

12 Analiza korelacji

12.1 Przykład

Przykład. Chcemy zbadać, czy istnieje zależność między miesięcznym dochodem rodziny na jedną osobę a miesięczną wartością wydatków na jedną osobę. Dane dotyczące tych dwóch cech dla dziesięciu rodzin podano w poniższej tabeli.

rodzina	1	2	3	4	5	6	7	8	9	10
przychody	210	270	290	310	370	400	450	480	510	520
wydatki	140	190	250	270	290	310	340	360	420	390

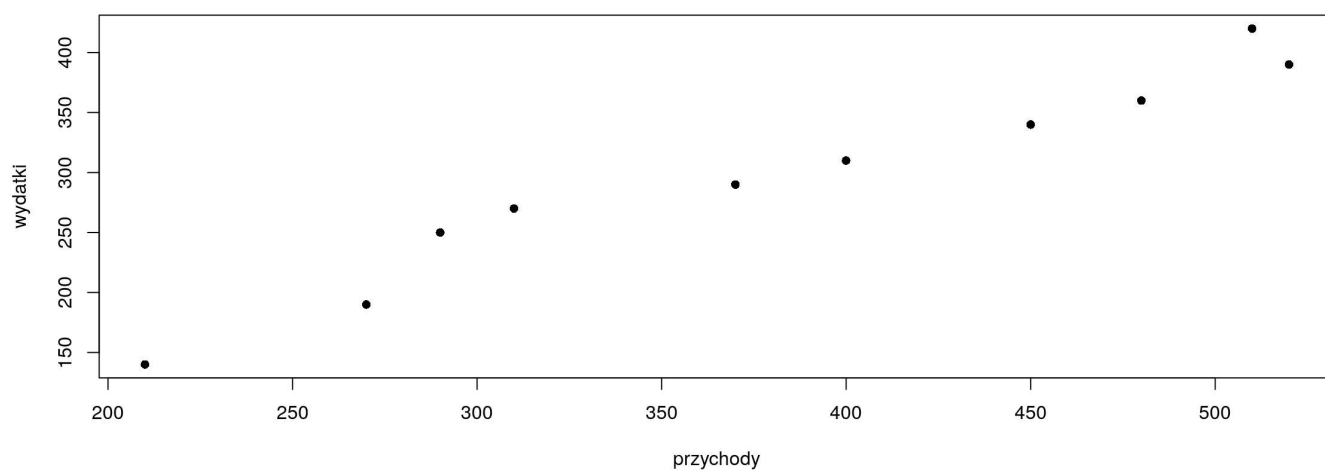
```
przychody <- c(210, 270, 290, 310, 370, 400, 450, 480, 510, 520)
wydatki <- c(140, 190, 250, 270, 290, 310, 340, 360, 420, 390)
data_set <- data.frame(przychody = przychody, wydatki = wydatki)
head(data_set)
```

```
##   przychody wydatki
## 1      210     140
## 2      270     190
## 3      290     250
## 4      310     270
## 5      370     290
## 6      400     310
```

```
# wykres rozrzutu
```

```
plot(data_set$przychody, data_set$wydatki, xlab = "przychody", ylab = "wydatki", pch = 1)
```





```
# założenia
```

```
shapiro.test(data_set$przychody)
```

```
##
```

```
## Shapiro-Wilk normality test
```

