



Projekt zaliczeniowy

Wydział: Wydział Zarządzania

Kierunek: Informatyka i Ekonometria

Autor: Mateusz Kucharz

Data: 16.06.2023

Spis treści

1. Cel Projektu oraz hipotezy badawcze	2
2. Dane	3
3. Statystyki opisowe	4
4. Korelacja i zależność.....	7
5. Szacowanie parametrów regresji.....	10
6. Redukcja ilości zmiennych	11
1. <i>Metoda Hellwiga</i>	<i>11</i>
2. <i>Metoda krokowo - wsteczna</i>	<i>11</i>
3. <i>Postać modelu.....</i>	<i>12</i>
7. Ocena modelu	13
1. <i>Współczynnik determinacji R^2.....</i>	<i>13</i>
2. <i>Katalizatory</i>	<i>13</i>
3. <i>Normalność rozkładu składnika losowego</i>	<i>14</i>
4. <i>Istotność zmiennych</i>	<i>15</i>
5. <i>Test dodanych zmiennych</i>	<i>17</i>
6. <i>Obserwacje odstające</i>	<i>18</i>
7. <i>Test liczby serii</i>	<i>18</i>
8. <i>Test Reset</i>	<i>19</i>
9. <i>Heteroskedastyczność</i>	<i>20</i>
10. <i>Test Chowa</i>	<i>21</i>
11. <i>Współliniowość.....</i>	<i>22</i>
8. Podsumowanie.....	25

1. Cel Projektu oraz hipotezy badawcze

Projekt ma za zadanie sprawdzić wpływ wybranych statystyk graczy NBA na ich zarobki. W tym celu postaram się zbudować odpowiedni model oraz przeprowadzę testy statystyczne.

Hipotezy:

1. Istnieje zależność między liczbą lat gry profesjonalnie a zarobkami gracza NBA.
2. Istnieje zależność między ilością zdobytych punktów na mecz a zarobkami gracza NBA.
3. Istnieje zależność między liczbą zdobytych zbiórek na mecz a zarobkami gracza NBA.
4. Istnieje zależność między liczbą zdobytych asyst na mecz a zarobkami gracza NBA.
5. Istnieje zależność między średnią ilością minut na boisku na grę a zarobkami gracza NBA.

2. Dane

Dane na których będę pracował to różne statystyki graczy NBA (National Basketball Association) z roku 1995. Zostały pobrane ze strony https://gretl.sourceforge.net/gretl_data.html, ale oryginalnie pochodzą z książki *Introductory Econometrics: A Modern Approach* autorstwa Jeffrey Wooldridge. Zawierają one 269 obserwacji oraz 22 zmiennych, jednak ja skupię się tylko na tych wybranych przeze mnie.

Zmienna objaśniana [Y]

1. Wage – roczna wypłata w tys. dolarów

Zmienne objaśniające [X]

1. Exper – ilość lat grania profesjonalnie
2. Points – ilość zdobywanych punktów na grę
3. Rebounds – ilość zdobywanych zbiórek na grę
4. Assists – liczba asyst na grę
5. Avgmin – średnia ilość minut na boisku na grę

Za poziom istotności przyjmuję $\alpha = 0.05$

3. Statystyki opisowe

\	Średnia	Mediana	Min	Max	Odchylenie
Y	1423,8	1186	150	5740	999,77
X1	5,119	4	1	18	4,4001
X2	10,26	9	1	30	5,8825
X3	4,4684	4	1	17	2,893
X4	2,4535	2	0	13	2,1481
X5	23,979	24,82	2,89	43,09	9,7311

Tabela 1. Statystyki opisowe

\	Zmienność	Skośność	Kurtoza	Kw. 5%	Kw. 95%	Q3-Q1
Y	0,70217	0,95681	0,84963	150	3250	1372,3
X1	0,66421	0,81601	0,24164	1	11	5,5
X2	0,57333	0,6507	0,041713	2	20,5	9
X3	0,64743	1,2099	1,3696	1	11	4
X4	0,87552	1,6512	3,5719	0	7	2
X5	0,40581	-0,20098	-1,017	7,23	37,515	16,605

Tabela 2. Statystyki opisowe cd.

Zmienna Y (zarobki) –

Średnie zarobki badanych graczy NBA wynoszą 1423,8 tysięcy dolarów. Wartości te są rozrzucone wokół średniej o 999,77 z wartością najmniejszą wynoszącą 150 tys. i największą 5740 tys.

Mediana jest równa 1186 więc 50% wartości zmiennej jest mniejsza od 1186 jak i 50% jest większa. Zmienna Y charakteryzuje się wysokim współczynnikiem zmienności co można zauważyć po odchyleniu standardowym.

Zmienna X1 (Lata gry profesjonalnie) –

Średnia ilość lat gry profesjonalnej badanych graczy NBA wynosi 5,119 lat. Wartości są rozrzucone wokół średniej o 4,4001 z wartością najmniejszą wynoszącą 1 rok. i największą 18 lat. Mediana jest równa 4 więc 50% wartości zmiennej jest mniejsza od 4 jak i 50% jest większa. Zmienna X1 charakteryzuje się wysokim współczynnikiem zmienności.

Zmienna X2 (Ilość punktów na grę) –

Średnia ilość zdobywanych punktów na grę badanych graczy NBA wynosi 10,26. Wartości są rozrzucone wokół średniej o 5,8825 z wartością najmniejszą wynoszącą 1 i największą 30. Mediana jest równa 9 więc 50% wartości zmiennej jest mniejsza od 9 jak i 50% jest większa. Zmienna X2 charakteryzuje się wysokim współczynnikiem zmienności.

