

Analiza cech fizycznych i osiągnięć sportowych uczniów

Kacper Omieliańczyk (459459)

2025-01-15

Wprowadzanie

Sport od zawsze odgrywał istotną rolę w moim życiu, szczególnie w okresie nastoletnim, kiedy to kształtował mój charakter, dyscyplinę oraz determinację. Aktywność fizyczna była nie tylko sposobem na spędzanie czasu, ale również przestrzenią do rozwijania umiejętności współpracy i rywalizacji. Wspomnienia tamtych lat niosą ze sobą ogromny sentyment, a myśl o możliwości dzielenia się wiedzą i doświadczeniem z innymi – być może jako trener – zawsze budziła we mnie szczególny rodzaj inspiracji. To właśnie ten sentyment i fascynacja światem sportu stanowią główną motywację do przeprowadzenia niniejszej analizy, która pozwala mi połączyć pasję do nauki i sportu w jedno spójne działanie.

Dataset został pobrany z repozytorium Mendeley Data, gdzie został udostępniony w ramach pracy naukowej dotyczącej czynników wpływających na wydolność fizyczną dzieci i młodzieży. Dane te zawierają informacje o zmiennych demograficznych, antropometrycznych oraz wynikach testów fizycznych, co czyni je idealnym źródłem do analizy zróżnicowania zdolności fizycznych w zależności od różnych czynników, takich jak płeć, wiek czy wskaźniki masy ciała. Szczegóły dotyczące zbioru można znaleźć pod adresem: Mendeley Data.

Analiza danych

Liczba obserwacji wynosi **1851**, natomiast liczba zmiennych to **8**.

Informacje o zmiennych

- **ID**: Unikalny identyfikator obserwacji.
- **Age**: Wiek dziecka w latach.
- **Gender**: Płeć (male/female).
- **Weight**: Waga w kilogramach.
- **Height**: Wzrost w centymetrach.
- **BMI**: Wskaźnik masy ciała (Body Mass Index).
- **Speed**: Szybkość mierzona w metrach na sekundę.
- **Jump**: Długość skoku w centymetrach.

Brakujące dane

Poniższa tabela przedstawia liczbę brakujących danych w każdej zmiennej.

Table 1: Brakujące dane w zbiorze danych.

ID	Age_Y	Gender	Weight_Kg	Height_cm	BMI	Speed_m/s	Jump_distance_cm
0	0	0	0	0	0	0	0

Podstawowe statystyki zmiennych

Statystyki dla zmiennych ciągłych

Podstawowe statystyki dla zmiennych ciągłych zostały przedstawione w tabeli:

Variable	min	max	mean	median	sd
Age	11.0	19.00	13.776877	14.0	1.5076114
Weight	26.0	96.00	48.466451	48.0	10.3517176
Height	135.0	190.00	160.260940	160.0	9.3993377
BMI	12.0	34.00	18.731496	18.0	3.0220410
Speed	3.6	8.06	5.377915	5.3	0.6462392
Jump	160.0	565.00	300.189627	292.0	79.8428065

Statystyki dla zmiennych jakościowych

Poniższa tabela przedstawia licznosc obserwacji w zależności od płci.

