

# **Przewidywanie czasu snu na podstawie innych czynników**

Adrianna Rosmus

# Dane

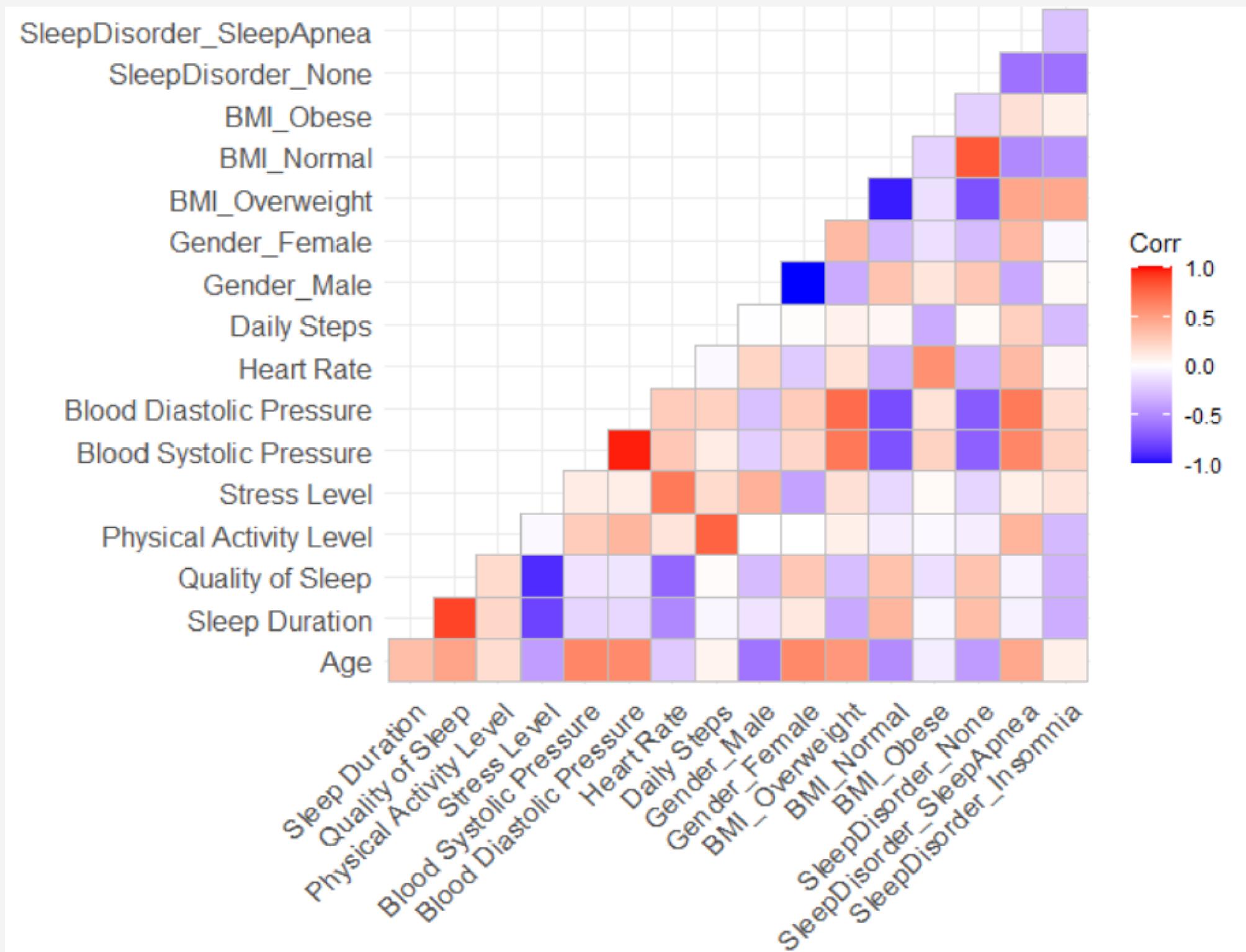
Kaggle

Person ID	Gender	Age	Occupation	Sleep Duration	Quality of Sleep	Physical Activity Level	Stress Level
<dbl>	<chr>	<dbl>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>
1	Male	27	Software Engineer	6.1	6	42	6
2	Male	28	Doctor	6.2	6	60	8
3	Male	28	Doctor	6.2	6	60	8
4	Male	28	Sales Representative	5.9	4	30	8
5	Male	28	Sales Representative	5.9	4	30	8
6	Male	28	Software Engineer	5.9	4	30	8

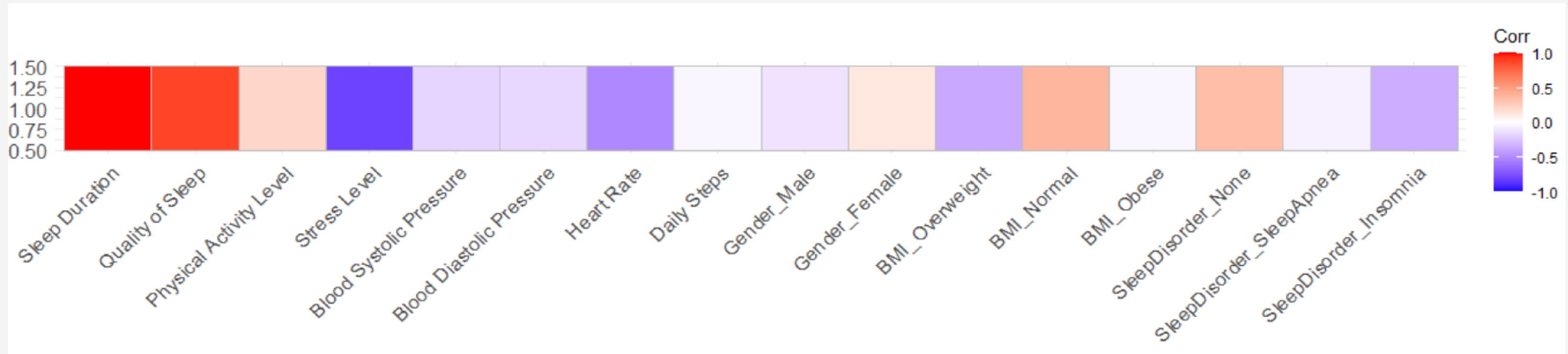
BMI Category	Blood Pressure	Heart Rate	Daily Steps	Sleep Disorder
<chr>	<chr>	<dbl>	<dbl>	<chr>
Overweight	126/83	77	4200	None
Normal	125/80	75	10000	None
Normal	125/80	75	10000	None
Obese	140/90	85	3000	Sleep Apnea
Obese	140/90	85	3000	Sleep Apnea
Obese	140/90	85	3000	Insomnia



