Karol Pichurski

321663

# Podstawy Analizy Danych – Lista 2

Residuals:

Min 1Q Median 3Q Max

-403.45 -167.25 -36.04 185.31 449.36

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -129.10235 140.12772 -0.921 0.366

obciazenie\_karty 0.76886 0.03045 25.248 <2e-16 \*\*\*

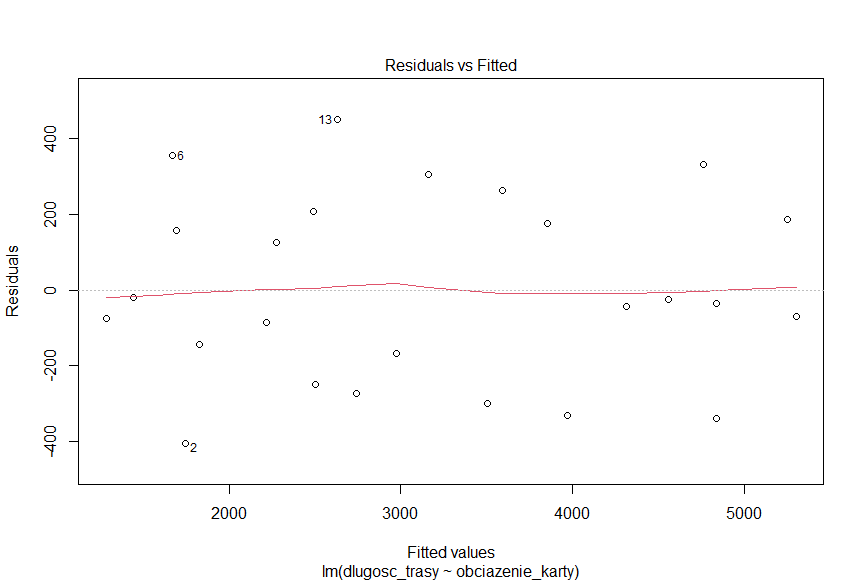
---

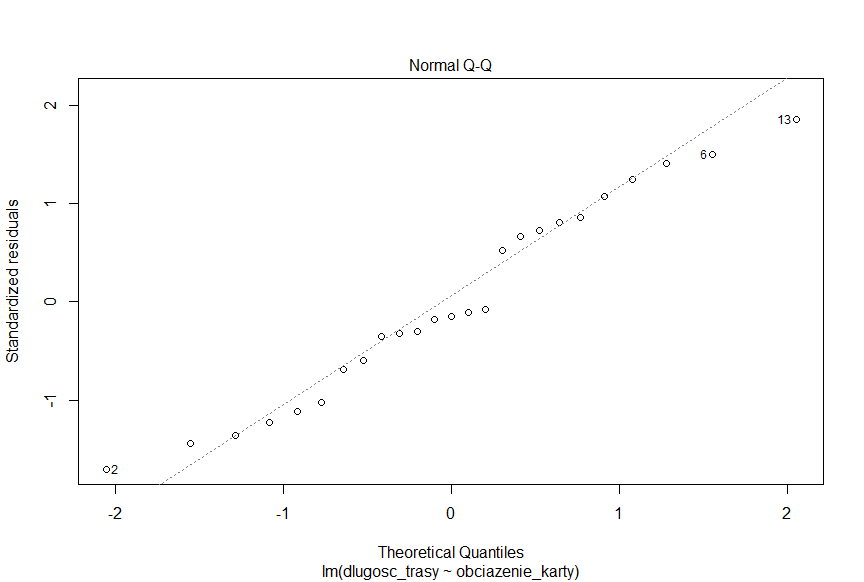
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

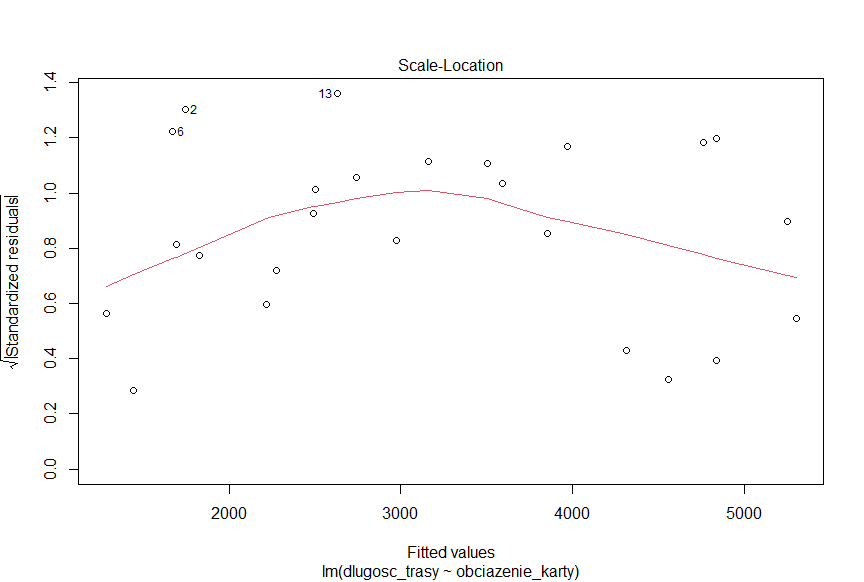
Residual standard error: 249 on 23 degrees of freedom