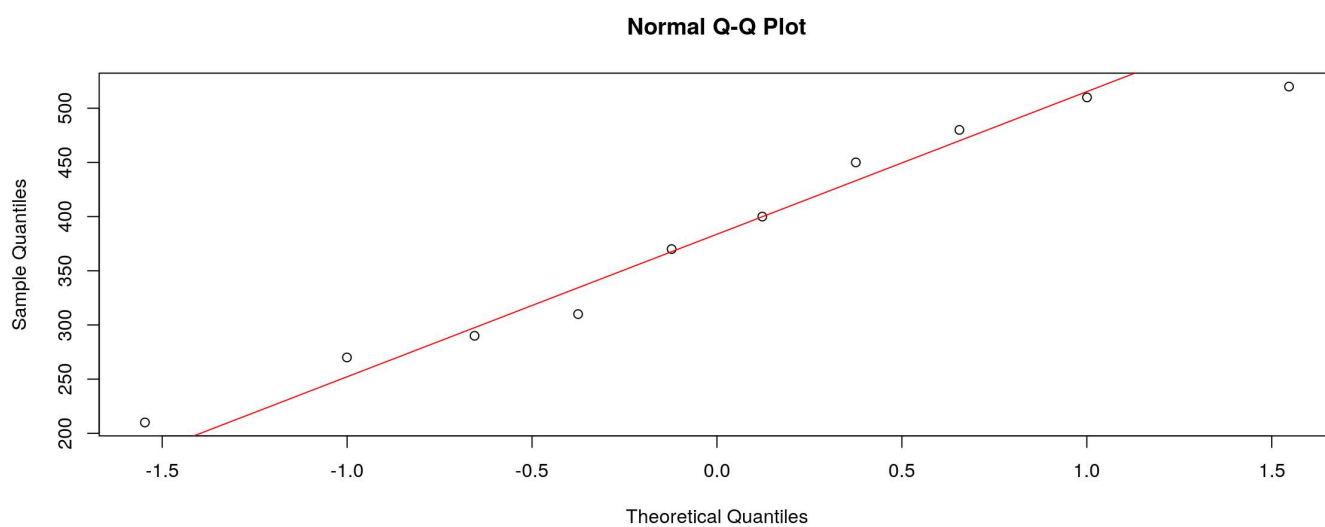
```
##
```

```
## data: data_set$przychody
```

```
## W = 0.94256, p-value = 0.5819
```

```
qqnorm(data_set$przychody)
```

```
qqline(data_set$przychody, col = "red")
```



```
shapiro.test(data_set$wydatki)
```

```
##
```

```
## Shapiro-Wilk normality test
```

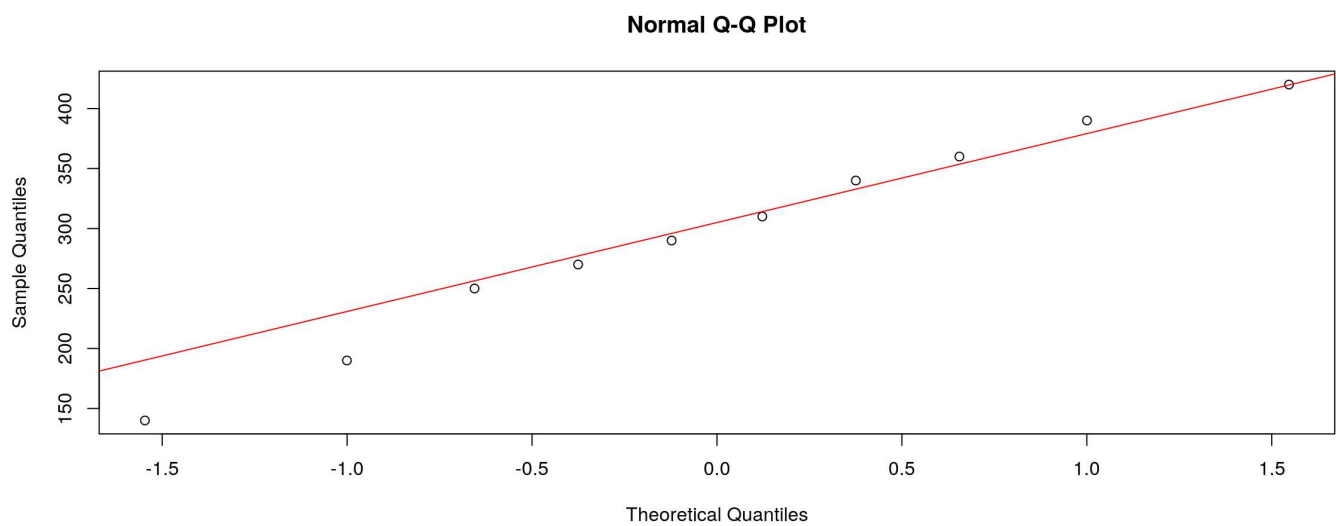
```
##
```

```
## data: data_set$wydatki
```

```
## W = 0.97753, p-value = 0.9506
```

```
qqnorm(data_set$wydatki)
```

```
qqline(data_set$wydatki, col = "red")
```



```
# testy
```

```
cor.test(data_set$przychody, data_set$wydatki, method = "pearson")
```

```
##
## Pearson's product-moment correlation
##
## data: data_set$przychody and data_set$wydatki
## t = 12.399, df = 8, p-value = 1.67e-06
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.8942997 0.9942521
## sample estimates:
## cor
## 0.9749541

cor.test(data_set$przychody, data_set$wydatki, method = "kendall")

##
## Kendall's rank correlation tau
##
## data: data_set$przychody and data_set$wydatki
## T = 44, p-value = 5.511e-06
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
## tau
## 0.9555556

cor.test(data_set$przychody, data_set$wydatki, method = "spearman")
```

