Katedra Elektrotechniki i Podstaw Informatyki

****

**Projekt ze statystycznej analizy danych – ludzie bezrobotni uprzednio pracujący w latach 2014 i 2024**

**Stechnij Damian**

Opiekun pracy:  
dr inż. Mariusz Startek

Rzeszów, 2024

Spis treści

[1. Wprowadzenie 3](#_Toc167123975)

[2. Importowanie danych 4](#_Toc167123976)

[3. Wyznaczenie parametrów opisowych 5](#_Toc167123977)

[3.1. Obliczanie parametrów 5](#_Toc167123978)

[3.1.1. Średnia 5](#_Toc167123979)

[3.1.2. Wariancja 5](#_Toc167123980)

[3.1.3. Odchylenie standardowe 5](#_Toc167123981)

[3.1.4. Współczynnik zmienności 5](#_Toc167123982)

[3.1.5. Kwantyle, mediana, minimum, maksimum 6](#_Toc167123983)

[3.1.6. Rozstęp 6](#_Toc167123984)

[3.1.7. Skośność 6](#_Toc167123985)

[3.1.8. Dominanta 6](#_Toc167123986)

[3.1.9. Kurtoza 6](#_Toc167123987)

[3.2. Użyte funkcje i biblioteki 6](#_Toc167123988)

[3.3. 6](#_Toc167123989)

[4. Graficzna prezentacja danych 7](#_Toc167123990)

[4.1. Wykres pudełkowy 7](#_Toc167123991)

[4.2. Histogram 7](#_Toc167123992)

[4.3. Wykres dystrybuanty 8](#_Toc167123993)

[4.4. Użyte funkcje i biblioteki 9](#_Toc167123994)

[5. Hipotezy statystyczne 9](#_Toc167123995)

[5.1. 9](#_Toc167123996)

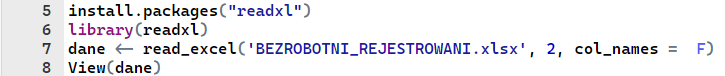
[5.2. 9](#_Toc167123997)

[6. Wnioski 10](#_Toc167123998)

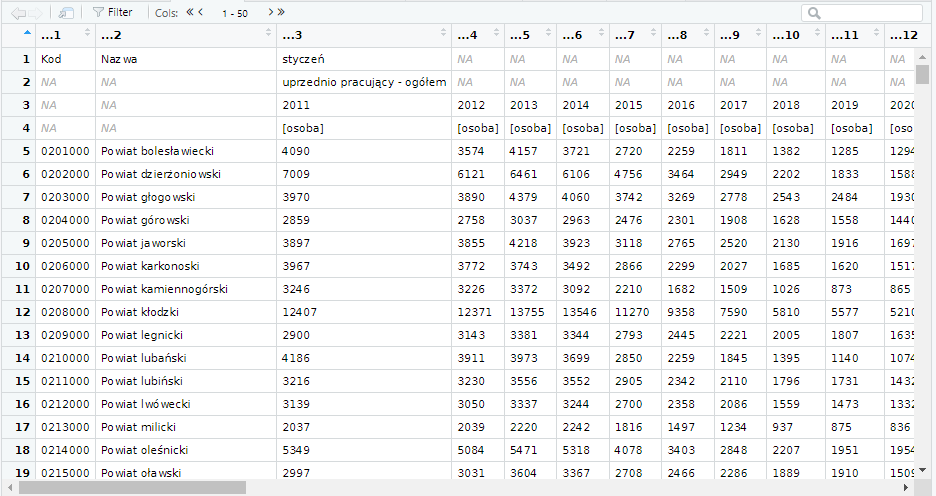
# Wprowadzenie

Projekt dotyczy statystycznej analizy danych ludzi bezrobotnych, którzy uprzednio pracowali. Dane zostały pobrane ze strony <https://bdl.stat.gov.pl>. Zawierają one lata od 2011 do 2024 z podziałem na miesiące oraz próbki ilości takowych ludzi pochodzą z powiatów w Polsce. Wybrano te dane, aby przeprowadzić analizę różnicy w próbkach pomiędzy rokiem 2014 a 2024 w miesiącu styczeń. W projekcie wyznaczono parametry opisowe, przedstawiono graficznie zebrane dane oraz zbadano hipotezy statystyczne.