Zmienna X3 (Ilość zbiórek na grę) –

Średnia ilość zdobywanych punktów na grę badanych graczy NBA wynosi 4,4684. Wartości są rozrzucone wokół średniej o 2,893 z wartością najmniejszą wynoszącą 1 i największą 17. Mediana jest równa 4 więc 50% wartości zmiennej jest mniejsza od 4 jak i 50%

jest większa. Zmienna X3 charakteryzuje się wysokim współczynnikiem zmienności.

Zmienna X4 (Ilość asyst na grę) –

Średnia ilość zdobywanych asyst na grę badanych graczy NBA wynosi 2,4535. Wartości są rozrzucone wokół średniej o 2,1481 z wartością najmniejszą wynoszącą 0 i największą 13. Mediana jest równa 2 więc 50% wartości zmiennej jest mniejsza od 2 jak i 50% jest większa. Zmienna X4 charakteryzuje się wysokim współczynnikiem zmienności.

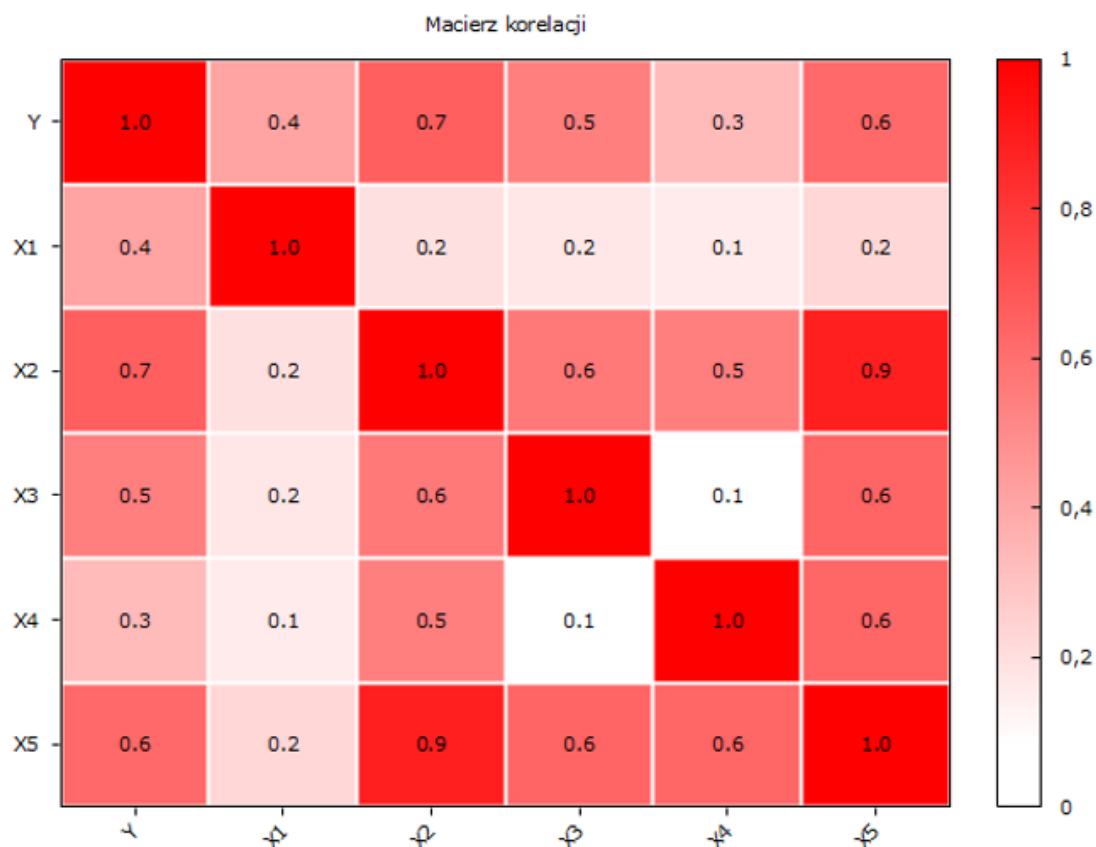
Zmienna X5 (Ilość minut na boisku na grę) –

Średni czas przebytych minut na boisku na grę badanych graczy NBA wynosi 23,979 minut. Wartości są rozrzucone wokół średniej o 9,7311 z wartością najmniejszą wynoszącą 2,89 i największą 43,09. Mediana jest równa 24,82 więc 50% wartości zmiennej jest mniejsza od tej wartości jak i 50% jest większa. Zmienna X5 również charakteryzuje się wysokim współczynnikiem zmienności.

4. Korelacja i zależność

W najlepszym wypadku korelacje pomiędzy zmienną objaśnianą a zmiennymi objaśniającymi powinny być wysokie, korelacje pomiędzy zmiennymi objaśniającymi powinny być niskie.

Współczynnik korelacji jest mierzony na skali, która zmienia się od +1, przez 0, do -1. Pełna korelacja między dwiema zmiennymi jest wyrażana jako +1 lub -1. Gdy jedna zmienna wzrasta wraz ze wzrostem drugiej, korelacja jest dodatnia; gdy jedna zmienna maleje wraz ze wzrostem drugiej, jest to korelacja ujemna. Brak jakiegokolwiek korelacji jest reprezentowany przez 0.¹