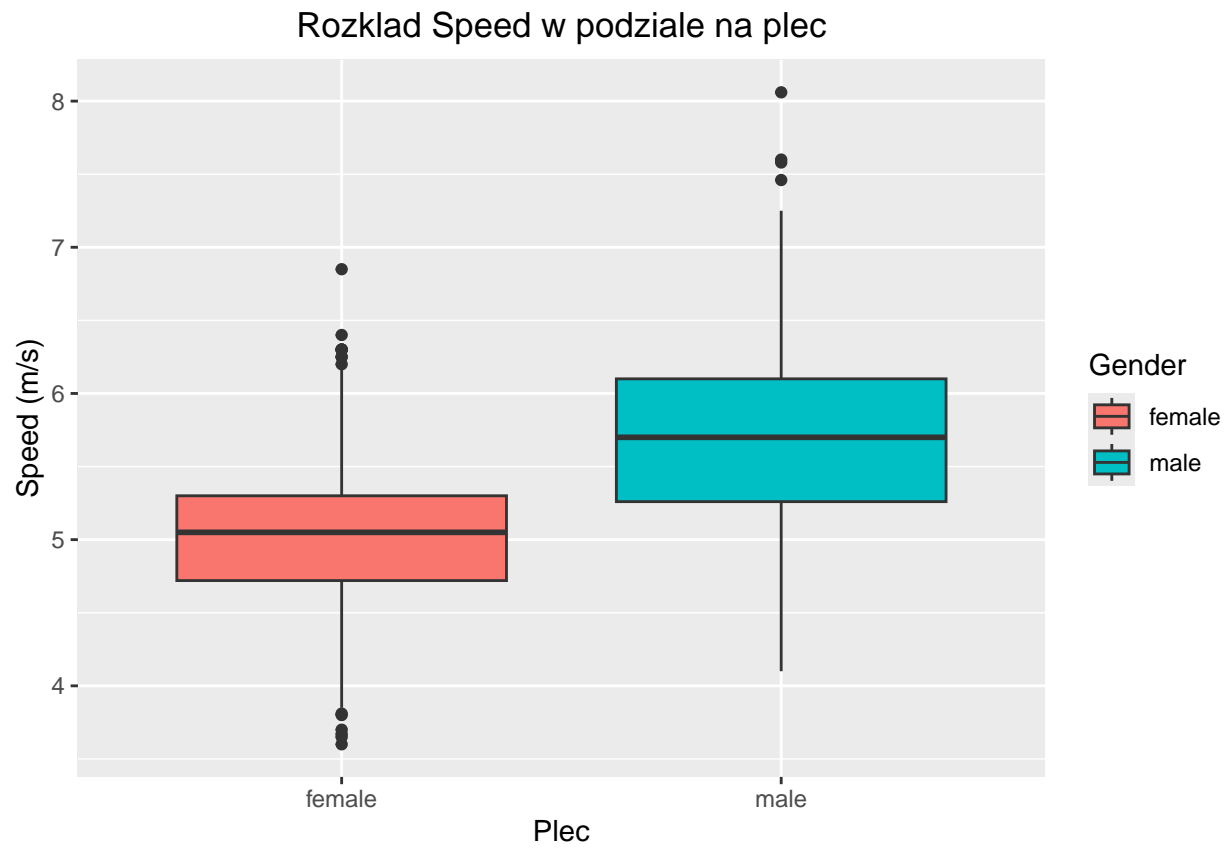
Table 3: Licznosci dla zmiennej Gender.

Gender	Liczebosc
female	867
male	984

Analiza porównawcza zmiennych ciągłych według płci

Wykresy rozkładów

Wykresy przedstawiają rozkłady zmiennych **Speed** i **Jump** w podziale na płeć.





Statystyki opisowe w podziale na płeć

Tabela przedstawia pełne statystyki opisowe zmiennych **Speed** i **Jump** w podziale na płeć. Dla lepszej czytelności podzielono statystyki na dwie tabele.

Statystyki dla zmiennej Speed_m/s

Gender	Średnia	Mediana	Min	Max	Odchylenie
female	5.024002	5.05	3.6	6.85	0.4637365
male	5.689746	5.70	4.1	8.06	0.6234594

Tabela 2: Statystyki zmiennej Speed w podziale na płeć.