# Macierz korelacji

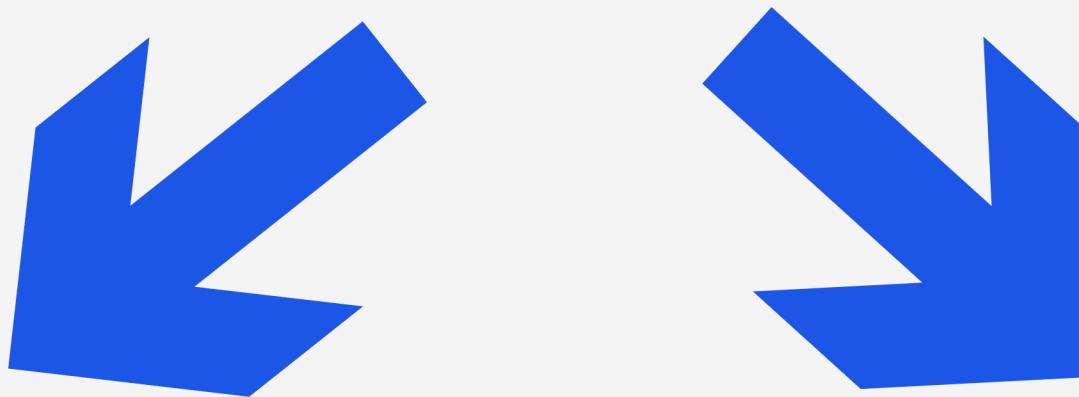


# Analiza korelacji między zmienną "Sleep Duration" a innymi zmiennymi



```
Sleep Duration vs Age: 0.34
Sleep Duration vs Sleep Duration: 1.00
Sleep Duration vs Quality of Sleep: 0.88
Sleep Duration vs Stress Level: -0.81
Sleep Duration vs Heart Rate: -0.52
Sleep Duration vs BMI_Overweight: -0.37
Sleep Duration vs BMI_Normal: 0.38
Sleep Duration vs SleepDisorder_None: 0.34
Sleep Duration vs SleepDisorder_Insomnia: -0.35
```