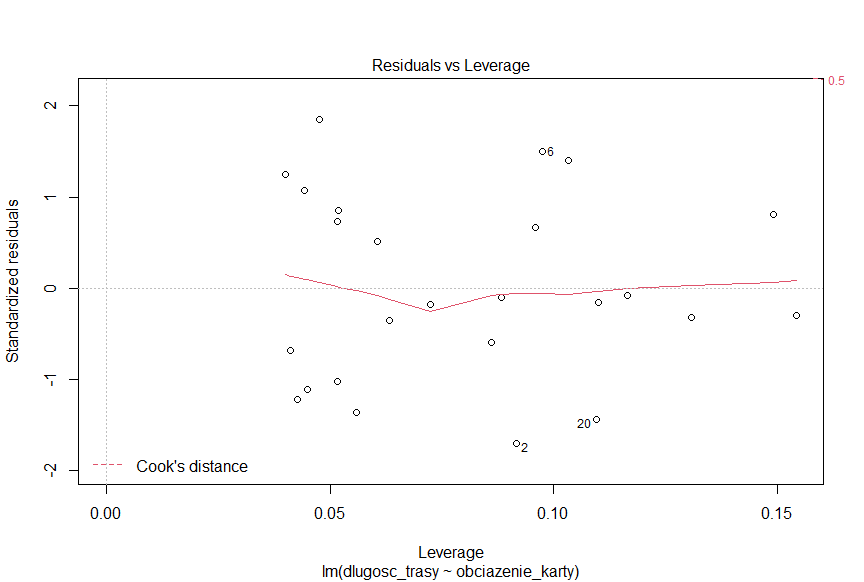
Multiple R-squared: 0.9652, Adjusted R-squared: 0.9637

F-statistic: 637.5 on 1 and 23 DF, p-value: < 2.2e-16









Całkowita suma kwadratów (SST): 66855898

Suma kwadratów błędów (SEE): 1425940.2430576

Regresyjna suma kwadratów (SSR): 71065566.5569424

Współczynnik determinacji R2: 1.06296630039944

Skorygowany współczynnik determinacji: 1.0657039656342

Minimalny znaczący współczynnik (MSR): 71065566.5569424

Nieobciążony estymator wariancji (MSE): 61997.4018720698

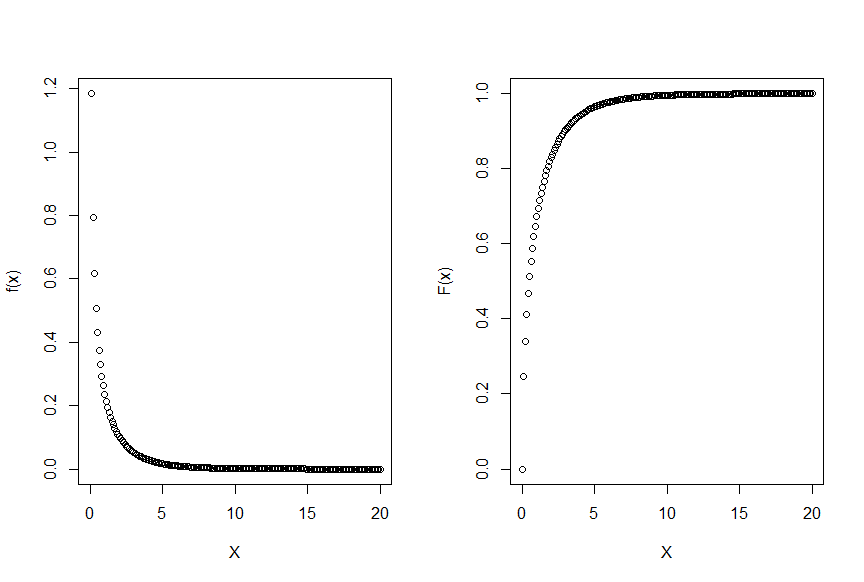
Graniczny poziom istotności: 3.97362008546336e-21

Statystyka F: 1146.2668500784

Obszar krytyczny wartości F: 4.27934430914465

Ponieważ F nie wpada do przedziału krytycznego, zatem hipotezę zerową należy odrzucić.

