

Raport

Porównanie wynagrodzeń kobiet i mężczyzn w
Polsce w maju 2024 r.

1 Wstęp

W tej pracy porównamy przeciętne wynagrodzenia kobiet i mężczyzn, a następnie spróbujemy odpowiedzieć na pytanie, czy faktycznie kobiety w Polsce zarabiają mniej. Do analizy wykorzystamy dane GUS zawierające przeciętne wynagrodzenie obu płci w poszczególnych powiatach w maju 2024 roku¹. Korzystając z różnych metryk statystycznych ocenimy, jak wygląda różnica płac w skali całego kraju.

2 Metryki statystyczne

Poniżej znajdują się wzory i definicje metryk, które zastosujemy do analizy danych.

2.1 Miary położenia

2.1.1 Średnia arytmetyczna

Średnią arytmetyczną wyrażamy wzorem

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

2.1.2 Średnia harmoniczna

Średnią harmoniczną wyrażamy wzorem

$$\bar{x}_h = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}.$$

2.1.3 Średnia geometryczna

Średnią geometryczną wyrażamy wzorem

$$\bar{x}_g = \sqrt[n]{\prod_{i=1}^n x_i}.$$

¹<https://stat.gov.pl/obszary-tematyczne/rynek-pracy/pracujacy-zatrudnieni-wynagrodzenia-koszty-pracy/roznice-w-wynagrodzeniach-kobiet-i-mezczyzn-w-gospodarce-narodowej-w-polsce-w-2024-r-,35,1.html>

2.1.4 Średnia ucinana

Średnią ucinaną uporządkowanej rosnąco próby x_1, \dots, x_n wyrażamy wzorem

$$\bar{x}_u = \frac{1}{n - 2k} \sum_{i=k+1}^{n-k} x_i.$$

Obliczymy dla $k = \lfloor 0,1n \rfloor$.

2.1.5 Średnia winsorowska

Średnią winsorowską uporządkowanej rosnąco próby x_1, \dots, x_n wyrażamy wzorem

$$\bar{x}_w = \frac{1}{n} \left[(k+1)x_{k+1} + \sum_{i=k+2}^{n-k-1} x_i + (k+1)x_{n-k} \right].$$

Obliczymy dla $k = \lfloor 0,1n \rfloor$.

2.1.6 Mediana

Medianę z uporządkowanej rosnąco próby x_1, \dots, x_n wyrażamy wzorem

$$\text{Med}(x) = \begin{cases} x_{\frac{n+1}{2}}, & \text{gdy } n \text{ jest liczbą nieparzystą} \\ \frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}), & \text{gdy } n \text{ jest liczbą parzystą.} \end{cases}$$

2.1.7 Kwartyle

Kwartyle dzielą zbiór danych na cztery grupy:

- drugi kwartył (Q_2) to mediana,
- pierwszy kwartył (Q_1) to mediana grupy obserwacji mniejszych od Q_2 ,
- trzeci kwartył (Q_3) to mediana grupy obserwacji większych od Q_2 .

2.1.8 Dominanta

Dominanta (D) to wartość, która występuje najczęściej w badanej próbie.

2.2 Miary rozproszone

2.2.1 Rozstęp z próby

Rozstęp z uporządkowanej rosnąco próby x_1, \dots, x_n wyrażamy wzorem

$$R = x_n - x_1.$$

2.2.2 Rozstęp międzykwartyłowy

Rozstęp międzykwartyłowy wyrażamy wzorem

$$IQR = Q3 - Q1.$$

2.2.3 Wariancja z próby

Wariancję z próby wyrażamy wzorem

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

2.2.4 Odchylenie standardowe z próby

Odchylenie standardowe z próby wyrażamy wzorem

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}.$$

2.2.5 Odchylenie przeciętne od wartości średniej

Odchylenie przeciętne od wartości średniej wyrażamy wzorem

$$d_1 = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|.$$

2.2.6 Współczynnik zmienności

Współczynnik zmienności wyrażamy wzorem

$$\nu = \frac{s}{\bar{x}} \cdot 100\%.$$

2.3 Miary kształtu

2.3.1 Współczynnik skośności

Współczynnik skośności wyrażamy wzorem

$$\alpha = \frac{n}{(n-1)(n-2)} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s} \right)^3.$$

