

# Determinanty zarobków graczy baseballowych

28-06-2021

Mateusz Smolnik

Analiza danych w R

## 1. Opis projektu

Celem projektu było zbadanie wpływu poszczególnych zmiennych na zarobki baseballistów. W projekcie zostały zbadane następujące tezy:

- Baseballiści ligi Amerykańskiej zarabiają średnio więcej od baseballistów ligi Narodowej,
- Zarobki graczy baseballowych zależą od ich doświadczenia oraz wieku,
- Liczba rozegranych meczów w poprzednich latach wpływa na zarobki baseballistów,
- Liczba uderzeń w poprzednich latach wpływa na zarobki baseballistów,
- Liczba uderzeń, po których baseballista zdobył wszystkie bazy wpływa na zarobki.

## 2. Opis danych

Badania zostało przeprowadzone na podstawie danych pochodzących z pakietu Lahman dla okresu od 2006 r. do 2016 r.

Poniżej zostały zaprezentowane statystyki opisowe dla zarobków baseballistów.

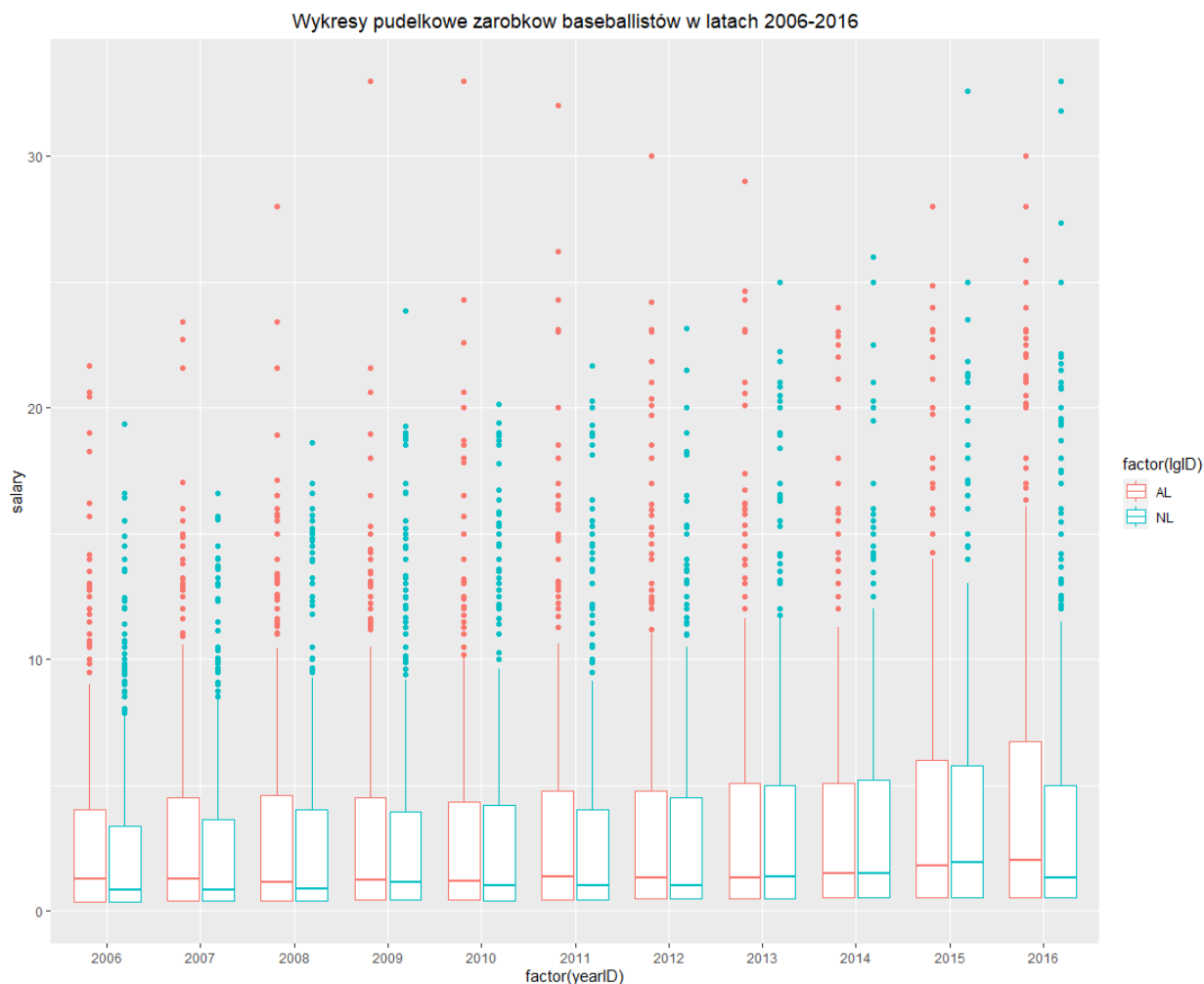
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
327000	497500	1200000	3511876	4650000	33000000

