

# Modele i Wnioskowanie Statystyczne

## Laboratorium 1

### Sprawozdanie

Igor Markiewicz

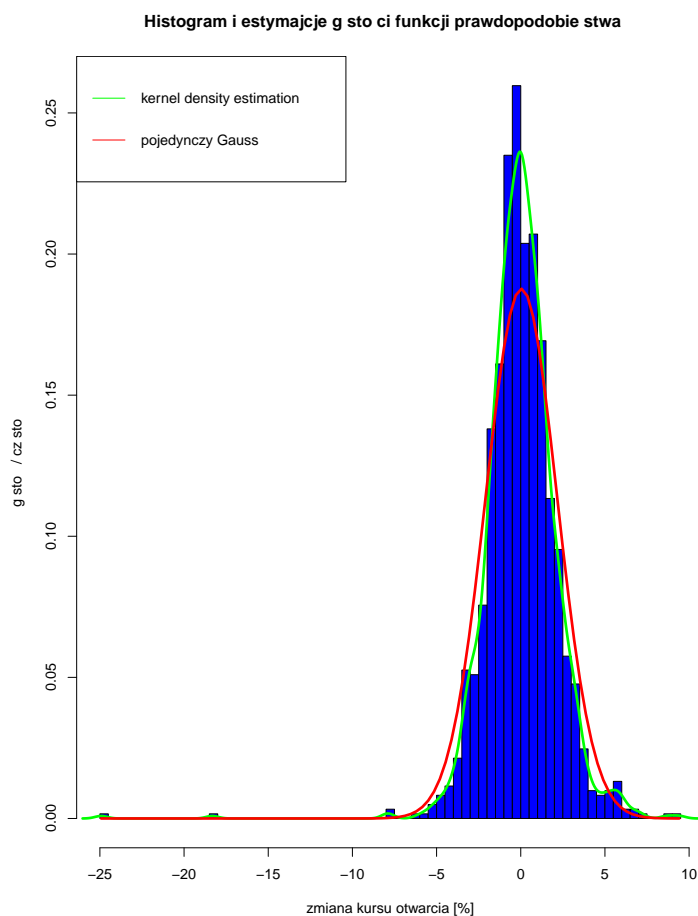
## Zadanie 1

a) Wybrano dane typu ALIOR. Po wczytaniu danych (o rozmiarze  $n$ ), procentowa zmiana kursu otwarcia została wyliczona w następujący sposób (warto zwrócić jednocześnie uwagę, że jest to miara niesymetryczna) :

$$diff = 100 \cdot \frac{x_{i+1} - x_i}{x_i} \quad \text{dla } i = 1, 2 \dots n - 1$$

Jako liczbę koszyków, przyjęto doświadczalnie liczbę 100 i stworzono histogram rozkładu zmiennej  $diff$ . Następnie wyestymowano funkcje gęstości prawdopodobieństwa dwiema metodami :

- *kernel density estimation*
- estymacja o rozkład Gaussa  $\mathcal{N}(\mu, \sigma^2)$ , gdzie  $\mu = \frac{1}{n} \sum_{i=1}^n x_i$  oraz  $\sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \mu)^2$