2.3.2 Kurtoza

Kurtozę wyrażamy wzorem

$$K = \frac{n-1}{(n-2)(n-3)} ((n+1)K_1 - 3(n-1)) + 3,$$

gdzie

$$K_1 = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^2}.$$

3 Analiza wyników

‘ Wartości obliczonych metryk zostały przedstawione w poniższej tabelce

Metryka	Mężczyźni	Kobiety
\bar{x}	7628,38	6955,66
\bar{x}_h	7530,22	6916,38
\bar{x}_g	7576,47	6934,94
\bar{x}_u	7475,07	6855,03
\bar{x}_w	7542,85	6890,07
$Q1$	7008,16	6630,69
$Q2$	7298,91	6786,28
$Q3$	7996,21	7086,83
D	6165,46	6072,07
R	6587,00	4678,09
IQR	988,06	456,14
σ^2	897752,01	325419,84
σ	947,50	570,46
d_1	691,63	379,58
ν	12,42	8,20
α	1,87	2,60
K	7,55	12,54

Na podstawie obliczonych miar możemy sformułować następujące wnioski:

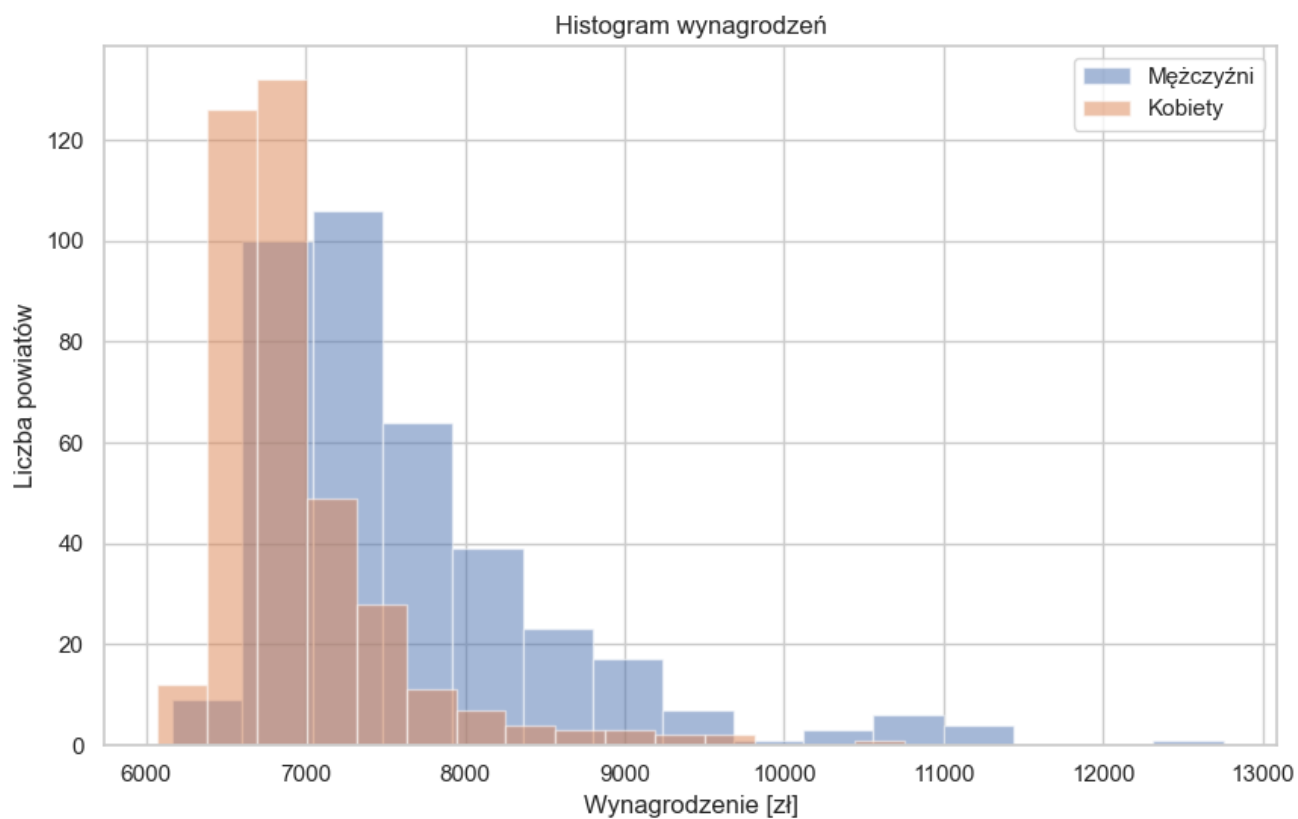
- Miary położenia:
 - Średnia arytmetyczna wynagrodzeń mężczyzn jest wyższa niż kobiet, co wskazuje na systematyczne różnice w płacach między płciami.
 - Mediana w obu przypadkach jest niższa od średniej, co sugeruje prawostronną asymetrię rozkładu (obecność powiatów z wyjątkowo wysokimi zarobkami).
 - Średnia ucinana i średnia winsorowska są niższe od klasycznej średniej, co wskazuje na wpływ obserwacji odstających (tzw. „outliers”) na wartość przeciętną.