Statystyki dla zmiennej Jump_distance_cm

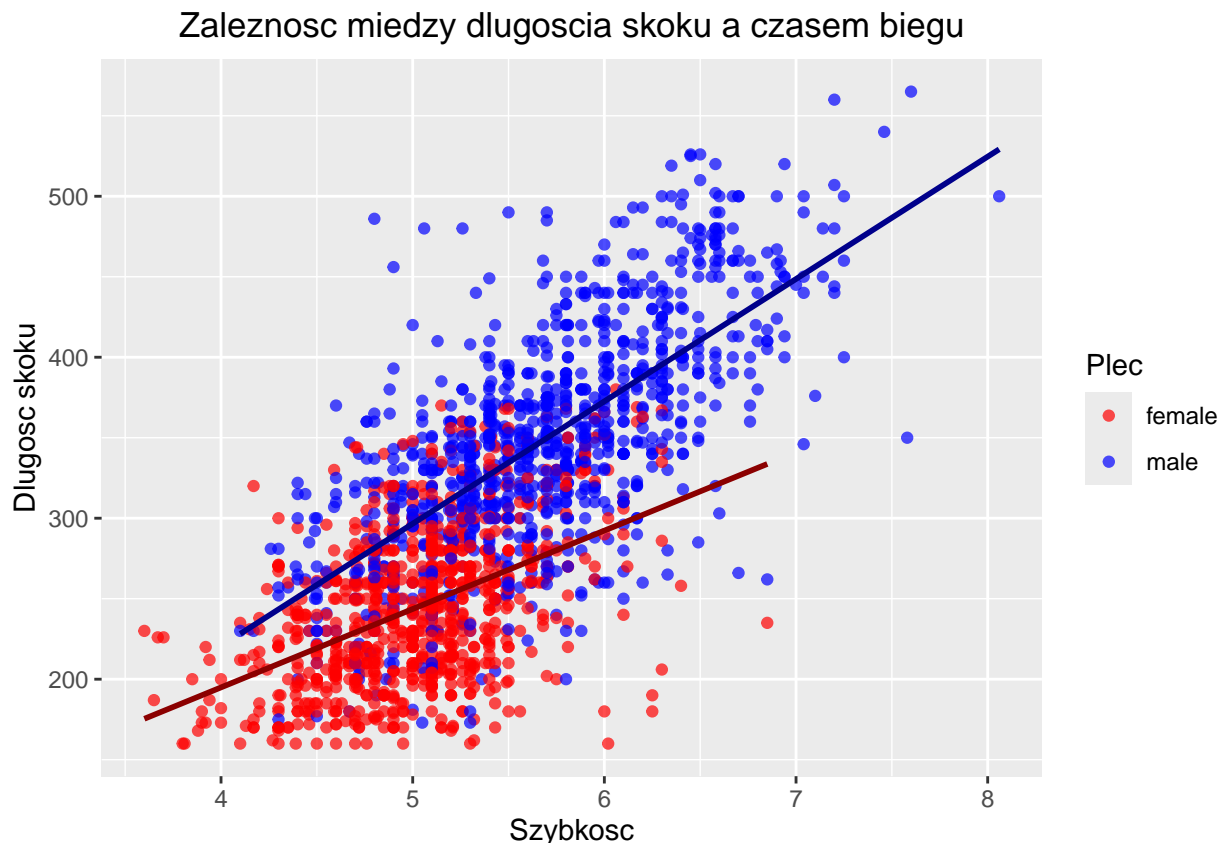
Gender	Średnia	Mediana	Min	Max	Odchylenie
female	244.7728	240	160	380	47.99935
male	349.0173	345	173	565	69.80377

Tabela 3: Statystyki zmiennej Jump w podziale na płeć.

Analiza zależności między długością skoku a czasem biegu

Wykres kropkowy z prostą regresji liniowej

Zależność między zmiennymi **Jump** (długość skoku) i **Speed** (szybkość biegu) została przedstawiona na wykresie poniżej. Oddzielnie zaznaczono dane oraz dopasowania regresji liniowej dla kobiet i mężczyzn.

