

Analiza wyników w trójboju

Emil Olszewski, Jakub Kempa

2024-01-17

1. Wstęp

Przedmiotem analizy są dane ze zbioru zawierającego informacje na temat trójboistów zrzeszonych w ramach federacji IPF. Dane zostały udostępnione na warunkach licencji **GNU AG-PLv3**. Głównymi zmiennymi, które będą nas interesować są **Age**, **Sex** (zmienne katégoryczne określające wiek i płeć zawodnika) oraz zmienne ciągłe **BodyweightKg**, **Best3BenchKg**, **Best3SquatKg**, **Best3DeadliftKg** i **TotalKg**, które wyrażają masę ciała zawodnika, wyniki w poszczególnych bojach (wyciskanie na ławce, przysiad ze sztangą oraz martwy ciąg), jak i wynik całkowity.

1.1 Pytania badawcze

W ramach analizy postaramy się odpowiedzieć na następujące pytania:

1. Czy istnieje zależność między wagą a wynikami w poszczególnych bojach?
2. Jakie parametry opisują rozkłady poszczególnych zmiennych?
3. W jaki sposób różnią się rozkłady wyników oraz wagi w zależności od płci?

Ładowanie danych

Wpierw przystąpimy do załadowania potrzebnych bibliotek i załadowania danych

```
library(tidyverse)
library(readxl)
dane <- read_excel("../dane.xlsx")

dim(dane)
```

```
[1] 158266      41
```

Jak widać mamy doczynienia z bardzo dużym zbiorem danych. Tym samym w celu dalszej analizy wygenerujemy losową próbkę o długości 5000

```
probka <- sample_n(dane, 5000)
```

Tak się prezentuje 10 początkowych obserwacji

```
head(probka)
```

```
# A tibble: 6 x 41
  Name      Sex  Event Equipment Age  AgeClass BirthYearClass Division
<chr>    <chr> <chr> <chr>    <chr> <chr>    <chr>        <chr>
1 Kurtis Waring M    SBD  Raw      <NA> <NA>    <NA>        retronov
2 Joshua Cary  M    SBD  Raw      28.5  24-34   24-39        MR-0
3 Keith Caton  M    SBD  Raw      41.0  40-44   40-49        Masters~
4 G. Esper     M    SBD  Raw      <NA> <NA>    <NA>        R Open
5 Daniel Loro Rodr~ M    SBD  Raw      17.5  18-19   14-18        Open
6 Parker Jayne M    SBD  Raw      21.5  20-23   19-23        MR-Jr
# i 33 more variables: BodyweightKg <chr>, WeightClassKg <chr>, Squat1Kg <chr>,
#   Squat2Kg <chr>, Squat3Kg <chr>, Squat4Kg <chr>, Best3SquatKg <chr>,
#   Bench1Kg <chr>, Bench2Kg <chr>, Bench3Kg <chr>, Bench4Kg <chr>,
#   Best3BenchKg <chr>, Deadlift1Kg <chr>, Deadlift2Kg <chr>,
#   Deadlift3Kg <chr>, Deadlift4Kg <chr>, Best3DeadliftKg <chr>, TotalKg <chr>,
#   Place <dbl>, Dots <chr>, Wilks <chr>, Glossbrenner <chr>, Goodlift <chr>,
#   Tested <chr>, Country <chr>, State <chr>, Federation <chr>, ...
```

Interesować nas będą tylko zmienne **Age**, **Sex**, **BodyWeightKg**, **Best3BenchKg**, **Best3DeadliftKg**, **Best3SquatKg**, **TotalKg**

```
probka <- probka[, c("Age", "Sex", "BodyweightKg", "Best3BenchKg", "Best3SquatKg", "Best3D
```

```
summary(probka)
```

Age	Sex	BodyweightKg	Best3BenchKg
Length:5000	Length:5000	Length:5000	Length:5000
Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character
Best3SquatKg	Best3DeadliftKg	TotalKg	

Length:5000	Length:5000	Length:5000
Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character

Napotykamy pierwszy problem. Wszystkie zmienne są typu character, gdy powinny być typu

- numeric dla **Age, BodyweightKg, Best3BenchKg, Best3SquatKg, Best3DeadliftKg, TotalKg**
- factor dla **Sex**

Dokonajmy więc konwersji

```

probka$Age      <- as.numeric(probka$Age)
probka$BodyweightKg <- as.numeric(probka$Age)
probka$Best3BenchKg <- as.numeric(probka$Best3BenchKg)
probka$Best3SquatKg <- as.numeric(probka$Best3SquatKg)
probka$Best3DeadliftKg <- as.numeric(probka$Best3DeadliftKg)
probka$TotalKg    <- as.numeric(probka$TotalKg)

probka$Sex      <- as.factor(probka$Sex)

```

```
summary(probka)
```

Age	Sex	BodyweightKg	Best3BenchKg	Best3SquatKg
Min. : 8.0	M:5000	Min. : 8.0	Min. : 16.78	Min. : 24.95
1st Qu.:19.0		1st Qu.:19.0	1st Qu.:100.00	1st Qu.:150.00
Median :23.0		Median :23.0	Median :117.93	Median :180.00
Mean :25.7		Mean :25.7	Mean :118.64	Mean :177.25
3rd Qu.:29.0		3rd Qu.:29.0	3rd Qu.:140.00	3rd Qu.:207.50
Max. :89.0		Max. :89.0	Max. :250.00	Max. :380.00
NA's :979		NA's :979	NA's :1	NA's :2
Best3DeadliftKg	TotalKg			
Min. : 38.56	Min. : 80.29			
1st Qu.:182.50	1st Qu.:435.00			
Median :210.00	Median :510.00			
Mean :207.75	Mean :503.61			
3rd Qu.:237.50	3rd Qu.:577.50			
Max. :362.50	Max. :915.00			
NA's :1				