.

Tabela 3.14 Wartość statystyk t-Studenta dla zmiennych objaśniających oraz p-value

| Zmienna | statystyka t-Studenta | p-value     |
|---------|-----------------------|-------------|
| Bramki  | 5,11495               | 0,00000158  |
| Podania | 2,39443               | 0,018568136 |

Źródło: opracowanie własne.

Wnioskując na podstawie tabeli 3.14, odrzucona została hipoteza zerowa na rzecz hipotezy alternatywnej, która mówi, że zmienna objaśniająca wpływa istotnie na zmienną objaśnianą. Wartości statystyki t-Studenta odczytana z tabeli 3.14 dla obu zmiennych objaśniających są wyższe niż wartość krytyczna statystyki t-Studenta. Również p-value dla obu zmiennych objaśniających znajduje się znacznie poniżej poziomu istotności 0,05 co potwierdza, że parametry istotnie różnią się od zera i zmienne wpływają istotnie na zmienną objaśnianą.

Następnym wykonanym testem był test t-Studenta symetrii reszt.. Stosując go sprawdzono czy reszty modelu są symetryczne. Test został wykonany na resztach modelu na poziomie istotności 0,05.

$H_0$  : reszty modelu są symetryczne

$H_1$  : reszty modelu nie są symetryczne

Tabela 3.15 Wartość statystyki t-Studenta oraz wartość krytyczna odczytana z tablic rozkładu t-Studenta

| Element testu  | Wartość |
|--|---------|
| Statystyka t-Studenta                                    | 2,6926  |
| Wartość krytyczna odczytana z tablic rozkładu t-Studenta | 1,984   |

Źródło: opracowanie własne.

Na bazie tabeli 3.15 widać, że wartość statystyki t-Studenta jest większa niż wartość krytyczna, więc leży w obszarze krytycznym. Na tej podstawie odrzucono hipotezę zerową na rzecz hipotezy alternatywnej. Reszty modelu nie są symetryczne. Najprawdopodobniej ma miejsce systematyczny błąd. Kolejnym z kroków weryfikacji modelu był test White'a, który sprawdza występowanie zjawiska heteroskedastyczności w modelu. Test pomógł określić czy reszty mają stałą wariancję. Został on wykonany na poziomie istotności 0,05. Hipotezy testu przedstawiają się następująco:

$H_0$ : homoskedastyczność wariancji reszt - stałość wariancji

$H_1$ : heteroskedastyczność wariancji reszt - zmienność wariancji



W tabeli 3.16 zaprezentowano elementy niezbędne do przeprowadzenia testu White'a.

Tabela 3.16 Wartość statystyki testowej testu White'a oraz wartość teoretyczna odczytana z tablic rozkładu Chi-kwadrat

| Element testu   | Wartość |
|---|---------|
| Statystyka testowa testu White'a                            | 24,82   |
| Wartość teoretyczna odczytana z tablic rozkładu Chi-kwadrat | 5,9915  |

Źródło: opracowanie własne.

Wnioskując na podstawie tabeli 3.16 wartość statystyki testowej testu White'a jest większa niż wartość teoretyczna odczytana z tablic. Zatem odrzucono hipotezę zerową na rzecz hipotezy alternatywnej. Występuje heteroskedastyczność wariancji reszt czyli zmienność wariancji. Jest to następny test, potwierdzający, że model nie spełnia założeń.

Ostatnim krokiem weryfikującym model jest analiza rozkładu reszt modelu. Aby sprawdzić, czy reszty modelu mają rozkład normalny posłużono się testem Kołmogorowa-Smirnowa na normalność reszt. Hipoteza zerowa zakłada, że reszty modelu posiadają rozkład normalny, zaś hipoteza alternatywna zakłada, że reszty modelu rozkładu normalnego nie posiadają.

$H_0$ : rozkład reszt jest rozkładem normalnym

$H_1$ : rozkład reszt jest różny od rozkładu normalnego

W tabeli 3.17 znajdują się wartości statystyki  $\lambda$  Kołmogorowa-Smirnowa oraz wartość krytyczna  $\lambda$ , które posłużyły do zweryfikowania hipotezy.

Tabela 3.17 Elementy testu Kołmogorowa-Smirnowa

| Element testu                  | Wartość statystyki |
|--------------------------------|--------------------|
| $\lambda$ Kołmogorowa-Smirnowa | 0,17045            |
| Wartość krytyczna $\lambda^*$  | 1,36               |

Źródło: opracowanie własne.

