



POLITECHNIKA RZESZOWSKA

im. Ignacego Łukasiewicza

WYDZIAŁ MATEMATYKI I FIZYKI STOSOWANEJ

Karolina Wolska

Projekt

Występowanie cukrzycy u kobiet

Opiekun pracy:

**Dr Mariusz Startek**

Rzeszów 2024

## SPIS TREŚCI

1. Opis danych .....	3
2. Wyznaczenie podstawowych parametrów .....	4
2.1. Wartość minimalna.....	4
2.2. Wartość maksymalna .....	4
2.3. Rozstęp.....	4
2.4. Średnia .....	5
2.5. Odchylenie standardowe .....	5
2.6. Kwantyle .....	5
2.7. Wariancja.....	6
2.8. Współczynnik zmienności .....	6
2.9. Korelacja.....	6
2.10. Współczynnik asymetrii.....	7
2.11. Wskaźnik asymetrii .....	7
2.12. Podsumowanie.....	8
3. Graficzna prezentacja danych .....	9
3.1. Histogramy .....	9
3.2. Wykres gęstości .....	11
3.3. Wykres dystrybucyjny .....	13
3.4. Wykres pudełkowy .....	15
4. Hipotezy statystyczne .....	17
4.1. Hipoteza 1 .....	17
4.2. Hipoteza 2.....	18
5. Listing komend .....	20
6. Opis użytych funkcji .....	30

## 1. Opis danych

Projekt ma na celu analizę statystyczną danych o występowaniu cukrzycy wśród kobiet. Zestaw danych zawiera informacje o pacjentkach, które spełniają określone kryteria: wszystkie są kobietami w wieku co najmniej 21 lat i są pochodzenia Pima Indian. Dane zawierają informacje takie jak: Liczba ciąż, poziom glukozy, ciśnienie krwi, grubość skóry, poziom insuliny, BMI, wiek i wynik diagnozy (Outcome): 1 – cukrzyca, 0 – brak cukrzycy.

Zestaw danych pochodzi z Narodowego Instytutu Cukrzycy oraz Chorób Układu Pokarmowego i Nerek. Pobrane zostały ze strony: [kaggle.com](https://www.kaggle.com).

Pobrane dane zostały wczytane do RStudio i przekształcono niektóre kolumny, aby zamiast wartości 0 zawierały NA.

```
3 dane <- read.csv("C://Users//Karolina//Desktop//SAD-projekt
//diabetes.csv")
4 dane$Glucose[dane$Glucose == 0] <- NA
5 dane$BloodPressure[dane$BloodPressure == 0] <- NA
6 dane$SkinThickness[dane$SkinThickness == 0] <- NA
7 dane$Insulin[dane$Insulin == 0] <- NA
8 dane$BMI[dane$BMI == 0.0] <- NA
```

Następnie podzielono dane na dwie tabele: W pierwszej kobiety z cukrzycą, a w drugiej kobiety bez cukrzycy.

```
10 osoby_z_cukrzycy <- dane[dane$Outcome == 1, ]
11 osoby_bez_cukrzycy <- dane[dane$Outcome == 0, ]
```

## 2. Wyznaczenie podstawowych parametrów

### 2.1. Wartość minimalna

```
15 #Parametry
16 #min
17 min_gluk_cukrzyk <- min(osoby_z_cukrzyca$Glucose, na.rm = T)
18 min_gluk_niecukrzyk <- min(osoby_bez_cukrzycy$Glucose, na.rm = T)
19 min_cisn_cukrzyk <- min(osoby_z_cukrzyca$BloodPressure, na.rm = T)
20 min_cisn_niecukrzyk <- min(osoby_bez_cukrzycy$BloodPressure, na.rm = T)

> paste("Minimalna glukoza we krwi kobiety z cukrzyca:", min_gluk_cukrzyk)
[1] "Minimalna glukoza we krwi kobiety z cukrzyca: 78"
> paste("Minimalna glukoza we krwi kobiety bez cukrzycy:", min_gluk_niecukrzyk)
[1] "Minimalna glukoza we krwi kobiety bez cukrzycy: 44"
> paste("Minimalne ciśnienie kobiety z cukrzyca:", min_cisn_cukrzyk)
[1] "Minimalne ciśnienie kobiety z cukrzyca: 30"
> paste("Minimalne ciśnienie kobiety bez cukrzycy:", min_cisn_niecukrzyk)
[1] "Minimalne ciśnienie kobiety bez cukrzycy: 24"
```

### 2.2. Wartość maksymalna

```
31 #max
32 max_gluk_cukrzyk <- max(osoby_z_cukrzyca$Glucose, na.rm = T)
33 max_gluk_niecukrzyk <- max(osoby_bez_cukrzycy$Glucose, na.rm = T)
34 max_cisn_cukrzyk <- max(osoby_z_cukrzyca$BloodPressure, na.rm = T)
35 max_cisn_niecukrzyk <- max(osoby_bez_cukrzycy$BloodPressure, na.rm = T)

> paste("Maksymalna glukoza we krwi kobiety z cukrzyca:", max_gluk_cukrzyk)
[1] "Maksymalna glukoza we krwi kobiety z cukrzyca: 199"
> paste("Maksymalna glukoza we krwi kobiety bez cukrzycy:", max_gluk_niecukrzyk)
[1] "Maksymalna glukoza we krwi kobiety bez cukrzycy: 197"
> paste("Maksymalne ciśnienie kobiety z cukrzyca:", max_cisn_cukrzyk)
[1] "Maksymalne ciśnienie kobiety z cukrzyca: 114"
> paste("Maksymalne ciśnienie kobiety bez cukrzycy:", max_cisn_niecukrzyk)
[1] "Maksymalne ciśnienie kobiety bez cukrzycy: 122"
```

### 2.3. Rozstęp

```
42 #rozstep
43 rozstep_gluk_cukrzyk <- max_gluk_cukrzyk - min_gluk_cukrzyk
44 rozstep_gluk_niecukrzyk <- max_gluk_niecukrzyk - min_gluk_niecukrzyk
45 rozstep_cisn_cukrzyk <- max_cisn_cukrzyk - min_cisn_cukrzyk
46 rozstep_cisn_niecukrzyk <- max_cisn_niecukrzyk - min_cisn_niecukrzyk

> paste("Rozstep glukozy u kobiet z cukrzyca:", rozstep_gluk_cukrzyk)
[1] "Rozstep glukozy u kobiet z cukrzyca: 121"
> paste("Rozstep glukozy u kobiet bez cukrzycy:", rozstep_gluk_niecukrzyk)
[1] "Rozstep glukozy u kobiet bez cukrzycy: 153"
> paste("Rozstep ciśnienia u kobiet z cukrzyca:", rozstep_cisn_cukrzyk)
[1] "Rozstep ciśnienia u kobiet z cukrzyca: 84"
> paste("Rozstep ciśnienia u kobiet bez cukrzycy:", rozstep_cisn_niecukrzyk)
[1] "Rozstep ciśnienia u kobiet bez cukrzycy: 98"
```

## 2.4. Średnia

```
51 #średnia
52 sred_gluk_cukrzyk <- mean(osoby_z_cukrzyca$Glucose, na.rm =T)
53 sred_gluk_niecukrzyk <- mean(osoby_bez_cukrzycy$Glucose, na.rm =T)
54 sred_cisn_cukrzyk <- mean(osoby_z_cukrzyca$BloodPressure, na.rm =T)
55 sred_cisn_niecukrzyk <- mean(osoby_bez_cukrzycy$BloodPressure, na.rm =T)
56
> paste("Średnia glukoza u kobiet z cukrzycą:", round(sred_gluk_cukrzyk,2))
[1] "Średnia glukoza u kobiet z cukrzycą: 142.32"
> paste("Średnia glukoza u kobiet bez cukrzycy:", round(sred_gluk_niecukrzyk,2))
[1] "Średnia glukoza u kobiet bez cukrzycy: 110.64"
> paste("Średnie ciśnienie u kobiet z cukrzycą:", round(sred_cisn_cukrzyk,2))
[1] "Średnie ciśnienie u kobiet z cukrzycą: 75.32"
> paste("Średnie ciśnienie u kobiet bez cukrzycy:", round(sred_cisn_niecukrzyk,2))
[1] "Średnie ciśnienie u kobiet bez cukrzycy: 70.88"
```

## 2.5. Odchylenie standardowe

```
94 #odchylenie standardowe
95 os_gluk_cukrzyk <- sd(osoby_z_cukrzyca$Glucose, na.rm=T)
96 os_gluk_niecukrzyk <- sd(osoby_bez_cukrzycy$Glucose, na.rm=T)
97 os_cisn_cukrzyk <- sd(osoby_z_cukrzyca$BloodPressure, na.rm=T)
98 os_cisn_niecukrzyk <- sd(osoby_bez_cukrzycy$BloodPressure, na.rm=T)
> paste("Odchylenie standardowe glukozy u kobiet z cukrzycą:", round(os_gluk_cukrzyk,2))
[1] "Odchylenie standardowe glukozy u kobiet z cukrzycą: 29.6"
> paste("Odchylenie standardowe glukozy u kobiet bez cukrzycy:", round(os_gluk_niecukrzyk,2))
[1] "Odchylenie standardowe glukozy u kobiet bez cukrzycy: 24.78"
> paste("Odchylenie standardowe ciśnienia u kobiet z cukrzycą:", round(os_cisn_cukrzyk,2))
[1] "Odchylenie standardowe ciśnienia u kobiet z cukrzycą: 12.3"
> paste("Odchylenie standardowe ciśnienia kobiet bez cukrzycy:", round(os_cisn_niecukrzyk,2))
[1] "Odchylenie standardowe ciśnienia kobiet bez cukrzycy: 12.16"
```