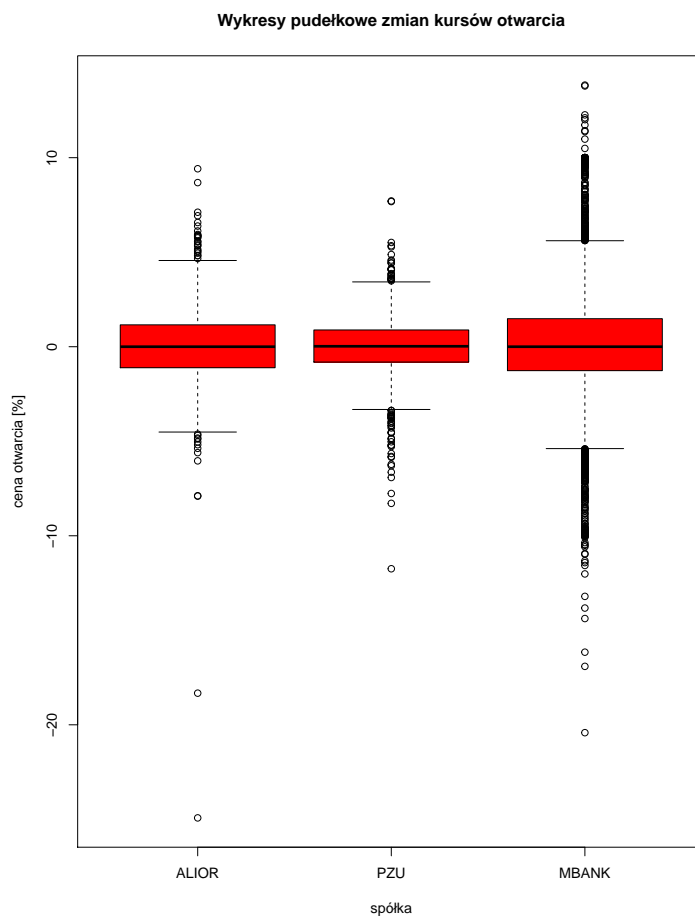


Rys. 1: Histogram i estymowane funkcje gęstości

Wnioski :

- Rozkład jest w przybliżeniu Gaussowski, choć nie jest do końca symetryczny
- metoda *kernel density estimation* lepiej przybliża rozkład
- Zmiany kursów otwarcia są skoncentrowane mniej, więcej wokół 0% z istotnym odchyleniem na poziomie ok 7,5%

b) Wybrano dane typu ALIOR, PZU oraz MBANK, a obliczanie procentowych zmian otwarcia przeprowadzono tak samo jak w podpunkcie a).



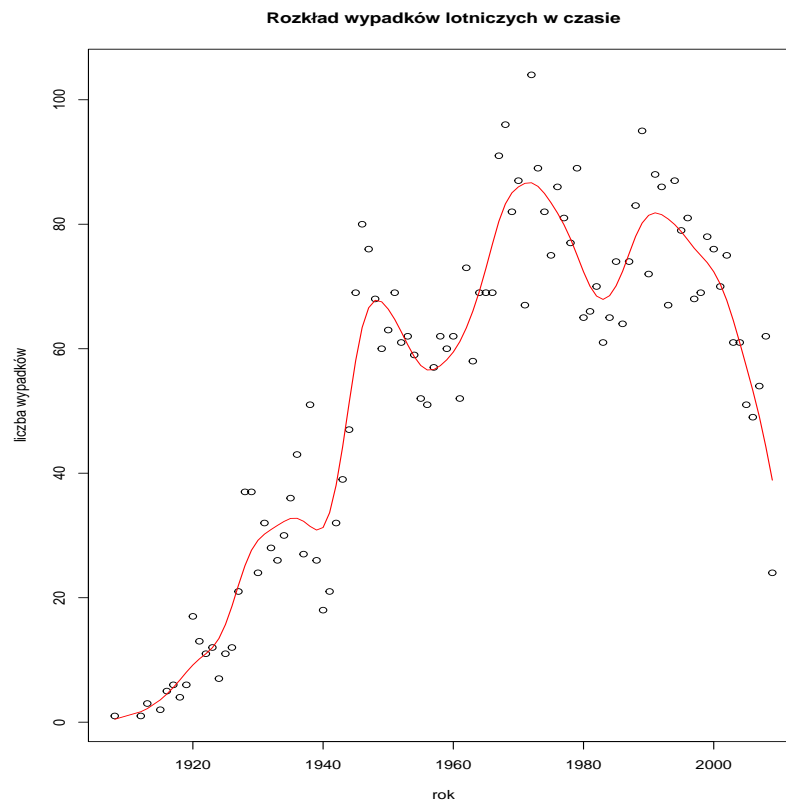
Rys. 2: Wykresy pudełkowe

Wnioski :