	item	group1	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
x11	1	2006	1	819	2834521	3778385	980000	1992637	966655.2	327000	21680727	21353727	2.04	4.23	132027.5
x12	2	2007	1	842	2941436	3810572	1000000	2099338	919212.0	380000	23428571	23048571	1.94	3.81	131321.0
x13	3	2008	1	856	3136517	4193269	1000000	2199649	904089.5	390000	28000000	27610000	1.94	3.78	143323.0
x14	4	2009	1	813	3277647	4371688	1225000	2300296	1221291.8	400000	33000000	32600000	2.08	5.03	153321.8
x15	5	2010	1	830	3278747	4461190	1093750	2251762	1025588.6	400000	33000000	32600000	2.17	5.34	154850.2
x16	6	2011	1	839	3318838	4541140	1200000	2270631	1162951.4	414000	32000000	31586000	2.24	5.63	156777.6
x17	7	2012	1	848	3458421	4710311	1100000	2370807	918526.3	480000	30000000	29520000	2.16	4.82	161752.8
x18	8	2013	1	815	3723344	4963773	1350000	2596613	1272812.1	480000	29000000	28520000	2.03	4.01	173873.4
x19	9	2014	1	802	3980446	5155339	1500000	2834304	1479634.8	500000	26000000	25500000	1.89	3.21	182041.4
x110	10	2015	1	817	4301276	5506178	1880000	3110279	2027455.5	507000	32571000	32064000	1.90	3.40	192636.8
x111	11	2016	1	853	4396410	5900036	1500000	3055566	1469627.2	507500	33000000	32492500	1.98	3.70	202013.4

Obserwując powyższe statystyki opisowe można zauważyć, że średnie wynagrodzenia zawodników baseballowych stale rosną. Dodatnia wartość kurtozy wskazuje na rozkład leptokurtyczny danych i większą częstotliwość występowania wartości skrajnych. Współczynnik skośności świadczy o tym, że

do czynienia mamy z danymi, które charakteryzują się asymetrią prawostronną, dominują niskie wartości wynagrodzeń.

### 3. Analiza danych – wykresy



Analizując powyższy wykres można zauważyć niewielką różnicę w zarobkach pomiędzy poszczególnymi ligami. Mediany zarobków są w większości lat do siebie zbliżone. Warto również zauważyć, że gdy widoczne są różnice w medianach to są one na korzyść ligi Amerykańskiej. Również w tej lidze obserwowane są wyższe wartości maksymalne zarobków. Największą różnicę można zaobserwować w roku 2016 na korzyść ligi amerykańskiej, kiedy to zarówno mediana jak i trzeci kwartyl były wyższe niż w lidze Narodowej. W następnym kroku zostały zbadane średnie zarobki w obu ligach przy użyciu testu t-studenta.

Test t-studenta dla istotności dwóch średnich (dla lat 2006-2016.)

```
> t.test(salariesAL$salary, salariesNL$salary, alternative="greater", mu=0,
+        var.equal=T)

Two Sample t-test

data:  salariesAL$salary and salariesNL$salary
t = 4.0132, df = 9132, p-value = 3.018e-05
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.2348811      Inf
sample estimates:
mean of x mean of y
 3.720176  3.322141
```

