

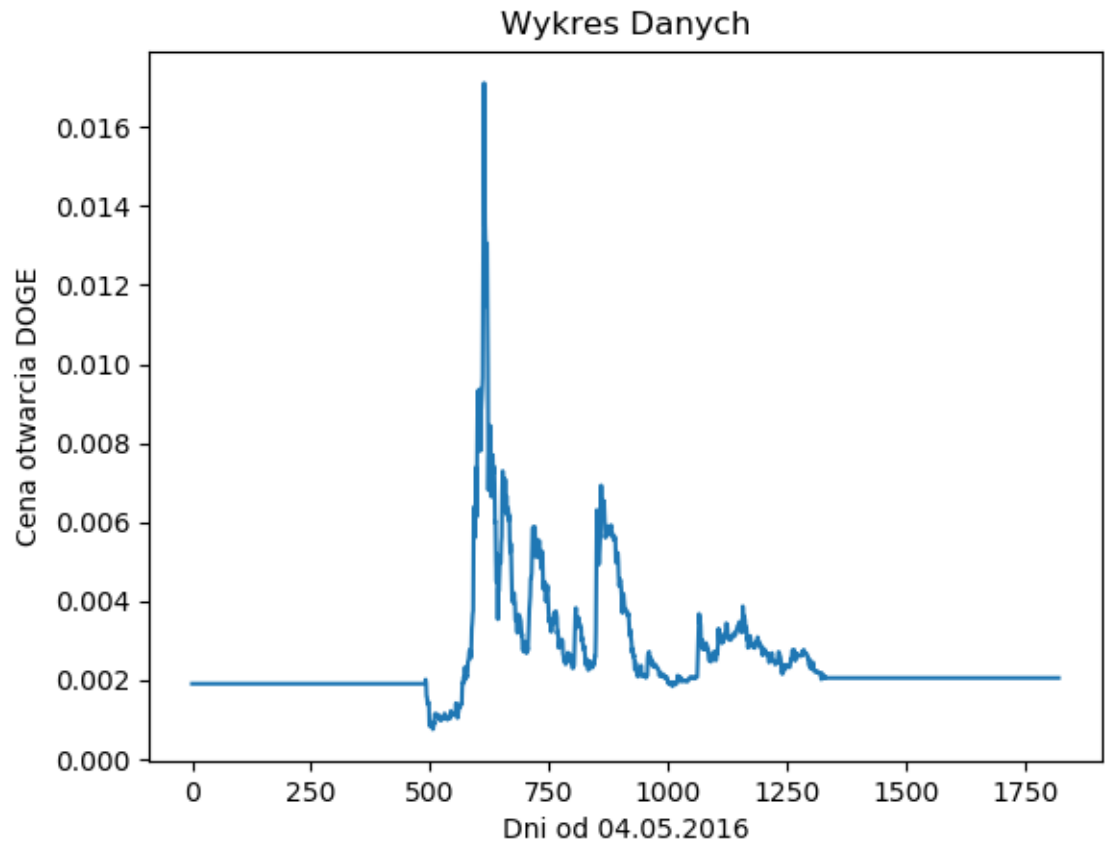
Sprawozdanie 1 ze statystyki

Kostiantyn Skopych, 255916

May 2021

1 Wstęp

Dane użyte w sprawozdaniu to zbiór cen otwarcia monety kryptograficznej Dogecoin (DOGE) za ostatnie 5 lat. Liczba obserwacji: 1827. Dane są wzięte ze strony <https://finance.yahoo.com/quote/DOGE-USD/history?period1=1462320000&period2=1620086400&interval=1d&filter=history&frequency=1d&includeAdjustedClose=true>. Przedstawiam wykres danych:



Rysunek 1: Wykres cen otwarcia DOGE za ostatnie 5 lat

2 Podstawowe statystyki

Średnia arytmetyczna: 0.008081425123422931

Średnia harmoniczna: 0.0008893944599099609

Średnia geometryczna: 0.002060121099394416

Mediana: 0.002059

Kwantyl rzędu 0.25: 0.001912

Kwantyl rzędu 0.5: 0.002059

Kwantyl rzędu 0.75: 0.0026475

Wariancja: 2.191291579337806e-06

Odchylenie standardowe: 0.0014803011785909672

Współczynnik skośności: 3.8112305538434765

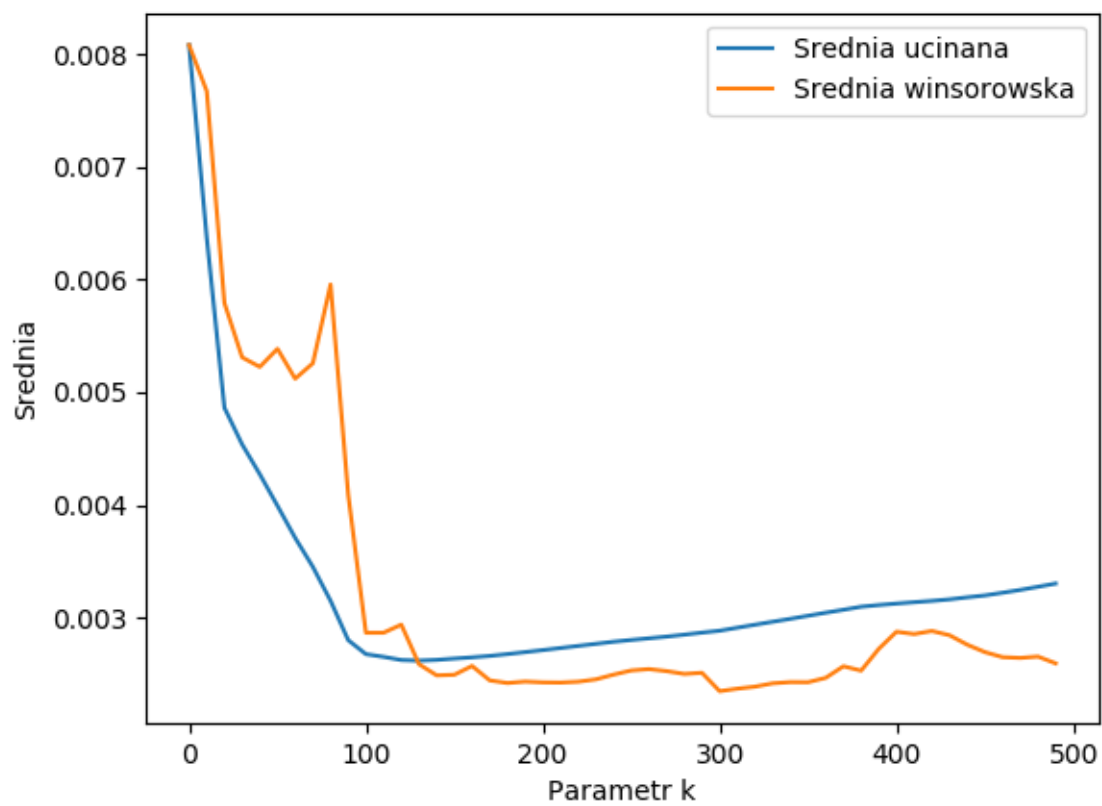
Kurtoza: 20.659248734341933

Eksces: 17.659248734341933

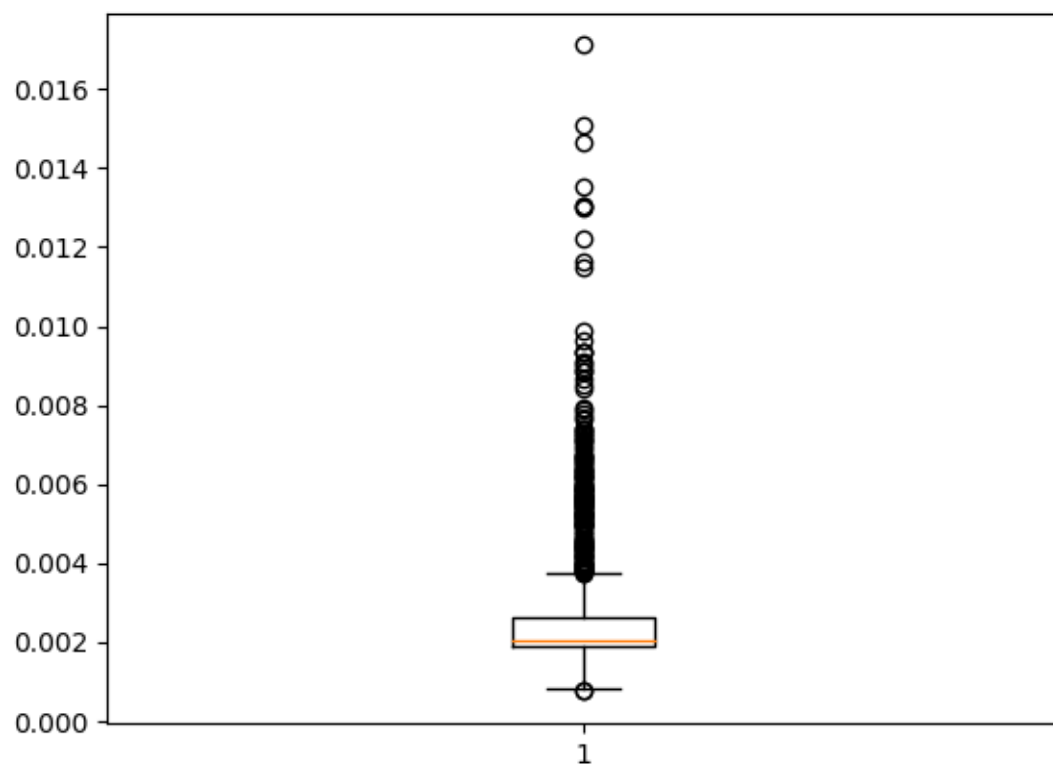
Z powyższych danych można wyciągnąć kilka wniosków. Rozkład danych jest dużo mniej spłaszczony niż rozkład normalny. Mówią nam o tym kurtoza i eksces. Mediana różni się znacząco od średnich, co świadczy o tym, że dane rozkładają się bardzo nierównomiernie. O tym też świadczą kwantyle, które są "prawie równe" sobie, czyli nie różnią się za bardzo. Oczywiście kwantyl rzędu 0.5 jest równy medianie. Zwracamy też uwagę na duże odchylenie standardowe. Powyższe wnioski potwierdza też współczynnik skośności, który mówi, że dane są niesymetryczne.