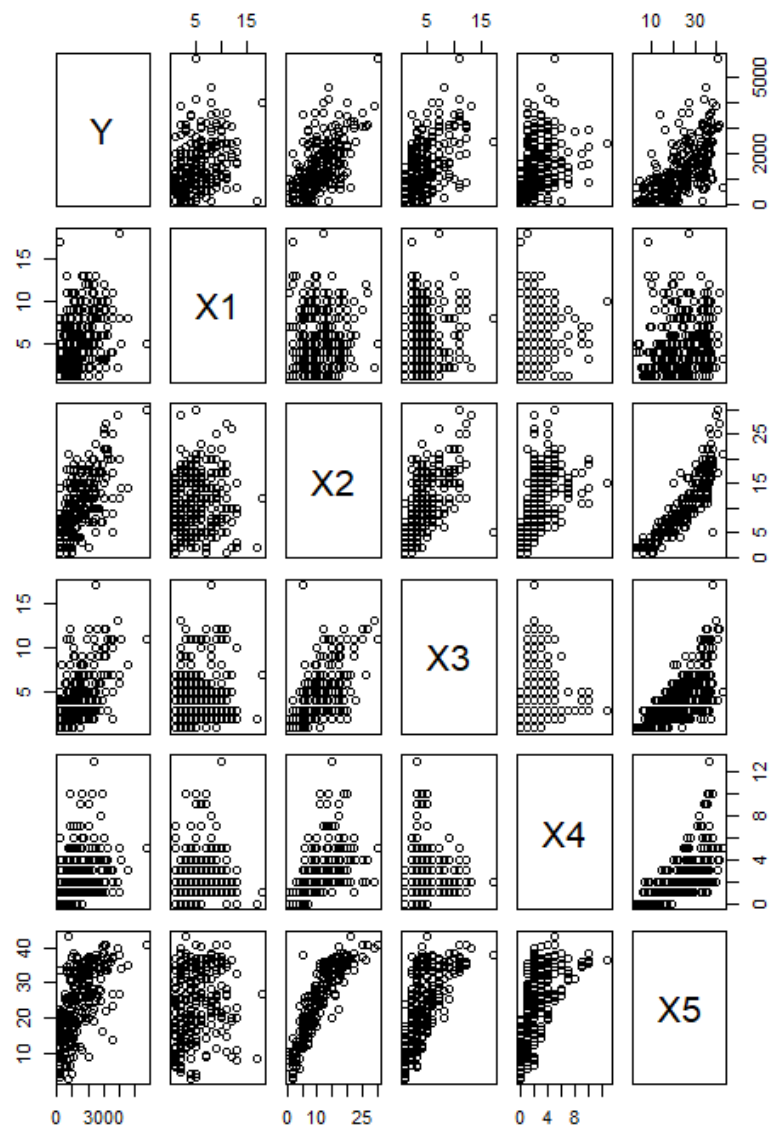
Rysunek 1. Macierz korelacji

¹ M.J. Campbell, Statistics at Square One 9th Edition, University of Southampton, London 1997, s.75.

Korelacje między zmienną objaśnianą a zmiennymi objaśnianymi są umiarkowane z zaznaczeniem, że największa występuje między Y i X2, a najslabsza dla Y i X4.

W przypadku korelacji między zmiennymi objaśniającymi można zauważyć bardzo słabą wartość dla zmiennej X1 co dobrze o niej świadczy. Reszta korelacji między zmiennymi jest umiarkowana z jednym wyjątkiem. Zmienna X2 oraz X5 mają silną zależność co może sugerować współliniowość.

Do dalszej weryfikacji zależności między zmiennymi posłużę się wykresem zależności zmiennych wygenerowanym w programie R.



Wykres 1. Wykres zależności

Jak można zauważyć w większości wypadków gdzie jest zależność między zmiennymi objaśniającymi wykresy są rozrzucone. Jedyne co przykuwa moją uwagę to zmienna X5, dla której wykresy nie wyglądają już tak dobrze. Domyślam się, że może to być problem związany ze współliniowością zmiennej.

W przypadku zmiennej Y wykresy są w miarę skupione co pokazuje umiarkowaną korelację między zmiennymi.

5. Szacowanie parametrów regresji

Za pomocą komendy w skrypcie *ols Y X1 X2 X3 X4 X5* przeprowadzam szacowanie parametrów regresji po przez MNK.

Model 1: Estymacja KMNK, wykorzystane obserwacje 1-269
Zmienna zależna (Y): Y

	współczynnik	błąd standardowy	t-Studenta	wartość p	
const	-121,632	127,356	-0,9551	0,3404	
X1	82,8619	12,4649	6,648	1,71e-010	***
X2	86,7911	15,2371	5,696	3,28e-08	***
X3	92,6760	22,5464	4,110	5,28e-05	***
X4	27,3695	30,3353	0,9022	0,3678	
X5	-10,4452	12,2534	-0,8524	0,3947	
Średn. aryt. zm. zależnej	1423,828	Odch. stand. zm. zależnej	999,7741		
Suma kwadratów reszt	1,20e+08	Błąd standardowy reszt	675,5416		
Wsp. determ. R-kwadrat	0,551955	Skorygowany R-kwadrat	0,543437		
F(5, 263)	64,79899	Wartość p dla testu F	6,71e-44		
Logarytm wiarygodności	-2131,334	Kryt. inform. Akaike'a	4274,668		
Kryt. bayes. Schwarza	4296,236	Kryt. Hannana-Quinna	4283,330		

Wyłączając stałą, największa wartość p jest dla zmiennej 6 (X5)

Gretl 1. Estymacja parametrów MNK

Funkcja regresji jest dopasowana do danych empirycznych w 55%. Test F sugeruje, że przynajmniej jedna zmienna jest statystycznie istotna. Wysokie wartości kryteriów informacyjnych mogą wskazywać na zbyt dużą złożoność modelu.