## 2.6. Kwantyle

```
76 #kwantyle
77 quantile(osoby_z_cukrzyca$Glucose, c(0.25, 0.5, 0.75), na.rm =T)
78 quantile(osoby_bez_cukrzycy$Glucose, c(0.25, 0.5, 0.75), na.rm =T)
79 quantile(osoby_z_cukrzyca$BloodPressure, c(0.25, 0.5, 0.75), na.rm =T)
80 quantile(osoby_bez_cukrzycy$BloodPressure, c(0.25, 0.5, 0.75), na.rm =T)
> quantile(osoby_z_cukrzyca$Glucose, c(0.25, 0.5, 0.75), na.rm =T)
25% 50% 75%
119 140 167
> quantile(osoby_bez_cukrzycy$Glucose, c(0.25, 0.5, 0.75), na.rm =T)
25% 50% 75%
93 107 125
> quantile(osoby_z_cukrzyca$BloodPressure, c(0.25, 0.5, 0.75), na.rm =T)
25% 50% 75%
68.0 74.5 84.0
> quantile(osoby_bez_cukrzycy$BloodPressure, c(0.25, 0.5, 0.75), na.rm =T)
25% 50% 75%
62 70 78
```

## 2.7. Wariancja

```
82 #wariancja
83 wariancja_glukozy_cukrzyk <- var(osoby_z_cukrzyca$Glucose, na.rm = T)
84 wariancja_glukozy_niecukrzyk <- var(osoby_bez_cukrzycy$Glucose, na.rm = T)
85 wariancja_cisnienia_cukrzyk <- var(osoby_z_cukrzyca$BloodPressure, na.rm = T)
86 wariancja_cisnienia_niecukrzyk <- var(osoby_bez_cukrzycy$BloodPressure, na.rm = T)

> paste("Wariancja glukozy u kobiet z cukrzycą:", round(wariancja_glukozy_cukrzyk,2))
[1] "Wariancja glukozy u kobiet z cukrzycą: 876.11"
> paste("Wariancja glukozy u kobiet bez cukrzycy:", round(wariancja_glukozy_niecukrzyk,2))
[1] "Wariancja glukozy u kobiet bez cukrzycy: 613.9"
> paste("Wariancja ciśnienia u kobiet z cukrzycą:", round(wariancja_cisnienia_cukrzyk,2))
[1] "Wariancja ciśnienia u kobiet z cukrzycą: 151.29"
> paste("Wariancja ciśnienia kobiet bez cukrzycy:", round(wariancja_cisnienia_niecukrzyk,2))
[1] "Wariancja ciśnienia kobiet bez cukrzycy: 147.9"
```

## 2.8. Współczynnik zmienności

```
105 #współczynnik zmienności
106 wz_gluk_cukrzyk <- os_gluk_cukrzyk/sred_gluk_cukrzyk*100
107 wz_gluk_niecukrzyk <- os_gluk_niecukrzyk/sred_gluk_niecukrzyk*100
108 wz_cisn_cukrzyk <- os_cisn_cukrzyk/sred_cisn_cukrzyk*100
109 wz_cisn_niecukrzyk <- os_cisn_niecukrzyk/sred_cisn_niecukrzyk*100

> paste("Współczynnik zmienności glukozy u kobiet z cukrzycą:", round(wz_gluk_cukrzyk,2), "%")
[1] "Współczynnik zmienności glukozy u kobiet z cukrzycą: 20.8 %"
> paste("Współczynnik zmienności glukozy u kobiet bez cukrzycy:", round(wz_gluk_niecukrzyk,2), "%")
[1] "Współczynnik zmienności glukozy u kobiet bez cukrzycy: 22.39 %"
> paste("Współczynnik zmienności ciśnienia u kobiet z cukrzycą:", round(wz_cisn_cukrzyk,2), "%")
[1] "Współczynnik zmienności ciśnienia u kobiet z cukrzycą: 16.33 %"
> paste("Współczynnik zmienności ciśnienia kobiet bez cukrzycy:", round(wz_cisn_niecukrzyk,2), "%")
[1] "Współczynnik zmienności ciśnienia kobiet bez cukrzycy: 17.16 %"
```

## 2.9. Korelacja

```
116 #Korelacja
117 cor(osoby_z_cukrzyca$Glucose, osoby_z_cukrzyca$BloodPressure, use = "complete.obs")
118 cor(osoby_z_cukrzyca$Glucose, osoby_z_cukrzyca$Pregnancies, use = "complete.obs")
119 cor(dane$Glucose, dane$Outcome, use = "complete.obs")

> cor(osoby_z_cukrzyca$Glucose, osoby_z_cukrzyca$BloodPressure, use = "complete.obs")
[1] 0.1013192
> cor(osoby_z_cukrzyca$Glucose, osoby_z_cukrzyca$Pregnancies, use = "complete.obs")
[1] -0.05300957
> cor(dane$Glucose, dane$Outcome, use = "complete.obs")
[1] 0.4946503
```

### Interpretacja:

- Korelacja 0.101 pomiędzy poziomem glukozy, a ciśnieniem krwi wskazuje na słabą pozytywną korelację. Oznacza to, że wyższy poziom glukozy jest nieznacznie związany z wyższym ciśnieniem krwi.
- Korelacja -0.053 pomiędzy poziomem glukozy, a liczbą ciąż wskazuje na bardzo słabą negatywną korelację. Oznacza to, że liczba ciąż nie ma istotnego wpływu na poziom glukozy we krwi.
- Korelacja 0.495 pomiędzy poziomem glukozy, a wystąpieniem cukrzycy wskazuje na umiarkowanie pozytywną korelację. Oznacza to, że wyższy poziom glukozy jest związany z wyższym prawdopodobieństwem wystąpienia cukrzycy.

## 2.10. Współczynnik asymetrii

```
123 #wsp asymetrii
124 wspas_gluk_cukrzyk <- skewness(osoby_z_cukrzyca$Glucose, na.rm=T)
125 wspas_gluk_niecukrzyk <- skewness(osoby_bez_cukrzycy$Glucose, na.rm=T)

> paste("Współczynnik asymetrii glukozy u kobiet z cukrzycą:", round(wspas_gluk_cukrzyk, 3) )
[1] "Współczynnik asymetrii glukozy u kobiet z cukrzycą: 0.09"
> paste("Współczynnik asymetrii glukozy u kobiet bez cukrzycy:", round(wspas_gluk_niecukrzyk, 3))
[1] "Współczynnik asymetrii glukozy u kobiet bez cukrzycy: 0.653"
```

Interpretacja:

- Współczynnik asymetrii glukozy u kobiet z cukrzycą jest bliski zera, co sugeruje, że rozkład poziomów glukozy w tej grupie jest niemal symetryczny
- Współczynnik asymetrii glukozy u kobiet bez cukrzycy ma dodatnią wartość, co sugeruje, że rozkład glukozy w tej grupie jest asymetryczny prawostronny

## 2.11. Wskaźnik asymetrii

```
130 #wsk asymetrii
131 wskas_gluk_cukrzyk <- kurtosis(osoby_z_cukrzyca$Glucose, na.rm=T)
132 wskaz_gluk_niecukrzyk <- kurtosis(osoby_bez_cukrzycy$Glucose, na.rm=T)

> paste("Wskaźnik asymetrii glukozy u kobiet z cukrzycą:", round(wskas_gluk_cukrzyk, 3))
[1] "Wskaźnik asymetrii glukozy u kobiet z cukrzycą: 2.037"
> paste("Wskaźnik asymetrii glukozy u kobiet bez cukrzycy:", round(wskaz_gluk_niecukrzyk, 3))
[1] "Wskaźnik asymetrii glukozy u kobiet bez cukrzycy: 3.643"
```

Interpretacja:

- Wskaźnik asymetrii glukozy u kobiet z cukrzycą jest mniejszy od 3, co wskazuje, że rozkład poziomów glukozy w tej grupie jest bardziej płaski niż rozkład normalny (mniejsza ilość wartości ekstremalnych)
- Wskaźnik asymetrii glukozy u kobiet bez cukrzycy jest większy niż 3, co sugeruje, że rozkład poziomów glukozy w tej grupie jest bardziej szpiczasty niż rozkład normalny (większa liczba wartości ekstremalnych)



## 2.12. Podsumowanie

```
137 #podsumowanie
138 summary(osoby_z_cukrzyca$Glucose)
139 summary(osoby_bez_cukrzycy$Glucose)
140 summary(osoby_z_cukrzyca$BloodPressure)
141 summary(osoby_bez_cukrzycy$BloodPressure)

> summary(osoby_z_cukrzyca$Glucose)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
  78.0   119.0   140.0   142.3   167.0   199.0     2

> summary(osoby_bez_cukrzycy$Glucose)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
  44.0   93.0   107.0   110.6   125.0   197.0     3

> summary(osoby_z_cukrzyca$BloodPressure)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
 30.00   68.00   74.50   75.32   84.00   114.00    16

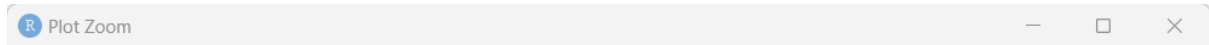
> summary(osoby_bez_cukrzycy$BloodPressure)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
 24.00   62.00   70.00   70.88   78.00   122.00    19
```