Na podstawie tabeli 3.17 nie ma podstaw do odrzucenia hipotezy zerowej. Wartość statystyki  $\lambda$  Kołmogorowa-Smirnowa jest znacznie mniejsza niż wartość krytyczna  $\lambda$ . Przyjęto, że reszty modelu są zgodne z rozkładem normalnym.

Po weryfikacji modelu następnym krokiem jest interpretacja oszacowań parametrów strukturalnych:

- Bramki – wraz ze wzrostem liczby strzelonych bramek o 1 przeciętna pensja roczna piłkarza zwiększy się o 698263,18\$ przy pozostałych czynnikach niezmiennych.
- Podania – wraz ze wzrostem podań o 1 przeciętna pensja roczna piłkarza zwiększy się o 4197,13\$ przy pozostałych czynnikach niezmiennych.

Na podstawie przeprowadzonej analizy można wnioskować, że strzelone bramki mają ogromny wpływ na pensje piłkarzy nożnych, znacznie większy niż wykonane podania na przestrzeni sezonu. Dobrze znanym faktem jest, że gole od zawsze były w piłce nożnej gloryfikowane. Nic dziwnego, ponieważ ich liczba wpływa bezpośrednio na wynik spotkania. Co za tym idzie mają również największy wpływ nie tylko na wynagrodzenie zawodników ale również na wszelakie rankingi oraz przyznawane nagrody za indywidualne osiągnięcia.

### 3.3. Analiza na podstawie wzorca rozwoju

Ostatnią metodą stosowaną w pracy była metoda wzorca rozwoju. Pomogła ona również sklasyfikować piłkarzy pod względem ich efektywności. Wskazana metoda pozwoliła na porównanie piłkarzy pod względem efektywności między sobą a następnie porównanie ich zarobków do przyporządkowanej klasy. Celem metody było uporządkowanie liniowe obiektów, w tym przypadku piłkarzy od najgorszego do najlepszego. Następnie zestawienie wyników z metodą analizy głównych składowych.

Pierwszym krokiem w metodzie wzorca rozwoju było ujednolicenie charakteru zmiennych. Wszystkie zmienne muszą być stymulantami lub destymulantami. W tabeli 3.18 zaprezentowano zmienne wraz z ich przyporządkowaniem.

Tabela 3.18 Zmienne podzielone na stymulanty i destymulanty

| Zmienna          | Oznaczenie | Charakter    |
|------------------|------------|--------------|
| Występy          | $X_1$      | stymulanta   |
| Rozegrane minuty | $X_2$      | stymulanta   |
| Bramki           | $X_3$      | stymulanta   |
| Asysty           | $X_4$      | stymulanta   |
| Podania celne    | $X_5$      | stymulanta   |
| Podania          | $X_6$      | stymulanta   |
| Strzały          | $X_7$      | stymulanta   |
| Strzały celne    | $X_8$      | stymulanta   |
| Popelnione faule | $X_9$      | destymulanta |
| Spalone          | $X_{10}$   | destymulanta |

Źródło: opracowanie własne.

Jak wskazują informacje zamieszczone w tabeli 3.18, w zbiorze zidentyfikowano osiem stymulant oraz dwie destymulanty. To przyporządkowanie było kluczowe w kolejnym kroku. Następnym działaniem w metodzie wzorca rozwoju było wyznaczenie wartości wzorca oraz antywzorca dla każdej zmiennej.

Tabela 3.19 przedstawia wartości wzorca i antywzorca dla każdej ze zmiennych. Posłuży ona do wyznaczenia odległości między poszczególnymi obiektami (piłkarzami) a wzorcem  $d_{i0}$  oraz między wzorcem i antywzorcem  $d_0$ . Im bardziej podobny do wzorca jest obiekt, tym lepszy poziom zjawiska dla tego obiektu.