6.Redukcja ilości zmiennych

1. Metoda Hellwiga

Metoda pojemności informacyjnej nośników (znana również jako metoda optymalnego wyboru predyktorów lub, według nazwiska jej autora, metoda Hellwiga) jest jedną z metod doboru zmiennych niezależnych dla modelu ekonometrycznego.²

Optymalna jest kombinacja X1 X2 X3, $H = 0,539134$

Gretl 2. Skrypt z zajęć

Model 1: Estymacja KMNK, wykorzystane obserwacje 1-269

Zmienna zależna (Y): Y

	współczynnik	błąd standardowy	t-Studenta	wartość p	
X1	72,0892	10,9194	6,602	2,19e-010	***
X2	74,9939	7,93864	9,447	1,88e-018	***
X3	70,7248	16,8016	4,209	3,51e-05	***
Średn.aryt.zm.zależnej	1423,828	Odch.stand.zm.zależnej	999,7741		
Suma kwadratów reszt	1,22e+08	Błąd standardowy reszt	677,0800		
Niecentrowany R-kwadr.	0,850047	Centrowany R-kwadrat	0,544778		
F(3, 266)	502,6308	Wartość p dla testu F	3,0e-109		
Logarytm wiarygodności	-2133,471	Kryt. inform. Akaike'a	4272,943		
Kryt. bayes. Schwarza	4283,727	Kryt. Hannana-Quinna	4277,274		

Gretl 3. Estymacja nowego modelu

2. Metoda krokowo - wsteczna

Ponieważ zmienna Y nie jest zmienną dyskretną nie mogą być użyte metody krokowo – wstecznej, użyje za to testu pominiętych zmiennych by sprawdzić postać modelu.

² Kowalik, Przemysław, On an implementation of the method of capacity of information bearers (the Hellwig method) in spreadsheets, Politechnika Lubelska, Lublin 2014, s. 31.

Sekwencyjna eliminacja nieistotnych zmiennych przy dwustronnym obszarze krytycznym, $\alpha = 0,05$

Wyeliminowano nieistotną zmienną: X5 (wartość p = 0,395)
Wyeliminowano nieistotną zmienną: X4 (wartość p = 0,622)

Test porównawczy z Modelem 3

Hipoteza zerowa: parametry regresji dla wskazanych zmiennych są równe zero
X4, X5

Statystyka testu: $F(2, 263) = 0,485$, wartość p 0,616247

Pominięcie zmiennych poprawiło 3 z 3 kryteriów informacyjnych (AIC, BIC, HQC).

Model 5: Estymacja KMNK, wykorzystane obserwacje 1-269

Zmienna zależna (Y): Y

	współczynnik	błąd standardowy	t-Studenta	wartość p	
const	-179,346	99,3985	-1,804	0,0723	*
X1	82,7281	12,3692	6,688	1,33e-010	***
X2	80,8070	8,53651	9,466	1,67e-018	***
X3	78,4607	17,2714	4,543	8,44e-06	***
Średn.aryt.zm.zależnej	1423,828	Odch.stand.zm.zależnej	999,7741		
Suma kwadratów reszt	1,20e+08	Błąd standardowy reszt	674,2275		
Wsp. determ. R-kwadrat	0,550303	Skorygowany R-kwadrat	0,545212		
F(3, 265)	108,0951	Wartość p dla testu F	9,96e-46		
Logarytm wiarygodności	-2131,829	Kryt. inform. Akaike'a	4271,658		
Kryt. bayes. Schwarza	4286,037	Kryt. Hannana-Quinna	4277,433		

Gretl 4. Test pominiętych zmiennych

3. Postać modelu

Obie metody pokazują ten sam zestaw zmiennych. Od teraz przybieram postać modelu ze zmiennym X1 X2 X3.

7. Ocena modelu

1. Współczynnik determinacji R^2 .

Ostateczna postać modelu estymowana za pomocą KMNK.

Model 5: Estymacja KMNK, wykorzystane obserwacje 1-269					
Zmienna zależna (Y): Y					
	współczynnik	błąd standardowy	t-Studenta	wartość p	
-----	-----	-----	-----	-----	-----
const	-179,346	99,3985	-1,804	0,0723	*
X1	82,7281	12,3692	6,688	1,33e-010	***
X2	80,8070	8,53651	9,466	1,67e-018	***
X3	78,4607	17,2714	4,543	8,44e-06	***
Średn. aryt. zm. zależnej	1423,828	Odch. stand. zm. zależnej	999,7741		
Suma kwadratów reszt	1,20e+08	Błąd standardowy reszt	674,2275		
Wsp. determ. R-kwadrat	0,550303	Skorygowany R-kwadrat	0,545212		
F(3, 265)	108,0951	Wartość p dla testu F	9,96e-46		
Logarytm wiarygodności	-2131,829	Kryt. inform. Akaike'a	4271,658		
Kryt. bayes. Schwarza	4286,037	Kryt. Hannana-Quinna	4277,433		

Gretl 5. Nowy model

Współczynnik R^2 wynosi 0.550303 i oznacza to, że dopasowanie funkcji regresji do danych empirycznych wynosi 55% tak jak w pierwszym modelu. Należy jednak zwrócić uwagę na to, że kryteria informacyjne się delikatnie poprawiły.