- Średnia geometryczna i harmoniczna są jeszcze niższe, co dodatkowo potwierdza niesymetryczny i rozciągnięty rozkład płac.
- Miary rozproszenia:
 - Rozstęp z próby i rozstęp międzykwartyłowy są wyraźnie większe u mężczyzn, co oznacza większe zróżnicowanie ich wynagrodzeń.
 - Odchylenie standardowe oraz wariancja są również wyższe u mężczyzn, co potwierdza większą niestabilność płac w tej grupie.
 - Odchylenie przeciętne i współczynnik zmienności również są wyższe u mężczyzn, co oznacza większą względną zmienność w stosunku do średniego wynagrodzenia.
- Miary kształtu:
 - Skośność rozkładów jest dodatnia dla obu grup, ale wyższa u mężczyzn – co oznacza, że rozkład płac ma długi „ogon” w prawo.
 - Kurtoza przekracza wartość 3 (czyli tzw. rozkład normalny) – mamy do czynienia z rozkładami leptokurtycznymi, bardziej spiczastymi, z większym udziałem wartości skrajnych.

3.1 Wykresy

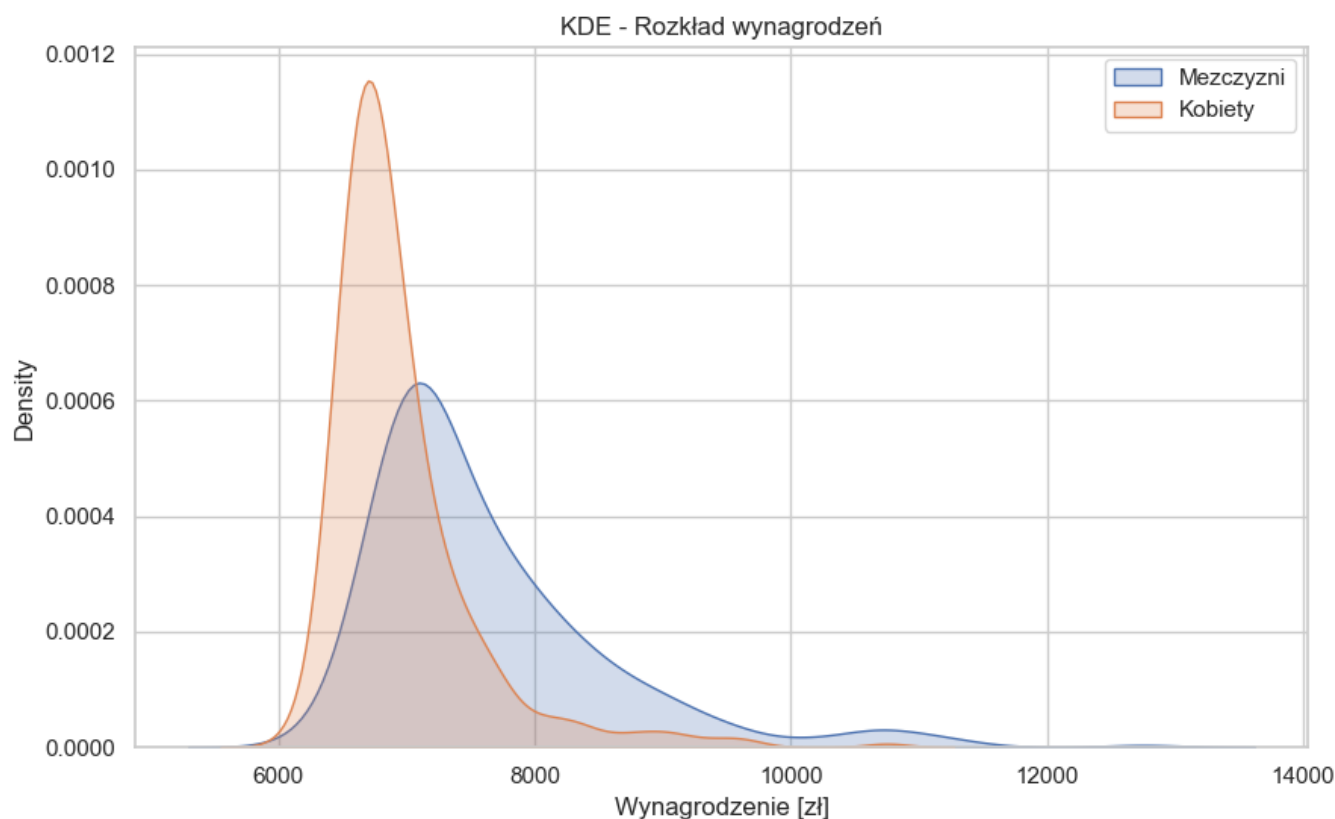
Wyniki przeprowadzonej analizy pokażemy na następujących wykresach.

3.1.1 Histogram wynagrodzeń



Histogram przedstawia rozkład przeciętnych wynagrodzeń mężczyzn i kobiet w poszczególnych powiatach. Pokazuje, ile powiatów mieści się w określonych przedziałach wynagrodzeń.