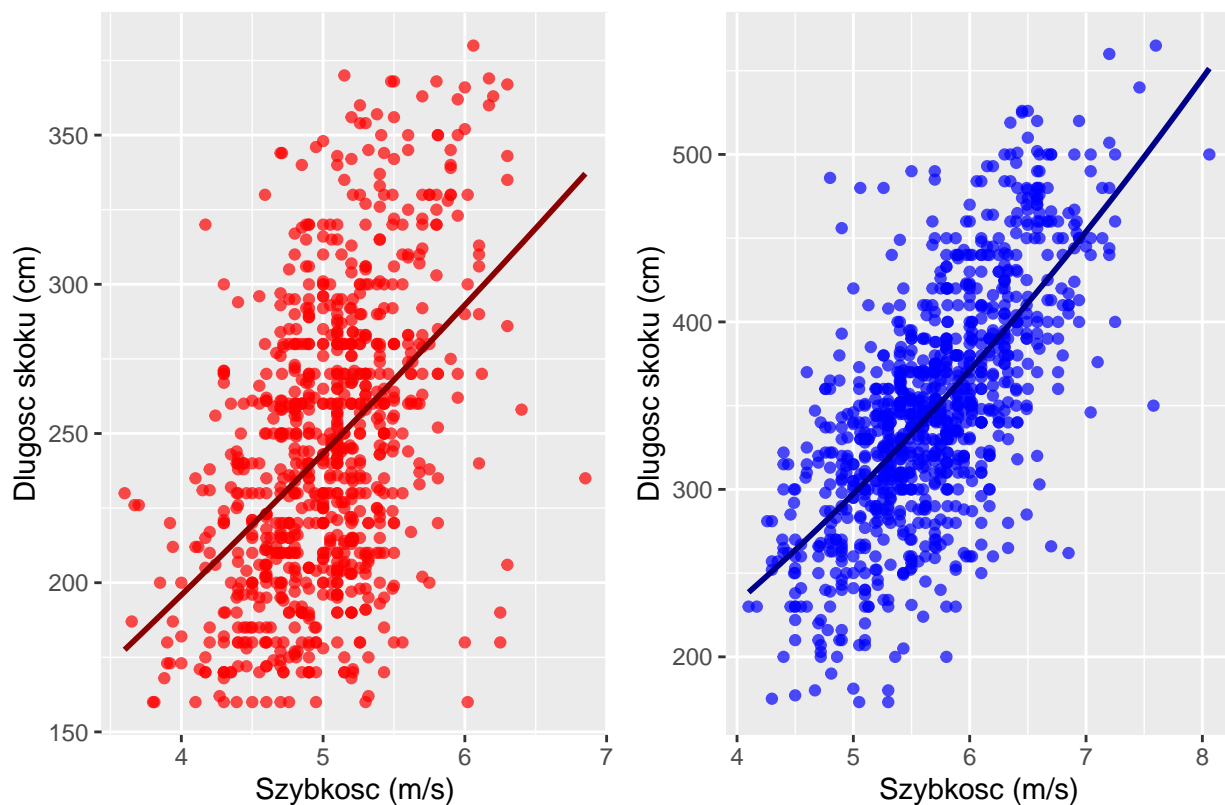


Czy dopasowanie wydaje się rozsądne? Prosta regresji dla mężczyzn wydaje się dość dobrze dopasowana do danych. Punkty na wykresie pokazują wyraźną liniową tendencję wzrostową. W przypadku kobiet dopasowanie prostej również oddaje ogólną tendencję wzrostową, ale jest bardziej rozproszone. Widać większy rozrzut punktów, co może sugerować, że prosty model liniowy nie jest idealny. W obu przypadkach rozrzut punktów poza linią wskazuje, że istnieje pewna nieliniowość w zależności.

Wykres kropkowy z dopasowaniem wielomianowym

Dla bardziej precyzyjnej analizy zależności między zmiennymi przedstawiono dopasowanie modelu wielomianowego drugiego stopnia dla każdej z grup (kobiety i mężczyźni) na osobnych wykresach.

