

Analiza wzrostu i wagi sportowców

Maria Nowacka i Bartłomiej Mielcarz

8 maja 2024

1 Cel ćwiczenia

Celem tego raportu jest statystyczna analiza dwóch zmiennych losowych wzrostu i wagi sportowców, aby zbadać czy istnieje między nimi korelacja, oraz żeby sprawdzić jaki rozkład mają te zmienne.

2 Źródło danych

Dane użyte w tej analizie pochodzą z:

www.kaggle.com/datasets/mysarahmadbhat/120-years-of-olympic-history?resource=download

Zostały zebrane w okresie 1896 - 2016 (większość danych z lat 1950 - 2016) i opisują pomiary wzrostu (X) i wagi (Y) dla 30181 sportowców.

3 Analiza danych

3.1 Statystyki opisowe

3.1.1 Średnia

Średnia to sposób na określenie typowej wartości w grupie danych, jak na przykład wzrost czy waga sportowców. Odpowiedni typ średniej pozwala lepiej zrozumieć rozkład i tendencje w grupie.

arytmetyczna	177.64
harmoniczna	176.97
geometryczna	177.31
ucinana	178.00
winsorowska	178.01
mediana	178.00

Table 1: średnie wartości dla wzrostu sportowców

arytmetyczna	73.75
harmoniczna	70.81
geometryczna	72.27
ucinana	73.00
winsorowska	73.00
mediana	73.00

Table 2: średnie wartości wagi sportowców

3.1.2 Mediana

Mediana to wartość środkowa, dzieląca zbiór na dwie równe części. Gdy analizujemy na przykład wzrost sportowców, mediana pokazuje wzrost środkowego sportowca.

3.1.3 Odchylenie standardowe

Odchylenie standardowe informuje nas o rozproszeniu danych wokół średniej. Im większe odchylenie standardowe, tym mniejsza jednorodność. W naszym przypadku odchylenie standardowe wynosi **10.92** dla wzrostu, a dla wagi **15.00**.

Takie odchylenie standardowe dla wzrostu, przy średniej arytmetycznej wynoszącej około 178 można uznać za w miarę duże. Z tego wynika, że występuje duże zróżnicowanie wzrostu. Większość sportowców ma wzrost bliski okolicy średniej, ale istnieje spora liczba osób o wzroście znacznie niższym lub wyższym. W przybliżeniu można założyć, że około 68,3% wszystkich sportowców ma wzrost w przedziale od 167 cm do 189 cm. Około 95,5% sportowców będzie miało wzrost w przedziale od 156 cm do 200 cm.

Odchylenie standardowe 15 kg przy średniej 73 kg jest stosunkowo większe niż w przypadku wzrostu. To świadczy o tym, że waga sportowców jest bardziej zróżnicowana niż ich wzrost. Około 68,3% sportowców waży między 58 kg a 88 kg, a około 95,5% sportowców waży między 43 kg a 103 kg, co pokazuje szeroki zakres wagowy. Większe odchylenie standardowe w kontekście wagi może wskazywać na większe zróżnicowanie w grupie sportowców, co może być wynikiem różnic w typach ciał lub specyfikach różnych dyscyplin sportowych. Na przykład, sportowcy uprawiający dyscypliny wymagające większej masy ciała lub siły mogą ważyć znacznie więcej. To, że wartości skrajne są bardziej powszechne w przypadku wagi niż wzrostu sugeruje, że waga jest cechą zmienną i jest bardziej podatna na takie czynniki jak dieta czy trening.

4 Wizualizacje danych

4.1 Histogramy

Histogramy dla wzrostu X i wagi Y , pokazujące rozkład danych. Poniższe histogramy i ich gęstości wskazują na fakt, że wartości skrajne są znacznie bardziej widoczne w przypadku wagi, niż w przypadku wzrostu. Możemy też zauważyć, że badane zmienne mają rozkład normalny (cechy dziedziczone genetycznie często wykazują tę właściwość).

