

Rozkład normalny

Bartosz Maćkiewicz

Dlaczego rozkład normalny jest ważny?

1. Bardzo wiele zmiennych zależnych, którymi zajmować się będą Państwo ma rozkład normalny lub można to przyjąć na potrzeby obliczeń.
2. Można pokazać, że rozkład średniej z próby w większości przypadków dąży do rozkładu normalnego (statystyczne testowanie hipotez)
3. Bardzo duża część procedur statystycznych zakłada, że populacja (lub błąd pomiaru) ma rozkład normalny.

Standaryzacja zmiennej losowej

Procedura w wyniku której zmienna uzyskuje średnią 0 oraz odchylenie standardowe 1.

$$z = \frac{x - \mu}{\sigma}$$

Wykres kwantylowy (wykres kwantyl-kwantyl, *Q-Q plot*)

W statystyce do sprawdzania, czy dana zmienna ma pewien rozkład (np. rozkład normalny) używa się tzw. wykresów kwantylowych (*Quantile to Quantile (Q-Q) plot*).

Wykres kwantylowy

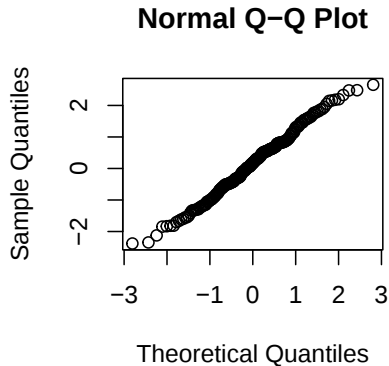
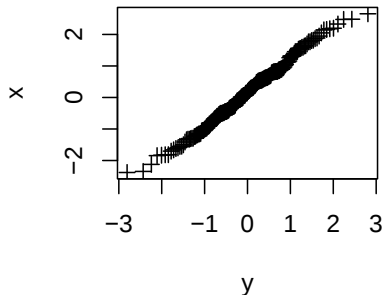
- ▶ Jeśli rozkład jest bardziej skoncentrowany na lewo od normalnego, to lewa część wykresu znajduje się ponad przekątną.
- ▶ Jeśli rozkład jest mniej skoncentrowany na lewo od normalnego, to lewa część wykresu znajduje się poniżej przekątnej.
- ▶ Jeśli rozkład jest bardziej skoncentrowany na prawo od normalnego, to prawa część wykresu znajduje się poniżej przekątnej.
- ▶ Jeśli rozkład jest mniej skoncentrowany na prawo od normalnego, to prawa część wykresu znajduje się ponad przekątną.
- ▶ Jeśli rozkład jest lewostronnie asymetryczny, to krzywa leży ponad przekątną w środku wykresu.
- ▶ Jeśli rozkład jest prawostronnie asymetryczny, to krzywa leży poniżej przekątnej w środku wykresu.
- ▶ Jeśli rozkład jest symetryczny, to krzywa przecina przekątną w środku wykresu.

Wykres kwantylowy - nasza własna implementacja

```
wykres_kwantylowy = function(x){  
  x = sort(x)  
  # Tworzymy kolejne teoretyczne obserwacje (np. dla n = 100:  
  # To będą kolejne wartości naszych percentyli  
  #  $(1 - 1/2)/100 = 0.005$   
  #  $(2 - 1/2)/100 = 0.015$   
  #  $(3 - 1/2)/100 = 0.025$   
  # ...  
  #  $(99 - 1/2)/100 = 0.985$   
  #  $(100 - 1/2)/100 = 0.995$   
  y = ((1:length(x)) - 1/2)/length(x)  
  y = qnorm(y) # tutaj obliczamy ile percentyle wynoszą  
  plot(y,x, pch =3) # i nanosimy to na wykres  
}
```

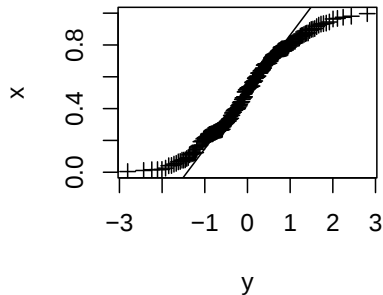
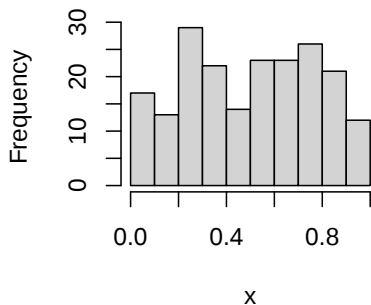
Wykres kwantylowy dla rozkładu normalnego

```
set.seed(12345)
x = rnorm(200)
par(mfrow=c(1,2))
wykres_kwantylowy(x)
qqnorm(x)
```



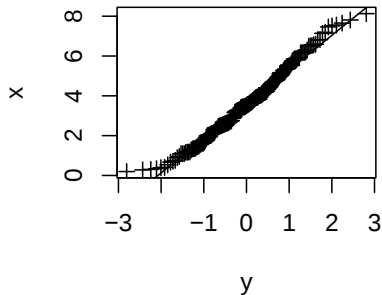
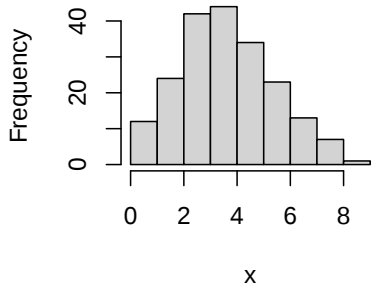
```
x = runif(200)
par(mfrow=c(1,2))
hist(x)
wykres_kwantylowy(x)
qqline(x)
```