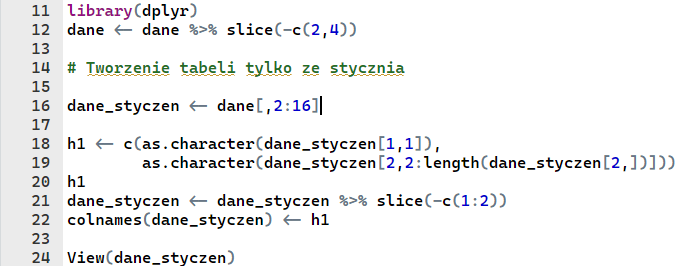
# Wczytanie danych

Wczytywany plik jest w formacie xlsx, więc trzeba użyć biblioteki readxl.

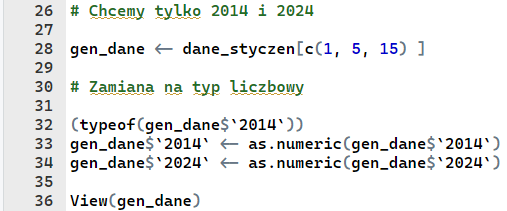
Zrzut ekranu 2.1 Wczytanie danych

Po wczytaniu danych otrzymujemy taką tabele z ilością osób bezrobotnych wcześniej pracujących z poszczególnych powiatów.

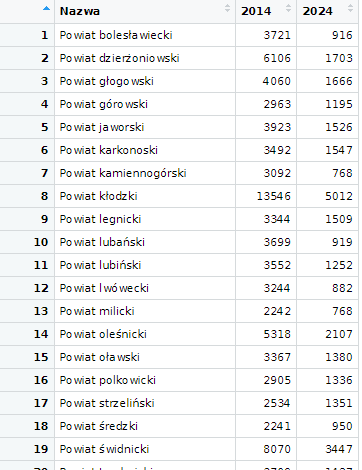
Zrzut ekranu 2.2 Tabela danych po wczytaniu

Postanowiono zebrać dane tylko ze stycznia, więc obcięto dane z pozostałych miesięcy oraz z pola z wartościami NA.

Zrzut ekranu 2.3 Obcięcie danych tylko do stycznia

Po obcięciu danych wybrano lata 2014 i 2024 by porównać 10 lat odstępu i zamieniono na typ liczbowy.

Zrzut ekranu 2.4 Wybranie tylko lat 2014 i 2024

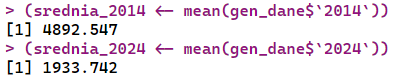
Gotowe dane po obróbce wyglądają następująco:

Zrzut ekranu 2.5 Dane po obróbce

# Wyznaczenie parametrów

## Średnia

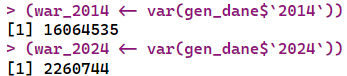
Obliczono średnią z obu lat. Użyto do tego funkcji mean, która oblicza średnią arytmetyczną. Średnia na przestrzeni lat zmieniła się ponad dwukrotnie, co oznacza, że o wiele mniej ludzi traci swoją pracę lub rzadziej zmieniają.



Zrzut ekranu 3.1 Obliczenie średniej

## Wariancja

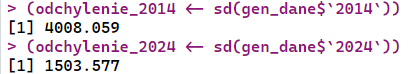
Następnie obliczono wariancję, która jest miarą rozproszenia danych wokół średniej wartości, dla obu kolumn.