Tabela 3.19 Wartości wzorca i antywzorca

| Zmienna          | Oznaczenie | Wzorzec $z_0$ | Antywzorzec $z_{-0}$ |
|------------------|------------|---------------|----------------------|
| Występy          | $X_1$      | 1,4535        | -2,7882              |
| Rozegrane minuty | $X_2$      | 3,9871        | -0,4175              |
| Bramki           | $X_3$      | 3,7290        | -1,1597              |
| Asysty           | $X_4$      | 3,2534        | -1,3418              |
| Podania celne    | $X_5$      | 3,0058        | -1,4551              |
| Podania          | $X_6$      | 2,7089        | -1,5648              |
| Strzały          | $X_7$      | 2,9919        | -1,4249              |
| Strzały celne    | $X_8$      | 3,9027        | -1,3313              |
| Popelnione faule | $X_9$      | -1,5125       | 3,8045               |
| Spalone          | $X_{10}$   | -1,2074       | 3,1214               |

Źródło: opracowanie własne.

Należy podkreślić na podstawie tabeli 3.19, że wszystkie wartości zostały zestandaryzowane oraz znajdują się w przedziale  $[-3,4]$ . Następnym etapem jest wyznaczenie dla każdego obiektu miary rozwoju według wzoru (2.31).

Wartość miary rozwoju znajduje się w przedziale  $[0,1]$ . Warto zaznaczyć, że im wyższa miara rozwoju tym wyższy poziom zjawiska, w tym przypadku tym lepiej piłkarz wypada przekrojowo pod względem statystyk w sezonie. Konsekwentnym krokiem było utworzenie 10 klas oraz przyporządkowanie 100 piłkarzy, po 10 do każdej klasy. Następnie przedstawiono zestawienie pierwszej oraz ostatniej klasy czyli odpowiednio dziesięciu najlepszych oraz porównano je z rankingiem według analizy głównych składowych znajdujących się w tabeli 3.7 oraz 3.8. W tabeli 3.20 znajdują się piłkarze z pierwszej, najlepszej klasy. Sklasyfikowani od najlepszego do dziesiątego najlepszego z kolei.

Tabela 3.20 Klasa 1 ustalona według metody wzorca rozwoju

| Zawodnik                  | Miejsce pod względem wzorca rozwoju | Miejsce pod względem zarobków | Zarobki roczne |
|---------------------------|-------------------------------------|-------------------------------|----------------|
| Lionel Messi              | 1                                   | 1                             | \$80 023 053   |
| Suso                      | 2                                   | 76                            | \$6 283 503    |
| Eden Hazard               | 3                                   | 23                            | \$15 396 581   |
| Hakan Calhanoglu          | 4                                   | 88                            | \$5 236 252    |
| Antoine Griezmann         | 5                                   | 3                             | \$51 845 682   |
| Cristiano Ronaldo         | 6                                   | 2                             | \$64 923 873   |
| Kevin Volland             | 7                                   | 83                            | \$5 635 474    |
| Pierre-Emerick Aubameyang | 8                                   | 41                            | \$12 348 929   |
| Ángel Di María            | 9                                   | 24                            | \$15 040 371   |
| Mohamed Salah             | 10                                  | 30                            | \$13 721 032   |

Źródło: opracowanie własne.

Zestawienie w tabeli 3.20 pozwala zauważyć następującą prawidłowość, że 6 na 10 zawodników będących w grupie 10 najlepszych zawodników według metody wzorca rozwoju również występuje w dziesiątce najlepszych piłkarzy w rankingu przeprowadzonego metodą analizy głównych składowych przedstawionego w tabeli 3.7. Na podstawie tabeli 3.20 można wywnioskować, że piłkarze znajdujący się w pierwszej trójce najlepiej zarabiających piłkarzy: Lionel Messi, Cristiano Ronaldo oraz Antoine Griezmann znajdują się w najlepszej szóstce rankingu. Oznacza to, że gigantyczne zarobki są powiązane z charakterystykami występów tej trójki. Aczkolwiek w pierwszej dziesiątce można zaobserwować takich piłkarzy jak: Mohamed Salah, Suso, Kevin Volland oraz Hakan Calhanoglu, których pensje są nieporównywalnie mniejsze od wynagrodzeń piłkarzy pierwszej trójki. Można zauważyć, że ci piłkarze byli na dalszych pozycjach pod względem zarobków a nadal byli w stanie zasłużyć sobie wynikami na boisku na wyniki w czołowej dziesiątce przeprowadzonego rankingu. Można zauważyć wstępną zależność pomiędzy prezentowanymi wynikami obu metod. W

tabeli 3.21 znajdują się piłkarze przyporządkowani do klasy 10. Jest to dziesięciu najgorszych zawodników według metody wzorca rozwoju. Opierając się o tabelę 3.8 przedstawiającą dziesięciu najgorszych piłkarzy według metody analizy głównych składowych zostanie dokonane porównanie.