3 Wizualizacja danych

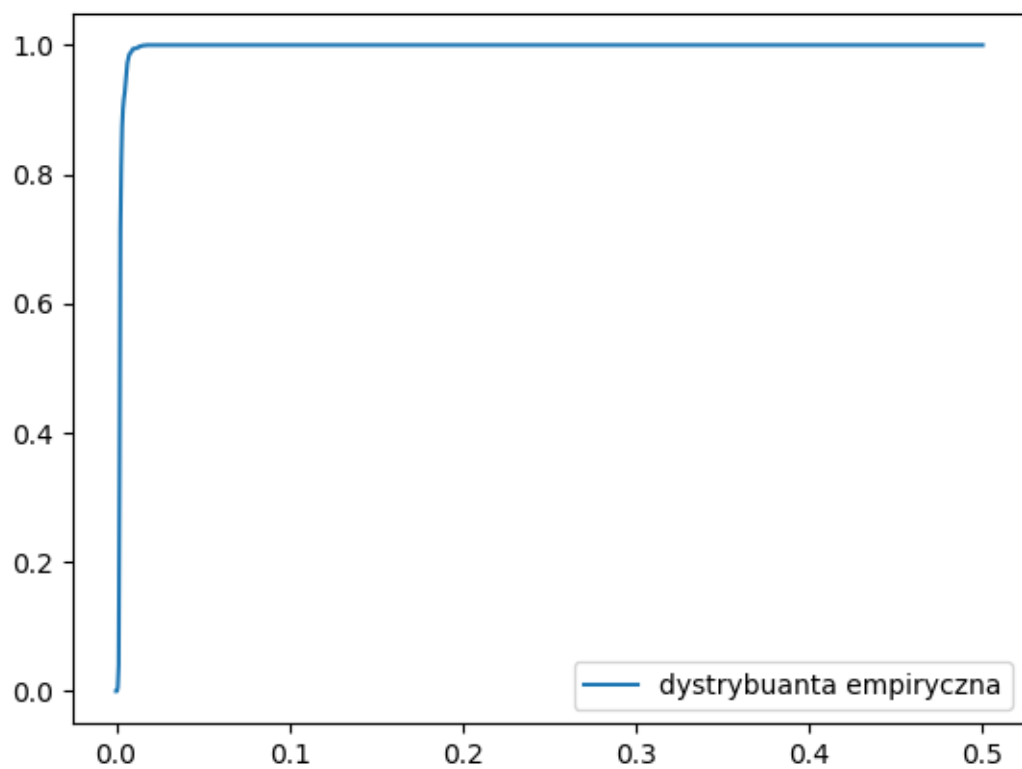
Poniżej są przedstawione kilka wykresów wizualizujących dane i pozwalających poprowadzić analizę danych dalej. Są przedstawione dystrybucja empiryczna rozkładu danych, histogram gęstości rozkładu danych, wykres zależności średniej winsorowskiej i średniej ucinanej danych od parametru k , oraz boxplot.



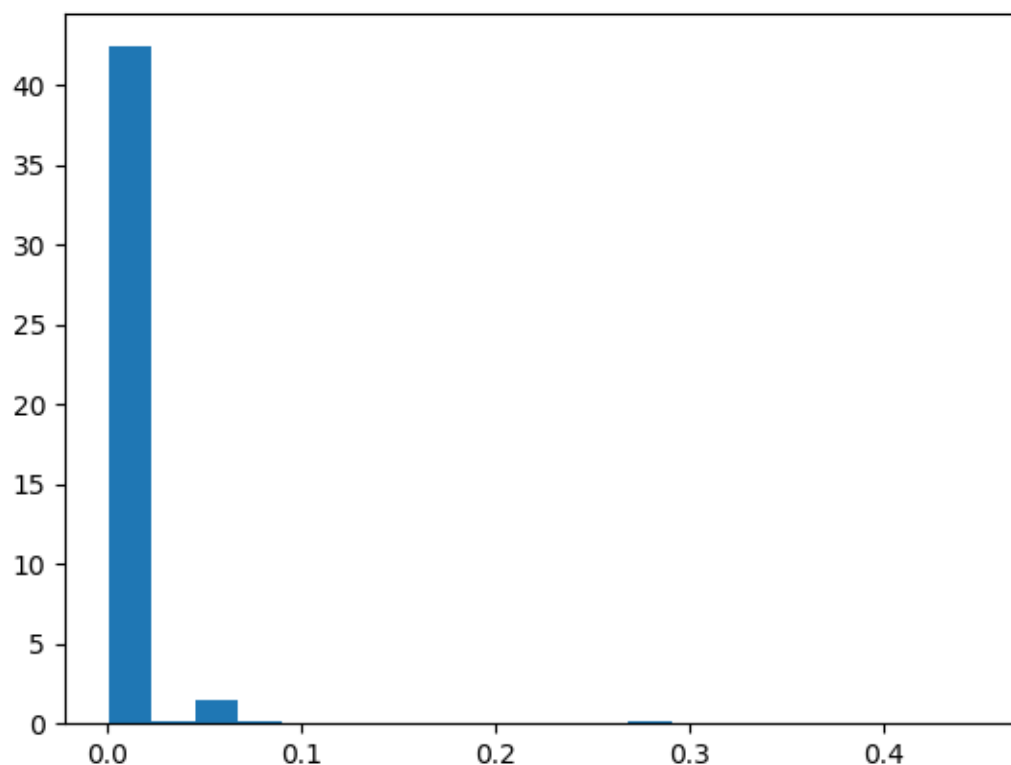
Rysunek 2: Wykres zależności średnich ucinanej oraz winsorowskiej od parametru k