Zrzut ekranu 3.2 Obliczenie wariancji

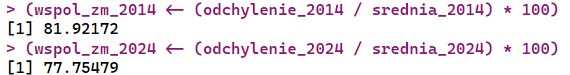
## Odchylenie standardowe

Odchylenie standardowe



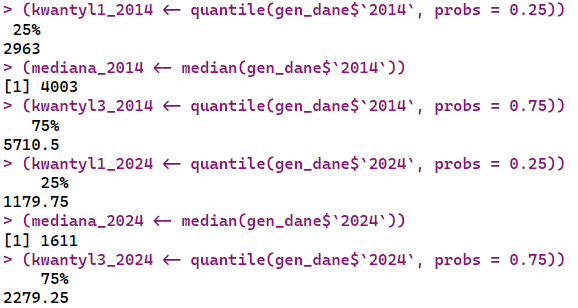
Zrzut ekranu 3.3 Obliczenie odchylenia standardowego

## Współczynnik zmienności

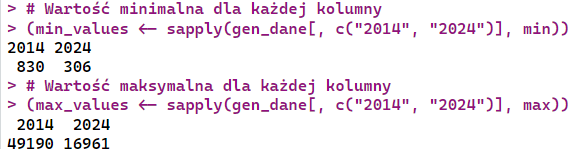


Zrzut ekranu 3.4 Obliczenie współczynnika zmienności

## Kwantyle, mediana, minimum, maksimum

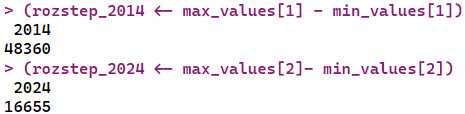


Zrzut ekranu 3.5 Obliczenie kwantyli



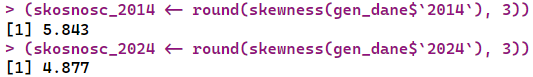
Zrzut ekranu 3.6 Znalezienie wartości maksymalnej i minimalnej

## Rozstęp



Zrzut ekranu 3.7 Obliczenie roztępu

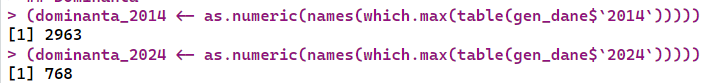
## Skośność



Zrzut ekranu 3.8 Obliczenie skośności

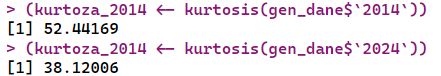
## Dominanta

Dominanta to wartość, którą wyznacza się poprzez znalezienie najczęściej występującej liczby w zbiorze danych.



