

# MWS Ćwiczenie 1

Elementy statystyki opisowej

Tomasz Korzeniowski, 265753

10 listopada 2017

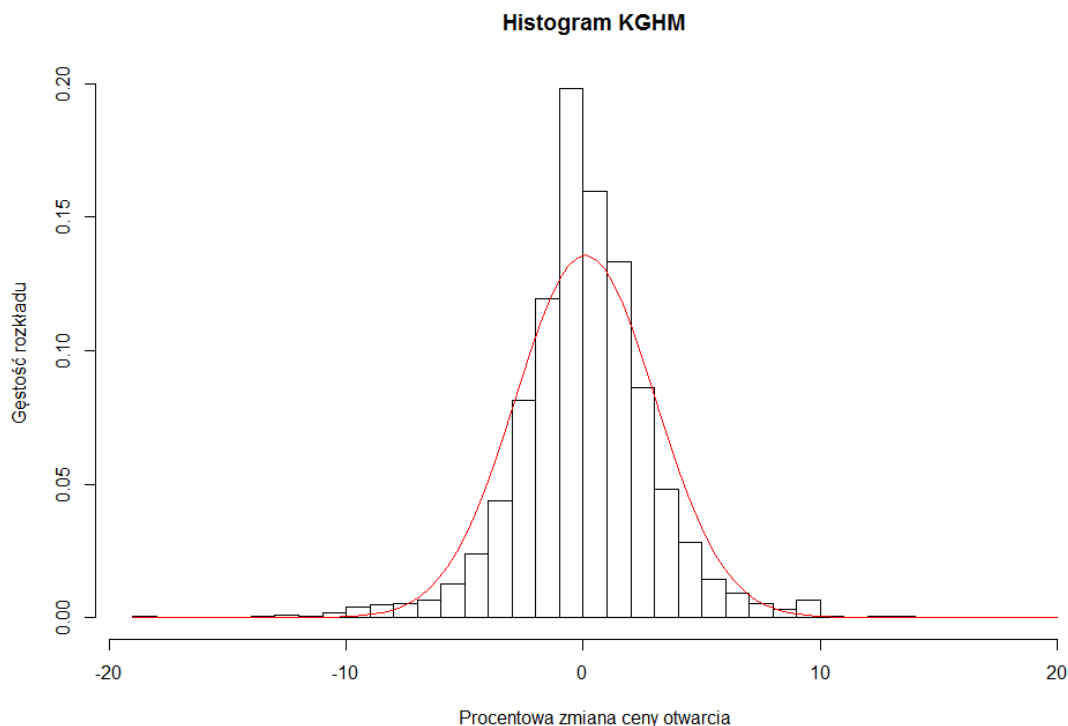
## 1 Zadanie 1

Procentowa zmiana ceny otwarcia wyraża się wzorem:

$$\frac{\text{cena nowa} - \text{cena stara}}{\text{cena stara}} * 100\%$$

W ramach ćwiczenia została stworzona funkcja wyznaczająca kolejne wartości zmiany w pętli. Histogram został narysowany z wykorzystaniem funkcji pakietu R: *hist*. Na niego został nałożony wykres gęstości rozkładu normalnego przez funkcję *curve*. Parametry rozkładu wyznaczono na podstawie funkcji *mean* oraz *sd*.

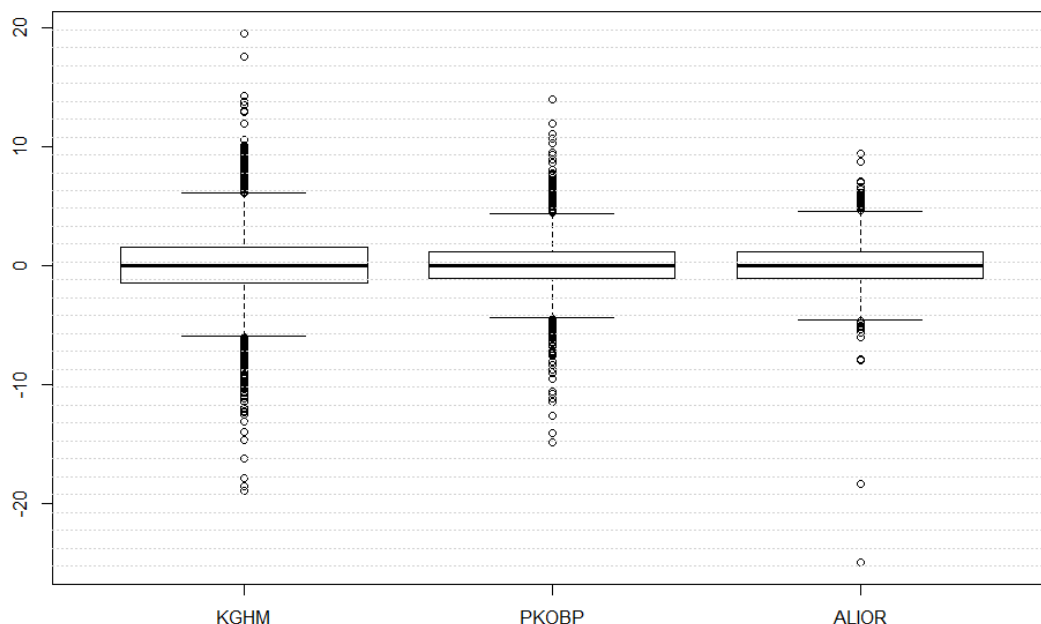
Dla firmy KGHM, wybranej podczas laboratorium, histogram procentowej został zaprezentowany na wykresie 1. Rozkład jaki został wyznaczony jest rozkładem  $N(0.0759249, 2.939349)$



