

Komputerowa analiza szeregów czasowych raport 1

Natalia Klepacka, Joanna Kołaczek

21.12.2022

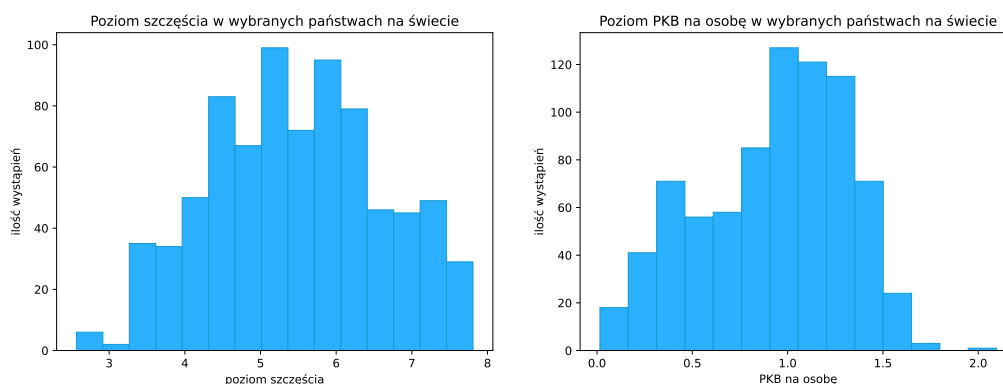
Spis treści

1	Opis danych	2
2	Analiza jednowymiarowa zmiennej zależnej oraz zmiennej niezależnej	2
3	Analiza zależności liniowej pomiędzy zmienną zależną a zmienną niezależną	4
3.1	Wykres rozproszenia i określenie zależności	4
3.2	Punktowa estymacja współczynników	5
3.3	Przedziałowa estymacja współczynników	6
3.4	Ocena poziomu zależności	6
3.5	Predykcja oraz jej przedziały ufności	6
3.6	Interpretacja wyników	6
4	Analiza residuów	6
5	Podsumowanie	7
6	Źródła	7

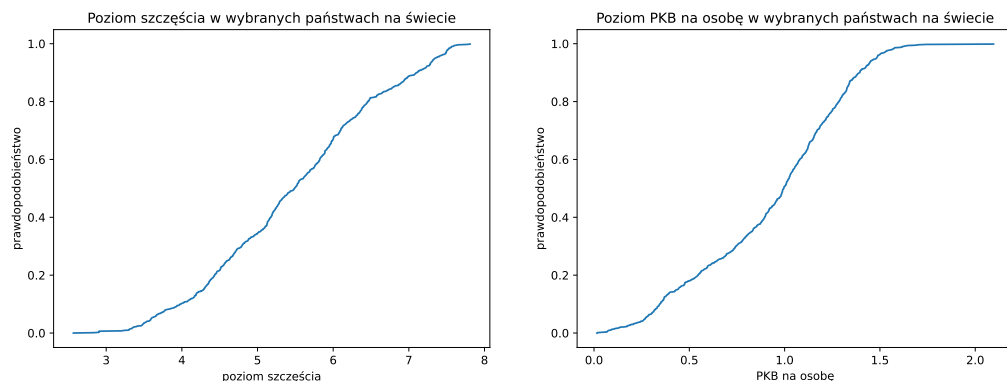
1 Opis danych

Niniejszy raport powstał na potrzeby realizacji laboratorium ze Statystyki Stosowanej, prowadzonych przez dr inż. Aleksandrę Grzesiek, do wykładu dr hab. inż. Krzysztofa Burneckiego. Będziemy analizować dane dotyczące długości ogonów mysołowów rdzawosternych (ang. red-tailed hawk) podane w milimetrach. Dysponujemy próbą o wielkości 577 pobraną z *tej strony*. Na jej podstawie przedstawimy statystyki podzielone według miar położenia, rozproszenia, asymetrii oraz spłaszczenia. Na koniec pierwszego rozdziału zebrane wielkości przedstawimy w formie tabeli. W drugiej części raportu zwizualizujemy dane za pomocą histogramu oraz wykresu pudełkowego. Wyznamy ponadto dystrybuantę empiryczną analizowanej próby. Na koniec wyniki badań omówimy w podsumowaniu. Życzymy Czytelnikowi miłej lektury.