Rozkład gęstości i dystrybuanta:



Wartość t: 25.2482109973585

Wartość graniczna t: 2.06865761041905

P-value: 2.85083541898162e-18

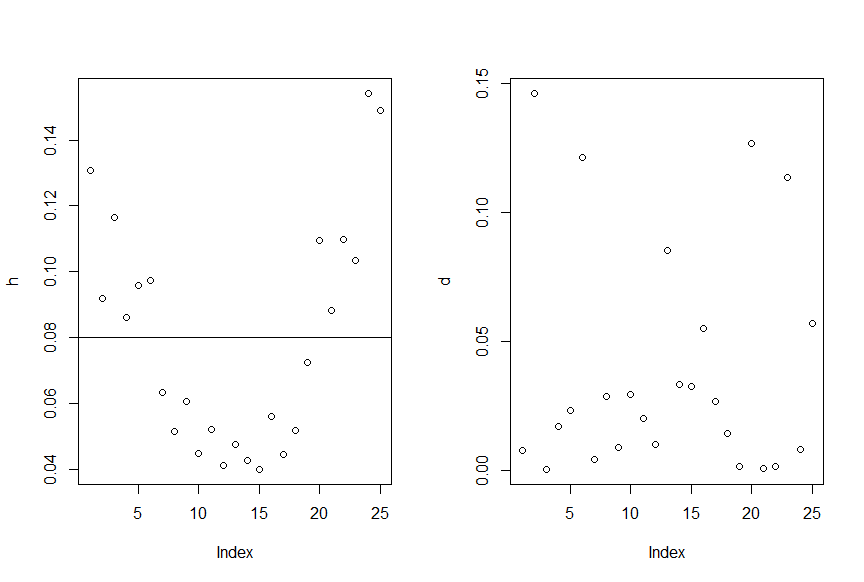
Ponieważ wartość statystyki t wpada do obszaru krytycznego, hipotezę zerową należy odrzucić. Oznacza, to że nie ma podstaw do wnioskowania, że parametr nieistotny.

Obserwacje odstające:

0.13067904 0.09170261 0.11634307 0.08597515 0.09584245 0.09741400

Obserwacje wpływowe:

0.0076020471 0.1459189700 0.0004297142 0.0169607788 0.0232086186 0.1211448157



Kod:

1. library(readxl)
3. dane **=** read\_excel("Zadanie\_domowe\_nr\_2\_2022\_2023\_KP.xlsx")
4. dlugosc\_trasy **=** dane$Długość\_trasy\_w\_milach
5. obciazenie\_karty **=** dane$`Obciążenie\_kart\_w USD`
6. model **=** lm(formula**=**dlugosc\_trasy~obciazenie\_karty, data**=**dane)
8. print(summary(model))
9. plot(model)
11. y\_mean **=** mean(obciazenie\_karty)
12. SST **=** sum((obciazenie\_karty **-** y\_mean) ^ 2)
13. my\_print("Całkowita suma kwadratów (SST)", SST)
15. SEE **=** sum(model$residuals ^ 2)
16. my\_print("Suma kwadratów błędów (SEE)", SEE)
18. SSR **=** sum((model$fitted.values **-** y\_mean) ^ 2)
19. my\_print("Regresyjna suma kwadratów (SSR)", SSR)
21. R2 **=** SSR **/** SST
22. my\_print("Współczynnik determinacji R2", R2)
24. n **=** nrow(dane)
25. p **=** 2
27. R2\_adj **=** 1 **-** (1 **-** R2) **\*** (n **-** 1) **/** (n **-** p)
28. my\_print("Skorygowany współczynnik determinacji", R2\_adj)
30. MSR **=** SSR **/** (1)
31. my\_print("Minimalny znaczący współczynnik (MSR)", MSR)
33. MSE **=** SEE **/** (n **-** p)
34. my\_print("Nieobciążony estymator wariancji (MSE)", MSE)
36. F **=** MSR **/** MSE
37. my\_print("Statystyka F", F)
39. par(mfrow**=**c(1, 2))
40. plot(seq(0, 20, 0.1), df(seq(0, 20, 0.1), p**-**1, n**-**p), xlab**=**"X", ylab**=**"f(x)")
41. plot(seq(0, 20, 0.1), pf(seq(0, 20, 0.1), p**-**1, n**-**p), xlab**=**"X", ylab**=**"F(x)")
43. obszar\_krytyczny **=** qf(0.95, p**-**1, n**-**p)
44. my\_print("Obszar krytyczny wartości F", obszar\_krytyczny)
46. graniczny\_poziom\_istotnosci **=** pf(F, p **-** 1, n **-** p, lower.tail**=**FALSE)
47. my\_print("Graniczny poziom istotności", graniczny\_poziom\_istotnosci)
49. alfa **=** model$coefficients
50. alfa\_se **=** sqrt(diag(vcov(model)))
51. t **=** alfa **/** alfa\_se
52. my\_print("Wartość t", t[2])
54. t\_gr **=** qt(1 **-** 0.05**/**2, df**=**n**-**p)
55. my\_print("Wartość graniczna t", t\_gr)
57. p\_value **=** 2**\***pt(**-**abs(t), df**=**n**-**p, lower.tail**=**TRUE)
58. my\_print("P-value", p\_value[2])

61. h **=** hatvalues(model)
62. print("Obserwacje odstające")
63. print(head(h))
64. plot(h)
65. abline(2 **/** n, 0)
67. d **=** cooks.distance(model)
68. print("Obserwacje wpływowe")
69. print(head(d))
70. plot(d)