# Podział na dane testowe i treningowe



**TRENINGOWE**

**80%**

**TESTOWE**

**20%**

# Model regresji prostej

## Sleep Duration ~ Quality of Sleep

```
Call:
lm(formula = `Sleep Duration` ~ `Quality of Sleep`, data = new_dataset)

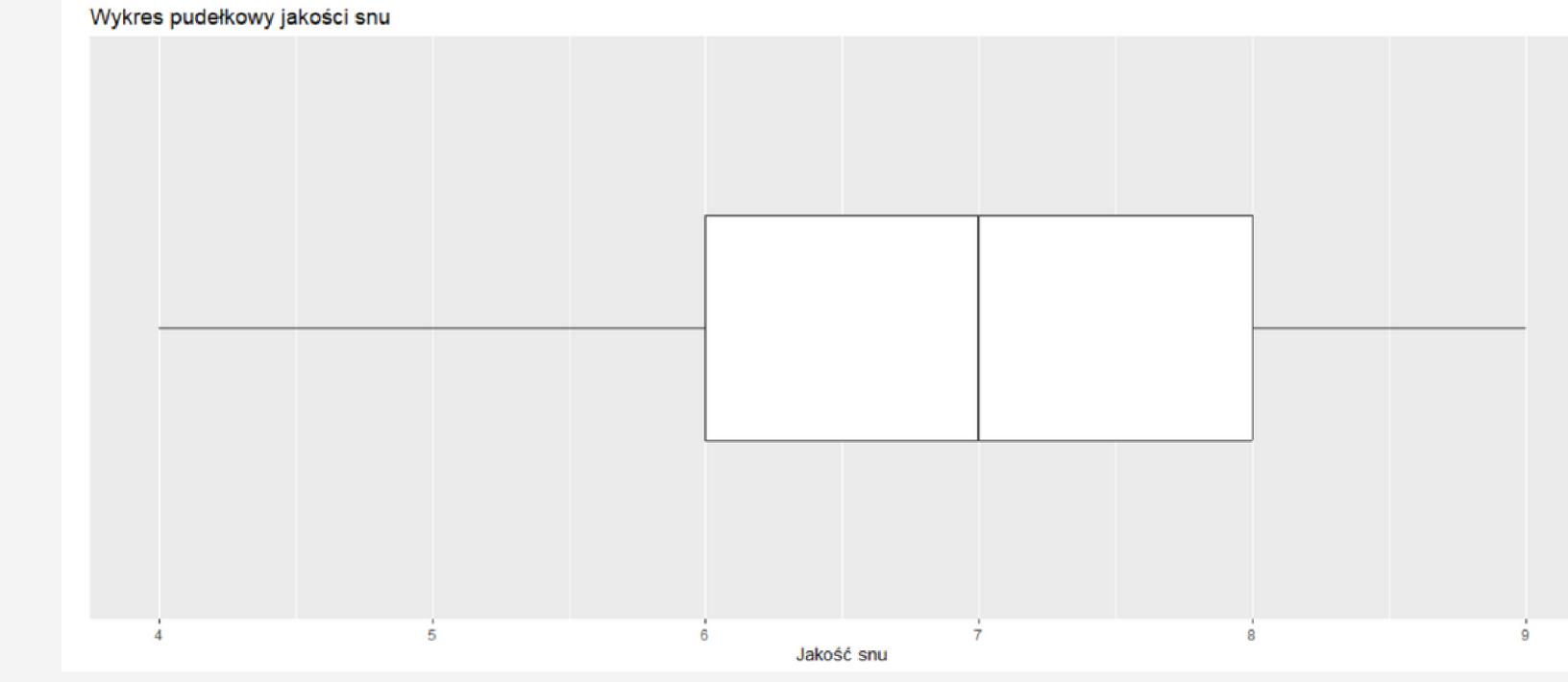
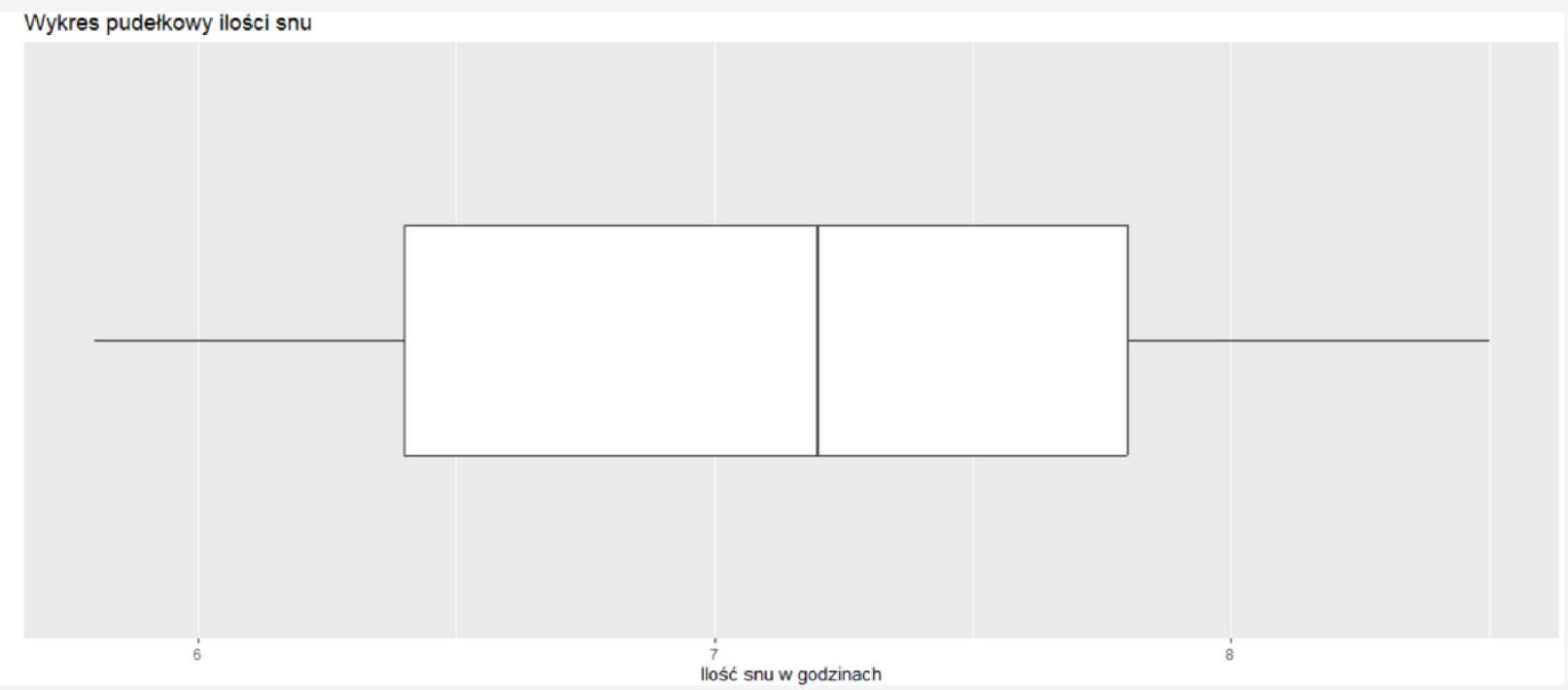
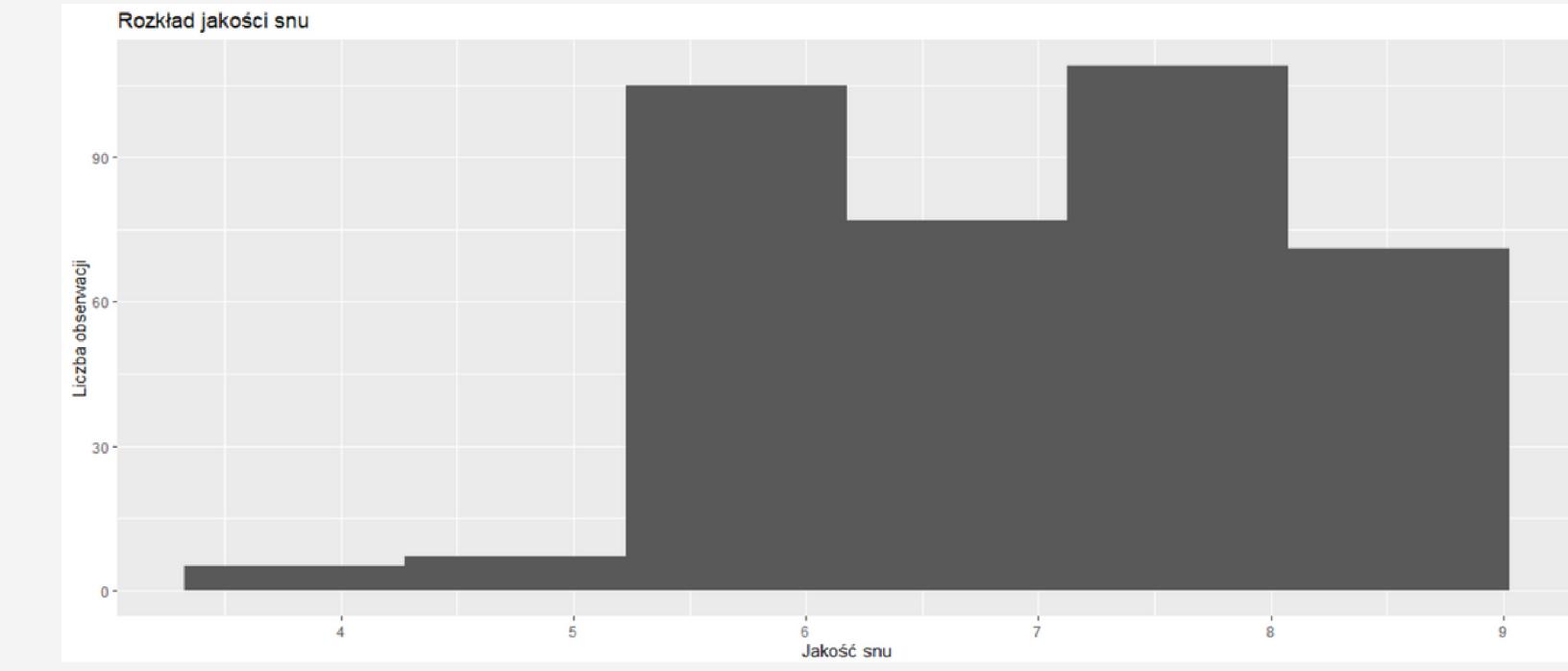
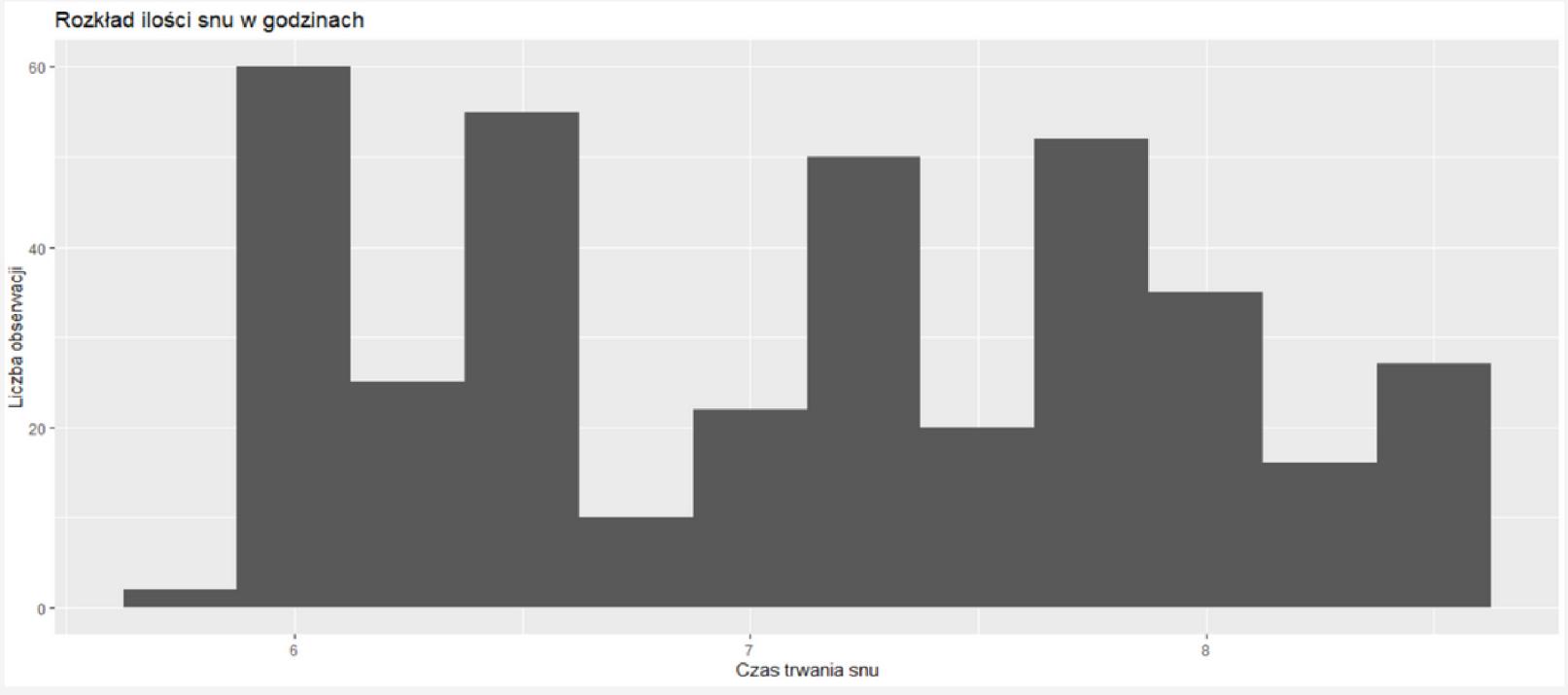
Residuals:
    Min      1Q  Median      3Q     Max 
-0.46132 -0.33552 -0.06132  0.26448  0.95158 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  2.83871   0.11977   23.70 <2e-16 ***
`Quality of Sleep` 0.58710   0.01616   36.32 <2e-16 ***
---