2. Katalizatory

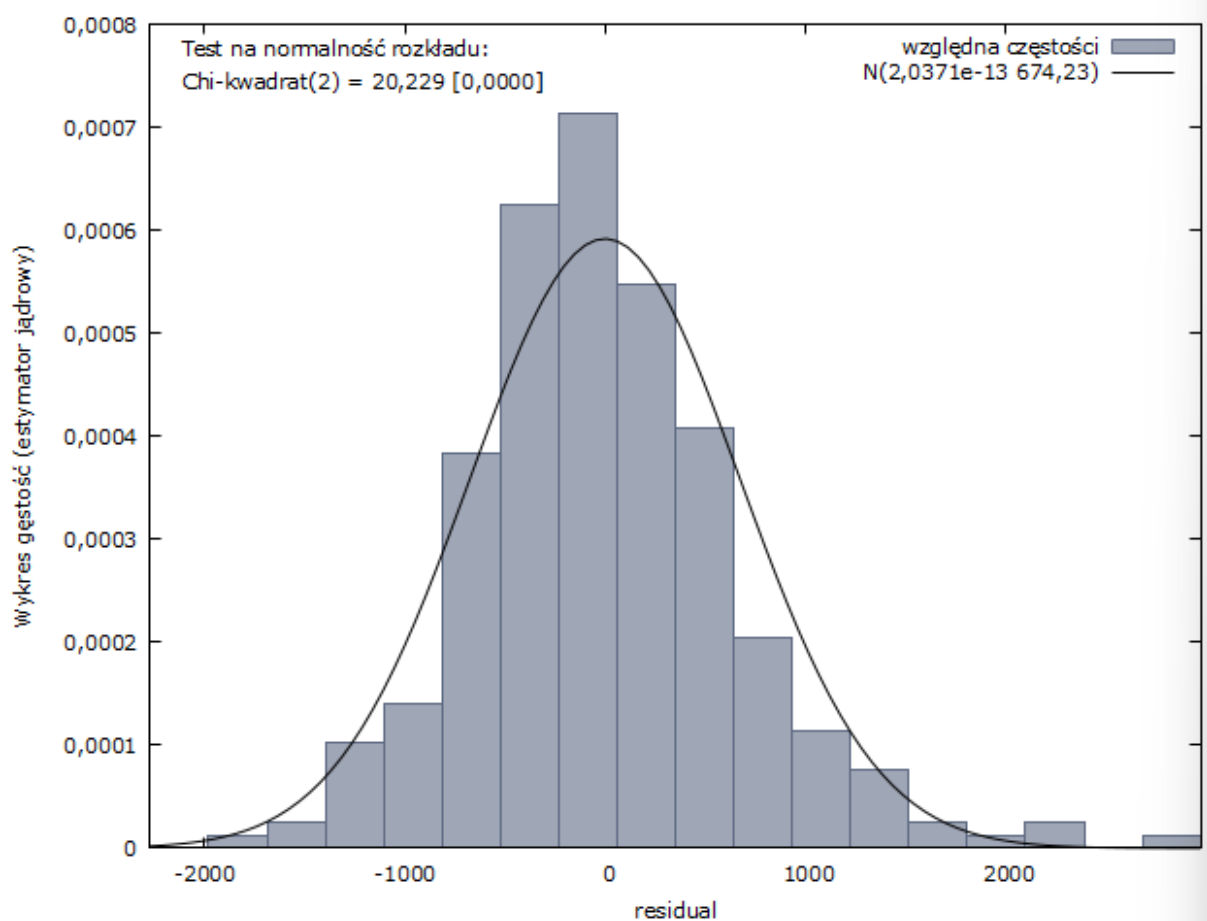
Według skryptu występuje brak katalizatorów a efekt katalizy wynosi 0,01116869, ale ze względu na brak katalizatorów nie interpretuje go.

3. Normalność rozkładu składnika losowego

H0: rozkład składnika losowego jest zgodny z rozkładem normalnym.

H1: rozkład składnika losowego nie jest zgodny z rozkładem normalnym.

Test Doornika- Hansena



Wykres 2. Wykres rozkładu

Hipoteza zerowa: dystrybuanta empiryczna posiada rozkład normalny.
Test Doornika-Hansena (1994) - transformowana skośność i kurtোza.:
Chi-kwadrat (2) = 20,229 z wartością p 0,00004

Gretl 6. Test Doornika-Hansena

Zapisuje reszty modelu i przeprowadzam na nich testy na normalność rozkładu:

```
Test na normalność rozkładu uhat12:  
  
Test Doornika-Hansena = 20,2288, z wartością p 4,04922e-005  
  
Test Shapiro-Wilka = 0,971235, z wartością p 3,06308e-005  
  
Test Lillieforsa = 0,0656858, z wartością p ~= 0,01  
  
Test Jarque'a-Bera = 50,9967, z wartością p 8,43747e-012
```

Gretl 7. Testy normalności rozkładu reszt

W obu przypadkach wychodzi na to, że występuje brak normalności rozkładu składnika losowego.

4. Istotność zmiennych

H0: Wszystkie parametry nieistotnie różnią się od 0.

H1: Co najmniej jeden parametr istotnie różni się od 0.

Istotność łączna

```
? ols Y const X1 X2 X3 --quiet  
? scalar R2 = $rsq  
Wygenerowano skalar R2 = 0,550303  
? scalar df = $df  
Wygenerowano skalar df = 265  
? scalar F_statystyka = (R2/(1-R2))*(df/2)  
Wygenerowano skalar F_statystyka = 162,143  
? pvalue F 2 df F_statystyka  
F(2, 265): prawostronny obszar krytyczny dla 162,143 = 1,02785e-046  
(lewostronny obszar krytyczny: 1)
```

Gretl 8. Skrypt z zajęć na test na istotność łączną

Na podstawie tych informacji można zinterpretować, że w analizowanym modelu regresji istnieje istotna statystycznie

zależność między zmiennymi objaśniającymi a zmienną zależną, ponieważ wartość testu F jest bardzo mała ($1,02785e-046$). Odrzucenie hipotezy zerowej wskazuje, że co najmniej jedna zmienna objaśniająca ma istotny wpływ na zmienną zależną.