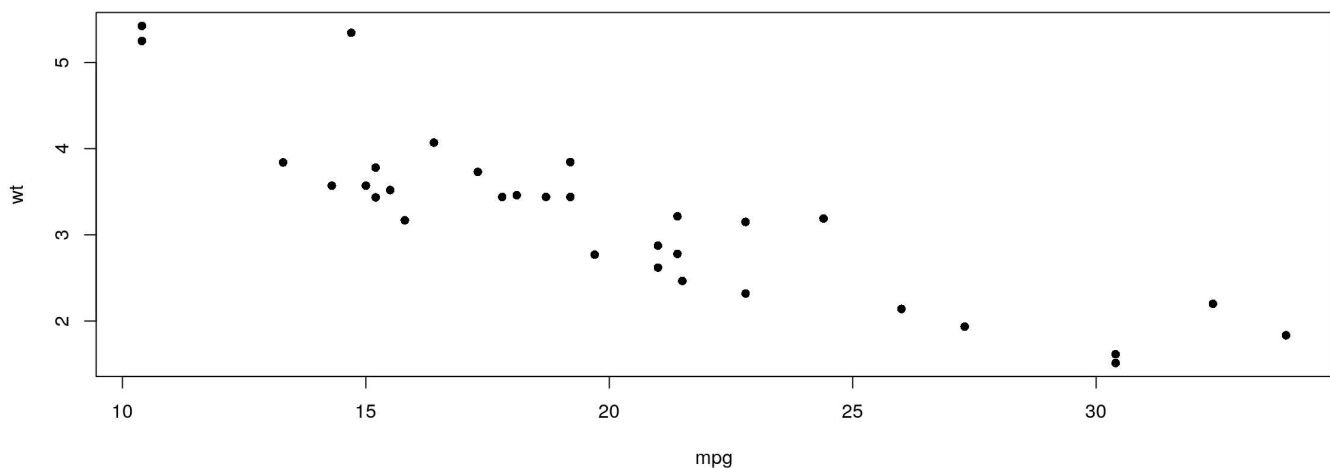
```
##
## Spearman's rank correlation rho
##
## data: data_set$przychody and data_set$wydatki
## S = 2, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.9878788
```

12.2 Zadania

Zadanie 1. Zbiór danych `mtcars` zawiera dane dotyczące pewnych cech samochodów. Interesuje nas zbadanie korelacji między zmiennymi `mpg` i `wt`.

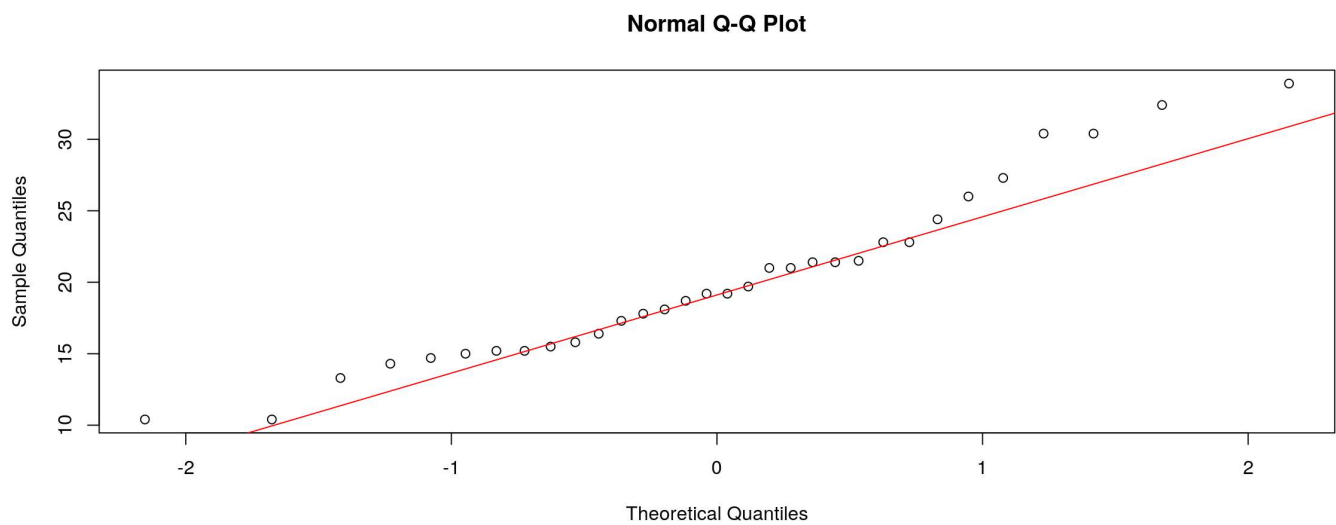
	mpg	cyl	displacement	horsepower	drat	weight	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225	105	2.76	3.460	20.22	1	0	3	1