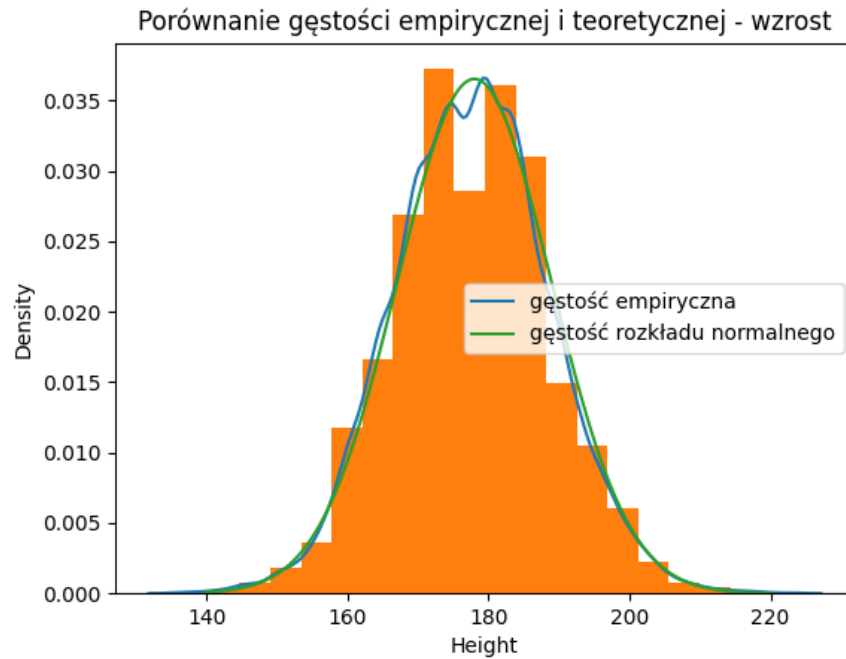


Figure 1: Histogram, gęstość empiryczna i teoretyczna wzrostu

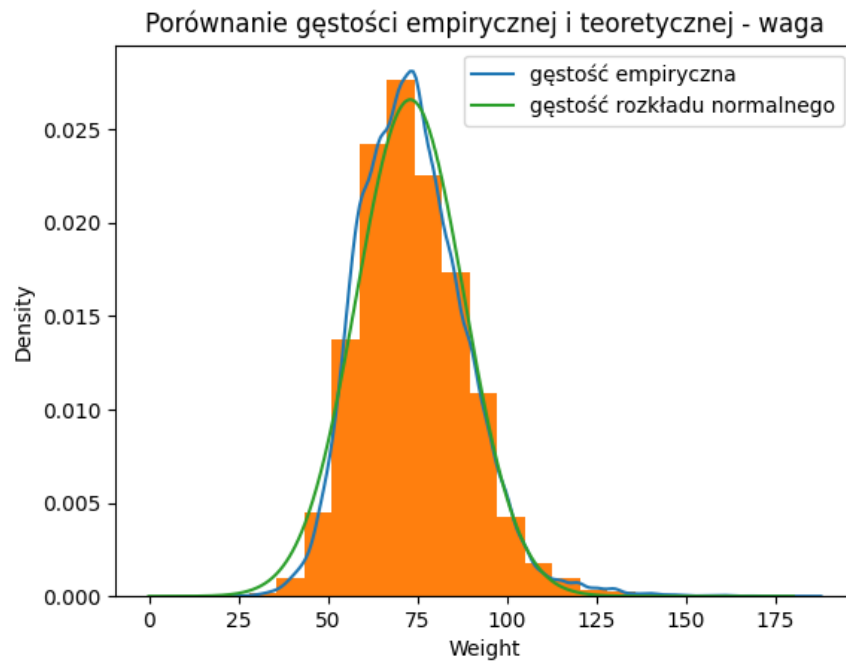


Figure 2: Histogram, gęstość empiryczna i teoretyczna wagi

4.2 Dystrybuanty

Dystrybuanty dla obu zmiennych oraz ich wartości teoretyczne, tzn. z rozkładu normalnego o odpowiednich parametrach.