Istotność pojedyncza

Zmienna	Statystyka t-studenta
X1	6,688
X2	9,466
X3	4,543

Tabela 3. Statystyki t-studenta zmiennych

Test T-studenta

$H_0: \alpha_j = 0$

$H_1: \alpha_j \neq 0$

Z tablic, $t\text{-student}(0.05, 265) = \sim 1.97$.

Każda zmienna ma statystykę większą od 1.97, więc przyjmują hipotezę alternatywną, która mówi, że w każdym przypadku zmienna jest istotna.

5. Test dodanych zmiennych

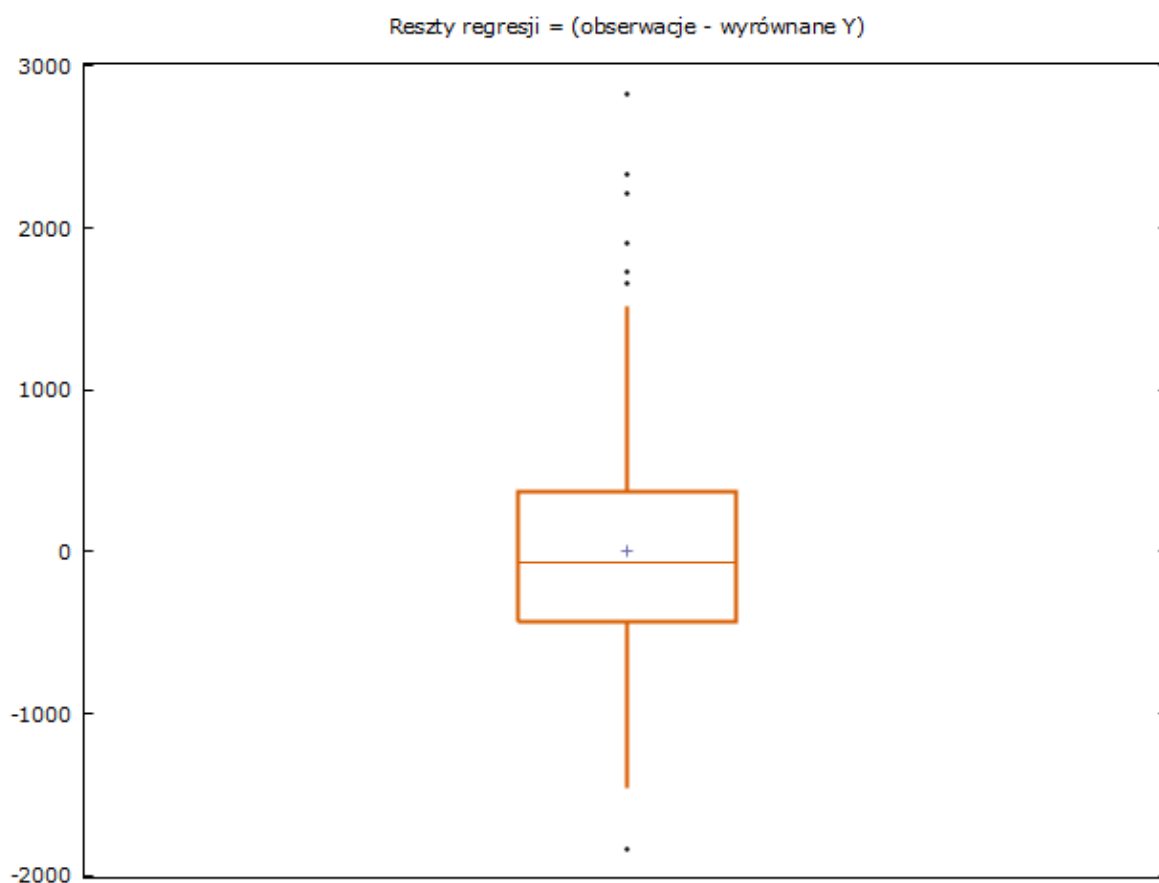
Hipoteza zerowa: parametry regresji dla wskazanych zmiennych są równe X4, X5				
Statystyka testu: $F(2, 263) = 0,485$, wartość p 0,616247				
Dodanie zmiennych poprawiło 0 z 3 kryteriów informacyjnych (AIC, BIC,				
Model 24: Estymacja KMNK, wykorzystane obserwacje 1-269				
Zmienna zależna (Y): Y				
	współczynnik	błąd standardowy	t-Studenta	wartość p
const	-121,632	127,356	-0,9551	0,3404
X1	82,8619	12,4649	6,648	1,71e-010 ***
X2	86,7911	15,2371	5,696	3,28e-08 ***
X3	92,6760	22,5464	4,110	5,28e-05 ***
X4	27,3695	30,3353	0,9022	0,3678
X5	-10,4452	12,2534	-0,8524	0,3947
Średn.aryt.zm.zależnej	1423,828	Odch.stand.zm.zależnej	999,7741	
Suma kwadratów reszt	1,20e+08	Błąd standardowy reszt	675,5416	
Wsp. determ. R-kwadrat	0,551955	Skorygowany R-kwadrat	0,543437	
F(5, 263)	64,79899	Wartość p dla testu F	6,71e-44	
Logarytm wiarygodności	-2131,334	Kryt. inform. Akaike'a	4274,668	
Kryt. bayes. Schwarza	4296,236	Kryt. Hannana-Quinna	4283,330	

Gretl 9. Test dodanych zmiennych

Jak można zauważyć dodanie zmiennych X4 i X5 nie poprawiła statystyk modelu, więc ostateczna postać została dobrze dobrana.

6. Obserwacje odstające

Do analizy obserwacji odstających posłużę się wykresem pudełkowym.



Wykres 3. Wykres pudełkowy reszt

Wykres pokazuje obecność obserwacji odstających.

7. Test liczby serii

Przeprowadzę teraz test liczby serii uprzednio sortując dane rosnąco według X1.