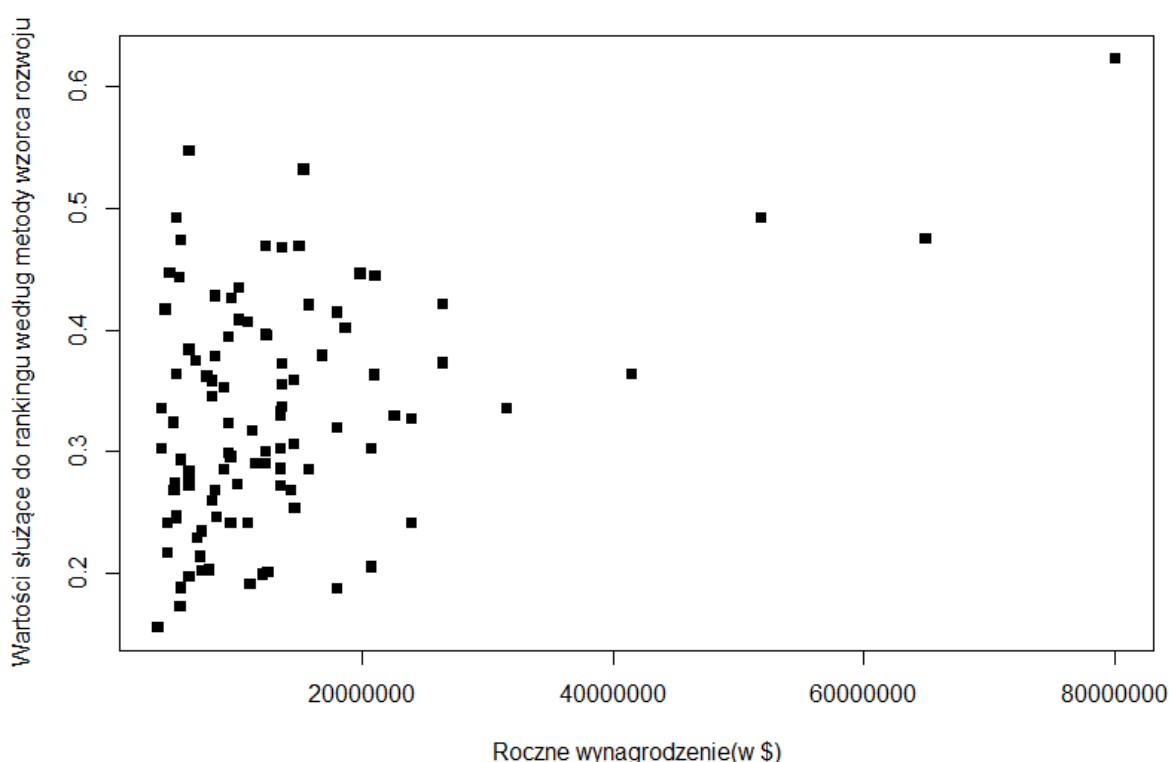
Tabela 3.21 Klasa 10 ustalona według metody wzorca rozwoju

| Zawodnik           | Miejsce pod względem wzorca rozwoju | Miejsce pod względem zarobków | Zarobki roczne |
|--------------------|-------------------------------------|-------------------------------|----------------|
| Arjen Robben       | 91                                  | 69                            | \$7 889 437    |
| Javier Pastore     | 92                                  | 71                            | \$7 329 622    |
| Douglas Costa      | 93                                  | 38                            | \$12 565 874   |
| Vitolo             | 94                                  | 45                            | \$12 172 307   |
| Antonio Candreva   | 95                                  | 77                            | \$6 283 503    |
| Álvaro Morata      | 96                                  | 48                            | \$11 136 366   |
| Yevhen Konoplyanka | 97                                  | 84                            | \$5 635 474    |
| Diego Costa        | 98                                  | 17                            | \$18 083 731   |
| Adrien Silva       | 99                                  | 85                            | \$5 568 749    |
| Stevan Jovetic     | 100                                 | 100                           | \$3 787 518    |

Źródło: opracowanie własne.