Signif. codes:  0 '****' 0.001 '***' 0.01 '**' 0.05 '*' 0.1 '.' 1

Residual standard error: 0.3736 on 372 degrees of freedom
Multiple R-squared:  0.7801,    Adjusted R-squared:  0.7795 
F-statistic: 1319 on 1 and 372 DF,  p-value: < 2.2e-16

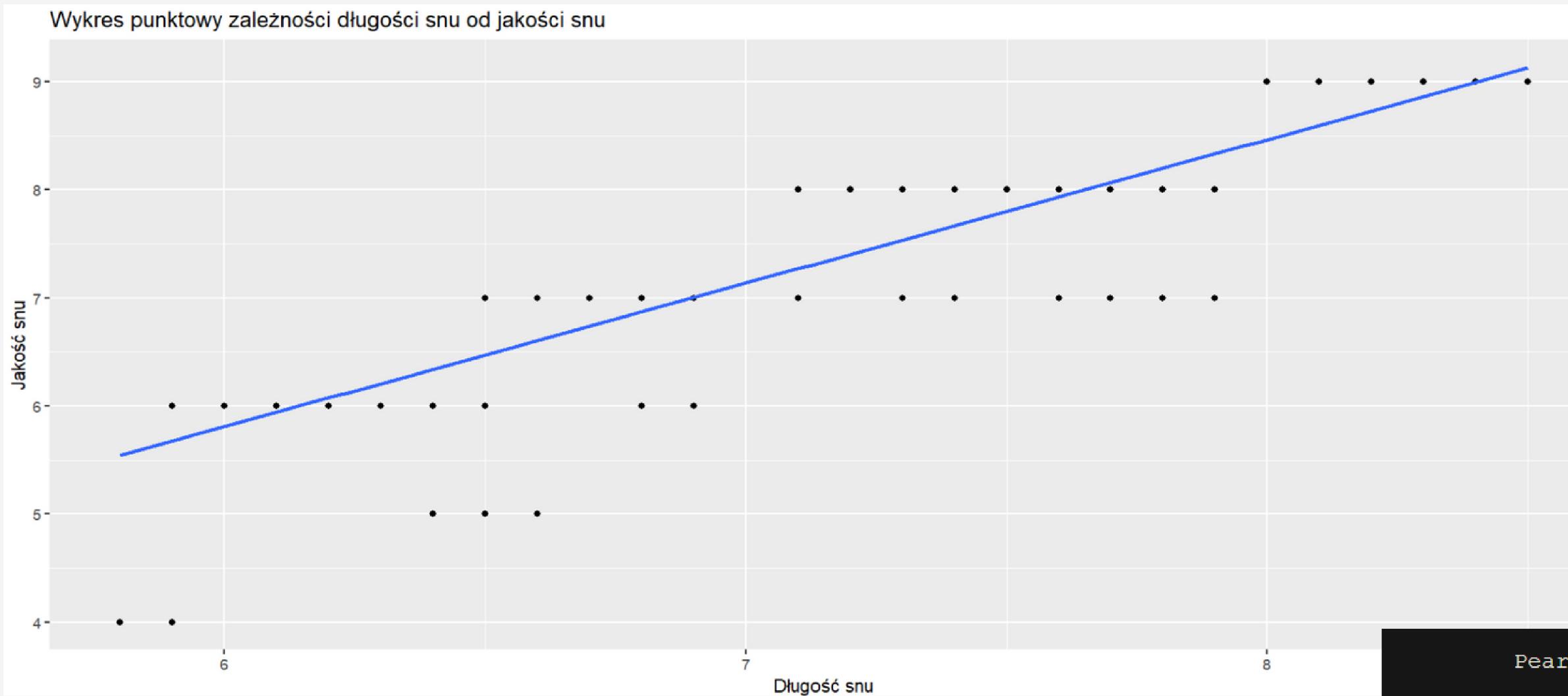
[1] "Wartości charakterystyk liczbowych modelu."
[1] -----
Treningowe R^2 wyniosło: 0.7800652
Treningowe "poprawione" R^2 wyniosło: 0.779474
Kryterium informacyjne Akaikego (AIC) wyniosło: 328.9867
-----
Charakterystyki "out-of-sample"
Charakterystyka | train | test |
RMSE wyniosło: | 0.373642 | 0.3993711 |
MAE wyniosło:  | 0.309407 | 0.3416302 |
MAPE wyniosło: | 4.4 %| 4.8 %|
```