1. Wykonaj wykres rozrzutu dla badanych cech.

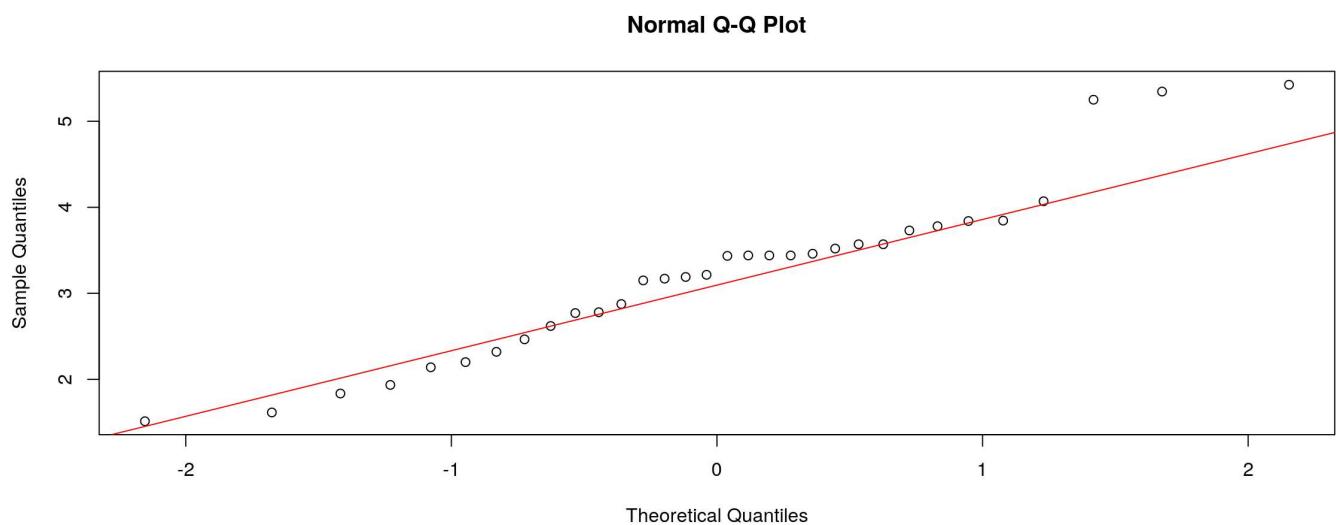


2. Sprawdź założenia testu istotności dla współczynnika korelacji.

```
## [1] 0.1228814
```



```
## [1] 0.09265499
```



3. Wykonaj test istotności dla współczynnika korelacji dla zmiennych `mpg` i `wg`. Oszacuj punktowo i przedziałowo współczynnik korelacji.

```
## [1] 1.293959e-10
```

```
## cor
```

```
## -0.8676594
```

```
## [1] -0.9338264 -0.7440872
## attr("conf.level")
## [1] 0.95
```

4. Wykonaj polecenie punktu 3 korzystając ze współczynników Kendalla i Spearmana.

```
## [1] 6.70577e-09
```

```
##          tau
## -0.7278321
```

```
## [1] 1.487595e-11
```

```
##          rho
## -0.886422
```