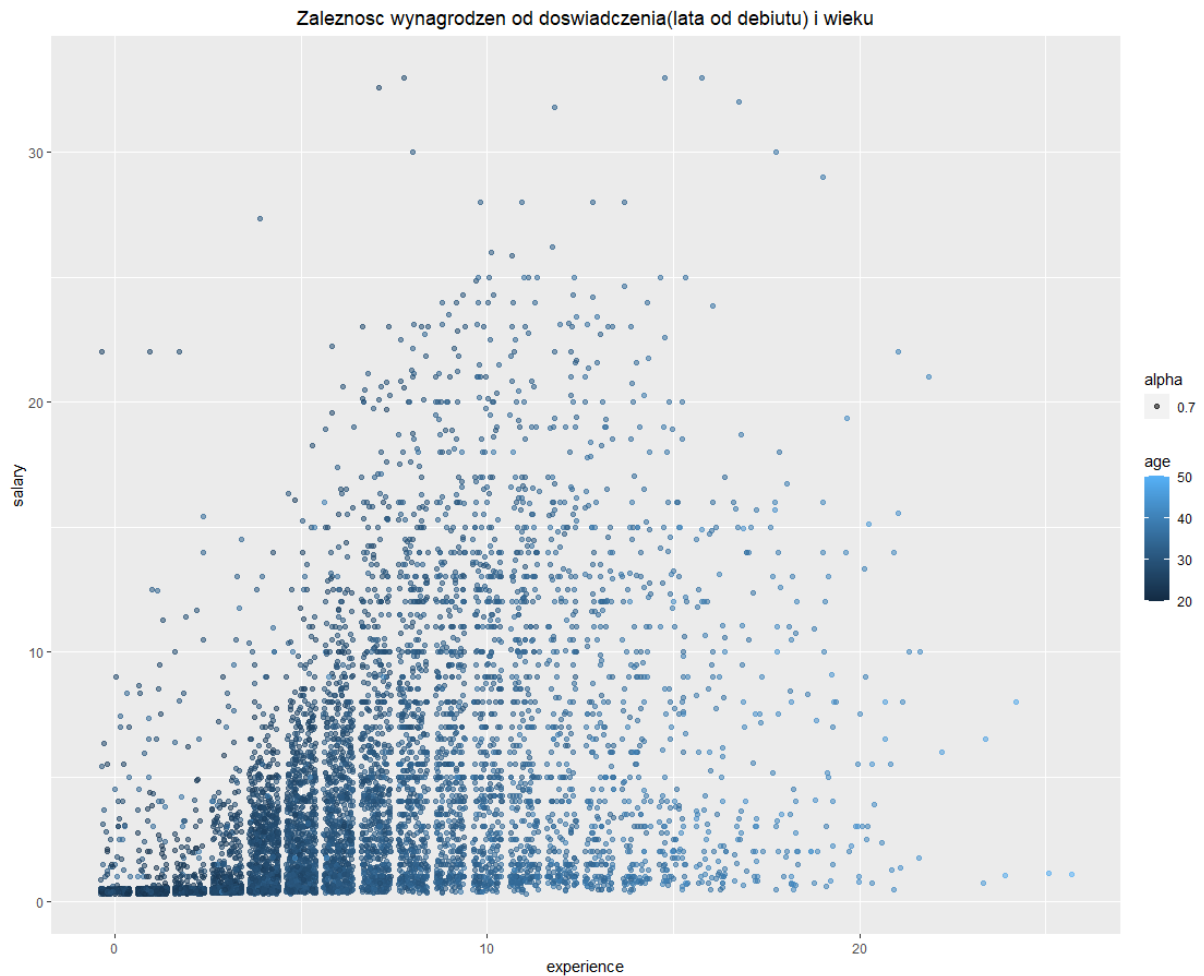
Test t-studenta dla istotności dwóch średnich (w roku 2016)

```
> t.test(salaries2016AL$salary, salaries2016NL$salary, alternative="greater", mu=0,
+        var.equal=T)

Two Sample t-test