# Sleep Duration



# Quality of Sleep

# Założenia modelu regresji liniowej prostej

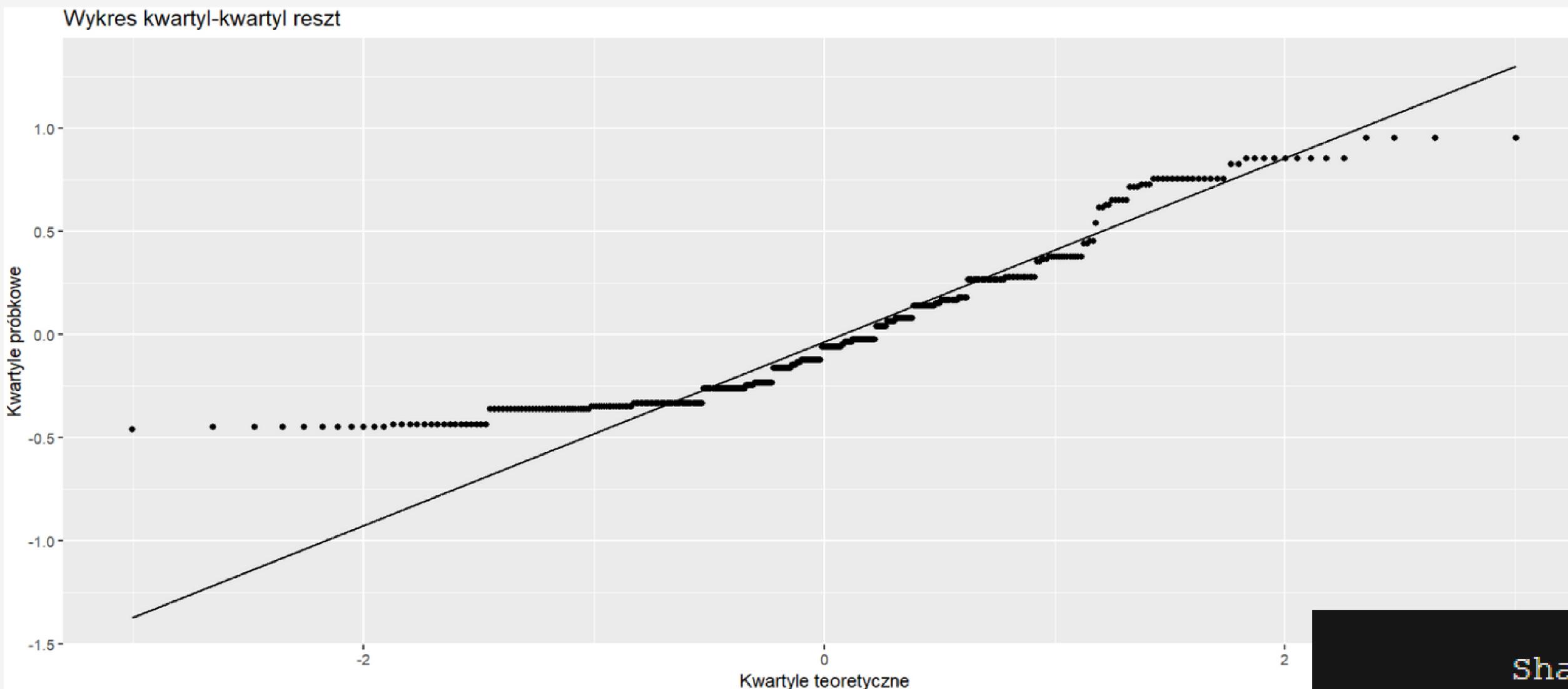


## Zależność liniowa

Pearson's product-moment correlation

```
data: new_dataset$`Sleep Duration` and new_dataset$`Quality of Sleep`  
t = 36.324, df = 372, p-value < 2.2e-16  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
 0.8587162 0.9036825  
sample estimates:  
 cor  
 0.883213
```

# Założenia modelu regresji liniowej prostej



## Rozkład reszt

Shapiro-Wilk normality test

```
data: model_Quality_of_Sleep$residuals  
W = 0.89686, p-value = 3.229e-15
```

# Założenia modelu regresji liniowej prostej

**Zerowa średnia  
reszt**

```
One Sample t-test

data: model_Quality_of_Sleep$residuals
t = -3.86e-15, df = 373, p-value = 1
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
-0.03793993 0.03793993
sample estimates:
mean of x
-7.447752e-17
```

**Niezależność reszt**

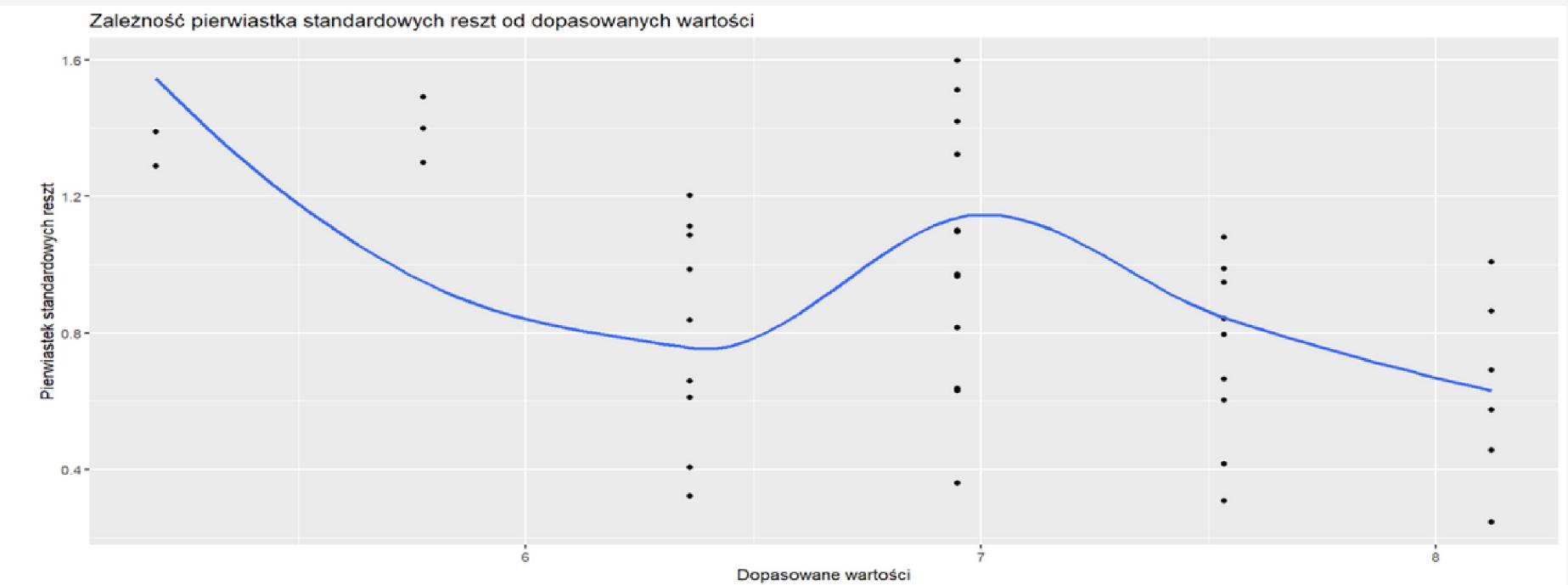
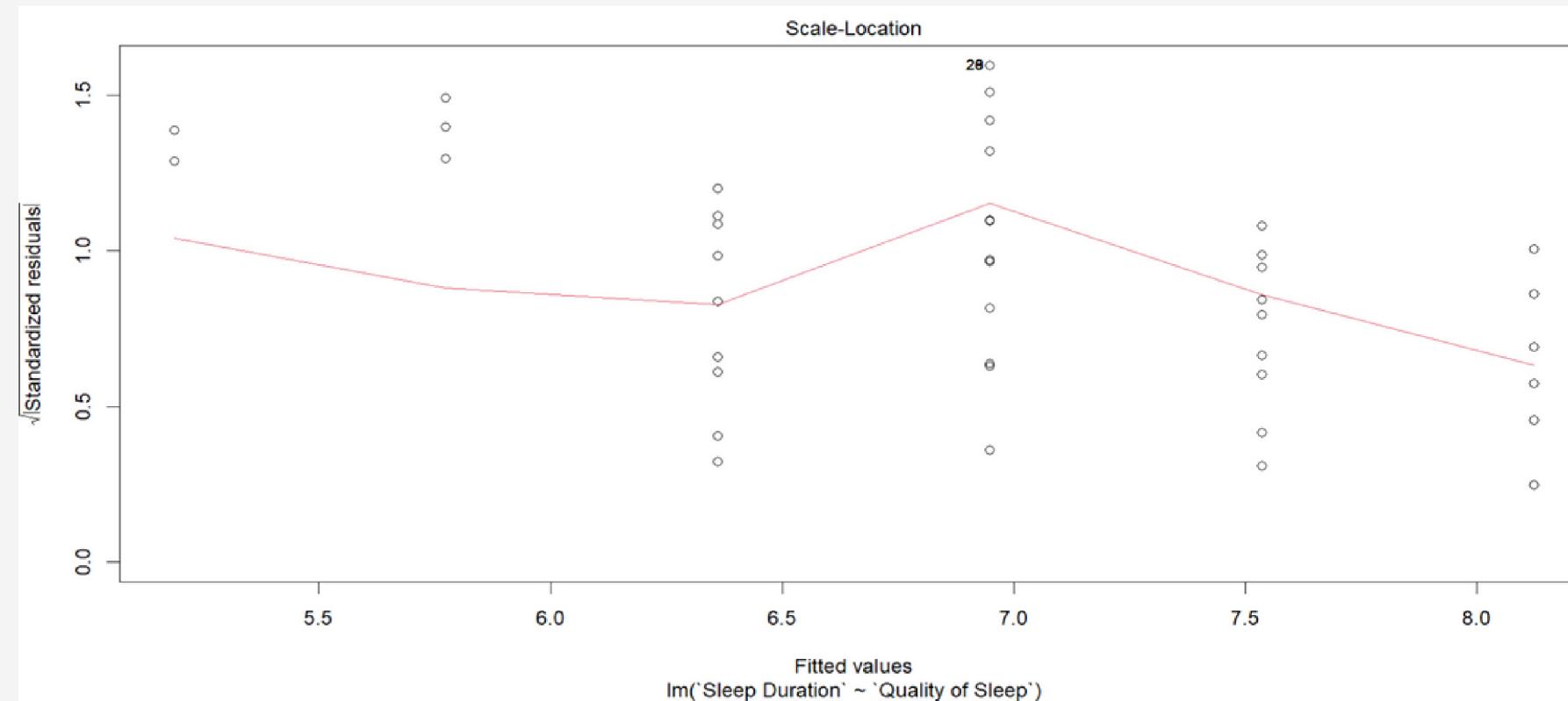
```
Durbin-Watson test

data: model_Quality_of_Sleep
DW = 0.82619, p-value < 2.2e-16
alternative hypothesis: true autocorrelation is greater than 0
```