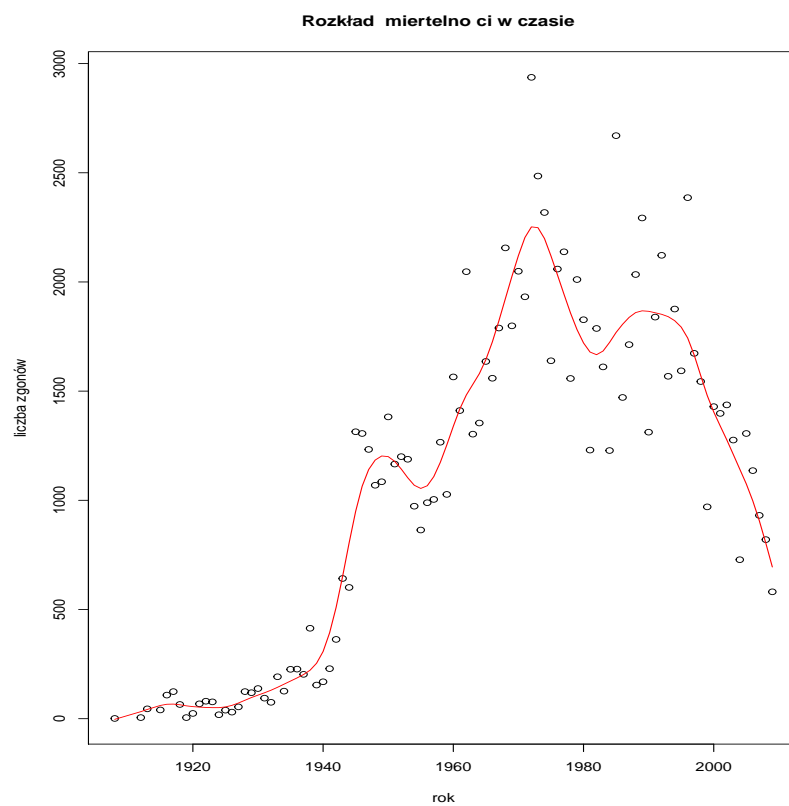
- mediany dla wszystkich spółek są w przybliżeniu równe 0%
- największe zróżnicowanie cechy (rozstęp kwartyłowy) ma MBANK, a najmniejsze PZU
- MBANK posiada najmniejszą jak i największą wartość nie będącą outlierem
- MBANK posiada największą liczbę outlierów
- Największy jednorazowy wzrost uzyskał MBANK, a największy jednorazowy spadek ALIOR

## Zadanie 2

a), b) W przypadku wartości NA, zdecydowano się na ich zastąpienie wartościami 0, co mogło zaniżyć nieco wartości danych (jednak liczba wartości NA nie jest na tyle duża, aby wpłynąć znacząco na wyniki). W przypadku wygładzania zdecydowano się użyć metod *smooth.spline* używającej do wygładzania wielomianów sklepanych trzeciego stopnia. Jako stopień wygładzania wybrano doświadczalnie  $\frac{1}{2} \in (0, 1]$



Rys. 3: Liczba wypadków w kolejnych latach



Rys. 4: Liczba zgonów w kolejnych latach

Wnioski :

- najmniejsza liczba katastrof i ofiar towarzyszyła na samym początku rozwoju lotnictwa
- istnieją ekstrema dla tych samych lat na obu wykresach
- największa liczba katastrof i ofiar śmiertelnych miała miejsce koło roku 1970 - miało to najprawdopodobniej związek z dynamicznym rozwojem lotnictwa, w szczególności pasażerskiego
- Od ok 1990 roku, zarówno liczba wypadków jak i ofiar śmiertelnych systematycznie maleje, co ma najprawdopodobniej przyczynę w rozwoju nowych technologii oraz procedur lotniczych