### 3. Graficzna prezentacja danych

#### 3.1. Histogramy

```
153 hist(osoby_z_cukrzycy$Glucose,  
154       main= "Liczba kobiet z cukrzycą, a poziom glukozy", xlab="poziom glukozy", ylab="liczba kobiet",col = "blue")  
155  
156 hist(osoby_bez_cukrzycy$Glucose,  
157       main= "Liczba kobiet bez cukrzycy, a poziom glukozy", xlab="poziom glukozy", ylab="liczba kobiet", col = "red")
```



Interpretacja:

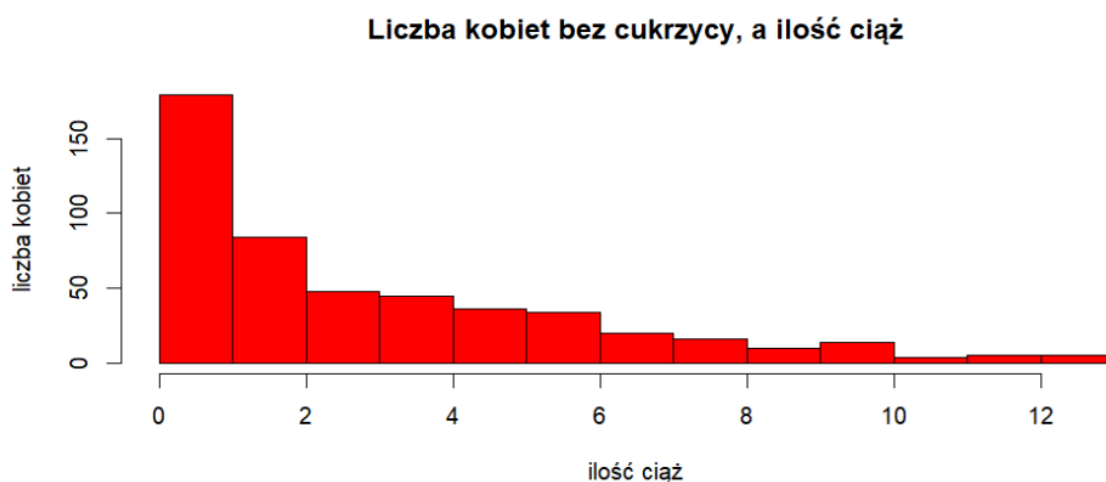
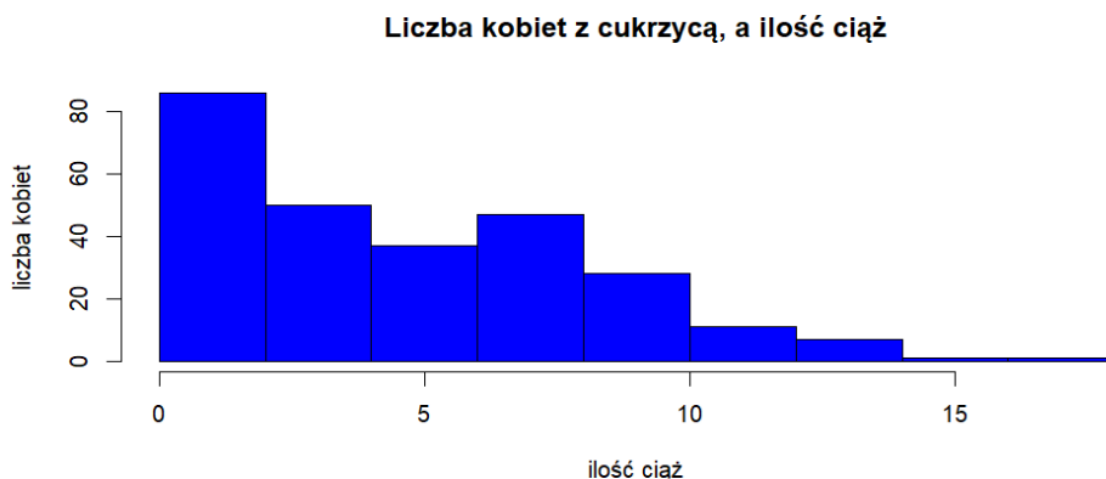
- Najwięcej kobiet z cukrzycą miało poziom glukozy w przedziale około 120-140. Poziomy glukozy poniżej 100 są mniej powszechne. Ogólnie rozkład jest dość symetryczny, z wyraźnym środkiem w okolicach 120-140
- Najwięcej kobiet bez cukrzycy miało poziom glukozy w przedziale około 100-120. Poziomy glukozy poniżej 80 oraz powyżej 150 są rzadsze. Rozkład poziomu glukozy jest bardziej skupiony w niższych wartościach w porównaniu do kobiet z cukrzycą.

```

147 hist(osoby_z_cukrzyca$Pregnancies,
148       main= "Liczba kobiet z cukrzycą, a ilość ciąż", xlab="ilość ciąż", ylab="liczba kobiet", col = "blue")
149
150 hist(osoby_bez_cukrzycy$Pregnancies,
151       main= "Liczba kobiet bez cukrzycy, a ilość ciąż", xlab="ilość ciąż", ylab="liczba kobiet", col = "red")

```

Plot Zoom

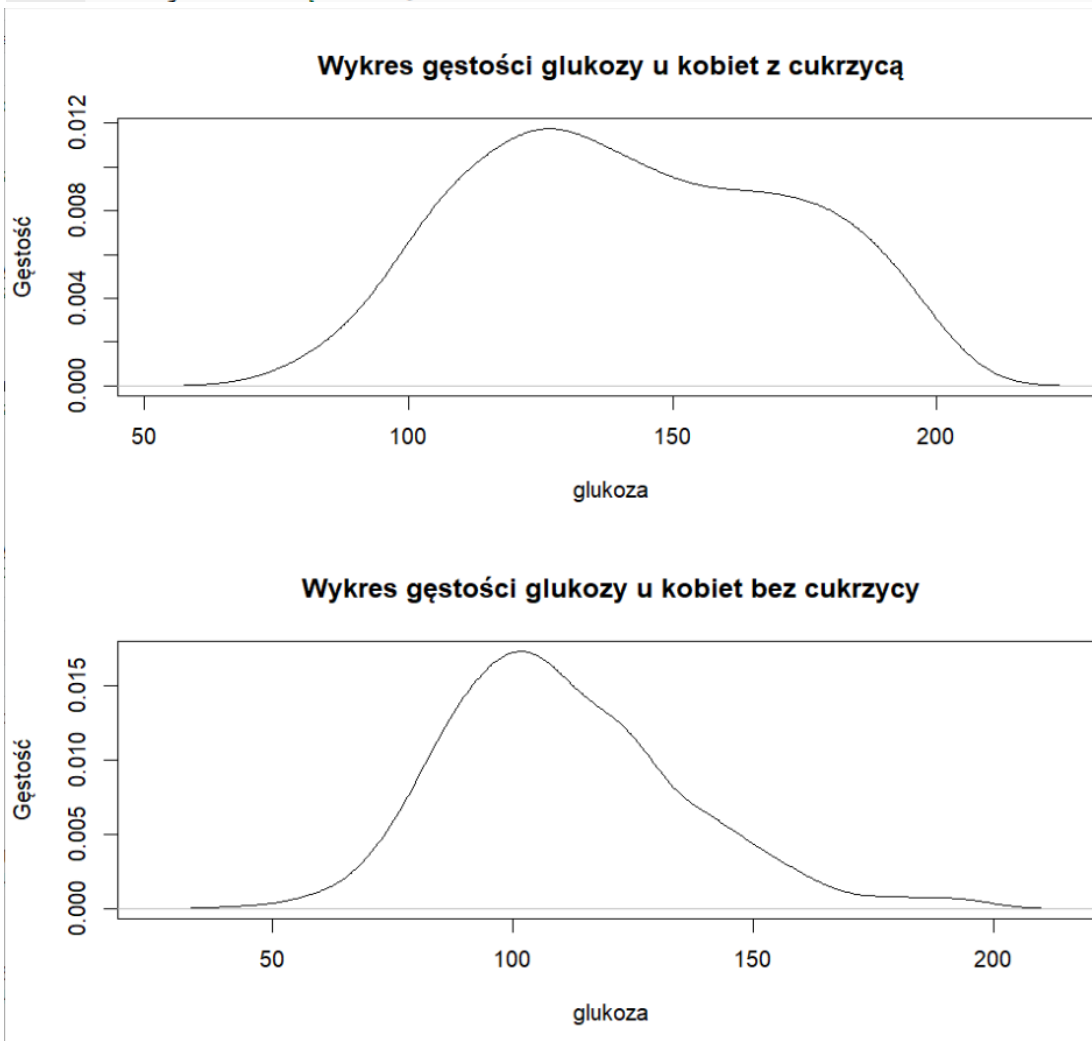


Interpretacja:

- Najwięcej kobiet z cukrzycą miało jedną lub dwie ciążę, co jest widoczne przez najwyższe słupki na wykresie. Liczba kobiet z cukrzycą maleje wraz z liczbą ciąż, jednak widać pewne odchylenia, np. nieco więcej kobiet miało około 5-6 ciąż.
- Najwięcej kobiet bez cukrzycy miało jedną ciążę. Z każdym kolejnym wzrostem liczby ciąż liczba kobiet maleje, przy czym najbardziej zauważalny jest spadek po pierwszej ciąży.