# Założenia modelu regresji liniowej prostej

## Homoskedatyczność

```
studentized Breusch-Pagan test  
data: model_Quality_of_Sleep  
BP = 25.199, df = 1, p-value = 5.17e-07
```



# Model regresji wielorakiej

```
Call:
lm(formula = `Sleep Duration` ~ Age + `Quality of Sleep` + `Stress Level` +
    `Heart Rate` + BMI_Overweight + BMI_Normal + SleepDisorder_None +
    SleepDisorder_Insomnia, data = new_dataset)

Residuals:
    Min      1Q  Median      3Q     Max 
-0.56161 -0.25876 -0.02199  0.19838  0.92171 

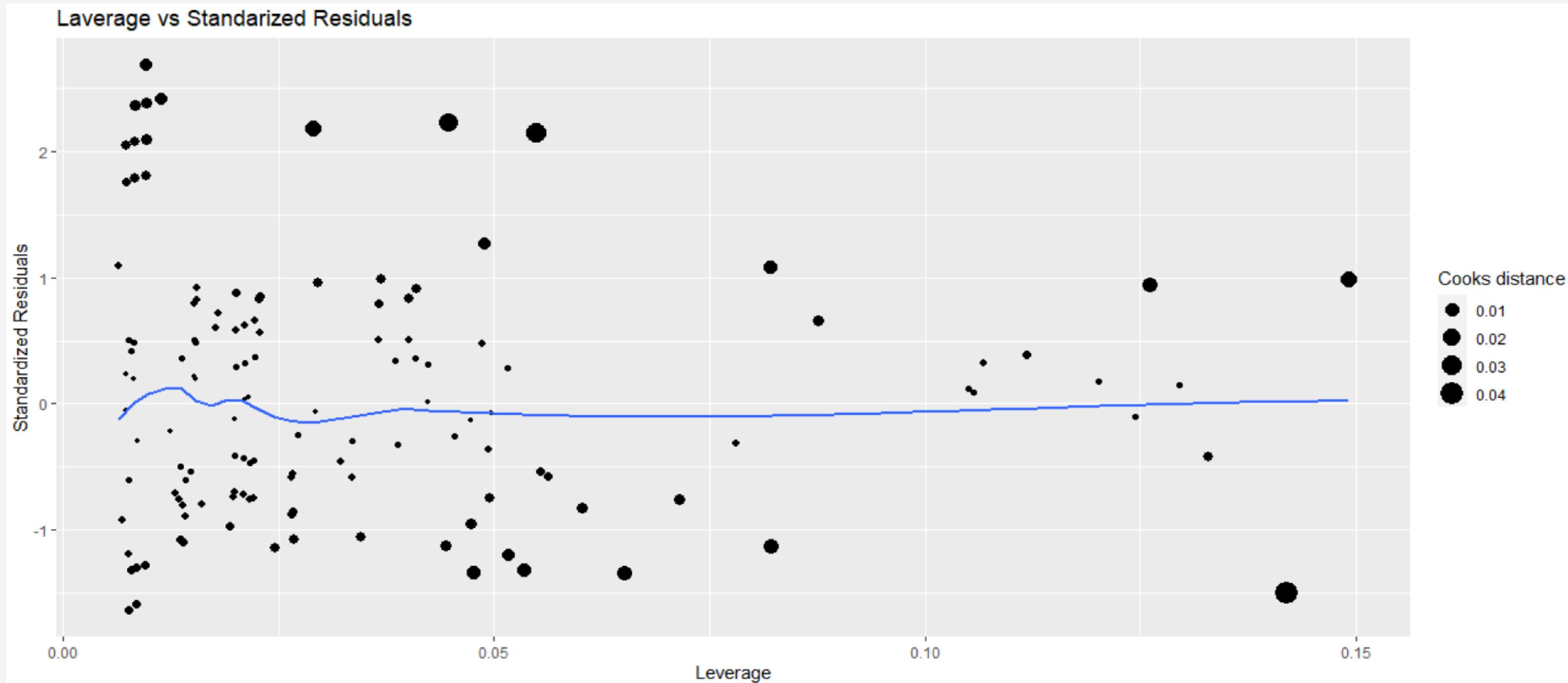
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 1.235437  0.901289  1.371  0.17130  
Age          0.007408  0.004314  1.717  0.08678 .  
`Quality of Sleep` 0.406134  0.053967  7.526 4.14e-13 *** 
`Stress Level`   -0.152995  0.030518 -5.013 8.37e-07 *** 
`Heart Rate`     0.044033  0.009839  4.476 1.02e-05 *** 
BMI_Overweight   0.126140  0.173010  0.729  0.46641  
BMI_Normal       0.508301  0.189772  2.678  0.00773 **  
SleepDisorder_None 0.005010  0.073184  0.068  0.94546  
SleepDisorder_Insomnia 0.009644  0.066558  0.145  0.88487  
---
Signif. codes:  0 '****' 0.001 '***' 0.01 '**' 0.05 '*' 0.1 '.' 1

Residual standard error: 0.3453 on 365 degrees of freedom
Multiple R-squared:  0.8157,    Adjusted R-squared:  0.8117 
F-statistic: 201.9 on 8 and 365 DF,  p-value: < 2.2e-16

[1] "Wartości charakterystyk liczbowych modelu."
[1] -----
Treningowe R^2 wyniosło: 0.8156996
Treningowe "poprawione" R^2 wyniosło: 0.8116601
Kryterium informacyjne Akaikego (AIC) wyniosło: 276.8769
-----
Charakterystyki "out-of-sample"
Charakterystyka |   train |   test  |
RMSE wyniosło: | 0.3453002 | 0.3709633 |
MAE wyniosło:  | 0.2770816 | 0.3059643 |
MAPE wyniosło: | 3.9 %| 4.29 %|
```

**Age**  
**Quality of Sleep**  
**Stress Level**  
**Heart Rate**  
**BMI\_Overweight**  
**BMI\_Normal**  
**SleepDisorder\_None**  
**SleepDisorder\_Insomnia**

