

Analiza 1

Estymacja parametrów rozkładu na podstawie danej próby

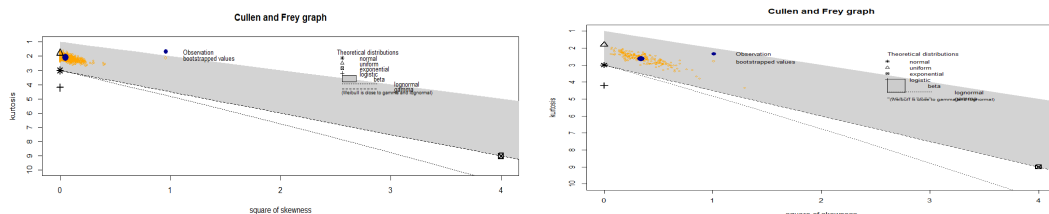
Antonina Brzeska 246824

1. Krótki opis wybranych danych. Pierwszy zestaw zawiera spis wagi i wysokości oraz płeć osoby. Wykonane były w 2017 roku. Zestaw został przefiltrowany, do badania wzięto 200 rekordów wag i przeliczono jednostkę na kilogramy. Drugi zestaw obrazuje ilość pasażerów linii lotniczej w latach 1949 - 1960.

2. Wstępna analiza danych.

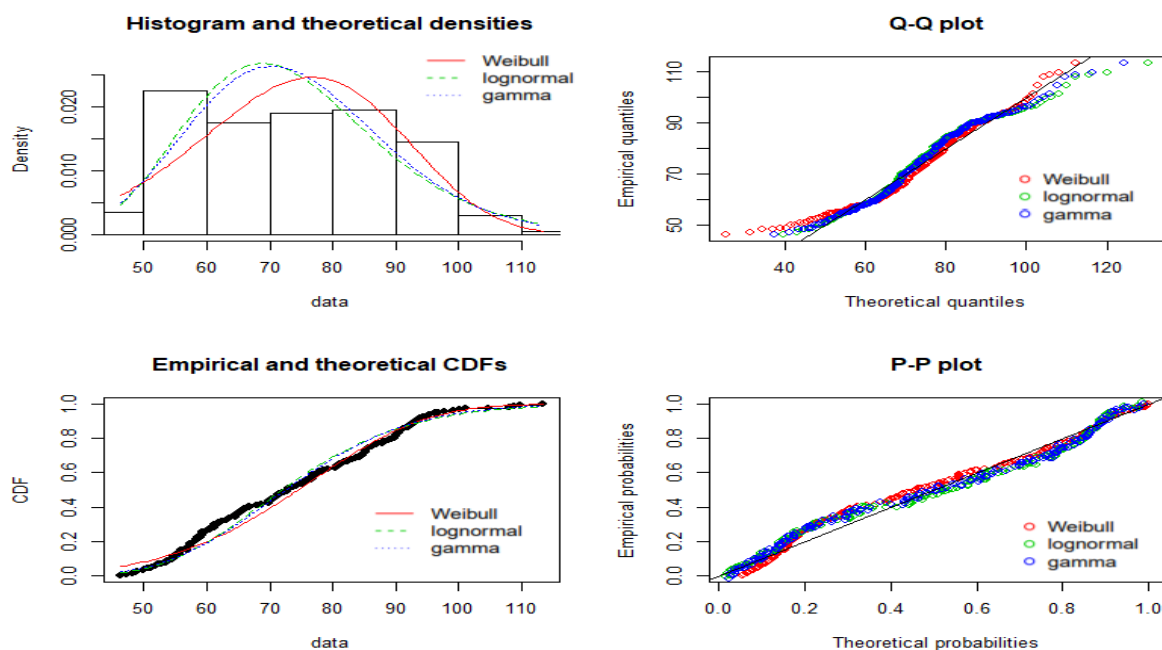
	\bar{x}	odch. st.	skośność	kurtoza
Wagi	73.55	15.41	0.22	2.09
Pasażerowie	280	120	0.58	2.6

Wykres skośność-kurtoza dla wag (po lewej) oraz pasażerów (po prawej):