3.1.2 Rozkład wynagrodzeń

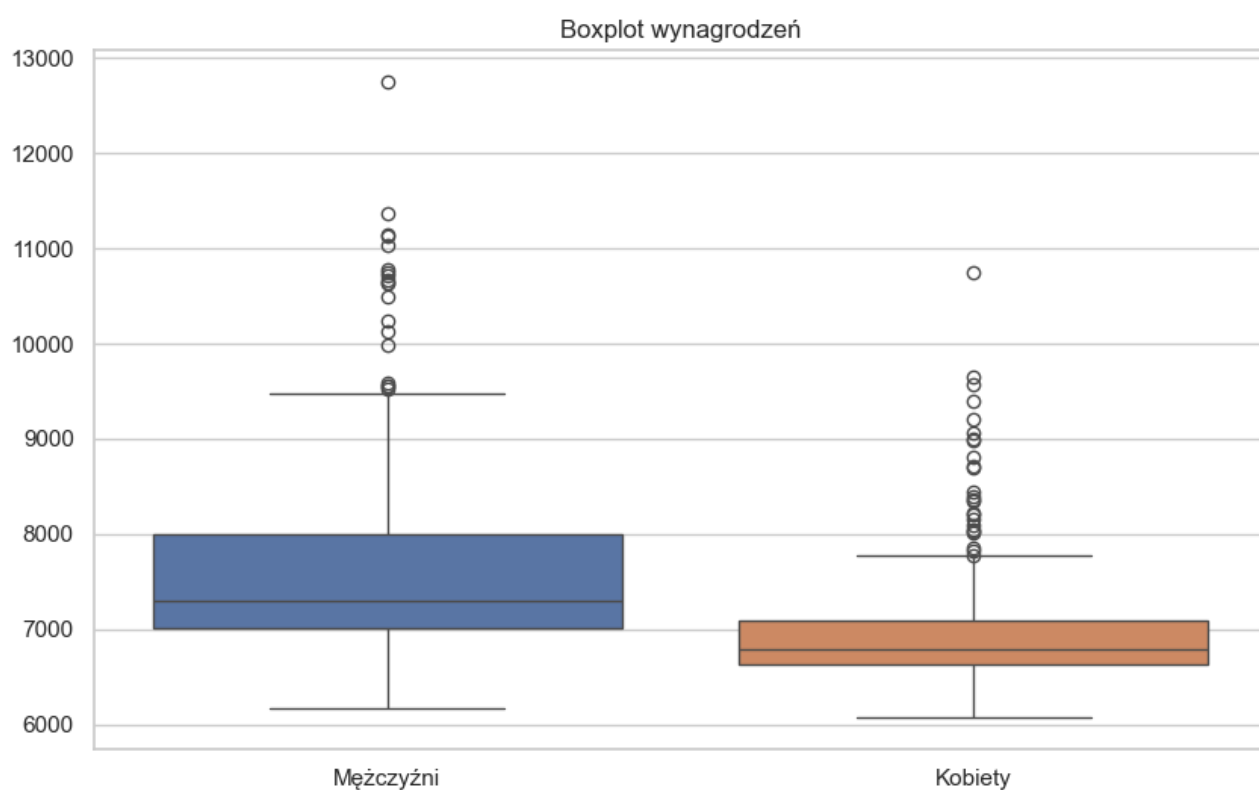


Wykres przedstawia rozkład przeciętnych wynagrodzeń kobiet i mężczyzn w poszczególnych powiatach na podstawie KDE. Oznacza to, że każda obserwacja na wykresie odpowiada średniemu wynagrodzeniu w jednym powiecie, osobno dla kobiet i dla mężczyzn.

Powyższy wykres oraz histogram dobrze obrazują podane wcześniej wnioski. Pomarańczowa krzywa jest wysoka i wąska, co oznacza, że większość powiatów ma zbliżony poziom przeciętnych płac kobiet, natomiast niebieska krzywa jest szersza z dłuższym ogonem w prawą stronę. Oznacza to, że zróżnicowanie płac wśród kobiet jest mniejsze, natomiast mężczyźni w niektórych

powiatach mogą liczyć na znacznie większe wynagrodzenie. Ponadto, zarówno na histogramie, jak i wykresie widać przesunięcie niebieskiego wykresu w prawo względem pomarańczowego, co wskazuje na systematyczną różnicę płac ze względu na płeć.

