

Sprawozdanie 1

Emil Olszewski, Jakub Kempa

18 stycznia 2024

Streszczenie

Przedmiotem analizy są dane ze zbioru zawierającego informacje na temat męskich trójboistów zrzeszonych w ramach federacji IPF. Dane zostały udostępnione na warunkach licencji GNU AGPLv3. Głównymi zmiennymi, które będą nas interesować będą *Age*, *Sex* (zmienne kategoryczne określające wiek i płeć zawodnika) oraz zmienne ciągłe *BodyweightKg* oraz *TotalKg*, które wyrażają odpowiednio masę ciała zawodnika oraz wynik całkowity, będący sumą wyników w trzech kategoriach: *wyciskanie na ławce*, *przysiad ze sztangą*, *martwy ciąg*.

Do analizy danych użyto języka R.

1 Opis danych

Pod uwagę wzięto tylko zawodników płci męskiej, dla których dostępny był pełen zestaw danych dotyczący wyników uzyskanych w każdym z trzech bojów. Ograniczono się dodatkowo do cenzusu wiekowego w przedziale od 16 do 40 lat oraz rozpatrywano tylko wyniki uzyskane w kategorii RAW (kategoria, która zabrania używania sprzętu dającego przewagę mechaniczną np. koszulek do wyciskania, kaftanów itd. Jest to klasyczna kategoria trójboju siłowego). Skoncentrowano się na sześciu kluczowych zmiennych:

- **Age (wiek zawodnika):** Ta zmienna kategoryczna reprezentuje wiek zawodnika. Choć intuicyjnie może się wydawać, że jest ona bardzo istotna (wraz z wiekiem vitalność sportowa powinna spadać), tak jednak w weightliftingu wiek nie jest kluczowy. Wykażą to późniejsze analizy w sprawozdaniu.
- **Sex (płeć zawodnika):** Płeć jako zmienna kategoryczna naturalnie istotnie wpływa na wyniki. Kobiety nie będą osiągać tak samo wysokich wyników, w tych samych kategoriach, co mężczyźni. Nasza analiza będzie skupiać się jedynie na mężczyznach.
- **BodyweightKg (masa ciała zawodnika):** Ta zmienna ciągła reprezentuje masę ciała zawodnika w kilogramach. Masa ciała jest istotnym parametrem w trójboju siłowym, ponieważ klasyfikuje zawodników w odpo-

wiednie kategorie wagowe i może wpływać na ich wydajność w zawodach.

- **TotalKg (Całkowity wynik):** Jako zmienna ciągła, całkowity wynik odnosi się do sumy maksymalnych ciężarów, które zawodnik podniósł w trzech próbach: przysiadzie, wyciskaniu leżąc i martwym ciągu. Jest to główny wskaźnik wydajności w trójboju siłowym, odzwierciedlający siłę i umiejętności zawodnika. W dalszej części raportu będziemy używać określeń takich jak **Wynik sumaryczny**, **Wynik total** czy po prostu **total**.

Wiersze, w których w naszych kategoriach pojawiają się braki danych usuwamy. Postanowiliśmy z otrzymanych danych wydzielić losową próbę o długości 1500. Poniżej zamieszczone są informacje o naszych danych.

```
## -- Attaching core tidyverse packages ----- tidyverse
2.0.0 --
## v dplyr      1.1.3      v readr      2.1.4
## v forcats    1.0.0      v stringr    1.5.0
## v ggplot2    3.4.3      v tibble     3.2.1
## v lubridate  1.9.3      v tidyr      1.3.0
## v purrr      1.0.2
## -- Conflicts ----- tidyverse_conflicts()
--
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to
force all conflicts to become errors
##
## Attaching package: 'dbplyr'
##
##
## The following objects are masked from 'package:dplyr':
##
##      ident, sql
```

	Age	BodyweightKg	TotalKg
## Min.	: 8.00	Min. : 29.85	Min. : 97.5
## 1st Qu.:	19.00	1st Qu.: 74.20	1st Qu.: 440.0
## Median	: 23.00	Median : 85.62	Median : 510.0
## Mean	: 25.61	Mean : 87.85	Mean : 511.5
## 3rd Qu.:	30.00	3rd Qu.: 98.00	3rd Qu.: 587.5
## Max.	: 72.00	Max. : 196.10	Max. : 925.0