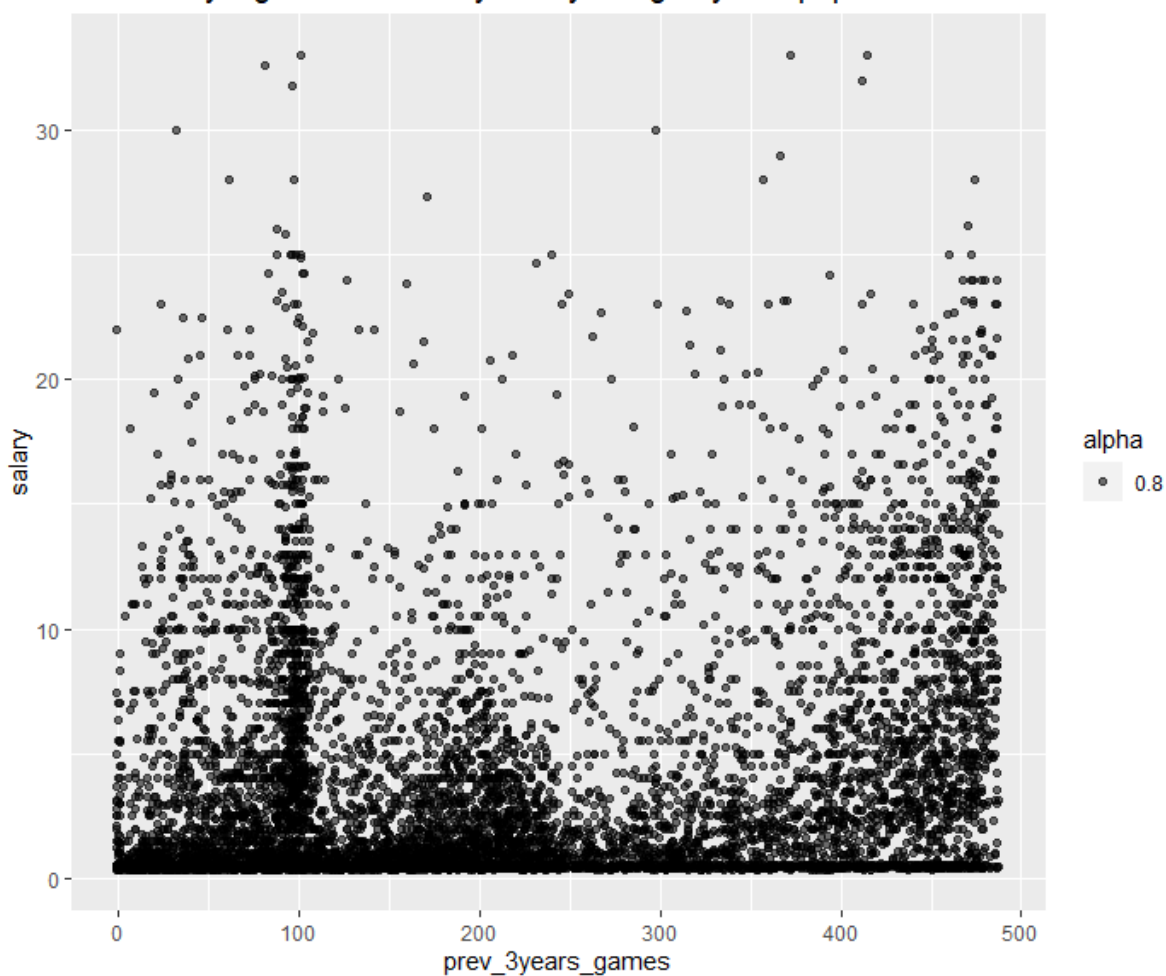
data:  salaries2016AL$salary and salaries2016NL$salary
t = 1.9211, df = 851, p-value = 0.02752
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.1107366      Inf
sample estimates:
mean of x mean of y
 4.788923  4.013890
```