Dopasowanie wielomianowe (stopień 2) między długością skoku a czasem biegu

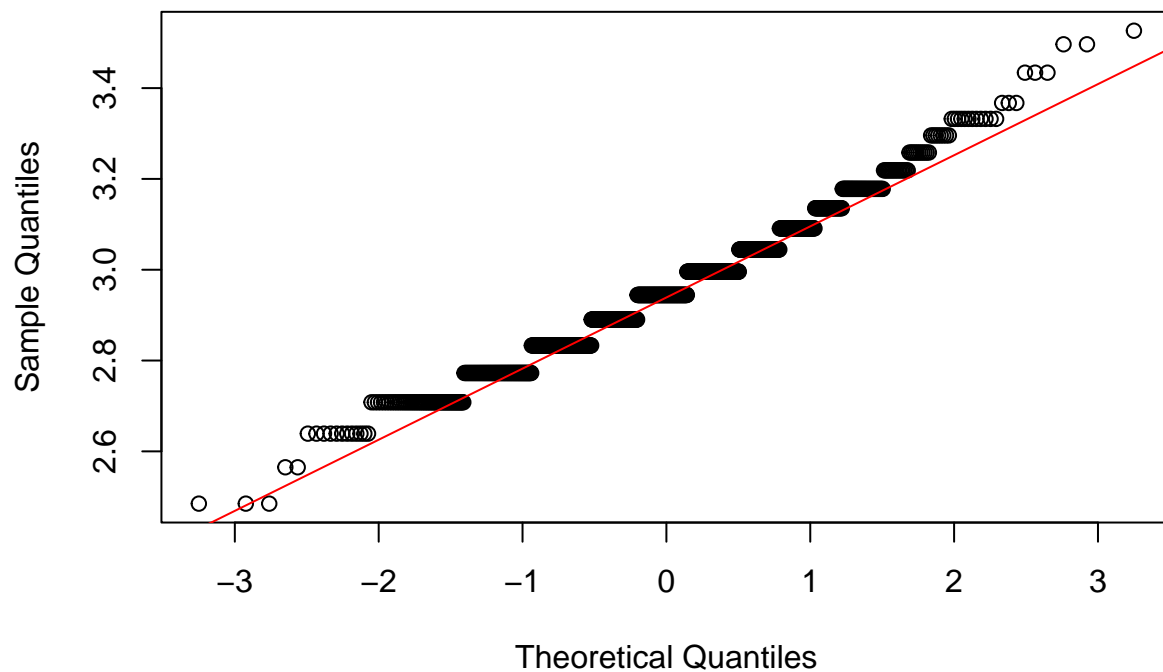


Dopasowanie trendu u mężczyzn wydaje się znaczące, wskazując na wyraźną zależność między analizowanymi zmiennymi. Natomiast w przypadku kobiet dane wykazują zbyt duże rozproszenie, co znacząco utrudnia stworzenie precyzyjnego modelu regresji.

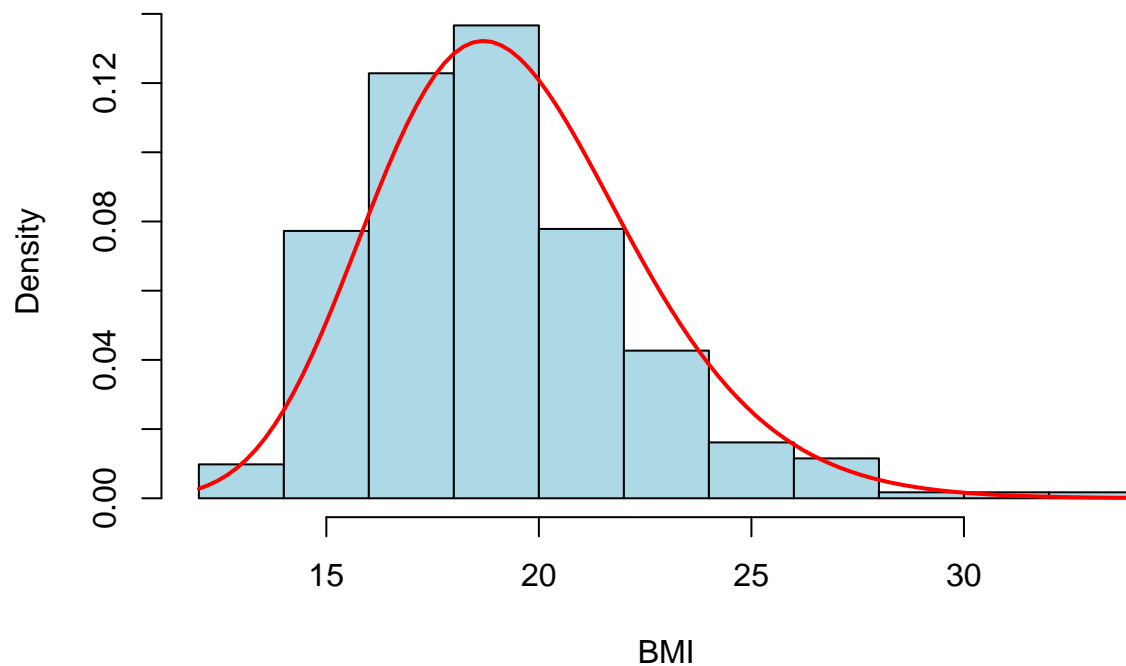
Dopasowanie rozkładu prawdopodobieństwa dla zmiennej BMI

Dla zmiennej **BMI** oddzielnie dla mężczyzn i kobiet dopasowano rozkład log-normalny metodą największej wiarygodności. Wyniki przedstawiono poniżej.