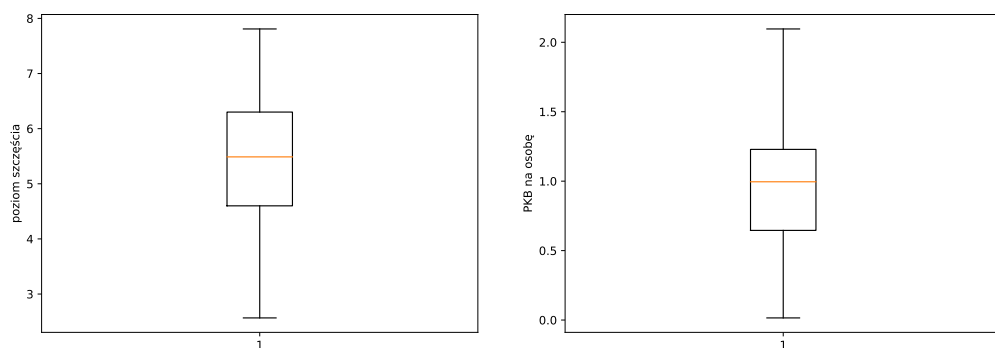
2 Analiza jednowymiarowa zmiennej zależnej oraz zmiennej niezależnej



Rysunek 1: Histogramy



Rysunek 2: Dystrybuanta empiryczna



Rysunek 3: Boxploty

Na histogramie [1] widzimy rozkład danych z badanej próby. Jest to dobry sposób wizualizacji danych, ponieważ można dzięki niemu łatwo oszacować niektóre statystyki takie jak średnia, skośność itp. Widzimy, że dla naszych danych najczęściej obserwujemy długości ogona zawarte w przedziale 220 do 225 mm.

Wykres pudełkowy (ang. *boxplot*) [3] jest to graficzna reprezentacja mediany, kwartyli oraz maksimum i minimum z danych (kropki oznaczają wartości odstające), wąsy mają długość półtorej wartości rozstępu międzykwartyłowego.

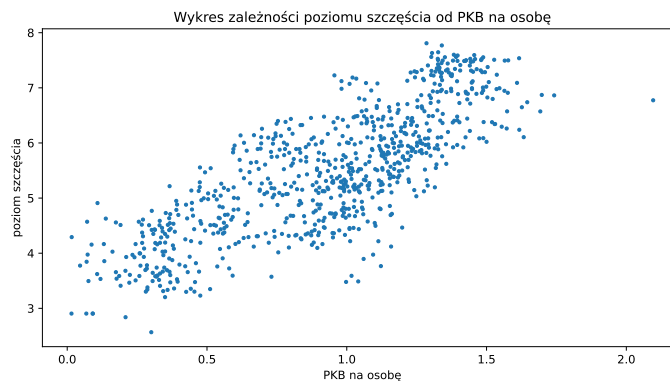
Dystrybuanta empiryczna [??] pokazuje z jakim prawdopodobieństwem natrafimy na myszołowa rdzawosternego o długości ogona mniejszej bądź równej od danej wartości. Widzimy, że szansa na to, aby spotkać osobnika z ogonem krótszym niż 200 mm lub dłuższym niż 250 mm jest nikła, natomiast długości między 200 a 250 mm występują najczęściej.

Miary		Poziom szczęścia	PKB na osobę
położenia	średnia arytmetyczna	5.47	0.93
	średnia geometryczna	5.23	0.58
	średnia harmoniczna	5.35	0.81
	średnia ucinana 10%	5.47	0.94
	mediana Q2	5.48	0.99
	Q1	4.59	0.64
	Q3	6.30	1.22
rozproszenia	rozstęp	5.24	2.08
	rozstęp międzykwartyłowy	1.70	0.58
	wariancja nieobciążona	1.26	0.14
	odchylenie standardowe	1.12	0.38
	współczynnik zmienności	20.52	41.33
splaszczania	kurtoza	2.25	2.31
skośności	skośność	-0.01	-0.36

Rysunek 4: Zestawienie omówionych statystyk

3 Analiza zależności liniowej pomiędzy zmienną zależną a zmienną niezależną

3.1 Wykres rozproszenia i określenie zależności



Rysunek 5: wykres rozproszenia

Na podstawie wykresu rozproszenia [5] możemy stwierdzić, że zależność pomiędzy zmiennymi prawdopodobnie jest liniowa.

3.2 Punkтова estymacja współczynników

Współczynniki regresji liniowej wyznaczamy ze wzorów

$$\begin{cases} \beta_1 = r \frac{S_y}{S_x} \\ \beta_0 = \bar{y} - r\beta_1\bar{x} \end{cases},$$

gdzie

$$r = \frac{1}{n-1} \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{S_x S_y},$$

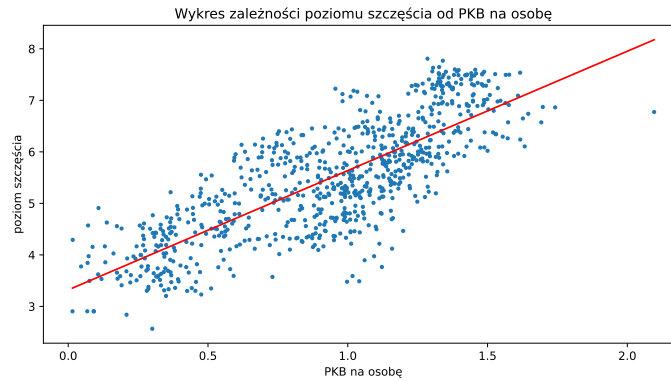
$$S_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2},$$

$$S_y = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2},$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i,$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i.$$

Wzory te zostały wyznaczone przy pomocy metody najmniejszych kwadratów.