W oparciu o tabelę 3.21 trzeba zwrócić uwagę, że aż 9 na 10 piłkarzy występuje w obu rankingach w najgorszej dziesiątce. Jest to tylko potwierdzeniem wcześniejszych obserwacji odnośnie małego przełożenia zarobków piłkarzy na ich występy sportowe. Tabela 3.21 jasno pokazuje prawidłowość wskazującą na małe powiązanie zarobków z wynikami sportowymi. W ostatniej dziesiątce rankingu znajdziemy piłkarzy zarabiających dużo ponad 10 milionów dolarów rocznie. Diego Costa znajduje się odpowiednio na 17 pozycji w odniesieniu do zarobków, a jednak znalazł się blisko ostatniego miejsca w zestawieniu sporządzonym na podstawie osiągnięć. Można zauważyć, że w najgorszej dziesiątce pod względem efektywności na boisku większość piłkarzy znajduje się w pierwszej pięćdziesiątce pod

względem zarobków. Po przeprowadzonej analizie można wywnioskować, że zarobki piłkarzy mają tylko wpływ na kilku czołowych z nich. Natomiast reszta sportowców przedstawiona w rankingu, dysponuje zbliżonymi do siebie umiejętnościami oraz statystykami. Poddaje to w wątpliwość wydawanie ogromnych sum pieniędzy na kontrakty przedstawionych piłkarzy. W celu potwierdzenia powyższych obserwacji dokonana została analiza korelacji. Na rysunku 3.3 znajduje się wykres korelacyjny pomiędzy rocznym wynagrodzeniem a wartościami służącymi do stworzenia rankingu według metody wzorca rozwoju.



Rysunek 3.3 Wykres korelacyjny pomiędzy rocznym wynagrodzeniem a wartościami służącymi do stworzenia rankingu według metody wzorca rozwoju

Źródło: opracowanie własne.

Na podstawie rysunku 3.3 można stwierdzić, że poprzednie obserwacje znajdują swoje potwierdzenie. Jedynie kilku zawodników z czołówki wielkości wynagrodzenia wyróżnia się na tle pozostałych. Następnym krokiem będzie obliczenie wartości współczynnika korelacji liniowej Pearsona dla wartości służących do stworzenia rankingu według metody wzorca rozwoju a wynagrodzeniem. Wartość współczynnika korelacji liniowej Pearsona pomiędzy wartościami rankingu a wynagrodzeniem wynosi 0,4080. Oznacza to, że korelacja jest przeciętna. Co oznacza, że istnieje zaledwie przeciętna zależność pomiędzy zarobkami

piłkarzy a ich występami. Jest to warte odnotowania, ponieważ pokrywa się z wcześniejszymi obserwacjami i wnioskami. Kolejnym a zarazem ostatnim testem został przeprowadzony test istotności korelacji, wszystkie wartości związane z przeprowadzeniem testu znajdują się w tabeli 3.22.

Tabela 3.22 Wartość statystyki testowej oraz wartość krytyczna t odczytana z tablic rozkładu t-Studenta

| Element testu       | Wartość |
|---------------------|---------|
| Statystyka testowa  | 4,4252  |
| Wartość krytyczna t | 1,984   |

Źródło: opracowanie własne.

Na podstawie tabeli 3.22 statystyka testowa znajduje się w obszarze krytycznym. Co za tym idzie odrzucono hipotezę zerową na rzecz hipotezy alternatywnej. Korelacja jest na pewno istotna. Co za tym idzie poprzednie rozważania były poprawne.

Podsumowując przeprowadzone analizy oraz ocenę relacji zarobków piłkarzy i ich efektywności w meczach piłkarskich zdecydowanie potwierdzają się wcześniejsze wnioski. Zarobki piłkarzy są wysokie i nie mają one dużego wpływu na ich wyniki sportowe, wyjątkiem jest zaledwie kilka indywidualnych przypadków. Interesującymi wynikami są obydwie przeprowadzone rankingi, które są wobec siebie bardzo podobne. Na potwierdzenie wszystkich wniosków policzono dwa współczynniki korelacji rang Spearmana rankingu zarobków wobec rankingu według pierwszej składowej oraz rankingu zarobków wobec rankingiem metody wzorca rozwoju. Na podstawie tego współczynnika udało się uwydatnić omawianą relację zarobków a efektywności na boisku. W tabeli 3.23 zostały przedstawione wartości współczynników korelacji Spearmana.