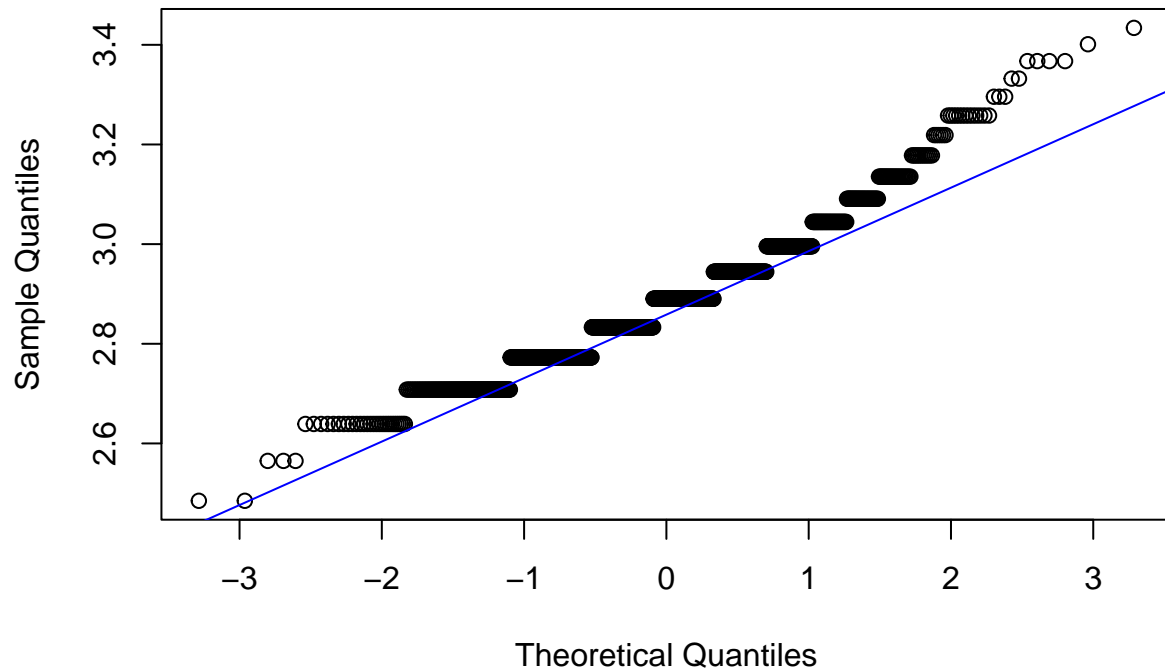
QQ-plot dla BMI kobiet (log-normal)



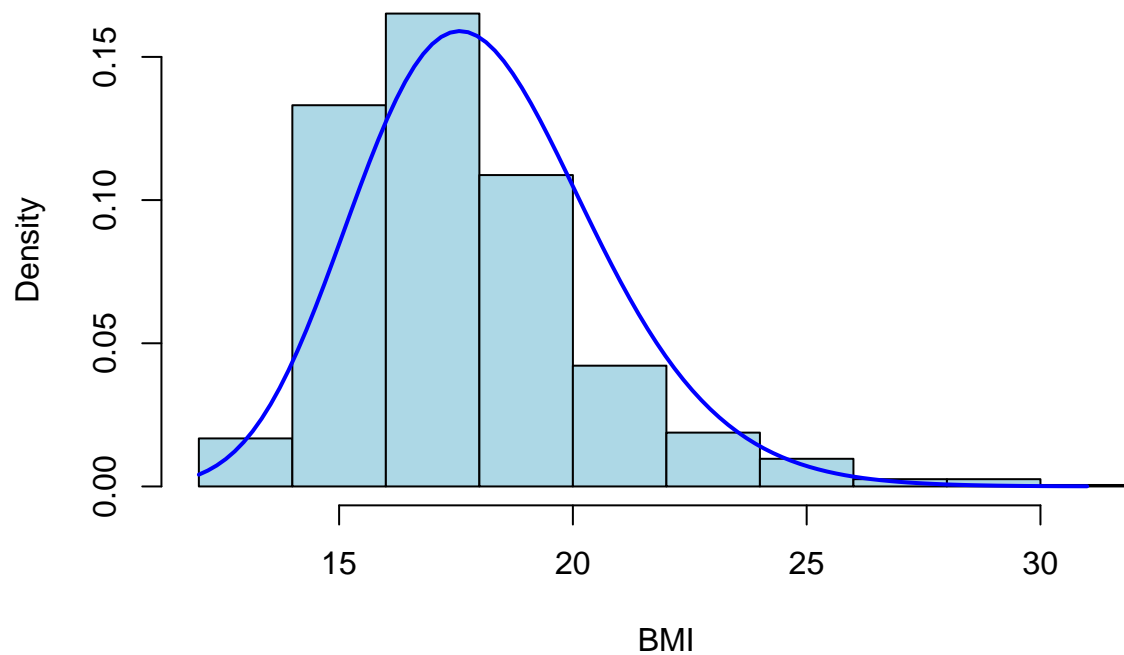
Histogram BMI kobiet z dopasowana gestoscia log-normalna



QQ-plot dla BMI mezczyzn (log-normal)



Histogram BMI mezczyzn z dopasowana gestoscia log-normalna



Komentarz: Dla testu kobiet uzyskano p -wartość: < 0.05 , oraz dla mężczyzn p -wartość: < 0.05 .

