

# Projekt Egzaminacyjny

Maria Koren

Listopad 2023

## Spis treści

<b>1</b>	<b>Wstęp</b>	<b>3</b>
<b>2</b>	<b>Analiza cen spółek</b>	<b>3</b>
<b>3</b>	<b>Część 1</b>	<b>3</b>
3.1	Wykres kursów zamknięcia pokazujący zmiany w czasie oraz histogram . . .	3
3.2	Statystyki opisowe . . . . .	4
3.2.1	Interpretacja wyników . . . . .	4
3.3	Estymacja parametrów trzech rozkładów korzystając z estymatora największej wiarygodności (MLE) . . . . .	5
3.4	Wykresy diagnostyczne . . . . .	6
3.4.1	Analiza wartości statystyk KS, CM i AD oraz kryteria informacyjne AIC i BIC . . . . .	7
3.5	Testowanie hipotezy o równości rozkładów, wykorzystując statystykę KS . .	7
<b>4</b>	<b>Podsumowanie</b>	<b>8</b>

# 1 Wstęp

W tym projekcie reprezentowana jest analiza danych spółki KMR.UK (Kenmare Resources Plc)

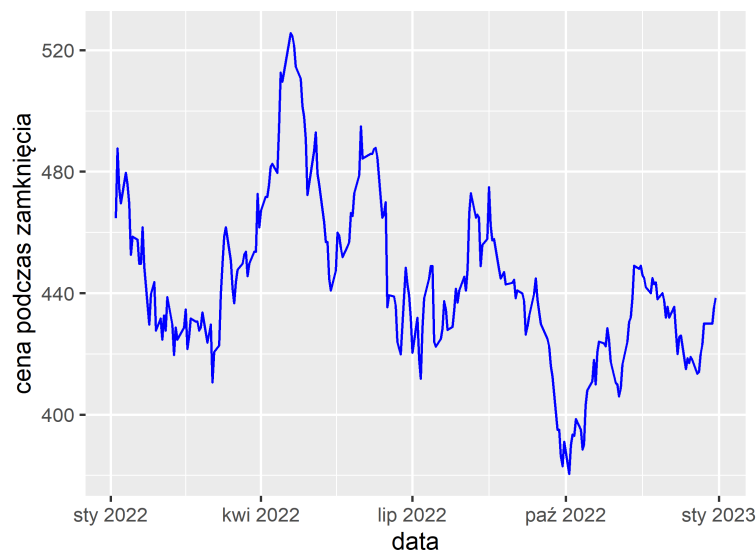
## 2 Analiza cen spółek

Kenmare Resources plc to uznana firma wydobywcza, która zarządza kopalnią minerałów tytanu Moma położoną na północno-wschodnim wybrzeżu Mozambiku. Kopalnia prowadzi produkcję komercyjną od 2009 roku i jest uznawana za głównego dostawcę produktów z piasku mineralnego dla klientów na całym świecie, działających w ponad 15 krajach.

## 3 Część 1

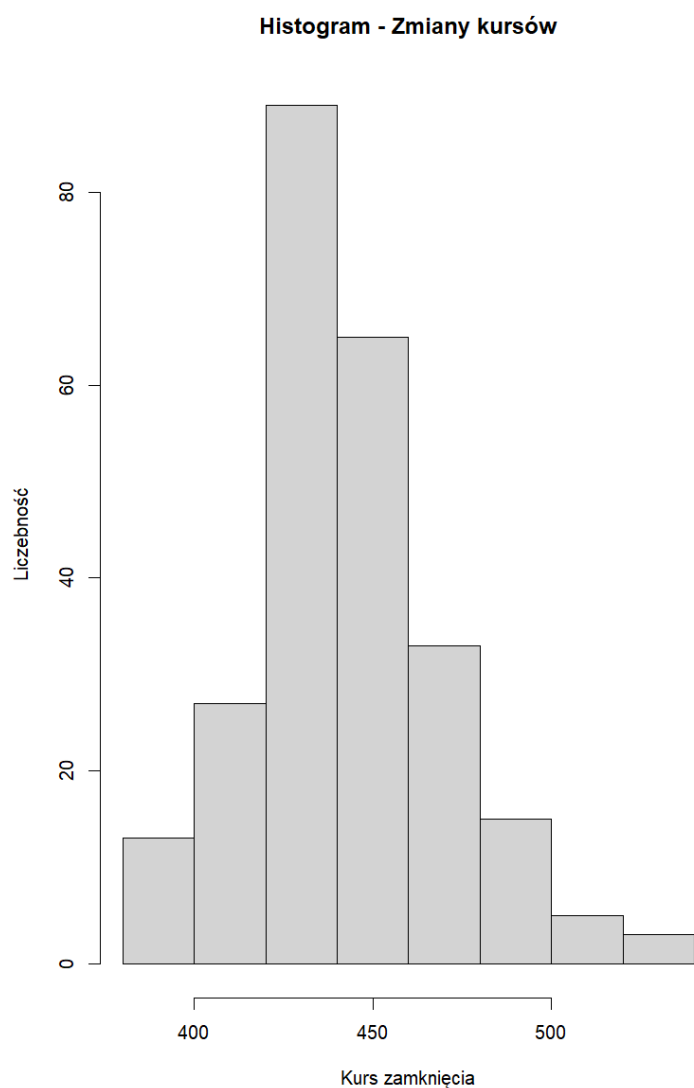
### 3.1 Wykres kursów zamknięcia pokazujący zmiany w czasie oraz histogram