3.1.3 Boxplot wynagrodzeń

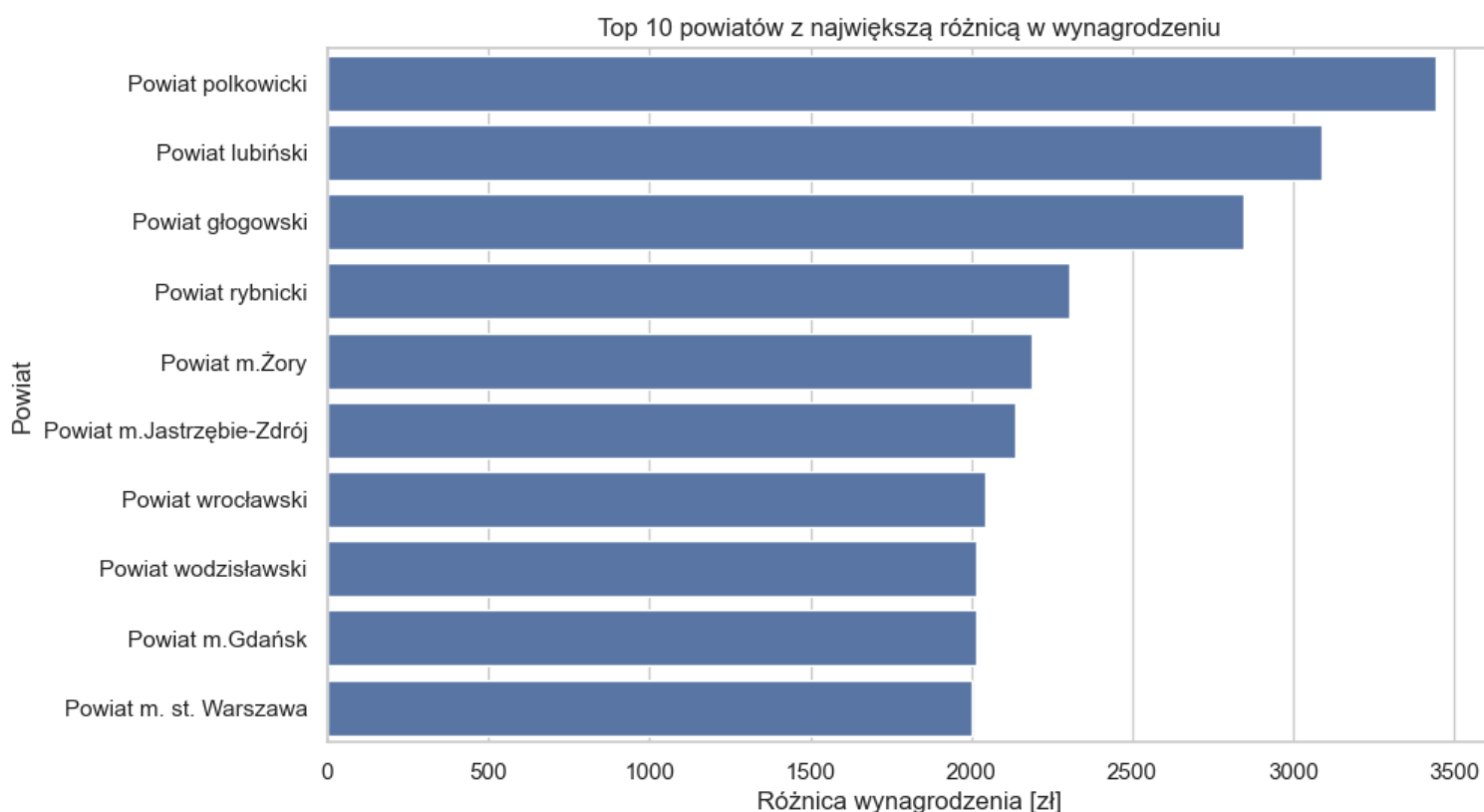


Wykres boxplot przedstawia rozkład przeciętnych wynagrodzeń kobiet i mężczyzn w poszczególnych powiatach. Dolna i górna krawędź pudełka oznacza pierwszy i trzeci kwartył, natomiast środkowa linia medianę. Długość pudełka to rozstęp międzykwartyłowy, a wąsy to zasięg wartości nieodbiegających

znacząco od reszty. Kropki ponad górnymi wąsami oznaczają wyjątkowo wysokie wartości wynagrodzeń.