Mimo wybrania rozkładu log-normalnego, dopasowanie nie spełnia założeń przyjętego poziomu istotności, co wynika z prawoskośnego charakteru rozkładu BMI. Niestety, nie udało się znaleźć dokładnego sposobu dopasowania tego typu danych. Przeszukując źródła, natrafiłem jednak na pracę, która porusza ten problem: [link](#).

Szacowanie prawdopodobieństwa występowania płci z przedziałem ufności

Poniżej oszacowano prawdopodobieństwo występowania każdej wartości zmiennej jakościowej **Gender** (płeć) wraz z 99% przedziałami ufności Wilsona.

Wyniki dla zmiennej Gender

Table 6: Prawdopodobieństwa występowania płci z 99% przedziałem ufności Wilsona.

Płeć	Liczność	Prawdopodobieństwo	Dolny 99% PU	Górny 99% PU
female	867	0.4683955	0.4983306	0.4386861
male	984	0.5316045	0.5613139	0.5016694

Regresja liniowa z uwzględnieniem zmiennej jakościowej

Przeprowadzono regresję liniową zmiennej **Speed** względem zmiennych **Age**, **BMI**, **Jump distance** oraz **Gender** (skonwertowanej do zmiennej jakościowej). Uzyskane wyniki zebrano w tabeli:

Wyniki analizy

Table 7: Współczynniki regresji dla modelu uwzględniającego płeć.

	Estymacja	Std. Błąd	t-wartość	Pr(> t)
(Intercept)	3.2576995	0.1043092	31.231168	0.0000000
Age	0.0487396	0.0076282	6.389353	0.0000000
BMI	-0.0076793	0.0035474	-2.164776	0.0305326
Jump	0.0051049	0.0001846	27.658719	0.0000000
Gendermale	0.1131380	0.0274140	4.127010	0.0000384

Uzyskany błąd średniokwadratowy wynosi 0.1872. Z tabeli wyczytujemy p -wartości, z których przy ustalonym poziomie istotności $\alpha = 0.01$ wynika, że wszystkie zmienne poza BMI są statystycznie istotne.

Testowanie hipotez statystycznych

W celu sprawdzenia zależności między zmiennymi sformułowano i przetestowano dwie hipotezy statystyczne.

Hipoteza 1: Czy średnia szybkość (Speed) różni się istotnie między płciami?

Istnieje pewna grupa społeczna, która stara się nie dostrzegać różnic w wydolności fizycznej między kobietami a mężczyznami. Z tego powodu ciekawym wydaje się przetestowanie, kto ma rację: oni czy zdrowy rozsądek?

Sformułowanie hipotezy

- Hipoteza zerowa (H_0): Średnia szybkość jest taka sama dla kobiet i mężczyzn.
- Hipoteza alternatywna (H_A): Średnia szybkość różni się między kobietami a mężczyznami.

Test t-Studenta Wyniki testu t-studenta: Statystyka: -26.2531 ; p -wartość: < 0.05 ; przedział ufności (99%): [-0.7155, -0.616]

Hipoteza 2: Czy długość skoku (Jump distance) koreluje istotnie z wiekiem (Age)?

Sformułowanie hipotezy

- Hipoteza zerowa (H_0): Nie ma korelacji między długością skoku a wiekiem.
- Hipoteza alternatywna (H_A): Istnieje istotna korelacja między długością skoku a wiekiem.

Test korelacji Pearsona Wyniki testu korelacji Pearsona: Estymacja korelacji: 0.3569 ; p -wartość: < 0.05 ; przedział ufności (99%): [0.3165, 0.396]

Podsumowanie wyników testów

- Płeć istotnie wpływa na średnią szybkość dzieci, co może być związane z różnicami biologicznymi lub innymi cechami płciowymi.
- Wiek jest umiarkowanie związany z długością skoku, co sugeruje, że starsze dzieci mają większe zdolności motoryczne lub siłę, umożliwiające dłuższe skoki.
- Oba wyniki wskazują na istotne statystycznie różnice i zależności, co warto uwzględnić w dalszych analizach.