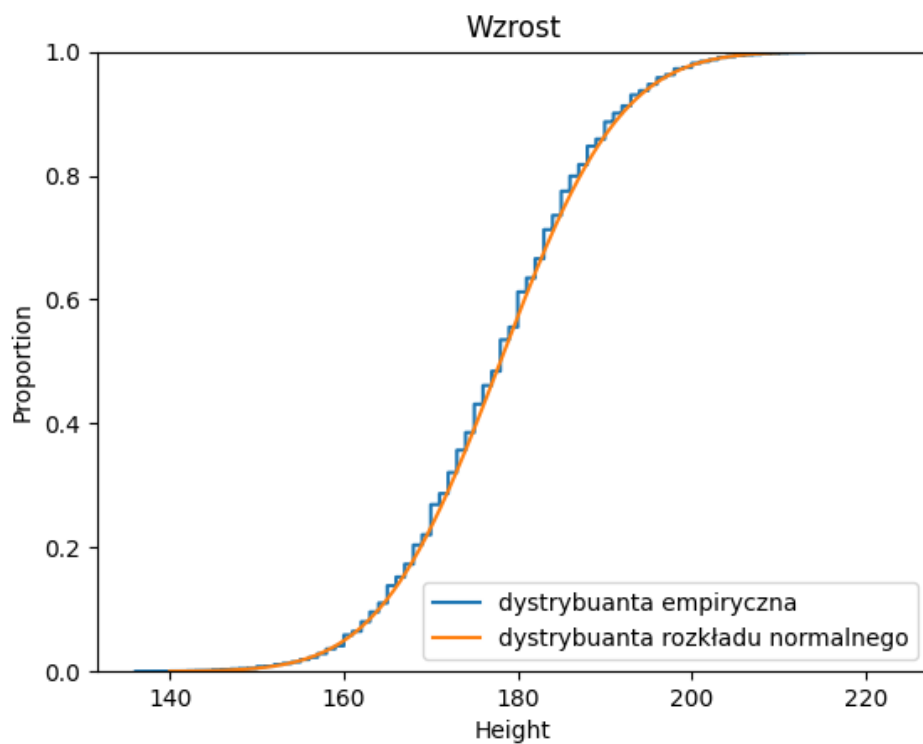


Figure 3: Dystrybuanta empiryczna i teoretyczna wzrostu

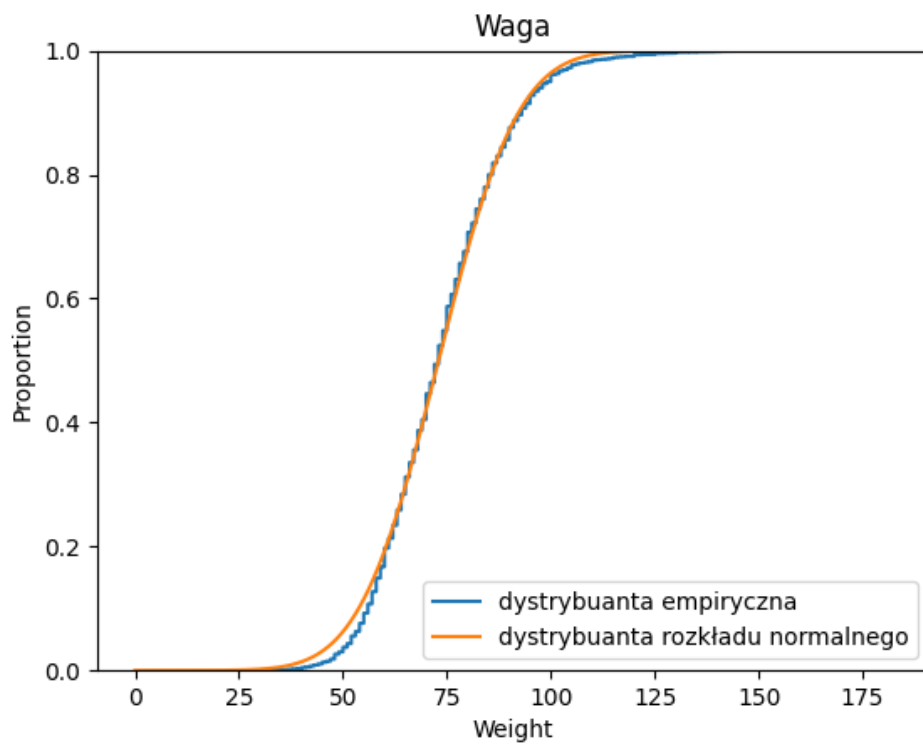


Figure 4: Dystrybuanta empiryczna i teoretyczna wagi

5 Porównanie zmiennych X i Y

5.1 Boxploty

Boxploty dla X i Y, przedstawiające rozkład danych i wartości odstające, z których również można wywnioskować, że odchylenie standardowe wagi jest większe niż odchylenie standardowe wzrostu.