Tabela 3.23 Wartość współczynnika korelacji Spearmana pomiędzy rankingiem zarobków a rankingiem według pierwszej składowej oraz pomiędzy rankingiem zarobków a rankingiem metody wzorca rozwoju.

| Współczynnik korelacji rang Spearmana                                | Wartość |
|--|---------|
| Pomiędzy rankingiem zarobków a rankingiem według pierwszej składowej | 0,2809  |
| Pomiędzy rankingiem zarobków a rankingiem metody wzorca rozwoju      | 0,3263  |

Źródło: opracowanie własne.

Na podstawie tabeli 3.23 warto zauważyć, że korelacja jest niska co tylko podkreśla coraz mocniej poprzednie rozważania, że wielkość wynagrodzenia ma niski wpływ na efektywność występów piłkarzy. Należy zaznaczyć, że korelacje są również do siebie zbliżone, pomimo stworzenia rankingów za pomocą dwóch innych metod. Można zauważyć wiele podobieństw, wskazujących na to, że piłkarze są mocno przepłaceni a zarobki nie mają odzwierciedlenia w ich poczynaniach boiskowych.

# Zakończenie

Celem pracy było zbadanie relacji pomiędzy wielkością kontraktów topowych piłkarzy grających w czołowych europejskich ligach a charakterystykami ich występów na boisku. W pierwszym rozdziale zawarto informacje dotyczące wpływu sportu na gospodarkę, został również przedstawiony rozwój na przestrzeni lat angielskiej ligi Premier League oraz wpływ Leo Messiego na cały klub FC Barcelona. Wskazano, że sport ma spory udział w gospodarkach wielu krajów. Tworzy również dużo miejsc pracy oraz ma swój wkład w PKB. W drugim rozdziale zostały przedstawione i wyjaśnione zastosowane w pracy metody oraz narzędzia użyte do przeprowadzania badania. W trzecim rozdziale dokonano analiz i zawarto wyniki badań. Stworzono rankingi efektywności za pomocą dwóch metod oraz ranking zarobków, które ze sobą zestawiono.

W pracy poszukiwano odpowiedzi na pytania badawcze zawarte we wstępie. Na podstawie przeprowadzonych działań udało się sformułować wnioski. Po pierwsze, pomiędzy zarobkami piłkarzy a efektywnością na boisku znaleziono związki, ale są one bardzo słabe. Po drugie, ustalono, że do syntetycznej oceny efektywności występów w meczach piłkarskich na pewno może służyć metoda wzorca rozwoju Hellwiga oraz modelowanie ekonometryczne. Natomiast analiza głównych składowych poprzez niewystarczająco mocno skorelowane zmienne nie spełniła swojej funkcji, tylko funkcję porównawczą. Po trzecie, sprawdzono, że wyniki otrzymane na podstawie rankingów stworzonych metodą wzorca rozwoju i analizą głównych składowych są relatywnie do siebie podobne. Po czwarte, za pomocą modelowania ekonometrycznego wskazano, że najważniejszym czynnikiem wpływającym na zarobki piłkarzy jest liczba strzelonych goli.

Wyniki wykazały, że istnieje słaby związek pomiędzy wynikami sportowymi a wysokością pensji piłkarzy. Należy również zwrócić uwagę na istotę goli strzelanych przez zawodników. To głównie one sprawiają, że kwoty zarabiane przez nich są bardzo wysokie. Warto dodać, że najlepiej opłacanymi zawodnikami są napastnicy, więc w nich pokładane są największe nadzieje na rozwój sportowy klubu. Presja na wyniki tworzona przez inwestorów i kibiców w największych klubach świata jest bardzo duża. Niestety, gdy zawodnicy nie spełniają pokrywanych w nich oczekiwań, stają się obciążeniem dla budżetu klubu. Z przeprowadzonego badania wynika, że tylko niewielka część najlepiej opłacanych piłkarzy jest w stanie wziąć presję na swoje barki i sprostać wszystkim oczekiwaniom.