### 3.2. Wykres gęstości

```
160 #glukoza
161 plot(density(na.omit(osoby_z_cukrzycy$Glucose)),
162      main = "Wykres gęstości glukozy u kobiet z cukrzycą",
163      xlab = "glukoza",
164      ylab = "Gęstość")
165
166 plot(density(na.omit(osoby_bez_cukrzycy$Glucose)),
167      main = "Wykres gęstości glukozy u kobiet bez cukrzycy",
168      xlab = "glukoza",
169      ylab = "Gęstość")
```



Interpretacja:

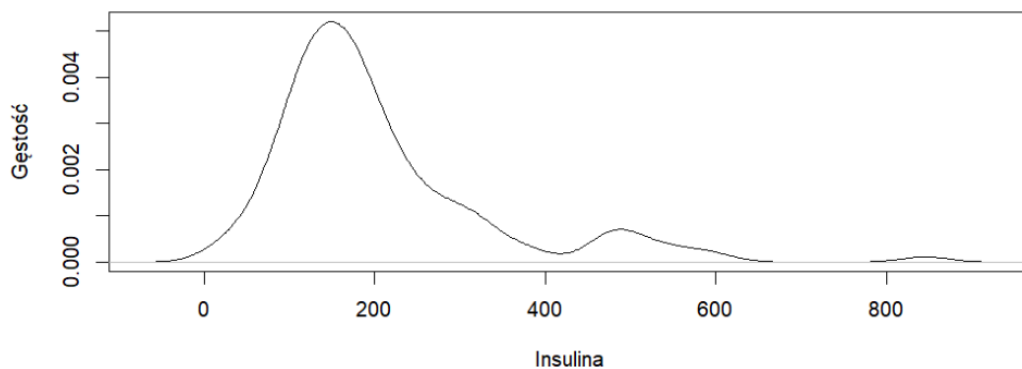
- Wykres gęstości pokazuje, że rozkład poziomów glukozy u kobiet z cukrzycą jest szeroko rozłożony, z głównym szczytem 125. Rozkład jest dość symetryczny, co jest zgodne z niskim współczynnikiem asymetrii (0.09).
- Rozkład poziomów glukozy u kobiet bez cukrzycy ma szczyt około 100. Rozkład jest bardziej skośny w prawo, co jest zgodne z wyższym współczynnikiem asymetrii

```

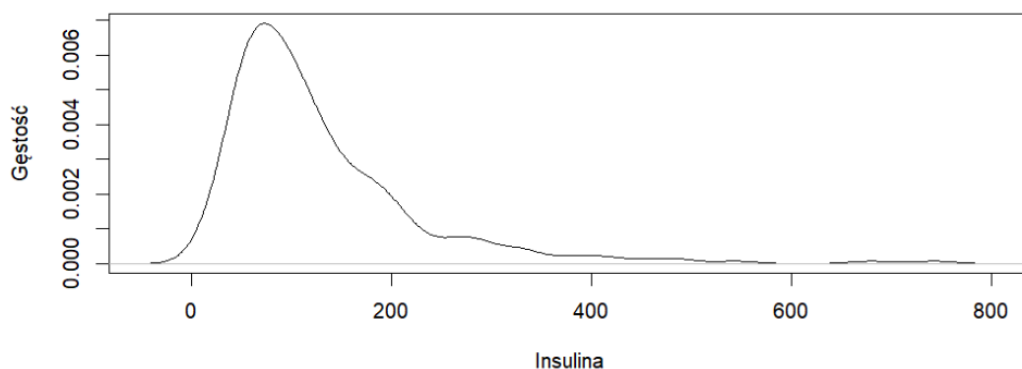
172 #Insulina
173 #gestosc
174 plot(density(na.omit(osoby_z_cukrzycy$Insulin)),
175       main = "Wykres gęstości insuliny u kobiet z cukrzycą",
176       xlab = "Insulina",
177       ylab = "Gęstość")
178
179 plot(density(na.omit(osoby_bez_cukrzycy$Insulin)),
180       main = "Wykres gęstości glukozy u kobiet bez cukrzycy",
181       xlab = "Insulina",
182       ylab = "Gęstość")
183

```

**Wykres gęstości insuliny u kobiet z cukrzycą**



**Wykres gęstości insuliny u kobiet bez cukrzycy**

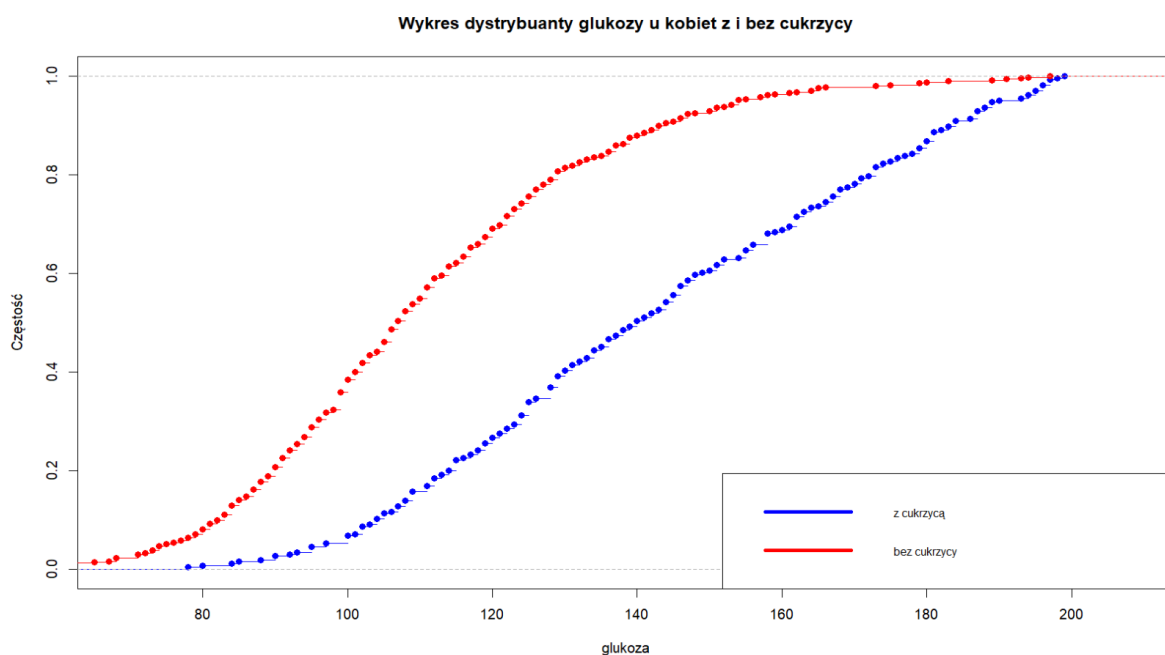


#### Interpretacja:

- Wykres pokazuje, że większość kobiet z cukrzycą ma poziom insuliny w przedziale około 100-200, co wskazuje na jeden wyraźny szczyt w tym zakresie. Poza głównym szczytem, wykres pokazuje kilka mniejszych szczytów w wyższych przedziałach insuliny, co sugeruje pewne odchylenia lub różnorodność w poziomach insuliny w tej grupie.
- Najwięcej kobiet bez cukrzycy ma poziom insuliny w przedziale 50-150, co jest widoczne przez wyraźny szczyt w tym zakresie. Poziom insuliny u kobiet bez cukrzycy jest bardziej skoncentrowany w niższym przedziale w porównaniu do kobiet z cukrzycą. Gęstość maleje dość szybko po osiągnięciu szczytu, co sugeruje, że wyższe poziomy insuliny są rzadsze w tej grupie.

### 3.3. Wykres dystrybuanty

```
184 #dystrybuanta
185 #glukoza
186 plot(ecdf(na.omit(osoby_z_cukrzyca$Glucose)),
187       main = "Wykres dystrybuanty glukozy u kobiet z i bez cukrzycy",|
188       xlab = "glukoza",
189       ylab = "Częstość", col = "blue")
190 plot(ecdf(na.omit(osoby_bez_cukrzycy$Glucose)), add=T, col="red")
191 # Dodanie legendy
192 legend("bottomright",
193       legend = c("z cukrzyca", "bez cukrzycy"),
194       col = c("blue", "red"),
195       lwd = 4,
196       cex = 0.8)
```