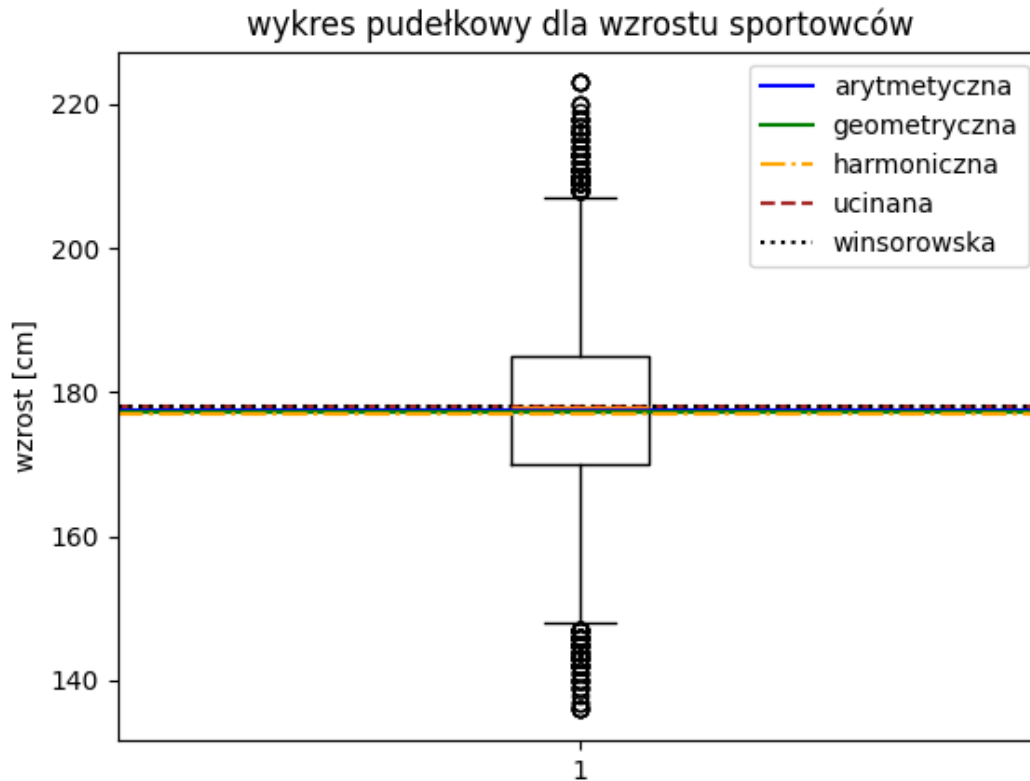


Figure 5: Wykres pudełkowy i średnie dla wzrostu sportowców

5.2 Macierz korelacji

Macierz korelacji między X a Y, wizualizacja zależności statystycznych. Wskazuje ona na korelację między wzrostem i wagą. Wartość bliska 1 świadczy o tym, że istnieje duże powiązanie między wzrostem a wagą wśród sportowców.

$$\begin{bmatrix} 1.0 & 0.8 \\ 0.8 & 1.0 \end{bmatrix}$$

Obliczony współczynnik korelacji Spearmana to **0.83**, co również wskazuje na dość silną zależność.

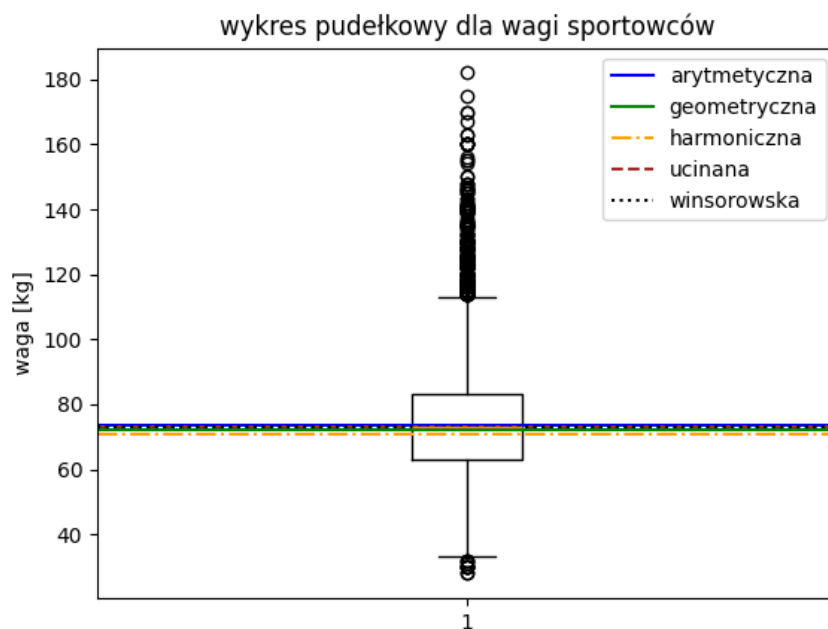


Figure 6: Wykres pudełkowy i średnie dla wagi sportowców

6 Podsumowanie

Po przeanalizowaniu i zwizualizowaniu danych możemy stwierdzić, że nasze zmienne losowe X i Y (wzrost i waga sportowców) są ze sobą dosyć silnie skorelowane oraz obie mają rozkład normalny. Potwierdza to wiedzę z podstaw genetyki - cechy ilościowe (mieralne) przyjmują wartości będące częścią rozkładu normalnego. Potwierdziliśmy to, generując na wykresach gęstości oraz dystrybuanty teoretyczne oraz empiryczne - linie znajdują się bardzo blisko siebie.

Wysoki współczynnik korelacji sugeruje, że waga rośnie wraz ze wzrostem. Jest to dosyć naturalne założenie, ponieważ zazwyczaj osoby wyższe mają wyższą wagę.

Kolejną obserwacją jest większe odchylenie standardowe dla wagi, niż dla wzrostu - może to wynikać z faktu, iż na wagę mamy większy wpływ, a także niektóre sporty wymagają konkretnej wagi u zawodników, przez co zmienna przyjmuje bardziej rozproszone wartości.