H0: Dobór jednostki do próby jest losowy, model jest liniowy.

H1: Dobór jednostki do próby nie jest losowy, model jest nieliniowy.

Test serii

```
Liczba serii (R) dla zmiennej 'e' = 138
Test niezależności oparty na liczbie dodatnich i ujemnych serii.
Hipoteza zerowa: próba jest losowa, dla R odpowiednio N(135,5, 8,18535),
test z-score = 0,305424, przy dwustronym obszarze krytycznym p = 0,760044
```

Gretl 10. Test liczby serii

Wartość p-value jest większa niż 0,05 więc nie ma podstaw do odrzucenia H0. Dobór jednostki do próby jest losowy a model liniowy.

8. Test Reset

H0: Postać modelu jest dobrze dobrana, model jest liniowy.

H1: Postać modelu jest źle dobrana, model jest nieliniowy.

```
Pomocnicze równanie regresji dla testu specyfikacji RESET
Estymacja KMNK, wykorzystane obserwacje 1-269
Zmienna zależna (Y): Y
```

	współczynnik	błąd standardowy	t-Studenta	wartość p
const	264,208	296,022	0,8925	0,3729
X1	25,3938	40,5219	0,6267	0,5314
X2	22,1925	39,7754	0,5579	0,5774
X3	20,9523	41,2370	0,5081	0,6118
yhat^2	0,000410121	0,000303806	1,350	0,1782
yhat^3	-6,40427e-08	5,54494e-08	-1,155	0,2491

```
Statystyka testu: F = 1,352096,
z wartością p = P(F(2,263) > 1,3521) = 0,26
```

Gretl 11. Test Reset

P-value jest na poziomie $0,26 > 0,05$ więc nie ma podstaw do odrzucenia hipotezy H_0 i postać modelu jest odpowiednio dobrana.

9. Heteroskedastyczność

Test White'a

H_0 : Reszty modelu są homoskedastyczne.

H_1 : Reszty modelu są heteroskedastyczne.

```
Statystyka testu:  $TR^2 = 18,920580$ ,  
z wartością  $p = P(\text{Chi-kwadrat}(9) > 18,920580) = 0,025877$ 
```

Gretl 12. Test White'a

Odrzucamy hipotezę zerową na rzecz alternatywnej, wychodzi na to, że występuje heteroskedastyczność.

Podjęmę się próby przekształcenia modelu w celu usunięcia z niego heteroskedastyczności a użyję do tego metody Cochrane-Orcut. Tą wersję UMNK stosuje się w przypadku wystąpienia autokorelacji pierwszego rzędu. Nazwa tej metody pochodzi od nazwisk dwóch statystyków: Donalda Cochranego i Guy'a Orcutta, a jej idea opiera się na transformacji autoregresyjnej pierwszego

rzędu zmiennej objaśnianej i zmiennych objaśniających, a następnie oszacowania wartości estymatorów KMNK.³

Transformuje zmienne poprzez dzielenie ich przez pierwiastek X3 oraz przeprowadzę testy na nowo.

Test White'a

Statystyka testu: $TR^2 = 10,605506$,
z wartością $p = P(\text{Chi-kwadrat}(13) > 10,605506) = 0,643828$

Gretl 13. Test White'a po transformacji zmiennych

Nowy model został pozbawiony heteroskedastyczności i tak on się teraz prezentuje.

	współczynnik	błąd standardowy	t-Studenta	wartość p	
nc	-71,1275	74,0841	-0,9601	0,3379	
x11	67,4425	10,8939	6,191	2,27e-09	***
x12	74,0574	8,31143	8,910	8,39e-017	***
x13	87,2513	20,4021	4,277	2,65e-05	***
Średn.aryt.zm.zależnej	692,8606	Odch.stand.zm.zależnej	421,2342		
Suma kwadratów reszt	31072472	Błąd standardowy reszt	342,4246		
Niecentrowany R-kwadr.	0,824140	Centrowany R-kwadrat	0,346578		
F(4, 265)	310,4697	Wartość p dla testu F	1,06e-98		
Logarytm wiarygodności	-1949,577	Kryt. inform. Akaike'a	3907,155		
Kryt. bayes. Schwarza	3921,533	Kryt. Hannana-Quinna	3912,929		

Gretl 14. Model po usunięciu heteroskedastyczności

10. Test Chowa

³ M.Kurpas, *Wybrane zagadnienia ekonometrii z wykorzystaniem programu Statistica*, Uniwersytet Śląski, Katowice, s. 59.

H0: Brak zmian strukturalnych (model charakteryzuje się stałością parametrów w czasie)

H1: Zmiany strukturalne występują (model nie charakteryzuje się stałością parametrów w czasie)

Test Chowa na zmiany strukturalne przy podziale próby w obserwacji 135
 $F(5, 260) = 2,11933$ z wartością p 0,0635

Gretl 15. Test Chowa

Nie zauważyłem punktu, który mógłby świadczyć o załamaniu strukturalnym dlatego dziele w połowie. P-value jest większe niż 0.05, więc nie ma podstaw do odrzucenia H0.

11. Współliniowość

Współliniowość oznacza dokładną lub bardzo wysoką korelację między regreso-rami. Dokładana korelacja jest błędem ekonometryka, który do zbioru zmiennych objaśniających wprowadził regresor lub regresory, będące kombinacją liniową innych regresorów.⁴

Przeprowadzę ocenę współliniowości, wartości >10 mogą wskazywać na jej występowanie.

VIF

X11	1,097
X12	1,047

⁴ B.R.Górecki, Ekonometria podstawy teorii i praktyki, Key Text, 2013,s.89

X13	6,919
-----	-------

Tabela 4. Wartości VIF dla zmiennych

Brak sygnałów wskazujących na współliniowość.