Rysunek 3: Boxplot



Rysunek 4: Dystrybuanta empiryczna



Rysunek 5: Histogram gęstości

Średnie ucinane i winsorowskie pokazują nam, że bardzo małych wartości jest dużo więcej niż dużych wartości. Wnioskujemy to z tego, że wykres szybko maleje. Dystrybuanta empiryczna i histogram gęstości najlepiej pokazują rozkład danych. Widać dobrze jak duża jest nierównomierność w rozkładzie danych. Boxplot też jest ciekawy. Widzimy tutaj nasze kwantyle, w tym medianę, minimum i maksimum oraz ogólny rozkład danych, o którym już dużo mówiono powyżej.

4 Użyte wzory oraz definicje

Średnia arytmetyczna – suma liczb podzielona przez ich liczbę. $Sr.ar. = \frac{a_1 + a_2 + \dots + a_n}{n}$
Średnią harmoniczną nazywamy liczbę podaną wzorem: $\frac{n}{\frac{1}{a_1} + \frac{1}{a_2} + \dots + \frac{1}{a_n}}$ Średnią geometryczną nazywamy liczbę podaną wzorem: $\sqrt[n]{a_1 \cdot a_2 \cdot \dots \cdot a_n}$ Mediana - wartość cechy w szeregu uporządkowanym, powyżej i poniżej której znajduje się

jednakowa liczba obserwacji. Mediana jest drugim kwartylem, czyli kwantylem rzędu $1/2$ Pierwszy kwartyl = dolny kwartyl = kwantyl rzędu $1/4 = 25$ Trzeci kwartyl (notacja: Q3) = górny kwartyl = kwantyl rzędu $3/4 = 75$ dzieli zbiór obserwacji na dwie części odpowiednio po 75

Wariancja – klasyczna miara zmienności. Intuicyjnie utożsamiana ze zróżnicowaniem zbiorowości; jest średnią arytmetyczną kwadratów odchyłeń (różnic) poszczególnych wartości cechy od wartości oczekiwanej. Wyznacza się wzorem: $D^2(X) = E(X^2) - E(X)^2$ Wartość oczekiwana - wartość określająca spodziewany wynik doświadczenia losowego. Wyraża się wzorem: $EX = \sum_{i=1}^n x_i p_i$ Odchylenie standardowe mówi, jak szeroko wartości jakiejś wielkości (na przykład wieku, inflacji, kursu walutowego) są rozrzucone wokół jej średniej. Współczynnik skośności rozkładu to miara asymetrii rozkładu wyznaczana według wzoru: $A = 3 \frac{\mu - m}{s}$, gdzie μ to średnia arytmetyczna, m to mediana, s to odchylenie standardowe. Kurtosis - jedna z miar kształtu rozkładu wartości cechy. Miara spłaszczenia. Definiuje się ją następującym wzorem: $Kurt = \frac{\mu_4}{\sigma^4}$, gdzie μ_4 to czwarty moment centralny, σ to odchylenie standardowe. Eksces jest równy Kurt-3. W praktyce wielkość spłaszczenia odnosimy do spłaszczenia rozkładu normalnego, którego kurtosis jest równa 3. Stąd we wzorze na eksces pojawia się -3.

5 Podsumowanie

Przeprowadziliśmy analizę danych ceny otwarcia monety kryptograficznej Dogecoin. Najważniejszym wnioskiem jest to, że większość czasu moneta była "spokojna", czyli jej cena znajdowała się na stabilnie niskim poziomie. Jednak było tak nie cały czas i ta moneta pokazywała czasem mocny wzrost.