

Spis treści

| | | |
|----------|--|----------|
| 1 | Cel | 2 |
| 2 | Zbiór danych | 2 |
| 2.1 | Opis zmiennych | 2 |
| 2.2 | Wybór zmiennych zależnych i niezależnych | 3 |
| 3 | Dopasowanie modelu | 4 |
| 3.1 | Model uwzględniający cechy motoryczne zawodników | 4 |
| 3.1.1 | Model WWW | 4 |
| 3.1.2 | Model WW | 4 |
| 3.1.3 | Model WB | 5 |
| 3.1.4 | Porównanie modeli | 5 |

1 Cel

Celem prowadzonych badań jest odszukanie występowania zależności między statystykami, w koszykówce, przy użyciu modeli regresji.

2 Zbiór danych

Przedmiotem badań są statystyki poszczególnych zawodników występujących na parkietach NBA w sezonie 2014/2015. W omawianym okresie, w lidze wystąpiło 490 zawodników, co przekłada się na taką samą długość próbki. Polecenie wymaga próbki o długości co najmniej 500 pomiarów, jednak w związku z atrakcyjnością omawianych danych zdecydowaliśmy się na ich użycie. Dane pochodzą ze

2.1 Opis zmiennych

Zmienne kategoryczne wyróżnione są poprzez nawiasy zawierające wartości, które zmienna może przyjąć. W sprawozdaniu zostały użyte następujące zmienne:

- „Name” – imię oraz nazwisko,
- „Age” – wiek,
- „Birth Place” – miejsce urodzenia (US, NONUS),
- „Height” – wzrost,
- „Pos” – pozycja na boisku (PG, SG, C, PF, SF),
- „Team” – ostatni zespół, w którym grał zawodnik,
- „Weight” – waga,
- „BMI” – wskaźnik Body Mass Indeks,
- „Games Played” - liczba rozegranych meczy,
- „MIN” – liczba minut na boisku,
- „PTS” – liczba zdobytych punktów w sezonie,
- „FGM” –
- „FGA” –
- „FG p” –
- „ThreePM” –

- „3PA” –
- „ThreeP p” –
- „FTM” –
- „FTA” –
- „FT p” –
- „OREB” –
- „DREB” –
- „REB” – liczba zbiórek,
- „AST” – liczba asyst,
- „STL” – liczba przejęć,
- „BLK” – liczba bloków,
- „TOV” – liczba strat,
- „PF” –
- „EFF” –
- „AST/TOV” –
- „STL/TOV” –
- „PPG” – liczba zdobytych punktów podzielona przez liczbę rozegranych meczy,
- „MPG” – liczba rozegranych minut podzielona przez liczbę rozegranych meczy.

2.2 Wybór zmiennych zależnych i niezależnych

Zmienną zależną jest liczba punktów zdobytych w sezonie podzielona przez liczbę rozegranych meczy. Do zmiennych niezależnych należą !!!! UZUPEŁNIC ZMIENNE NIEZALEŻNE !!!!. Decydujemy się na taki wybór, gdyż chcemy poznać wpływ poszczególnych statystyk na liczbę zdobywanych punktów. (?) Odnalezienie odpowiedniego modelu będzie podstawą do wyznaczania zawodników pretendujących do bycia najlepiej punktującymi w kolejnych sezonach. (?)

3 Dopasowanie modelu

3.1 Model uwzględniający cechy motoryczne zawodników

Podjęliśmy trzy próby dobrania najlepszego modelu, który uzależniony jest od fizyczności zawodników. Oczywiście model musi być rozszerzony o ilość rozegranych minut na mecz, gdyż jest niezbędną informacją do analizy modelu.

3.1.1 Model WWW

Pierwszy z dopasowywanych modeli uwzględnia wzrost, wiek oraz wagę zawodników. Zmienne istotne dla dobranego modelu to minuty na mecz (jak zakładaliśmy wyżej, jest to najważniejszy czynnik) oraz waga. W tabeli 1 przedstawione zostały wartości współczynników oraz pozostałe statystyki dla omawianego modelu.