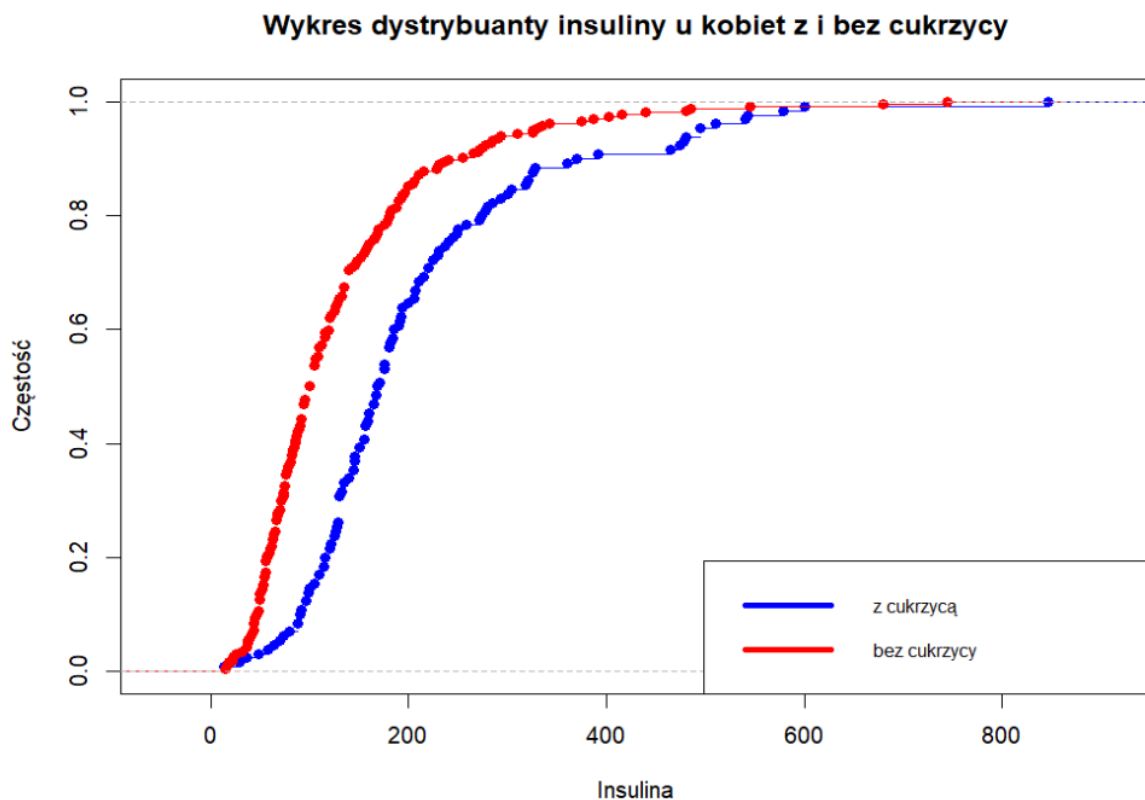
Interpretacja:

- Kobiety z cukrzyca mają wyższe poziomy glukozy w porównaniu do kobiet bez cukrzycy, co jest widoczne poprzez położenie krzywych.
- Położenie na osi X, gdzie dystrybuanta osiąga 0.5, jest wyższe dla niebieskiej linii niż dla czerwonej. Oznacza to, że mediana poziomu glukozy jest wyższa dla kobiet z cukrzyca.

```

202 #insulina
203 plot(ecdf(na.omit(osoby_z_cukrzyca$Insulin)),
204       main = "Wykres dystrybuanty insuliny u kobiet z i bez cukrzycy",
205       xlab = "Insulina",
206       ylab = "Częstość", col = "blue")
207 plot(ecdf(na.omit(osoby_bez_cukrzycy$Insulin)), add=T, col="red")
208 # Dodanie legendy
209 legend("bottomright",
210       legend = c("z cukrzycą", "bez cukrzycy"),
211       col = c("blue", "red"),
212       lwd = 4,
213       cex = 0.8)

```

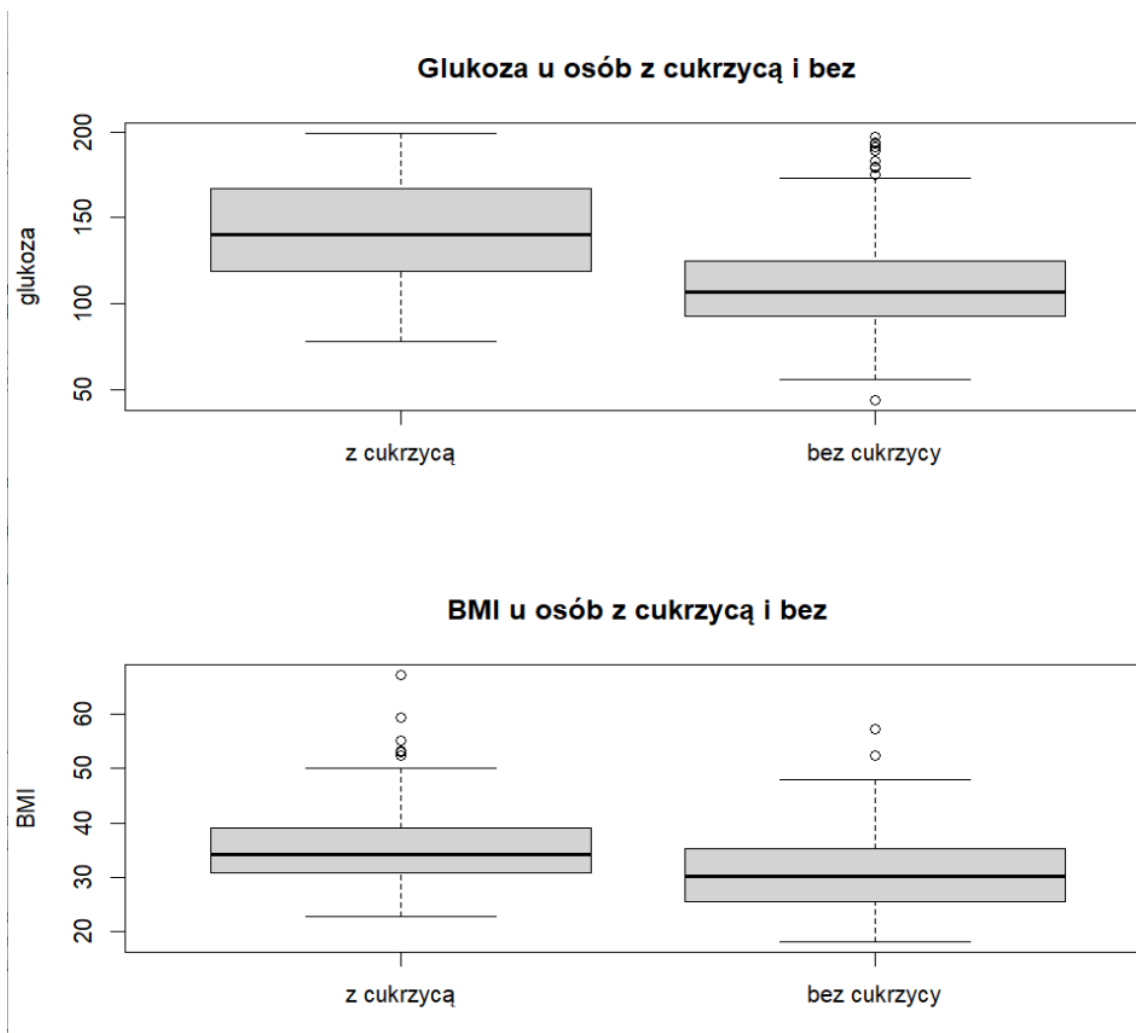


Interpretacja:

- Kobiety z cukrzycą mają wyższe stężenia insuliny we krwi niż kobiety bez cukrzycy. Na wykresie niebieska linia jest po prawej stronie czerwonej linii, co wskazuje, że dla każdego poziomu dystrybuanty, wartości insuliny są wyższe dla kobiet z cukrzycą.
- Położenie na osi X, gdzie dystrybuanta osiąga 0.5, jest wyższe dla niebieskiej linii niż dla czerwonej. Oznacza to, że mediana poziomu insuliny jest wyższa dla kobiet z cukrzycą.

### 3.4. Wykres pudełkowy

```
201 #glukoza
202 boxplot(ylab="glukoza", main = "Glukoza u osób z cukrzycą i bez",
203         osoby_z_cukrzycą$Glucose, osoby_bez_cukrzycy$Glucose, names=c("z cukrzycą", "bez cukrzycy"))
204
205 boxplot(ylab="BMI", main = "BMI u osób z cukrzycą i bez",
206         osoby_z_cukrzycą$BMI, osoby_bez_cukrzycy$BMI, names=c("z cukrzycą", "bez cukrzycy"))
```



Interpretacja:

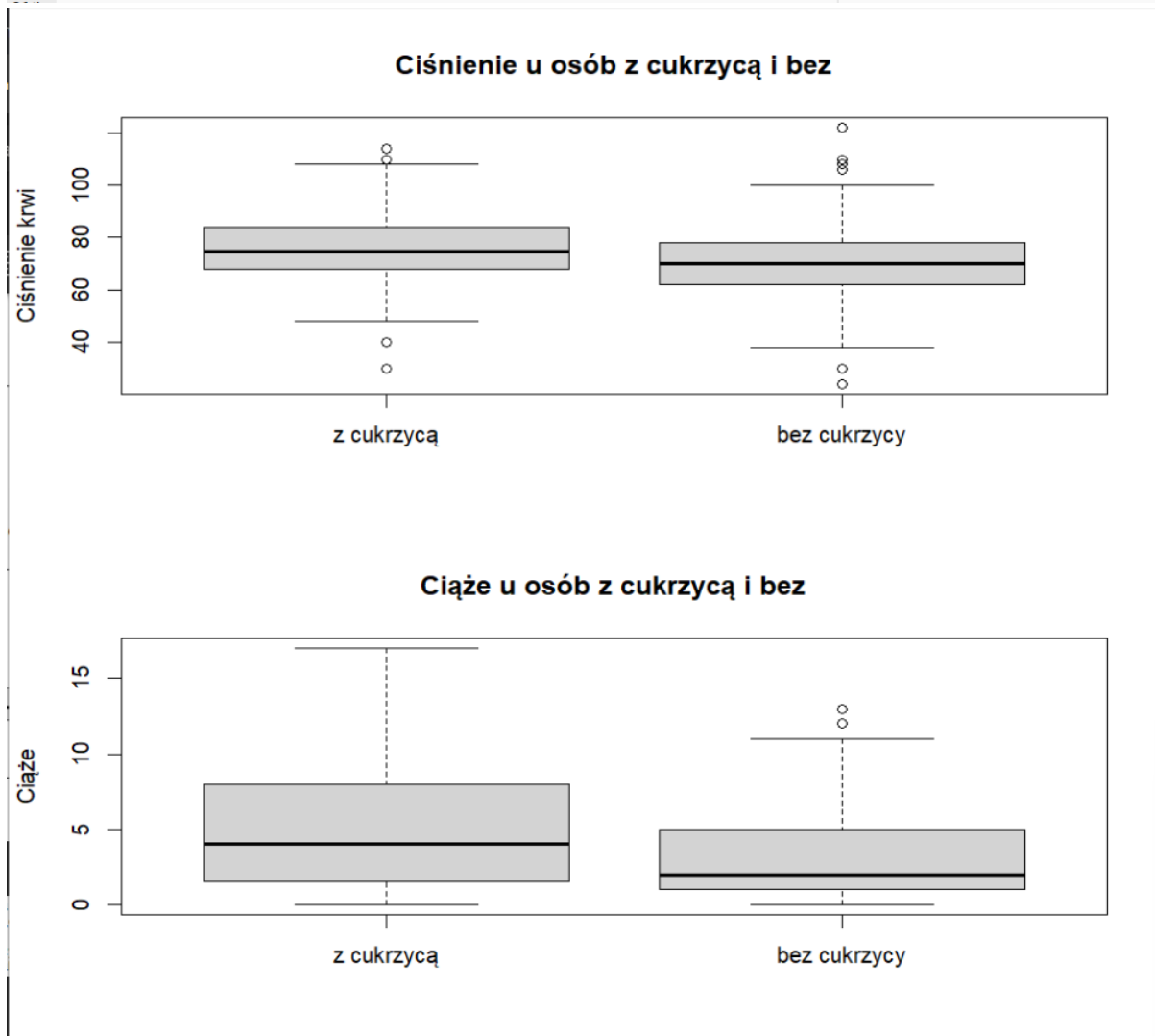
- Mediana poziomu glukozy jest wyższa u kobiet z cukrzycą. Rozkład poziomów glukozy u kobiet z cukrzycą jest bardziej rozłożony. U kobiet bez cukrzycy rozkład jest bardziej skupiony z wartościami głównie poniżej 150.
- Średnie BMI jest nieco wyższe u kobiet z cukrzycą. U kobiet z cukrzycą widocznych jest więcej wartości ekstremalnych w wyższych wartościach BMI.



```

209 boxplot(ylab="Ciśnienie krwi", main = "Ciśnienie u osób z cukrzycą i bez",
210         osoby_z_cukrzycy$BloodPressure, osoby_bez_cukrzycy$BloodPressure, names=c("z cukrzycą", "bez cukrzycy"))
211
212 boxplot(ylab="Cięża", main = "Cięża u osób z cukrzycą i bez",
213         osoby_z_cukrzycy$Pregnancies, osoby_bez_cukrzycy$Pregnancies, names=c("z cukrzycą", "bez cukrzycy"))

```



Interpretacja:

- Mediana ciśnienia krwi jest podobna w obu grupach, delikatnie wyższa w grupie kobiet z cukrzycą. Występuje więcej wartości odstających w grupie bez cukrzycy, co może sugerować większą zmienność ciśnienia krwi w tej grupie.
- Mediana liczby ciąż u kobiet z cukrzycą jest wyższa niż w grupie kobiet bez cukrzycy. Liczba ciąż jest również bardziej rozłożona w grupie z cukrzycą.

## 4. Hipotezy statystyczne

### 4.1. Hipoteza 1

```
248 #HIPOTEZY
249 #1
250 #H0: srednia glukozy u kobiet z cukrzyca jest wieksza niz 120
251 #H1: średnia glukoza <120
252 t.test(osoby_z_cukrzyca$Glucose, mu=120, alternative="less")
253
254 #H0: srednia glukozy u kobiet z cukrzyca jest mniejsza niz 120
255 #H1: średnia glukoza > 120
256 t.test(osoby_z_cukrzyca$Glucose, mu=120, alternative="greater")
257
258 #H0: srednia glukozy u kobiet z cukrzyca jest równa niz 120
259 #H1: średnia glukoza jest różna od 120
260 t.test(osoby_z_cukrzyca$Glucose, mu=120, alternative="two.sided")
261
```

```
> #H0: srednia glukozy u kobiet z cukrzyca jest wieksza niz 120
> #H1: średnia glukoza <120
> t.test(osoby_z_cukrzyca$Glucose, mu=120, alternative="less")
```

#### One Sample t-test

```
data: osoby_z_cukrzyca$Glucose
t = 12.298, df = 265, p-value = 1
alternative hypothesis: true mean is less than 120
95 percent confidence interval:
 -Inf 145.3152
sample estimates:
mean of x
142.3195
```

Nie można odrzucić  $H_0$ , ponieważ p-value wynosi 1. Przedział ufności (95%)  $[-\infty, 145.3152]$  obejmuje wartość 120, co oznacza, że średnia może być większa niż 120.

```
> #H0: srednia glukozy u kobiet z cukrzyca jest mniejsza niz 120
> #H1: średnia glukoza > 120
> t.test(osoby_z_cukrzyca$Glucose, mu=120, alternative="greater")
```

#### One Sample t-test

```
data: osoby_z_cukrzyca$Glucose
t = 12.298, df = 265, p-value < 2.2e-16
alternative hypothesis: true mean is greater than 120
95 percent confidence interval:
 139.3239      Inf
sample estimates:
mean of x
142.3195
```

p-value jest ekstremalnie małe, co pozwala nam odrzucić  $H_0$ . Przedział ufności (95%)  $[139.3239, \infty]$  wskazuje, że dolna granica przedziału jest znacznie powyżej 120, co oznacza, że średnia glukozy jest większa od 120.

```
> #H0: srednia glukozy u kobiet z cukrzyca jest równa niż 120
> #H1: średnia glukoza jest różna od 120
> t.test(osoby_z_cukrzyca$Glucose, mu=120, alternative="two.sided")
```

#### One Sample t-test

```
data: osoby_z_cukrzyca$Glucose
t = 12.298, df = 265, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 120
95 percent confidence interval:
 138.7462 145.8929
sample estimates:
mean of x
 142.3195
```

Wynik testu wskazuje, że średnia glukozy jest statystycznie różna od 120, ponieważ p-value jest ekstremalnie małe. Przedział ufności (95%) [138.7462, 145.8929] nie obejmuje wartości 120, co dodatkowo potwierdza, że średnia jest różna od 120.

## 4.2. Hipoteza 2

```
264 #H0: wariancje obu próbek są równe
265 #H1: wariancje obu próbek nie są równe
266 var.test(osoby_z_cukrzyca$BMI, osoby_bez_cukrzycy$BMI)
267 var.test(osoby_z_cukrzyca$Glucose, osoby_bez_cukrzycy$Glucose)
```

```
> #H0: wariancje obu próbek są równe
> #H1: wariancje obu próbek nie są równe
> var.test(osoby_z_cukrzyca$BMI, osoby_bez_cukrzycy$BMI)
```

#### F test to compare two variances

```
data: osoby_z_cukrzyca$BMI and osoby_bez_cukrzycy$BMI
F = 1.0166, num df = 265, denom df = 490, p-value = 0.8702
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.8254077 1.2605924
sample estimates:
ratio of variances
 1.016605
```

p-value wynosi 0.8702, co jest znacznie większe niż typowy poziom istotności (np. 0.05). Oznacza to, że nie mamy podstaw do odrzucenia hipotezy zerowej. Możemy stwierdzić, że nie ma statystycznie istotnej różnicy między wariancjami BMI w obu grupach (kobiety z cukrzyca i bez cukrzycy). Przedział ufności dla stosunku wariancji [0.8254077, 1.2605924] obejmuje wartość 1, co dodatkowo potwierdza, że wariancje są równe.

```
> var.test(osoby_z_cukrzyca$Glucose, osoby_bez_cukrzycy$Glucose)

      F test to compare two variances

data:  osoby_z_cukrzyca$Glucose and osoby_bez_cukrzycy$Glucose
F = 1.4271, num df = 265, denom df = 496, p-value = 0.0007627
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 1.159308 1.768942
sample estimates:
ratio of variances
 1.427137
```

P-value wynosi 0.0007627, co jest znacznie mniejsze niż typowy poziom istotności (np. 0.05). Oznacza to, że mamy podstawy do odrzucenia hipotezy zerowej. Możemy stwierdzić, że istnieje statystycznie istotna różnica między wariancjami poziomu glukozy w obu grupach. Przedział ufności dla stosunku wariancji [1.159308, 1.768942] nie obejmuje wartości 1, co dodatkowo potwierdza, że wariancje nie są równe.