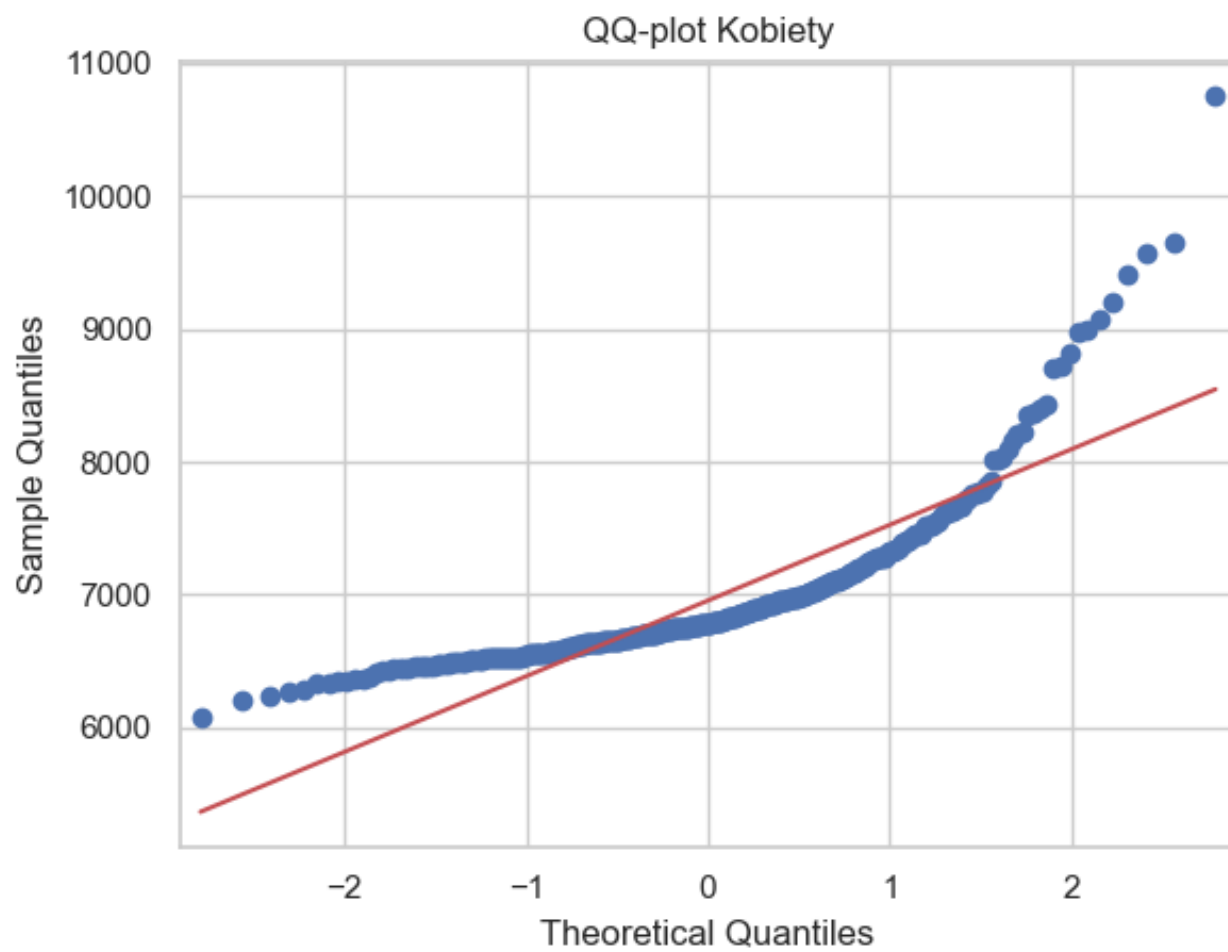
Podobnie jak z wykresu gęstości i histogramu możemy odczytać z boxplotu informacje o większej rozpiętości oraz wysokości płac mężczyzn. Ponadto na tym wykresie można bardzo dobrze zauważyć, że najwyższa przeciętna płaca mężczyzn jest wyższa od najwyższej przeciętnej płacy kobiet aż o około 2000 zł.