Wyniki obu testów t-studenta wykazały, że na poziomie istotności 0,05, zarobki w lidze Amerykańskiej są istotnie wyższe niż w lidze Narodowej.



Na powyższym wykresie można zauważyć wyraźną zależność wysokości wynagrodzenia od lat doświadczenia. Im większe doświadczenie tym wyższe są zarobki danego zawodnika. Warto również zauważyć, że wraz z wiekiem zawodnika w pewnym momencie wysokość wynagrodzeń zaczyna spadać. Analizując te dwa czynniki można wnioskować, że na najwyższe zarobki mogą liczyć młodzi i doświadczeni zawodnicy w szczycie kariery. Natomiast wraz ze zbliżającym się końcem kariery ich zarobki maleją.

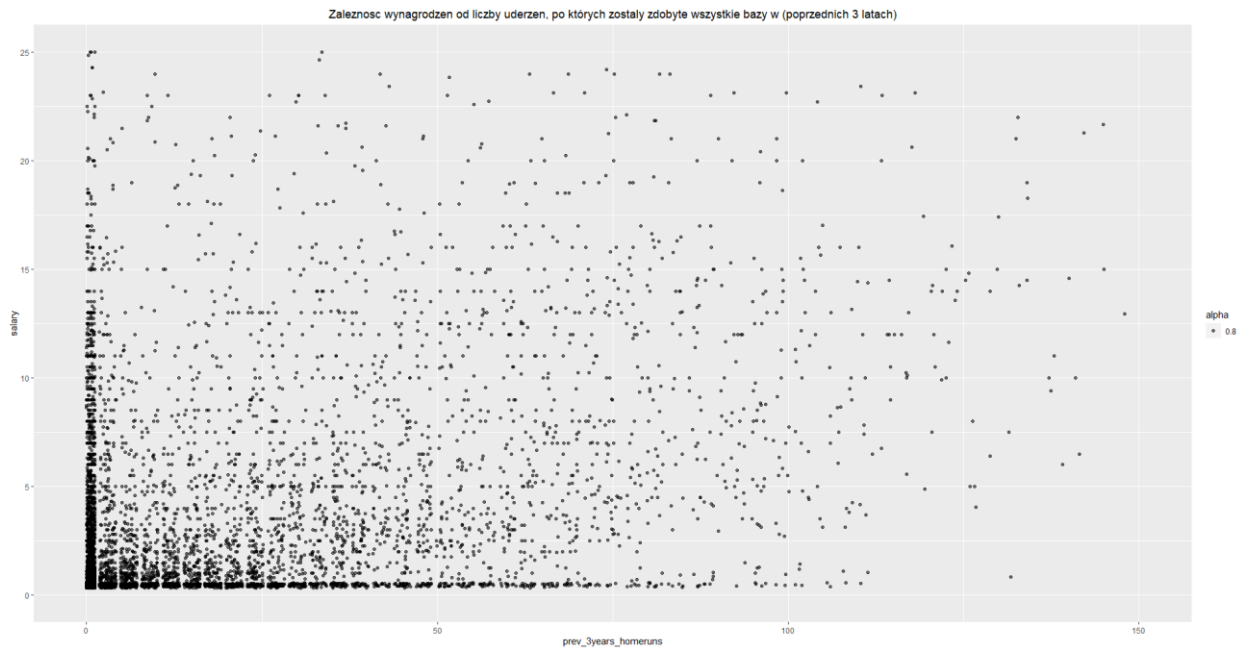
Zależność wynagrodzeń od liczby meczy rozegranych w poprzednich 3 latach



Analizując wykres powyżej ciężko jest jednoznacznie wnioskować czy liczba rozegranych spotkań przez gracza wpływa na jego wynagrodzenia. Nieoczekiwanie obserwujemy tu duże skupienie wysokich wynagrodzenie, dla graczy, którzy rozegrali około 100 lub 200 spotkań w ciągu ostatnich lat. Natomiast wraz z wzrostem liczby rozegranych meczy powyżej 300 zaobserwować możemy wyraźny spadek skupienia niskich wynagrodzeń, widać tu również pewną zależność, im więcej rozegranych meczy tym większa jest szansa na wyższe zarobki.



Na powyższym wykresie można zaobserwować wpływ liczby uderzeń w poprzednich latach na wysokość wynagrodzeń. Zależność ta jest widoczna szczególnie po przekroczeniu 200 uderzeń.



Obserwując powyższy wykres nie widać wyraźniej zależności między liczbą uderzeń, po których zostały zdobyte wszystkie bazy a wynagrodzeniami. Uznaję, że zmienna ta nie ma wpływu na wysokość pensji baseballisty.

## 4. Wnioski końcowe

Podsumowując powyższe obserwacje, można stwierdzić, że wynagrodzenie baseballistów jest zależne od wielu czynników. Na średnio większe zarobki mogą liczyć baseballiści ligi Amerykańskiej. Na zarobki zawodników wpływa również ich doświadczenie i wiek, a najwięcej zarabiają młodzi zawodnicy z kilkuletnim doświadczeniem. Zaobserwowano również wpływ liczby uderzeń w poprzednich 3 latach na zarobki zawodników, wraz ze wzrostem liczby uderzeń rośnie średnie wynagrodzenie, a zależność ta jest szczególnie widoczna powyżej 200 uderzeń. Niewątpliwie na pensje graczy baseballowych wpływa również liczba meczy rozegranych w poprzednich sezonach, co ciekawe, nie tylko gracze, którzy rozegrali większość meczy w ostatnich 3 latach mogą liczyć na wyższe zarobki, ale również zawodnicy, którzy przez ostatnie 3 lata rozegrali około 100 lub 200 spotkań. Nie stwierdzono natomiast zależności między liczbą uderzeń, po których baseballista zdobył wszystkie bazy a wysokością wynagrodzenia.

Przy budowie modelu powinny zostać uwzględnione następujące zmienne:

- liga,
- doświadczenie oraz wiek,
- liczba rozegranych meczy w poprzednich latach,

- liczba uderzeń w poprzednich latach.