| Zmienna | Estymacja | Błąd standardowy | T-wartość | $\Pr(< t)$ |
|----------------|-----------|------------------|-----------|--------------|
| Minuty na mecz | 0.54219 | 0.01242 | 43.638 | $<2e-16$ |
| Wzrost | -0.03394 | 0.02154 | -1.576 | 0.1156 |
| Wiek | -0.01660 | 0.02667 | -0.622 | 0.5340 |
| Waga | 0.02721 | 0.01462 | 1.861 | 0.0634 |
| Stała | 1.63120 | 3.35589 | 0.486 | 0.6271 |

Tabela 1: Model WWW - współczynniki modelu

3.1.2 Model WW

Drugi z dopasowywanych modeli uwzględnia wiek oraz wagę zawodników. Odjęcie wzrostu ma na celu sprawdzenie jak zachowa się model po odjęciu jednej zmiennej nie będącej zmienną istotną. W tabeli 2 przedstawione zostały współczynniki modelu WW. Jak można zauważyć zmienne istotne dla modelu nie są już takie same, do minut na mecz dołączyła wartość stała. Ponadto waga zawodnika przestała należeć do istotnych zmiennych.

| Zmienna | Estymacja | Błąd standardowy | T-wartość | $\Pr(< t)$ |
|----------------|-----------|------------------|-----------|--------------|
| Minuty na mecz | 0.54356 | 0.01241 | 43.78 | $<2e-16$ |
| Wiek | -0.01304 | 0.02661 | -0.490 | 0.62428 |
| Waga | 0.00852 | 0.00857 | 0.994 | 0.32059 |
| Stała | -3.35012 | 1.13018 | -2.964 | 0.00318 |

Tabela 2: Model WW - współczynniki modelu

3.1.3 Model WB

Trzeci z dopasowywanych modeli uwzględnia wiek oraz wskaźnik BMI. Omawiany wskaźnik jest popularną statystyką mówiącą o stanie zdrowia fizycznego i wyliczana jest na podstawie poniższego wzoru:

$$BMI = \frac{waga}{wzrost^2}.$$

W związku z tym niejawnie do modelu włączamy wagę i wzrost zawodników ligi. Wyniki tego zabiegu przedstawione są w tabeli 3. Podobnie jak w modelu WW istotne zmienne to liczba minut na boisku oraz stała.

| Zmienna | Estymacja | Błąd standardowy | T-wartość | Pr(< t) |
|----------------|-----------|------------------|-----------|-----------|
| Minuty na mecz | 0.54259 | 0.01238 | 43.838 | <2e-16 |
| Wiek | -0.01608 | 0.02667 | -0.603 | 0.54676 |
| BMI | 0.09639 | 0.05878 | 1.640 | 0.10166 |
| Stała | -4.84181 | 1.60703 | -3.013 | 0.00272 |

Tabela 3: Model WB - współczynniki modelu

3.1.4 Porównanie modeli

Znając współczynniki dobranych modeli wyznaczyliśmy:

- skorygowany współczynnik determinacji, oznaczany jako R^2 ,
- kryterium Akaikego, oznaczane jako AIC,
- logarytm wskaźnika wiarygodności, oznaczany jako LogLik,
- wartość w z testu Shapiro-Wilka, oznaczana jako W.

| Model | R^2 | AIC | LogLik | W |
|-------|--------|----------|-----------|---------|
| WWW | 0.7977 | 2282.972 | -1135.486 | 0.95495 |
| WW | 0.7971 | 2283.475 | -1136.737 | 0.95403 |
| WB | 0.7978 | 2281.766 | -1135.883 | 0.95493 |

Tabela 4: Porównanie modeli

Każdy z dopasowanych modeli jest „poprawnym modelem”, gdyż ich residua mają rozkład normalny. Sprawdzenia tej hipotezy dokonaliśmy przy użyciu testu Shapiro-Wilka (długość naszej próbki nie przekracza 5000). Wartość W jest mniejsza od liczby 0.947 (wartość zaczerpnięta z tablic), co oznacza, że nie ma podstaw by odrzucić hipotezę o normalności residuów. W tabeli 4 widać, że model WB cechuje się najlepszym dopasowaniem do

danych według współczynnika R^2 . Kryterium informacyjne Akaikego również przemawia za wybraniem trzeciego modelu jako najlepiej opisującego badany zbiór wartości. Jedyną statystyką przemawiającą za innym modelem niż WB jest logarytm wskaźnik wiarygodności, który wskazuje, że to pierwszy model jest najlepszy. Na podstawie zaprezentowanych wyników należy wybrać model uwzględniający wiek oraz BMI zawodników.