Wykres 1: Histogram procentowej zmiany ceny otwarcia firmy KGHM

Na podstawie powyższego wykresu można stwierdzić, że prawdą jest, że wszelkiego rodzaju zjawiska, na które ma wpływ wiele niezależnych czynników opisuje się rozkładem normalnym.

W dalszej części zadania zostały porównane zmiany procentowe cen otwarcia trzech firmy. Oprócz KGHM, wybrano także PKOBP i ALIOR. Do narysowania wykresu pudełkowego wykorzystano funkcję *boxplot*. Wyniki porównania prezentuje wykres 2.



Wykres 2: Zmiany procentowe trzech wybranych firm

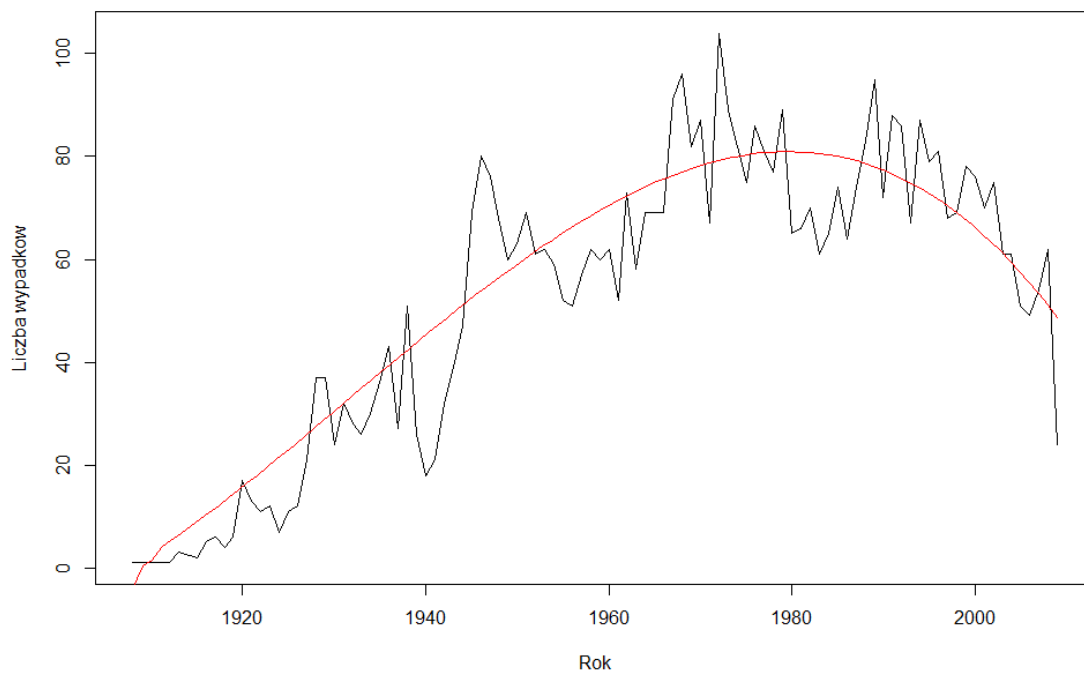
Na powyższym wykresie można zauważyć, że mediana zmiany procentowej cen otwarcia każdej w wybranych firm jest bliska zero. Wąsy wykresu pudełkowego pokazują, że zmiany procentowe są nie większe niż  $\pm 4 - 6\%$ . Widać także zmiany procentowe w firmie KGHM są większe niż w przypadku pozostałych firm. Wskazuje na to większa długość pudełka, a także dłuższe wąsy. ALIOR ma najmniej obserwacji odstających, lecz istnieje wśród nich obserwacja najbardziej odległa w porównaniu z innymi firmami.