Histogram of x



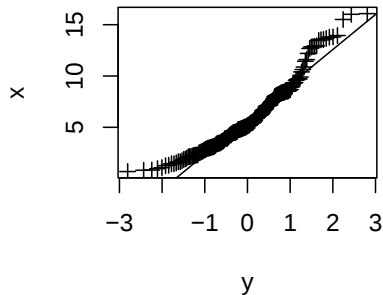
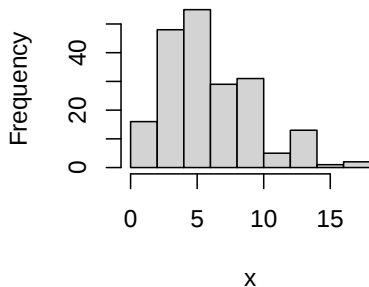
```
x = rpois(200,3) + runif(200)
par(mfrow=c(1,2))
hist(x)
wykres_kwantylowy(x)
qqline(x)
```

Histogram of x



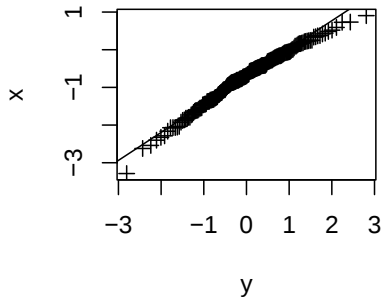
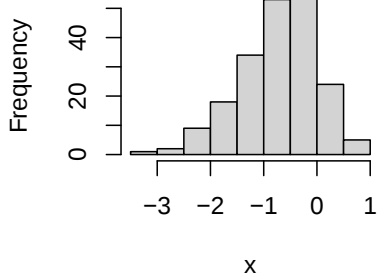
```
x = rgamma(200, 3, 1/2)
par(mfrow=c(1,2))
hist(x)
wykres_kwantylowy(x)
qqline(x)
```

Histogram of x



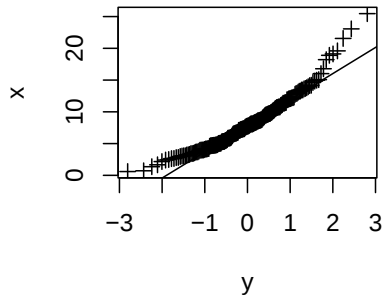
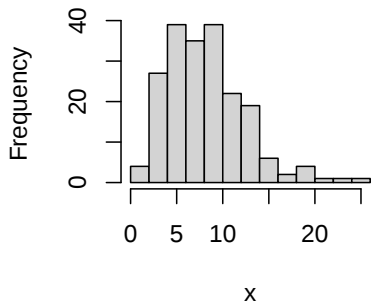
```
x = rsn(200, alpha = -2)
par(mfrow=c(1,2))
hist(x)
wykres_kwantylowy(x)
qqline(x)
```

Histogram of x



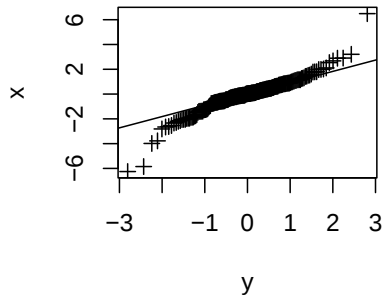
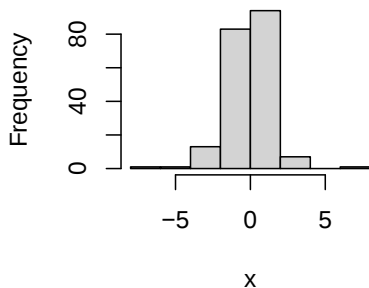
```
x = rchisq(200, 8)
par(mfrow=c(1,2))
hist(x)
wykres_kwantylowy(x)
qqline(x)
```

Histogram of x



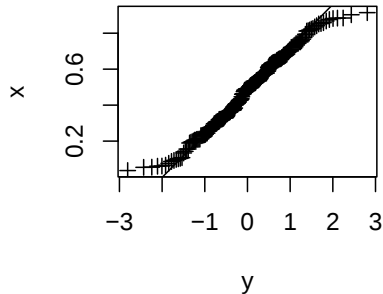
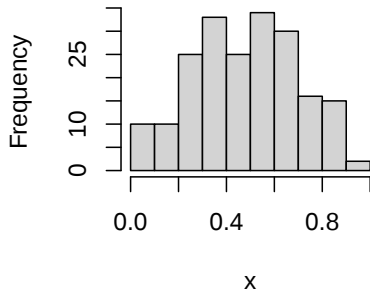
```
x = rt(200, 4)
par(mfrow=c(1,2))
hist(x)
wykres_kwantylowy(x)
qqline(x)
```

Histogram of x



```
x = rbeta(200,2,2)
par(mfrow=c(1,2))
hist(x)
wykres_kwantylowy(x)
qqline(x)
```

Histogram of x



```
x = rcauchy(200); x = x[x>-10 & x<10]
par(mfrow=c(1,2))
hist(x)
wykres_kwantylowy(x)
qqline(x)
```

Histogram of x