# Bibliografia

1. Aczel A., *Statystyka w zarządzaniu*, Wydawnictwo Naukowe PWN, Warszawa, 2000, s. 742-743
2. Baran, Lis, Magda, *Ocena korzyści społecznych inwestycji w sport w odniesieniu do ponoszonych kosztów. Raport analityczny.*, Warszawa 2016
3. Bąk A., *Porządkowanie liniowe obiektów metodą Hellwiga, TOPSIS – analiza porównawcza*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu nr 426, Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław 2016, s. 25.
4. Borkowski B., Dudek H., Szczęsny W., *Ekonometria. Wybrane zagadnienia.*, Wydawnictwo Naukowe PWN, Warszawa, 2004, s. 39-42
5. Deloitte, *Annual Review of Football Finance 2020*, <https://www2.deloitte.com/uk/en/pages/sports-business-group/articles/annual-review-of-football-finance.html>, [dostęp: 27.04.2021]
6. Dziechciarz J., *Ekonometria. Metody, przykłady, zadania*. Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław 2003, s. 287
7. Górecki Brunon R. *Ekonometria podstawy teorii i praktyki*, Wydawnictwo Key Text, Warszawa 2010, s. 14-15.
8. Górniak J., *Analiza czynnikowa i analiza głównych składowych*, ASK, 1998, nr. 7, s. 93.
9. Kaczmarek M., *Zastosowanie analizy głównych składowych w ewaluacji skali pomiaru użyteczności serwisu internetowego*, *Studia Oeconomica Posnaniensia*, 2016, vol. 4, no. 1, s. 131.
10. Komisja Europejska (2018), *Study on the economic impact of sport through sport satellite accounts*, <https://op.europa.eu/en/publication-detail/-/publication/865ef44c-5ca1-11e8-ab41-01aa75ed71a1/language-en/format-PDF/source-71256399> [dostęp: 25.05.2021].
11. Krzyśko M., *Podstawy wielowymiarowego wnioskowania statystycznego*, Wydawnictwo Naukowe, Poznań, 2009, s. 241.

12. Kutwa K., Rafał M., (2019), *Polski rynek sportu*, Polski Instytut Ekonomiczny, Warszawa, [https://pie.net.pl/wp-content/uploads/2019/08/Raport\\_PIE-Sport-gospodarka.pdf](https://pie.net.pl/wp-content/uploads/2019/08/Raport_PIE-Sport-gospodarka.pdf) [dostęp: 25.04.2021]
13. Llimós A., *Barça i Nike, un acord rècord per aquest juny*, [https://www.ara.cat/esports/nike-acord-barca\\_1\\_1641892.html](https://www.ara.cat/esports/nike-acord-barca_1_1641892.html) [dostęp: 20.05.2021]
14. Marketing Week, *Premier League has become a data minefield for brands*, <https://www.marketingweek.com/premier-league-has-become-a-data-minefield-for-brands/> [dostęp: 28.04.2021]
15. McMahon B., *La Liga Distributes €1.4B In TV Money; Barcelona And 17 Other Teams Receive Record Payouts*, <https://www.forbes.com/sites/bobbymcmahon/2019/12/15/la-liga-distributes-14b-in-tv-money-barcelona-and-17-other-teams-receive-record-payouts/?sh=376aca2027ab>, [dostęp: 26.04.2021]
16. Namyślak B., *Zastosowanie metody wzorca rozwoju Hellwiga do badania sektora kultury w miastach wojewódzkich*, Wiadomości Statystyczne nr 3/2015, s. 42-43
17. Navarro C., *The millions that Messi brings in at Barcelona, Barcelona 2021*, <https://www.marca.com/en/football/barcelona/2021/02/02/6019282446163fb3af8b45df.html>, [dostęp: 27.04.2021]
18. Ostasiewicz S., Siedlecka U., Rusnak S., *Statystyka: elementy teorii i zadania*, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław 2006
19. Planet Football, *Club-by-club: A breakdown of the 2019-20 Premier League prize money*, <https://www.planetfootball.com/quick-reads/club-by-club-a-breakdown-of-the-2019-20-premier-league-prize-money/> [dostęp: 24.04.2021]
20. *Premier League in record £5.14bn TV rights deal* <https://www.bbc.com/news/business-31379128> [dostęp: 20.05.2021]
21. Statista, *Soccer (football) shirt sales worldwide in 2018/2019*, <https://www.statista.com/statistics/1118294/football-shirt-sales-by-club/>, [dostęp: 29.04.2021]
22. Śniadek J., *Wpływ mega eventów sportowych na społeczność i gospodarkę kraju goszczącego*, Ekonomiczne Problemy Usług nr 84, 693-710, Poznań 2013