## 2 Zadanie 2

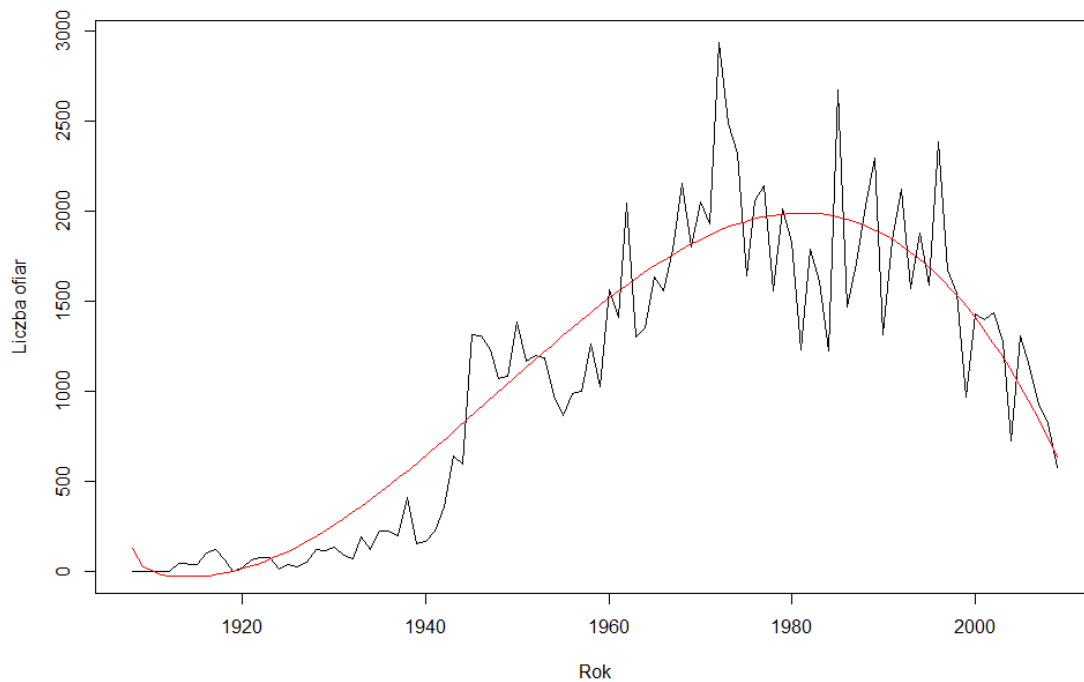
Zliczanie liczby wypadków i ich ofiar śmiertelnych w kolejnych latach przyjmuje pewne założenia. W sytuacji, gdy nieznana jest liczba ofiar w wypadku, owe nieznane wartości są pomijane przy wyliczaniu sumy liczby ofiar w danym roku. Oznacza to możliwe zaniżenie liczby ofiar. Ponadto, jeśli w danym roku nie było wypadków, rok ten jest pomijany.

Zależności liczby wypadków i ich ofiar w kolejnych latach prezentują wykresy 3 i 4 odpowiednio. W celu "wygładzenia" wykresów została zastosowana funkcja *lm*, która pozwala na aproksymację funkcji wielomianem. Przyjęto wielomian trzeciego stopnia, gdyż wyższe stopnie wielomianu wprowadzają błędy numeryczne, a niższe stopnie wielomianu nie pozwalają dostatecznie dobrze "wygładzić" wykresów. Dodatkowo konieczne jest użycie funkcji *predict*, która pozwala narysować wyznaczony wielomian w zadanym przedziale przez wyliczenie wartości funkcji otrzymanej z *lm*.

Na poniższych wykresach widać, że "wygładzanie" zostało zrealizowane zgodnie z oczekiwaniami. W początkowych latach można zauważyć nietypowe "ogonki" wielomianu aproksymującego. Nie wynikają one z błędu aproksymacji, lecz najprawdopodobniej z kwestii numerycznych związanych z ustaleniem kroku metody przybliżającej wielomian.



Wykres 3: Liczba wypadków w kolejnych latach



Wykres 4: Liczba ofiar w kolejnych latach