Rysunek 6: wykres rozproszenia z zaznaczoną prostą regresji

Z powyższych wzorów otrzymaliśmy

$$\begin{cases} \beta_1 \approx 2.32 \\ \beta_0 \approx 3.32 \end{cases}.$$

Współczynniki te opisują prostą widoczną na wykresie [6].

3.3 Przedziałowa estymacja współczynników

Z założenia o normalności rozkładu $\{\varepsilon\}_{i=1}^n$, przy braku znajomości jego wariancji możemy stwierdzić, że unormowane parametry β_0, β_1 mają rozkład t-studenta z $n-2$ stopniami swobody. Przedziały ufności przyjmują wtedy postać

$$\begin{cases} P\left(\hat{\beta}_0 - t_{n-2}(1 - \frac{\alpha}{2})S\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} < \beta_0 < \hat{\beta}_0 + t_{n-2}(1 - \frac{\alpha}{2})S\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}\right) = 1 - \alpha \\ P\left(\hat{\beta}_1 - t_{n-2}(1 - \frac{\alpha}{2})\frac{S}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}} < \beta_1 < \hat{\beta}_1 + t_{n-2}(1 - \frac{\alpha}{2})\frac{S}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}\right) = 1 - \alpha \end{cases},$$

gdzie

$$S = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - 2}}.$$

Przyjmując $\alpha = 0.05$ otrzymujemy zatem, że z 95% prawdopodobieństwem

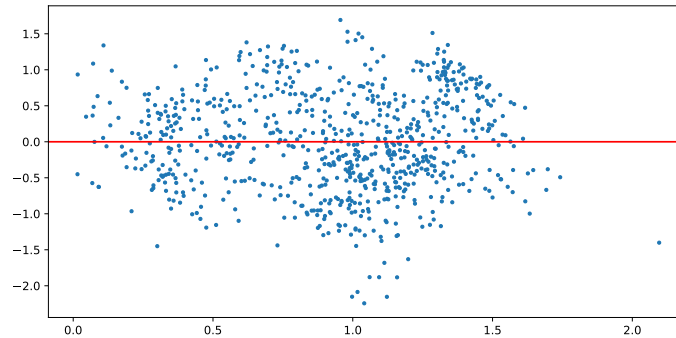
$$\begin{cases} \beta_1 \in (2.19, 2.44) \\ \beta_0 \in (3.20, 3.44) \end{cases}.$$

3.4 Ocena poziomu zależności

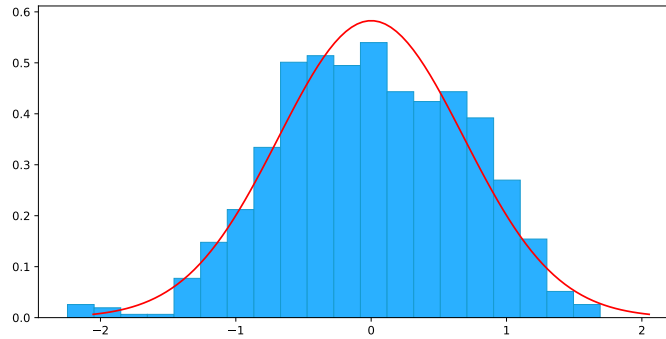
3.5 Predykcja oraz jej przedziały ufności

3.6 Interpretacja wyników

4 Analiza residuów



Rysunek 7: Wy



Rysunek 8: Wy

5 Podsumowanie

Analizując przedstawione w raporcie statystyki możemy sformułować następujące wnioski i przypuszczenia dotyczące długości ogonów w populacji myszów rdzawosternych. Z histogramu widzimy, że badany rozkład przypomina rozkład normalny, jednak po obliczeniu skośności okazuje się, że różni się ona od skośności rozkładu normalnego, która wynosi zero. Podejrzewamy, iż może to być spowodowane liczebnością próby. Statystyczny mysz rdzawosterny powinien mieć ogon o długości od 207,638 do 236.66 mm, (średnia arytmetyczna \pm odchylenie standardowe). Biorąc średnią z populacji przewidujemy długość około 222,15 mm. Wartości skrajne nie wpływają znacząco na wartość średniej - wiemy to po obliczeniu średniej Winsorowskiej (222,182 mm) i ucinanej (222,104 mm).

6 Źródła

- Wykłady
- <https://vincentarelbundock.github.io/Rdatasets/csv/Stat2Data/HawkTail.csv>
- <https://www.geo.fu-berlin.de/en/v/soga/Basics-of-statistics/Descriptive-Statistics/Measures-of-Position/index.html>
- https://www.investopedia.com/terms/w/winsorized_mean.asp