12. *Koincydencja*

Sprawdzę koincydencję poprzez sprawdzenie znaków parametrów modelu oraz korelacji między zmiennymi objaśniającymi a zmienną objaśnianą.

	Znak parametru	Znak korelacji
X11	67,4425 (+)	0,3355 (+)
X12	74,0574 (+)	0,4959 (+)
X13	87,2513 (+)	0,1132 (+)

Tabela 5. Sprawdzanie koincydencji porównując znaki

Model jest koincydentny, znaki się zgadzają.

13. *Parametry modelu*

Ostateczna postać modelu wygląda następująco:

$$\frac{Y}{\sqrt{X_3}} = -71,1275 * \frac{const}{\sqrt{X_3}} + 67,4425 * \frac{X_1}{\sqrt{X_3}} + 74,0574 * \frac{X_2}{\sqrt{X_3}} + 87,2513 * \frac{X_3}{\sqrt{X_3}}$$

1. Postać modelu z parametrami

67,4425

Dla każdej jednostkowej zmiany w ilości lat gry profesjonalnie w NBA ($X_1/\sqrt{X_3}$), oczekujemy wzrostu $Y/\sqrt{X_3}$ o 67.4425 jednostek, przy założeniu stałych wartości pozostałych zmiennych.

74,0574

Współczynnik 74.0574 wskazuje, że dla każdej jednostkowej zmiany $X_2/\sqrt{X_3}$, oczekujemy wzrostu $Y/\sqrt{X_3}$ o 74.0574 jednostek, przy założeniu stałych wartości pozostałych zmiennych.

87,2513

Dla każdej jednostkowej zmiany w liczbie zbiórek przez osobę na mecz ($X_3/\sqrt{X_3}$), oczekujemy wzrostu $Y/\sqrt{X_3}$ o 87.2513 jednostek, przy założeniu stałych wartości pozostałych zmiennych. Ponieważ X_3 jest również obecne w mianowniku, to oznacza, że ten efekt jest uwzględniony poprzez pierwiastek.

14. Predykcja

Wykonam predykcje punktową dla średnich wartości zmiennych i przedziału ufności 95%.

Wartość prognozy	661,759
Wariancja prognozy	119080
Błąd prognozy	345,08
Przedział ufności	[-17,6883 ; 1341,21]

Tabela 6. Wartości parametrów predykcji

Prognoza punktowa dla wartości średnich zmiennych wyniosła 661,759 i tyle wynoszą zarobki zawodnika NBA przy średnich wartościach zmiennych objaśniających.

Przedział ufności wynosi [-17,6883 ; 1341,21]. Z prawdopodobieństwem 95%, zarobki znajdą się w tym przedziale.

8. Podsumowanie

Ostateczna postać modelu wygląda w następujący sposób:

	współczynnik	błąd standardowy	t-Studenta	wartość p	
nc	-71,1275	74,0841	-0,9601	0,3379	
x11	67,4425	10,8939	6,191	2,27e-09	***
x12	74,0574	8,31143	8,910	8,39e-017	***
x13	87,2513	20,4021	4,277	2,65e-05	***
Średn.aryt.zm.zależnej	692,8606	Odch.stand.zm.zależnej	421,2342		
Suma kwadratów reszt	31072472	Błąd standardowy reszt	342,4246		
Niecentrowany R-kwadr.	0,824140	Centrowany R-kwadrat	0,346578		
F(4, 265)	310,4697	Wartość p dla testu F	1,06e-98		
Logarytm wiarygodności	-1949,577	Kryt. inform. Akaike'a	3907,155		
Kryt. bayes. Schwarza	3921,533	Kryt. Hannana-Quinna	3912,929		

Gretl 16. Parametry modelu

Ostatecznie model został z 3 transformowanymi zmiennymi objaśniającymi oraz transformowanym wyrazem wolnym i transformowana zmienną objaśnianą. Parametry modelu mają umiarkowane błędy standardowe.

Rzuca się w oczy fakt, że brak tu standardowego wyrazu wolnego stąd pojawia się niecentrowany współczynnik R^2 , który wynosi 0,824140 co oznacza, że zmienne egzogeniczne mają duży wpływ na wyjaśnianie zmienności zmiennej objaśnianej. Centrowany R^2 za to wynosi 0,346578.

Kryteria informacyjne są mniejsze niż w przypadku modelu na początku co jest dobrą informacją.

Weryfikacja hipotez:

1. Istnieje zależność między liczbą lat gry profesjonalnie a zarobkami gracza NBA.

Hipoteza jest prawdziwa, występuje zależność między liczbą lat gry profesjonalnie a zarobkami gracza NBA.

2. Istnieje zależność między ilością zdobytych punktów na mecz a zarobkami gracza NBA.

Hipoteza jest prawdziwa, występuje zależność między ilością zdobytych punktów na mecz a zarobkami gracza NBA.

3. Istnieje zależność między liczbą zdobytych zbiórek na mecz a zarobkami gracza NBA.

Hipoteza jest prawdziwa, występuje zależność między liczbą zdobytych zbiórek na mecz a zarobkami gracza NBA.

4. Istnieje zależność między liczbą zdobytych asyst na mecz a zarobkami gracza NBA.

Hipoteza nie jest prawdziwa, nie występuje zależność między liczbą zdobytych asyst na mecz a zarobkami gracza NBA.

5. Istnieje zależność między średnią ilością minut na boisku na grę a zarobkami gracza NBA.

Hipoteza nie jest prawdziwa, nie występuje zależność między ilością minut na boisku na grę a zarobkami gracza NBA.