## 5. Listing komend

```
library(moments) #wsk+wsp asymterii
```

```
dane <- read.csv("C://Users//Karolina//Desktop//SAD-projekt//diabetes.csv")
```

```
dane$Glucose[dane$Glucose == 0] <- NA
```

```
dane$BloodPressure[dane$BloodPressure == 0] <- NA
```

```
dane$SkinThickness[dane$SkinThickness == 0] <- NA
```

```
dane$Insulin[dane$Insulin == 0] <- NA
```

```
dane$BMI[dane$BMI == 0.0] <- NA
```

```
osoby_z_cukrzyca <- dane[dane$Outcome == 1, ]
```

```
osoby_bez_cukrzycy <- dane[dane$Outcome == 0, ]
```

```
#Parametry
```

```
#min
```

```
min_gluk_cukrzyk <- min(osoby_z_cukrzyca$Glucose, na.rm = T)
```

```
min_gluk_niecukrzyk <- min(osoby_bez_cukrzycy$Glucose, na.rm = T)
```

```
min_cisn_cukrzyk <- min(osoby_z_cukrzyca$BloodPressure, na.rm = T)
```

```
min_cisn_niecukrzyk <- min(osoby_bez_cukrzycy$BloodPressure, na.rm = T)
```

```
paste("Minimalna glukoza we krwi kobiety z cukrzycą:", min_gluk_cukrzyk)
```

```
paste("Minimalna glukoza we krwi kobiety bez cukrzycy:", min_gluk_niecukrzyk)
```

```
paste("Minimalne ciśnienie kobiety z cukrzycą:", min_cisn_cukrzyk)
```

```
paste("Minimalne ciśnienie kobiety bez cukrzycy:", min_cisn_niecukrzyk)
```

```
#max
```

```
max_gluk_cukrzyk <- max(osoby_z_cukrzyca$Glucose, na.rm = T)
```

```
max_gluk_niecukrzyk <- max(osoby_bez_cukrzycy$Glucose, na.rm = T)
```

```
max_cisn_cukrzyk <- max(osoby_z_cukrzyca$BloodPressure, na.rm = T)
```

```
max_cisn_niecukrzyk <- max(osoby_bez_cukrzycy$BloodPressure, na.rm = T)
```

```
paste("Maksymalna glukoza we krwi kobiety z cukrzycą:", max_gluk_cukrzyk)
```

```
paste("Maksymalna glukoza we krwi kobiety bez cukrzycy:", max_gluk_niecukrzyk)
```

```
paste("Maksymalne ciśnienie kobiety z cukrzycą:", max_cisn_cukrzyk)
```

```
paste("Maksymalne ciśnienie kobiety bez cukrzycy:", max_cisn_niecukrzyk)
```

```
#rozstep
```

```
rozstep_gluk_cukrzyk <- max_gluk_cukrzyk - min_gluk_cukrzyk
```

```
rozstep_gluk_niecukrzyk <- max_gluk_niecukrzyk - min_gluk_niecukrzyk
```

```
rozstep_cisn_cukrzyk <- max_cisn_cukrzyk - min_cisn_cukrzyk
```

```
rozstep_cisn_niecukrzyk <- max_cisn_niecukrzyk - min_cisn_niecukrzyk
```

```
paste("Rozstep glukozy u kobiet z cukrzycą:", rozstep_gluk_cukrzyk)
```

```
paste("Rozstep glukozy u kobiet bez cukrzycy:", rozstep_gluk_niecukrzyk)
```

```
paste("Rozstep ciśnienia u kobiet z cukrzycą:", rozstep_cisn_cukrzyk)
```

```
paste("Rozstep ciśnienia u kobiet bez cukrzycy:", rozstep_cisn_niecukrzyk)
```

```
#średnia
```

```
sred_gluk_cukrzyk <- mean(osoby_z_cukrzyca$Glucose, na.rm = T)
```

```
sred_gluk_niecukrzyk <- mean(osoby_bez_cukrzycy$Glucose, na.rm = T)
```

```
sred_cisn_cukrzyk <- mean(osoby_z_cukrzyca$BloodPressure, na.rm = T)
```

```
sred_cisn_niecukrzyk <- mean(osoby_bez_cukrzycy$BloodPressure, na.rm = T)
```

```
paste("Średnia glukoza u kobiet z cukrzycą:", round(sred_gluk_cukrzyk,2))
paste("Średnia glukoza u kobiet bez cukrzycy:", round(sred_gluk_niecukrzyk,2))
paste("Średnie ciśnienie u kobiet z cukrzycą:", round(sred_cisn_cukrzyk,2))
paste("Średnie ciśnienie u kobiet bez cukrzycy:", round(sred_cisn_niecukrzyk,2))
```

```
#mediana
```

```
med_gluk_cukrzyk <- median(osoby_z_cukrzyca$Glucose, na.rm = T)
med_gluk_niecukrzyk <- median(osoby_bez_cukrzycy$Glucose, na.rm = T)
med_cisn_cukrzyk <- median(osoby_z_cukrzyca$BloodPressure, na.rm = T)
med_cisn_niecukrzyk <- median(osoby_bez_cukrzycy$BloodPressure, na.rm = T)
```

```
paste("Mediana glukozy u kobiet z cukrzycą:", med_gluk_cukrzyk)
paste("Mediana glukozy u kobiet bez cukrzycy:", med_gluk_niecukrzyk)
paste("Mediana ciśnienia u kobiet z cukrzycą:", med_cisn_cukrzyk)
paste("Mediana ciśnienia kobiet bez cukrzycy:", med_cisn_niecukrzyk)
```

```
#kwantyle
```

```
quantile(osoby_z_cukrzyca$Glucose, c(0.25, 0.5, 0.75), na.rm = T)
quantile(osoby_bez_cukrzycy$Glucose, c(0.25, 0.5, 0.75), na.rm = T)
quantile(osoby_z_cukrzyca$BloodPressure, c(0.25, 0.5, 0.75), na.rm = T)
quantile(osoby_bez_cukrzycy$BloodPressure, c(0.25, 0.5, 0.75), na.rm = T)
```

```
#wariancja
```

```
wariancja_glukozy_cukrzyk <- var(osoby_z_cukrzyca$Glucose, na.rm = T)
wariancja_glukozy_niecukrzyk <- var(osoby_bez_cukrzycy$Glucose, na.rm = T)
wariancja_ciśnienia_cukrzyk <- var(osoby_z_cukrzyca$BloodPressure, na.rm = T)
```



```
wariancja_cisnienia_niecukrzyk <- var(osoby_bez_cukrzycy$BloodPressure, na.rm = T)
```

```
paste("Wariancja glukozy u kobiet z cukrzycą:", round(wariancja_glukozy_cukrzyk,2))
```

```
paste("Wariancja glukozy u kobiet bez cukrzycy:",  
round(wariancja_glukozy_niecukrzyk,2))
```

```
paste("Wariancja ciśnienia u kobiet z cukrzycą:", round(wariancja_cisnienia_cukrzyk,2))
```

```
paste("Wariancja ciśnienia kobiet bez cukrzycy:",  
round(wariancja_cisnienia_niecukrzyk,2))
```

```
#odchylenie standardowe
```

```
os_gluk_cukrzyk <- sd(osoby_z_cukrzyca$Glucose, na.rm=T)
```

```
os_gluk_niecukrzyk <- sd(osoby_bez_cukrzycy$Glucose, na.rm=T)
```

```
os_cisn_cukrzyk <- sd(osoby_z_cukrzyca$BloodPressure, na.rm=T)
```

```
os_cisn_niecukrzyk <- sd(osoby_bez_cukrzycy$BloodPressure, na.rm=T)
```

```
paste("Odchylenie standardowe glukozy u kobiet z cukrzycą:",  
round(os_gluk_cukrzyk,2))
```

```
paste("Odchylenie standardowe glukozy u kobiet bez cukrzycy:",  
round(os_gluk_niecukrzyk,2))
```

```
paste("Odchylenie standardowe ciśnienia u kobiet z cukrzycą:",  
round(os_cisn_cukrzyk,2))
```

```
paste("Odchylenie standardowe ciśnienia kobiet bez cukrzycy:",  
round(os_cisn_niecukrzyk,2))
```

```
#współczynnik zmienności
```

```
wz_gluk_cukrzyk <- os_gluk_cukrzyk/sred_gluk_cukrzyk*100
```

```
wz_gluk_niecukrzyk <- os_gluk_niecukrzyk/sred_gluk_niecukrzyk*100
```

```
wz_cisn_cukrzyk <- os_cisn_cukrzyk/sred_cisn_cukrzyk*100
```

```
wz_cisn_niecukrzyk <- os_cisn_niecukrzyk/sred_cisn_niecukrzyk*100
```

```
paste("Współczynnik zmienności glukozy u kobiet z cukrzycą:",  
round(wz_gluk_cukrzyk,2), "%")
```

```
paste("Współczynnik zmienności glukozy u kobiet bez cukrzycy:",  
round(wz_gluk_niecukrzyk,2), "%")
```

```
paste("Współczynnik zmienności ciśnienia u kobiet z cukrzycą:",  
round(wz_cisn_cukrzyk,2), "%")
```

```
paste("Współczynnik zmienności ciśnienia kobiet bez cukrzycy:",  
round(wz_cisn_niecukrzyk,2), "%")
```

```
#Korelacja
```

```
cor(osoby_z_cukrzyca$Glucose,osoby_z_cukrzyca$BloodPressure, use =  
"complete.obs")
```

```
cor(osoby_z_cukrzyca$Glucose,osoby_z_cukrzyca$Pregnancies, use =  
"complete.obs")
```

```
cor(dane$Glucose,dane$Outcome, use = "complete.obs")
```

```
#wsp asymetrii
```

```
wspas_gluk_cukrzyk <- skewness(osoby_z_cukrzyca$Glucose, na.rm=T)
```

```
wspas_gluk_niecukrzyk <- skewness(osoby_bez_cukrzycy$Glucose, na.rm=T)
```

```
paste("Współczynnik asymterii glukozy u kobiet z  
cukrzycą:",round(wspas_gluk_cukrzyk,3) )
```

```
paste("Współczynnik asymetrii glukozy u kobiet bez cukrzycy:",  
round(wspas_gluk_niecukrzyk,3))
```

```
#wsk asymterii
```

```
wskas_gluk_cukrzyk <- kurtosis(osoby_z_cukrzyca$Glucose, na.rm=T)
```

```
wskaz_gluk_niecukrzyk <- kurtosis(osoby_bez_cukrzycy$Glucose, na.rm=T)
```

```
paste("Wskaźnik asymterii glukozy u kobiet z cukrzycą:",round(wskas_gluk_cukrzyk,3))
```

```
paste("Wskaźnik asymetrii glukozy u kobiet bez cukrzycy:",  
round(wskaz_gluk_niecukrzyk,3))
```

```
#podsumowanie
```

```
summary(osoby_z_cukrzyca$Glucose)
```

```
summary(osoby_bez_cukrzycy$Glucose)
```

```
summary(osoby_z_cukrzyca$BloodPressure)
```

```
summary(osoby_bez_cukrzycy$BloodPressure)
```

```
#GRAFICZNA PREZENTACJA DANYCH
```

```
#Histogramy
```

```
par(mfrow = c(2, 1))
```

```
hist(osoby_z_cukrzyca$Pregnancies,
```

```
      main= "Liczba kobiet z cukrzycą, a ilość ciąż", xlab="ilość ciąż", ylab="liczba  
kobiet",col = "blue")
```

```
hist(osoby_bez_cukrzycy$Pregnancies,
```

```
      main= "Liczba kobiet bez cukrzycy, a ilość ciąż", xlab="ilość ciąż", ylab="liczba  
kobiet", col = "red")
```

```
hist(osoby_z_cukrzyca$Glucose,
```

```
      main= "Liczba kobiet z cukrzycą, a poziom glukozy", xlab="poziom glukozy",  
ylab="liczba kobiet",col = "blue")
```

```
hist(osoby_bez_cukrzycy$Glucose,
```

```
      main= "Liczba kobiet bez cukrzycy, a poziom glukozy", xlab="poziom glukozy",  
ylab="liczba kobiet", col = "red")
```

```
#Gęstość
#glukoza
par(mfrow = c(1, 1))
plot(density(na.omit(osoby_z_cukrzyca$Glucose)),
      main = "Wykres gęstości glukozy u kobiet z cukrzycą",
      xlab = "glukoza",
      ylab = "Gęstość")

plot(density(na.omit(osoby_bez_cukrzycy$Glucose)),
      main = "Wykres gęstości glukozy u kobiet bez cukrzycy",
      xlab = "glukoza",
      ylab = "Gęstość")
```

```
#Gęstość
#Insulina
plot(density(na.omit(osoby_z_cukrzyca$Insulin)),
      main = "Wykres gęstości insuliny u kobiet z cukrzycą",
      xlab = "Insulina",
      ylab = "Gęstość")

plot(density(na.omit(osoby_bez_cukrzycy$Insulin)),
      main = "Wykres gęstości insuliny u kobiet bez cukrzycy",
      xlab = "Insulina",
      ylab = "Gęstość")
```

```
#Dystrubuantą
#glukoza
plot(ecdf(na.omit(osoby_z_cukrzyca$Glucose)),
```

```
main = "Wykres dystrybuanty glukozy u kobiet z i bez cukrzycy",
xlab = "glukoza",
ylab = "Częstość", col = "blue")
plot(ecdf(na.omit(osoby_bez_cukrzycy$Glucose)), add=T, col="red")
# Dodanie legendy
legend("bottomright",
      legend = c("z cukrzycą", "bez cukrzycy"),
      col = c("blue", "red"),
      lwd = 4,
      cex = 0.8)
```

```
#Dystrybuanta
```

```
#insulina
```

```
plot(ecdf(na.omit(osoby_z_cukrzyca$Insulin)),
      main = "Wykres dystrybuanty insuliny u kobiet z i bez cukrzycy",
      xlab = "Insulina",
      ylab = "Częstość", col = "blue")
plot(ecdf(na.omit(osoby_bez_cukrzycy$Insulin)), add=T, col="red")
# Dodanie legendy
legend("bottomright",
      legend = c("z cukrzycą", "bez cukrzycy"),
      col = c("blue", "red"),
      lwd = 4,
      cex = 0.8)
```

```
#pudełkowe
```

```
#glukoza
```

```
par(mfrow = c(2, 1))
```

```
boxplot(ylab="glukoza", main = "Glukoza u osób z cukrzycą i bez",
```

```
      osoby_z_cukrzyca$Glucose, osoby_bez_cukrzycy$Glucose, names=c("z cukrzycą",  
"bez cukrzycy"))
```

```
#BMI
```

```
boxplot(ylab="BMI", main = "BMI u osób z cukrzycą i bez",
```

```
      osoby_z_cukrzyca$BMI, osoby_bez_cukrzycy$BMI, names=c("z cukrzycą", "bez  
cukrzycy"))
```

```
#Ciśnienie krwi
```

```
boxplot(ylab="Ciśnienie krwi", main = "Ciśnienie u osób z cukrzycą i bez",
```

```
      osoby_z_cukrzyca$BloodPressure, osoby_bez_cukrzycy$BloodPressure,  
names=c("z cukrzycą", "bez cukrzycy"))
```

```
#Cięża
```

```
boxplot(ylab="Cięża", main = "Cięża u osób z cukrzycą i bez",
```

```
      osoby_z_cukrzyca$Pregnancies, osoby_bez_cukrzycy$Pregnancies, names=c("z  
cukrzycą", "bez cukrzycy"))
```

```
#HIPOTEZY
```

```
#1
```

```
#H0: srednia glukozy u kobiet z cukrzycą jest większa niż 120
```

```
#H1: średnia glukoza <120
```

```
t.test(osoby_z_cukrzyca$Glucose, mu=120, alternative="less")
```

```
#H0: srednia glukozy u kobiet z cukrzycą jest mniejsza niż 120
```

```
#H1: średnia glukoza > 120
```

```
t.test(osoby_z_cukrzyca$Glucose, mu=120, alternative="greater")
```