Zrobiono wykres kursu zamknięcia pokazujący zmiany w czasie, rysunek 1



Rysunek 1: Cena podczas zamknięcia

Oraz histogram, pokazujący liczebność danych, rysunek 2



Rysunek 2: Histogram danych

### 3.2 Statystyki opisowe

Zostały obliczone następujące statystyki opisowe: średnia, odchylenie standardowe, skośność oraz kurtoza

Wyniki statystyk znajdują się w poniższej tabeli 1

	$\bar{x}$	odch. st.	skośność	kurtoza
akcje	442.8741	26.7262	3.574323	0.52783

Tabela 1: Statystyki opisowe

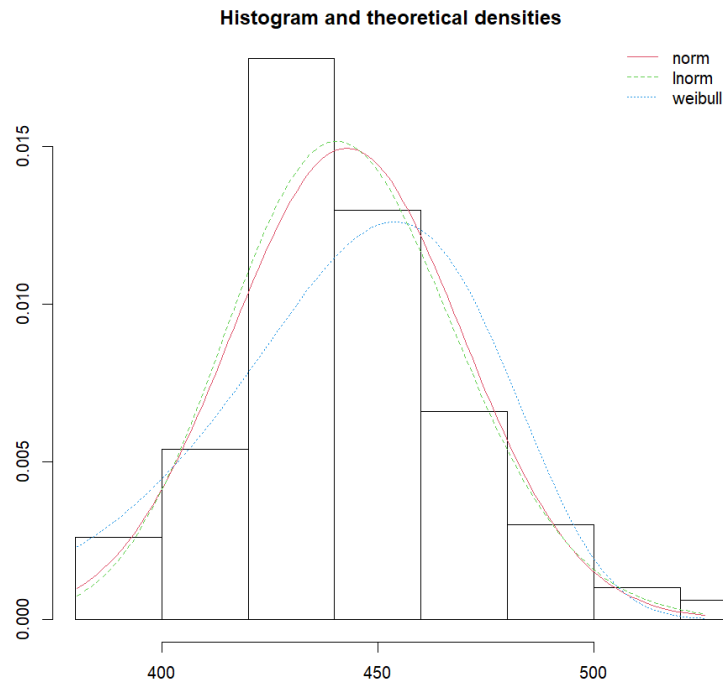
#### 3.2.1 Interpretacja wyników

- Otrzymana skośność mówi o przewadze wartości wyższych (wartość skośności powyżej zera)

- Otrzymana kurtoza mówi o cieńszych ogonach niż rozkład normalny (bardziej płaski) (wartość kurtozy mniej niż 3)

### 3.3 Estymacja parametrów trzech rozkładów korzystając z estymatora największej wiarygodności (MLE)

Wystymowano wyniki trzech rozkładów: normalnego, log-normalnego oraz rozkładu Weibulla za pomocą estymatora MLE. Wyżej wymienione wykresy dodano do wcześniejszego histogramu, co widać za rysunku 3



Rysunek 3: Histogram wraz z estymowanymi rozkładami

Wyniki estymacji parametrów są przedstawione w tabeli 2

	$\mu$	$\sigma$
normalny	442.8741	26.67269
	$\mu$	$\sigma$
log-normalny	6.091499	0.05957788
	$a$	$\sigma$
weibulla	15.61307	455.8914