# Odległość Cook'a



# Najlepszy z modeli

```
Call:
lm(formula = `Sleep Duration` ~ Age + `Quality of Sleep` + `Stress Level` +
    `Heart Rate` + BMI_Overweight + BMI_Normal + SleepDisorder_None +
    SleepDisorder_Insomnia, data = new_dataset_2)

Residuals:
    Min      1Q  Median      3Q     Max 
-0.35133 -0.13694  0.01775  0.12854  0.54393 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -1.640977  0.638708 -2.569   0.01072 *  
Age          0.028214  0.002722 10.366  < 2e-16 *** 
`Quality of Sleep` 0.383707  0.036754 10.440  < 2e-16 *** 
`Stress Level`  -0.201264  0.021872 -9.202  < 2e-16 *** 
`Heart Rate`    0.070353  0.007253  9.700  < 2e-16 *** 
BMI_Overweight  0.425075  0.145501  2.921  0.00377 ** 
BMI_Normal     0.974704  0.150354  6.483  4.18e-10 *** 
SleepDisorder_None 0.092948  0.054149  1.717  0.08720 .  
SleepDisorder_Insomnia 0.257626  0.040369  6.382  7.44e-10 *** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.178 on 274 degrees of freedom
Multiple R-squared:  0.9565,    Adjusted R-squared:  0.9552 
F-statistic: 752.6 on 8 and 274 DF,  p-value: < 2.2e-16

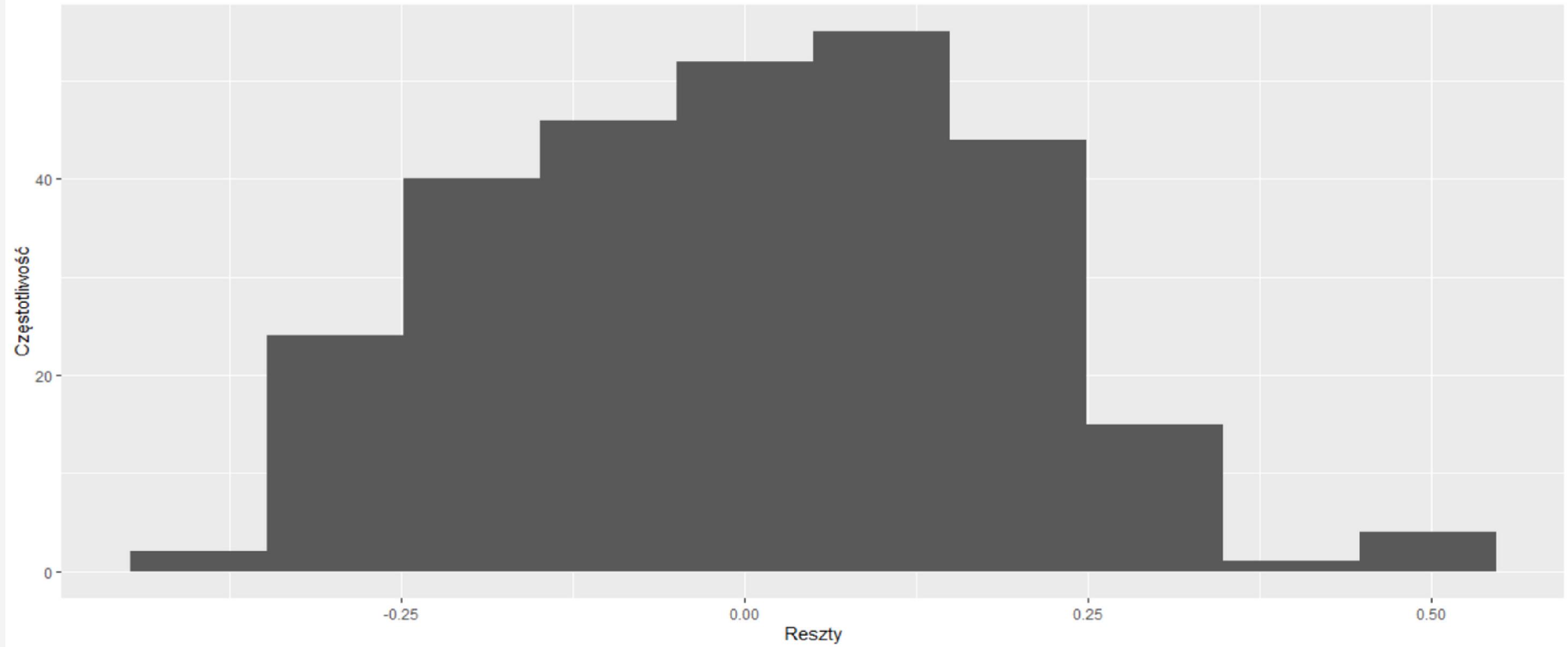
[1] "Wartości charakterystyk liczbowych modelu."
[1] "-----"
Treningowe R^2 wyniosło: 0.9564715
Treningowe "poprawione" R^2 wyniosło: 0.9552006
Kryterium informacyjne Akaikego (AIC) wyniosło: -162.878
-----
Charakterystyki "out-of-sample"
Charakterystyka |   train   |   test   |
RMSE wyniosło: | 0.1780154 | 0.1579953 |
MAE wyniosło:  | 0.143318  | 0.1280896 |
MAPE wyniosło: | 2.04 %| 1.83 %|
```

# Założenia

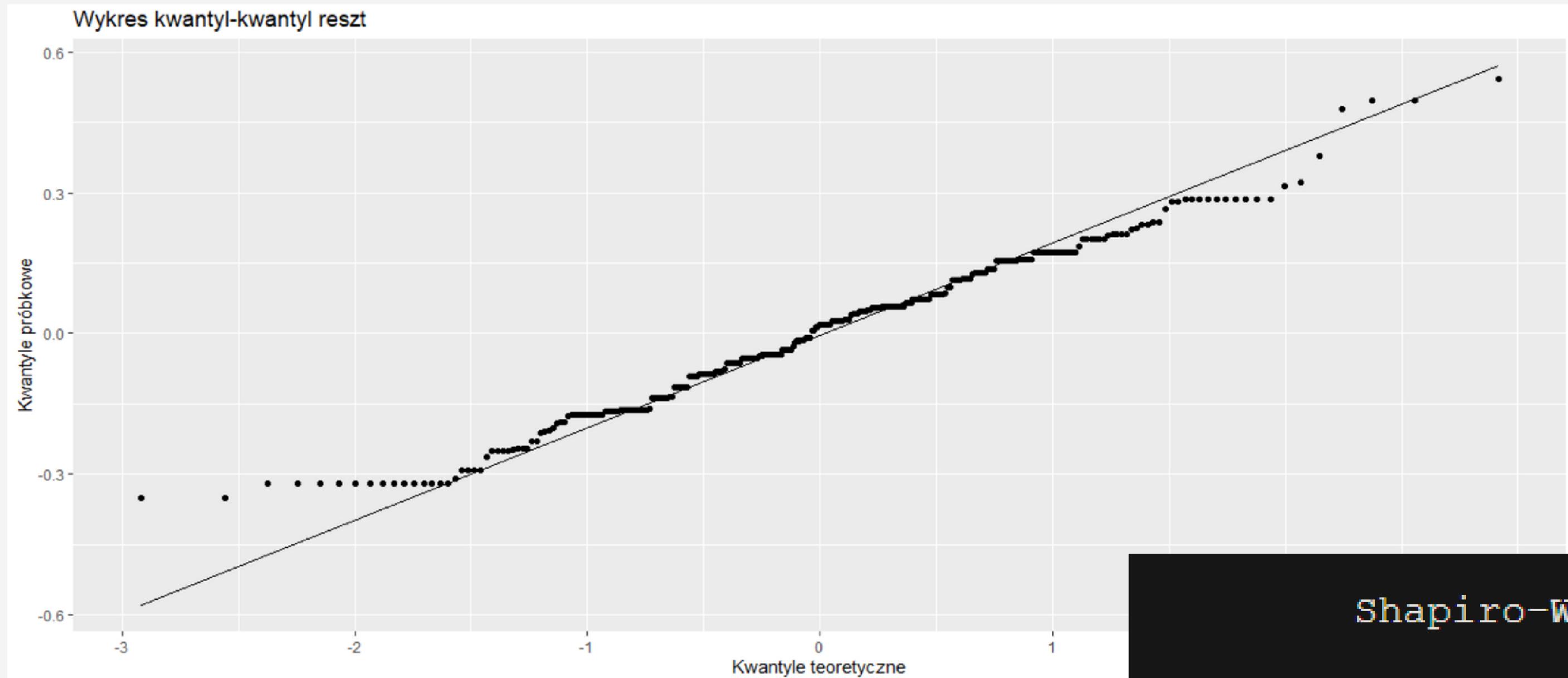
1. Normalność reszt
2. Zerowość średniej reszt
3. Niezależność reszt
4. Homoskedastyczność
5. Liniowa niezależność