Zrzut ekranu 3.9 Obliczenie dominanty

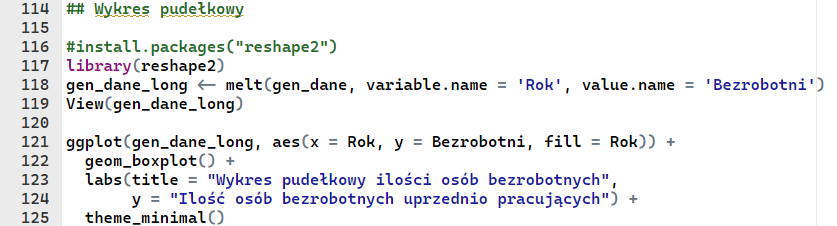
## Kurtoza

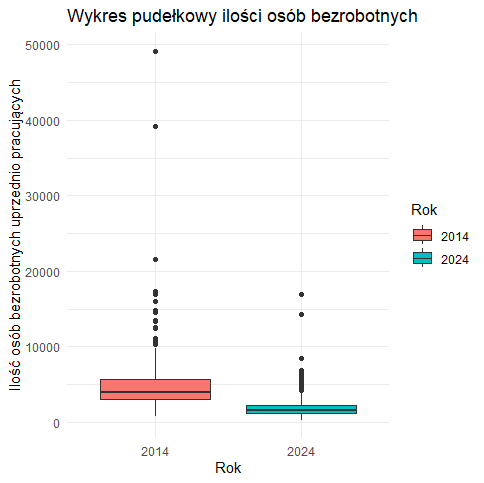


Zrzut ekranu 3.10 Wyniki dla kurtozy

# Graficzna prezentacja danych

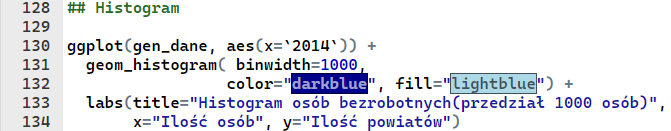
## Wykres pudełkowy



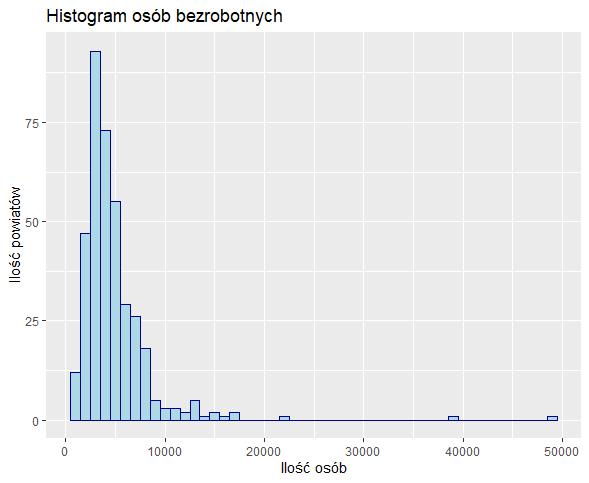
Zrzut ekranu 4.1 Kod tworzenia wykresu pudełkowego

Rysunek 4‑1 Wykres pudełkowy

## Histogram

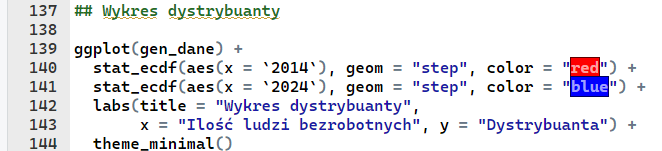


Zrzut ekranu 4.2 Kod tworzenia histogramu

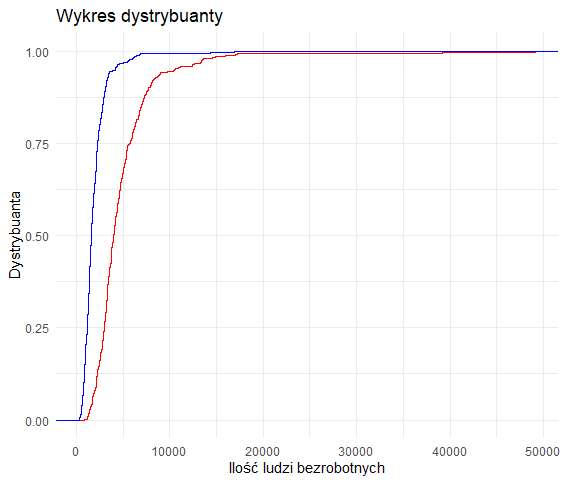


Rysunek 4‑2 Histogram

## Wykres dystrybuanty



Zrzut ekranu 4.3 Kod tworzenia wykresu dystrybuanty



Rysunek 4‑3 Wykres dystrybuanty

# Hipotezy statystyczne

## Hipoteza dotycząca mediany liczby bezrobotnych

H0: Średnia ludzi bezrobotnych nie zmieniła się

H1: Średnia ludzi bezrobotnych zmniejszyła się

Wartość p jest mniejsza niż 0,05 co oznacza, że hipoteza zerowa jest nieprawdziwa

## Hipoteza dotycząca wariancji liczby bezrobotnych

# Użyte biblioteki i polecenia

# Wnioski