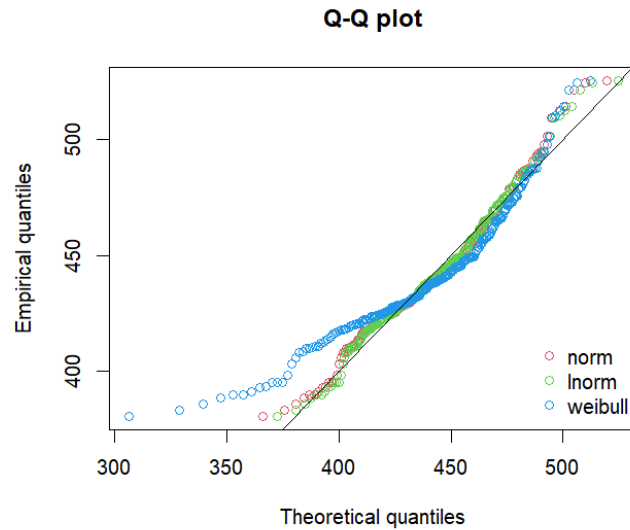
Tabela 2: Wyniki estymacji parametrów

Te wyniki oznaczają, że zostały dopasowane następujące rozkłady:

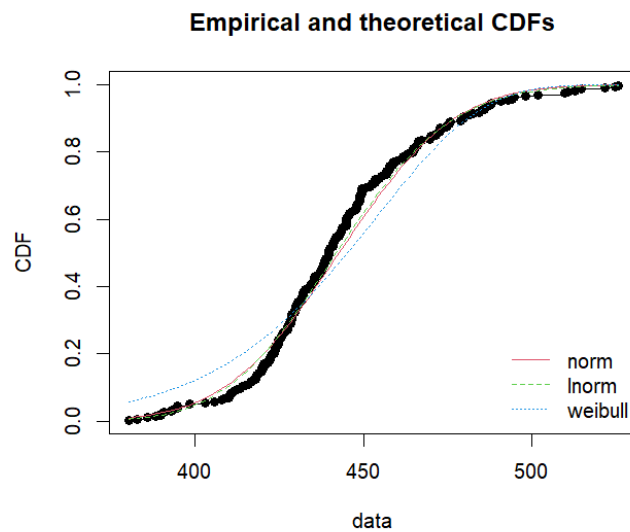
- $X \sim N(442.87, 26.67)$
- $X \sim LN(6.09, 0.059)$
- $X \sim W(15.61, 455.89)$

### 3.4 Wykresy diagnostyczne

Zostały zrobione wykresy diagnostyczne qq-plot (rysunek 4) oraz cdf (rysunek 5)



Rysunek 4: Wykres qq-plot



Rysunek 5: Wykres cdf

- Wykres qq-plot

Jest to wykres kwantyl-kwantyl, na osi pionowej są kwantyle teoretyczne, na osi poziomej są kwantyle empiryczne. Kwantyl rzędu  $\alpha \in (0, 1)$  zmiennej losowej ciągłej  $X$  to taka liczba  $q$ , dla której prawdopodobieństwo, że zmienna  $X$  przyjmuje wartości mniejsze lub równe  $q$  jest równe  $\alpha$ .

Najlepiej jest gdy te kwantyle są takie same bądź bardzo bliskie siebie. Dlatego najlepszym rozkładem jest najbliższy do prostej  $y = x$ . W rozważanym przykład takim jest wykres log-normalny  $X \sim LN(6.09, 0.059)$

- Wykres CDF

Funkcja rozkładu kumulacyjnego (CDF, Cumulative Distribution Function) to graficzna reprezentacja kumulatywnej dystrybuanty danej zmiennej losowej. CDF dla danej wartości  $x$  to prawdopodobieństwo, że zmienna losowa przyjmuje wartość mniejszą lub równą  $x$ . Czarnym zaznaczone są dane empiryczne. Najlepszym wykresem jest mający teoretyczne dane najbliższe do danych empirycznych. W rozważanym przykładzie takim wykresem jest log-normalny  $X \sim LN(6.091, 0.059)$

Na podstawie wykresów diagnostycznych najlepszym rozkładem jest rozkład logarytmiczno-normalny

### 3.4.1 Analiza wartości statystyk KS, CM i AD oraz kryteria informacyjne AIC i BIC

Bazując wyłącznie na wykresach diagnostycznych, nie jest możliwe wybranie jednego najlepszego wykresu. Dlatego skorzystano ze statystyk Kołmogorowa-Smirnowa, Cramera-von-Misesa, Andersona-Darlinga, a także z kryteriów informacyjnych AIC (Akaike's Information Criterion) oraz BIC (Bayesian Information Criterion)