# Normalność reszt

Histogram reszt z modelu



# Normalność reszt



Shapiro-Wilk normality test

```
data: WIELORAKA_2$residuals  
W = 0.98484, p-value = 0.004351
```

# Zerowość średniej reszt

```
One Sample t-test

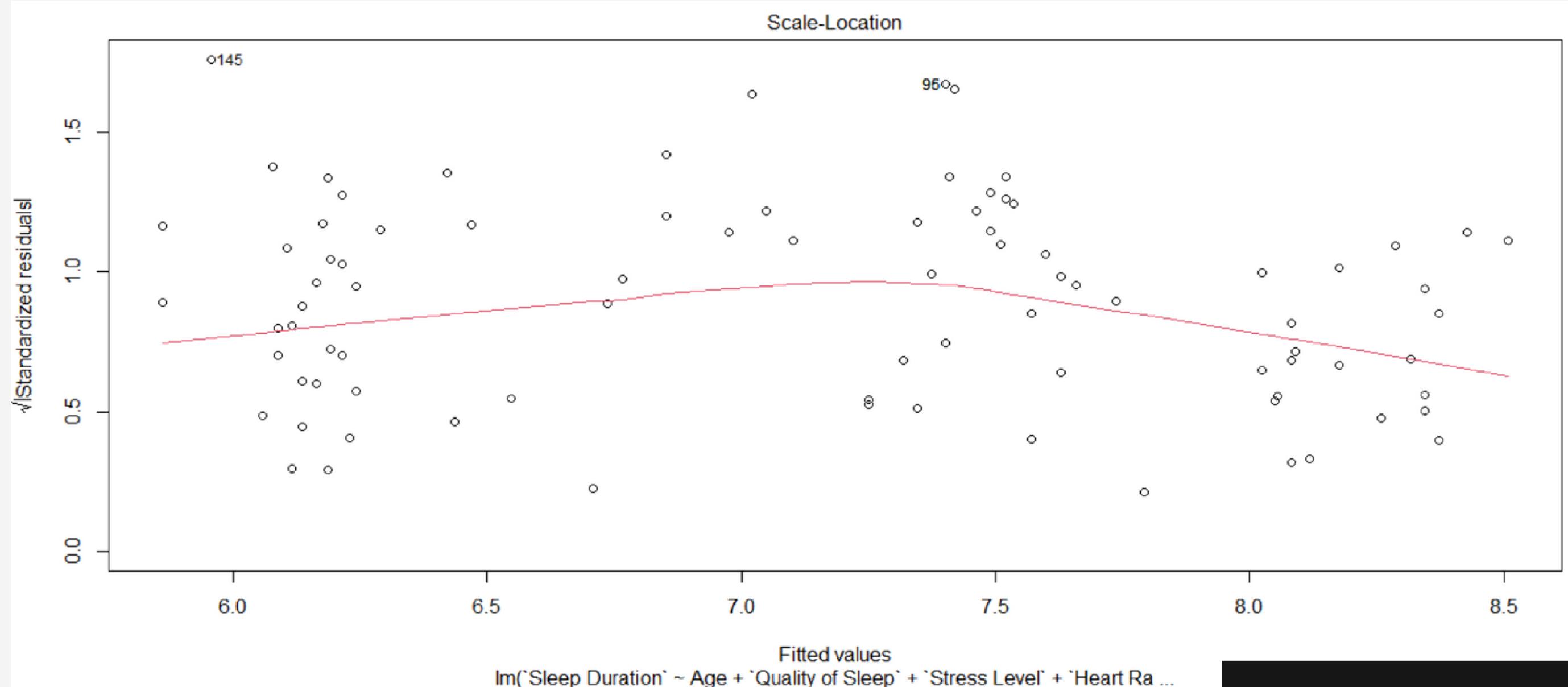
data: WIELORAKA_2$residuals
t = -7.8246e-16, df = 282, p-value = 1
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
-0.02053199 0.02053199
sample estimates:
mean of x
-8.161626e-18
```

# Niezależność reszt

```
Durbin-Watson test
```

```
data: WIELORAKA_2
DW = 0.96945, p-value < 2.2e-16
alternative hypothesis: true autocorrelation is greater than 0
```

# Homoskedastyczność



studentized Breusch-Pagan test

```
data: WIELORAKA_2  
BP = 23.934, df = 8, p-value = 0.002351
```

# Liniowa niezależność

	Sleep Duration	Age	Quality of Sleep	Stress Level	Heart Rate	BMI_Overweight	BMI_Normal
Sleep Duration	1.00	0.49	0.95	-0.90	-0.64	-0.32	0.31
Age	0.49	1.00	0.48	-0.44	-0.20	0.54	-0.54
Quality of Sleep	0.95	0.48	1.00	-0.93	-0.73	-0.29	0.30
Stress Level	-0.90	-0.44	-0.93	1.00	0.78	0.19	-0.19
Heart Rate	-0.64	-0.20	-0.73	0.78	1.00	0.26	-0.34
BMI_Overweight	-0.32	0.54	-0.29	0.19	0.26	1.00	-0.98
BMI_Normal	0.31	-0.54	0.30	-0.19	-0.34	-0.98	1.00
SleepDisorder_Insomnia	-0.33	0.01	-0.39	0.19	0.14	0.46	-0.49
SleepDisorder_None	0.30	-0.46	0.33	-0.21	-0.39	-0.86	0.88
	SleepDisorder_Insomnia	SleepDisorder_None					
Sleep Duration	-0.33	0.30					
Age	0.01	-0.46					
Quality of Sleep	-0.39	0.33					
Stress Level	0.19	-0.21					
Heart Rate	0.14	-0.39					
BMI_Overweight	0.46	-0.86					
BMI_Normal	-0.49	0.88					
SleepDisorder_Insomnia	1.00	-0.59					
SleepDisorder_None	-0.59	1.00					