#H0: srednia glukozy u kobiet z cukrzycą jest równa niż 120

#H1: średnia glukoza jest różna od 120

```
t.test(osoby_z_cukrzyca$Glucose, mu=120, alternative="two.sided")
```

#2

#H0: wariancje obu próbek są równe

#H1: wariancje obu próbek nie są równe

```
var.test(osoby_z_cukrzyca$BMI, osoby_bez_cukrzycy$BMI)
```

```
var.test(osoby_z_cukrzyca$Glucose, osoby_bez_cukrzycy$Glucose)
```



## 6. Opis użytych funkcji

Biblioteki:

library(moments) - znajdują się tutaj funkcje użyte do obliczenia współczynnika i wskaźnika asymetrii

Funkcje:

- **read.csv()**: Importuje dane z pliku CSV
- **na.rm** - argument używany w różnych funkcjach (np. min(), max(), mean()) do usunięcia wartości NA przed wykonaniem operacji.
- **dane\$Glucose[dane\$Glucose == 0] <- NA** - zamienia wartości 0 na NA w kolumnie "Glucose"
- **dane[dane\$Outcome == 1, ]** - wybiera wiersze z ramki danych, gdzie kolumna "Outcome" ma wartość 1.
- **min()** - wyświetla wartość minimalną.
- **max()** - wyświetla wartość maksymalną.
- **paste()** - łączy łańcuchy znaków.
- **mean()** - wyświetla wartość średnią.
- **median()** - wyświetla wartość mediany.
- **quantile()** - wyświetla wartości kwantyli.
- **var()** - wyświetla wartość wariancji.
- **cor()** - oblicza współczynnik korelacji.
- **sd()** - wyświetla wartość odchylenia standardowego.
- **Skewness()** - wyświetla współczynnik asymetrii
- **Kurtosis()** - wyświetla wskaźnik asymetrii
- **summary()** - podsumowuje dane, wyświetlając podstawowe statystyki opisowe.
- **hist()** - wykres-histogram, ogólna funkcja hist oblicza histogram podanych wartości danych.
- **plot()** - podstawowy wykres punktowy.
- **density()** - oszacowanie gęstości.
- **ecdf()** - dystrybunta, empiryczna funkcja dystrybucji skumulowanej.
- **legend()** - dodaje legendę do wykresów.
- **par()** - ustawia lub odczytuje parametry graficzne.
- **boxplot()** - tworzy wykres pudełkowy z podanych wartości.
- **t.test()** - wykonuje test t na wektorach danych.
- **var.test()** - wykonuje test F w celu porównania wariancji dwóch próbek z normalnych populacji.