Wartości ze statystyk KS, CM, AD są umieszczone w tabeli 3. Wartości kryteriów AIC, BIC w tabeli 4

	normalny	log-normalny	weibull
Kolmogorov-Smirnov	0.09168955	0.0798396	0.138442
Cramer-von Mises	0.3613063	0.2485924	1.257677
Anderson-Darling	2.005469	1.412547	7.416593

Tabela 3: Statystyki

	normalny	log-normalny	weibull
Akaike's Information Criterion	2355.289	2348.983	2415.692
Bayesian Information Criterion	2362.332	2356.026	2422.735

Tabela 4: Kryteria informacyjne

Ponieważ statystyki są oparte na porównaniu odległości dystrybuant, najlepszym rozkładem jest ten, który jest najbliższy do danych teoretycznych (ma najmniejszą odległość), czyli ma najmniejszą wartość statystyki. W kryteriach informacyjnych za najlepszy rozkład również jest uważany rozkład, mający najmniejszą wartość kryterium. W rozważanym przykładzie takim rozkładem jest rozkład log-normalny  $X \sim LN(6.09, 0.059)$

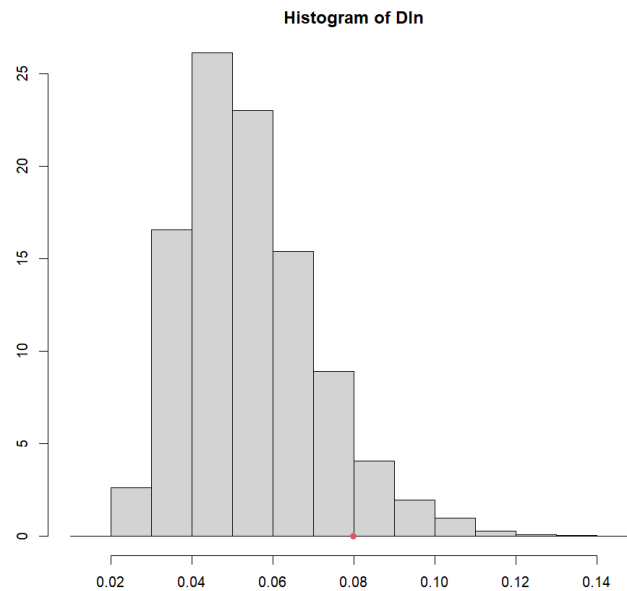
### 3.5 Testowanie hipotezy o równości rozkładów, wykorzystując statystykę KS

Zrobiona hipoteza  $H_0: F = LN(6.09, 0.0595)$  przeciwko hipotezie  $H_1: F$  nie jest równy  $LN(6.09, 0.059)$

Zgenerowano  $N = 10000$  próbek licznosci  $n$  (równej ilości danych) z rozkładu  $F0 = LN(6.09, 0.059)$  wybranego wcześniej jako najlepszego rozkładu i obliczono odległość dystrybuant empirycznych od rozkładu  $F0$  (wartosc statystyki  $Dn$ )

Obliczona również wartość statystyki dla rzeczywistych danych

Rysowany jest histogram statystyk testu KS uzyskanych z danych losowych, a także dodany jest punkt dla statystyki testu KS uzyskanej z rzeczywistych danych dla porównania danych rzeczywistych z danymi losowymi tego rozkładu



Rysunek 6: Histogram danych teoretycznych

Z wykresu widać, że wynik z danych rzeczywistych jest umieszczony w miejscu, gdzie dane losowe jeszcze są

Ta statystyka zwraca 2 dane: odległość (wartość *statistic*) oraz prawdopodobieństwo że wartości statystyki KS są takie same lub większe, gdy hipoteza zerowa jest prawdziwa

Wyniki tego testowania zostaną umieszczone w tabeli 5:

statistic	p-value
0.0798396	0.0827

Tabela 5: Wartość statystyki KS w testowaniu hipotezy

P-wartość informuje o prawdopodobieństwie uzyskania takiej samej lub bardziej ekstremalnej statystyki testu, niż ta, którą otrzymaliśmy z danych rzeczywistych (zakładając że dane pochodzą z tego samego rozkładu)

Ponieważ p-wartość  $p = 0.0827 > 0.05$  zatem nie ma powodów odrzucenia hipotezy. Uzyskane wyniki potwierdzają wybraną hipotezę o logarytmiczno-normalnym  $LN(6.09, 0.059)$  rozkładzie naszych danych

## 4 Podsumowanie

Badając dane spółki Kenmare Resources Plc za 2022 rok, wychodzi że dane kursu zamknięcia mieli rozkład logarytmiczno-normalny  $LN(6.09, 0.059)$