23. The Business Research Company (2019), *Sports Market By Type (Participatory Sports, Spectator Sports, Sports Team & Clubs, Racing & Individual Sports And Spectator Sports), By Competitive Landscape And By Geography – Global Forecast To 2022 - Sample*, <https://www.thebusinessresearchcompany.com/report/sports-market> [dostęp: 24.04.2021]
24. Total Sportek, *25 World's Most Popular Sports (Ranked by 13 factors)*, <https://www.totalsportek.com/popular-sports/> [dostęp: 20.04.2021]
25. White H., *A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity*, „Econometrica” 48 (4), s. 817–838.
26. Wołowski D., *Ile Leo Messi zarabia dla Barcelony?* <https://sport.interia.pl/primeradivision/news-ile-leo-messi-zarabia-dla-barcelony,nId,5025112> [dostęp: 20.05.2021]

## Spis tabel

|  |    |
|--|----|
| Tabela 3.1 Wartości własne badanych składowych.....  | 33 |
| Tabela 3.2 Wektory własne macierzy korelacji.....  | 34 |
| Tabela 3.3 Zasoby zmienności wspólnej korelacji .....  | 36 |
| Tabela 3.4 Zmienne wybrane do przeprowadzenia rankingu.....  | 37 |
| Tabela 3.5 Współrzędne czynnikowe zmiennych .....  | 38 |
| Tabela 3.6 Skorygowane ładunki czynnikowe zmiennych .....  | 39 |
| Tabela 3.7 Zestawienie 10 najlepszych zawodników według rankingu .....   | 40 |
| Tabela 3.8 Zestawienie 10 najgorszych zawodników według rankingu.....  | 41 |
| Tabela 3.9 Wartość statystyki testowej oraz wartość krytyczna t odczytana z tablic rozkładu t-Studenta.....  | 43 |
| Tabela 3.10 wartość t-Studenta oraz współczynnika korelacji $r^*$ .....  | 44 |
| Tabela 3.11 Wektor współczynników korelacji $R^2$ pomiędzy zmienną objaśnianą a potencjalnymi zmiennymi objaśniającymi.....  | 45 |
| Tabela 3.12 Macierz współczynników korelacji zmiennych objaśniających .....  | 46 |
| Tabela 3.13 Wartości miar oceny dopasowania modelu do danych.....  | 47 |
| Tabela 3.14 Wartość statystyk t-Studenta dla zmiennych objaśniających oraz p-value.....  | 47 |
| Tabela 3.15 Wartość statystyki t-Studenta oraz wartość krytyczna odczytana z tablic rozkładu t-Studenta .....  | 48 |
| Tabela 3.16 Wartość statystyki testowej testu White'a oraz wartość teoretyczna odczytana z tablic rozkładu Chi-kwadrat .....   | 49 |
| Tabela 3.17 Elementy testu Kołmogorowa-Smirnowa.....   | 49 |
| Tabela 3.18 Zmienne podzielone na stymulanty i destymulanty .....  | 51 |
| Tabela 3.19 Wartości wzorca i antywzorca.....  | 52 |
| Tabela 3.20 Klasa 1 ustalona według metody wzorca rozwoju .....  | 53 |
| Tabela 3.21 Klasa 10 ustalona według metody wzorca rozwoju .....   | 54 |
| Tabela 3.22 Wartość statystyki testowej oraz wartość krytyczna t odczytana z tablic rozkładu t-Studenta .....  | 56 |
| Tabela 3.23 Wartość współczynnika korelacji Spearmana pomiędzy rankingiem zarobków a rankingiem według pierwszej składowej oraz pomiędzy rankingiem zarobków a rankingiem metody wzorca rozwoju..... | 57 |

## Spis rysunków

|   |    |
|---|----|
| Rysunek 1.1 Wielkość rynku globalnego sportu w miliardach dolarów .....   | 7  |
| Rysunek 1.2 PKB generowane przez sport w 2012 roku w EU-28 (oprócz Słowenii – B/D)...   | 8  |
| Rysunek 1.3 procentowy udział PKB generowany przez sport w całości PKB w roku 2012 w<br>EU-28 (oprócz Słowenii – B/D) .....                         | 9  |
| Rysunek 3.1 Wykres osypiska.....  | 34 |
| Rysunek 3.2 Wykres korelacyjny pomiędzy wynagrodzeniem a wartościami służącymi do<br>stworzenia rankingu według pierwszej składowej .....           | 42 |
| Rysunek 3.3 Wykres korelacyjny pomiędzy rocznym wynagrodzeniem a wartościami<br>służącymi do stworzenia rankingu według metody wzorca rozwoju ..... | 55 |