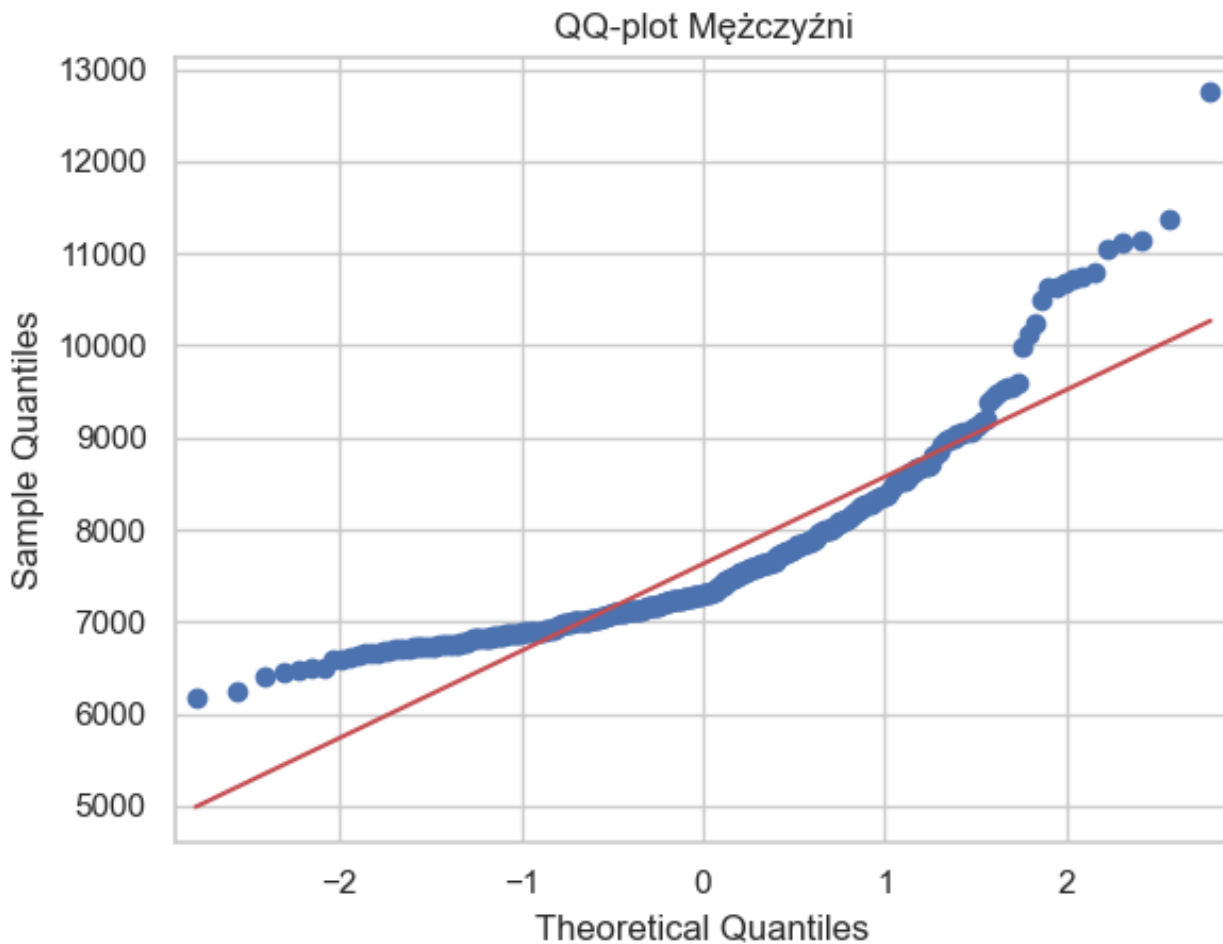
3.1.4 Powiaty z największą różnicą w wynagrodzeniu



Wykres przedstawia dziesięć powiatów z największą różnicą w wynagrodzeniu między mężczyznami a kobietami. Wynika z niego, że różnice przeciętnych płac płci mogą się różnić nawet o prawie 3500 zł w obrębie jednego powiatu.

3.1.5 QQ-plot wynagrodzeń





Powyższe wykresy QQ-plot porównują kwantyle danych empirycznych, czyli w tym przypadku wynagrodzenia kobiet i mężczyzn z różnych powiatów, z kwantylami rozkładu normalnego. Chcemy w ten sposób sprawdzić, czy rozkłady wynagrodzeń są podobne do rozkładu normalnego.

Na obydwu wykresach możemy zauważyć, że dane empiryczne odbiegają od czerwonej linii, zatem rozkłady wynagrodzeń mężczyzn i kobiet znacząco odbiegają od rozkładu normalnego. Obydwa wykresy są także wyraźnie zakrzywione do góry po prawej stronie, co sugeruje, że dane mają długi ogon po prawej stronie, czyli jest znaczna liczba wynagrodzeń wyższych od normy.

4 Podsumowanie

Analiza wskazuje, że mężczyźni nie tylko zarabiają więcej, ale ich wynagrodzenia są również bardziej zróżnicowane. W przypadku kobiet rozkład płac jest bardziej spłaszczony i mniej zróżnicowany. Istnieją także powiaty, w których różnica między wynagrodzeniami płci jest stosunkowo duża. Powyższe rezultaty mogą być wynikiem wielu różnych czynników społecznych, kulturowych, ekonomicznych, a także politycznych, jednakże z danych statystycznych wynika, że ciężko jest podważyć tezę, że kobiety w Polsce zarabiają mniej od mężczyzn.