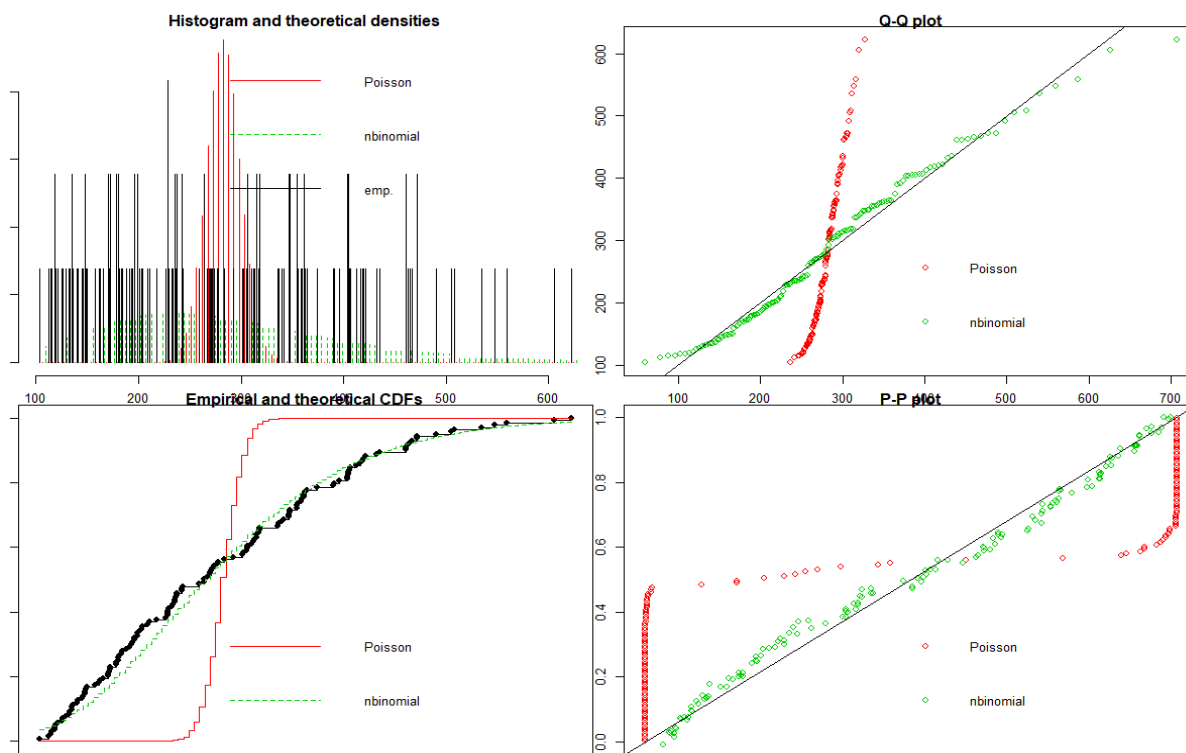
3. Estymacja parametrów wybranych rozkładów – analiza wstępna.

Wykresy diagnostyczne dla rozkładu ciągłego (wagi):



”Gołym okiem” ciężko dostrzec znaczącą różnicę, jednak rozkład gamma wydaje się najdokładniejszy.

Wykresy diagnostyczne dla rozkładu dyskretnego (pasażerowie):



Z obserwacji widać, że rozkład dwumianowy uzyskał lepszą dokładność.

4. Analiza pogłębiona. Po obliczeniu kryteriów AIC oraz BIC można zauważyć, że dla rozkładu ciągłego najlepszy okazuje się rozkład gamma. Dla rozkładu dyskretnego różnica jest oczywista i tu rozkład dwumianowy uzyskuje znacznie lepszą dokładność.

Kryteria rozkładu ciągłego (wagi):

	Weibull	log-normal	gamma
AIC	1670.183	1660.335	1659.826
BIC	1676.78	1666.932	1666.423

Kryteria rozkładu dyskretnego (pasażerowie):

	Rozkład Poissona	Dwumianowy
AIC	8281.635	1772.221
BIC	8284.605	1778.16

5. Zdarzenia rzadkie. Analizując obliczone kwantyle można zauważyć, że mało prawdopodobne jest wylosowanie osoby ważącej 100 kg oraz rzadkością było przekroczenie miesięcznej ilości pasażerów liczbą powyżej 500.

	Wagi	Pasażerowie
kwantyle rzędu 0.95	96.39	488
kwantyle rzędu 0.98	104.82	550

Źródła danych:

- <https://www.kaggle.com/mustafaali96/weight-height>
- <https://www.kaggle.com/rakannimer/air-passengers>