

Lista 4 - Modele liniowe - Raport

Erwin Jasic

28 stycznia 2021

Zadanie 1

(a)

[1] 27

(b)

[1] 13

(c)

[1] 1.890184

[1] 6.109816

W zadaniu 1 mieliśmy, do policzenia prognozę dla Y , następnie obliczyliśmy $\sigma^2(pred)$, na podstawie danych oraz obliczyliśmy przedział ufności w podpunkcie c), pierwszy wynik w tym podpunkcie oznacza lewy kraniec przedziału, a drugi prawy. Wszystkie obliczenia zostały wykonane na kartce.

Zadanie 2

(a)

[1] 20

(b)

[1] 1.5

[1] 4.351244

(c)

[1] 1.5

[1] 3.492828

(d)

[1] 6

[1] 3.098391

(e)

[1] 14.34783

(f)

[1] 0.6282809

W zadaniu 2 też mieliśmy policzyć kilka rzeczy na kartce. W podpunkcie (a) wynik jest równy 20, ponieważ jest to ostatnia wartość w tabelce, a wtedy SS1 i SS2 dla takiej zmiennej są takie same. W podpunkcie (b) mieliśmy przetestować hipotezę czy β_1 ma istotny wpływ na Y, nie odrzucamy hipotezy zerowej. W podpunkcie (c) testujemy czy β_2 lub β_3 ma istotny wpływ na Y, nie odrzucamy hipotezy zerowej. W podpunkcie (d) testujemy, czy jakkolwiek z bet ma wpływ na Y, odrzucamy hipotezę zerową. W podpunkcie (e) testujemy czy β_1 ma istotny wpływ w modelu zredukowanym, odrzucamy hipotezę zerową. W ostatnim podpunkcie liczymy korelację między X1 a Y.

Zadanie 3

```
##
## Call:
## lm(formula = Y ~ X[, 1])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.88718 -0.48185  0.00427  0.61123  2.71962
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.0332     0.1040   0.319   0.7503
## X[, 1]        2.4170     1.1251   2.148   0.0342 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.038 on 98 degrees of freedom
## Multiple R-squared:  0.04497,    Adjusted R-squared:  0.03522
## F-statistic: 4.615 on 1 and 98 DF,  p-value: 0.03417
##
## Call:
## lm(formula = Y ~ X[, 1] + X[, 2])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.84616 -0.62415  0.01469  0.60560  2.86142
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.0384     0.1040   0.369   0.713
## X[, 1]       -0.2638     2.5943  -0.102   0.919
## X[, 2]        2.9342     2.5595   1.146   0.254
##
## Residual standard error: 1.036 on 97 degrees of freedom
## Multiple R-squared:  0.05774,    Adjusted R-squared:  0.03831
## F-statistic: 2.972 on 2 and 97 DF,  p-value: 0.05589
##
##           2.5 %    97.5 %
## X[, 1] 0.1841547 4.649775
##
##           2.5 %    97.5 %
## X[, 1] -5.412678 4.885041
```

W zadaniu 3 (podpunkty a i b) generujemy macierz sigma, macierz X, wektor Y oraz wektor błędów losowych. Następnie tworzymy model pełny oraz zredukowany. Liczymy przedziały ufności. Widzimy, że w pierwszym przypadku nie mamy podstaw do odrzucenia hipotezy zerowej a w drugim odrzucamy.

(c) Odchylenia standardowe dla zredukowanego i pełnego:

```
## [1] 1.265948
```

```
## [1] 2.594246
```

Moce dla zredukowanego i pełnego:

```
## [1] 0.7824441
```

```
## [1] 0.2205107
```

(d) Doświadczalne odchylenia standardowe:

```
## [1] 0.770994
```

```
## [1] 0.2319962
```

```
## [1] 1.170541
```

```
## [1] 2.654002
```

Zadanie 4

(a)

```
## [1] 0.003391774 0.012969732 0.064216354 0.081603208 1.082822188
```

```
## [6] 99.663188916 -1.333292407 -1.198560191
```

```
## [1] 0 0 0 0 1 4 28 14
```

```
## [1] 1434.69869 1318.22517 1057.50831 1056.24645 1008.50606 964.07375 531.86130
```

```
## [8] 60.95778
```

```
## [1] 3204.832 3122.163 2907.793 2916.599 2950.347 3005.290 3210.505 1944.303
```

```
## [ ,1] [ ,2]
```

```
## [1,] 0.00000000000000003295099994 NA
```

```
## [2,] 0.0000000000000000440273118 0.000000000000000004097951556
```

```
## [3,] 0.00000000000000001903198 0.000000000000000000011721551
```

```
## [4,] 0.00000000000000002998843 0.000000000000000000030929351
```

```
## [5,] 0.00000000000000003163317 0.000000000000000000002370725
```

```
## [6,] 0.00000000000000002060343 0.0000000000000000000134171082
```

```
## [7,] 0.00000001566163212766797 0.000000000000000487864691086476
```

```
## [8,] 0.03205108549434355219576 0.00029078455253626818996851022
```

(b)

```
## [1] 0.0009706218 0.0006948973 0.2376224964 0.2528806810 0.4873634712
```

```
## [6] 0.6447479270 10.5991749122 375.9669906441
```

```
## [1] 0 0 0 0 9 14 295 14
```

```
## [1] 1402.15895 1304.60373 1301.08046 1289.30164 1078.30217 1014.27161 318.35795
```

```
## [8] 60.95778
```

```
## [1] 3181.890 3111.776 3115.072 3115.978 3017.265 3056.048 2697.298 1944.303
```

```
## [ ,1] [ ,2]
```

```
## [1,] 0.00000000000000000300517906661250977420 NA
```

```
## [2,] 0.0000000000000000005559344446118341544 0.000000000000000023082709714470
```

```
## [3,] 0.00000000000000000050934494809163752640 0.000000000000000014341151525783
```

```
## [4,] 0.00000000000000000013781601565299649119 0.000000000000000023588183115457
```

```
## [5,] 0.0000000000000000000315815916732777951 0.00000000000000000912587569299
```

```
## [6,] 0.000000000000000000042477903440459083 0.000000000000000003186095815051
```

[illegible]

Zadanie 5

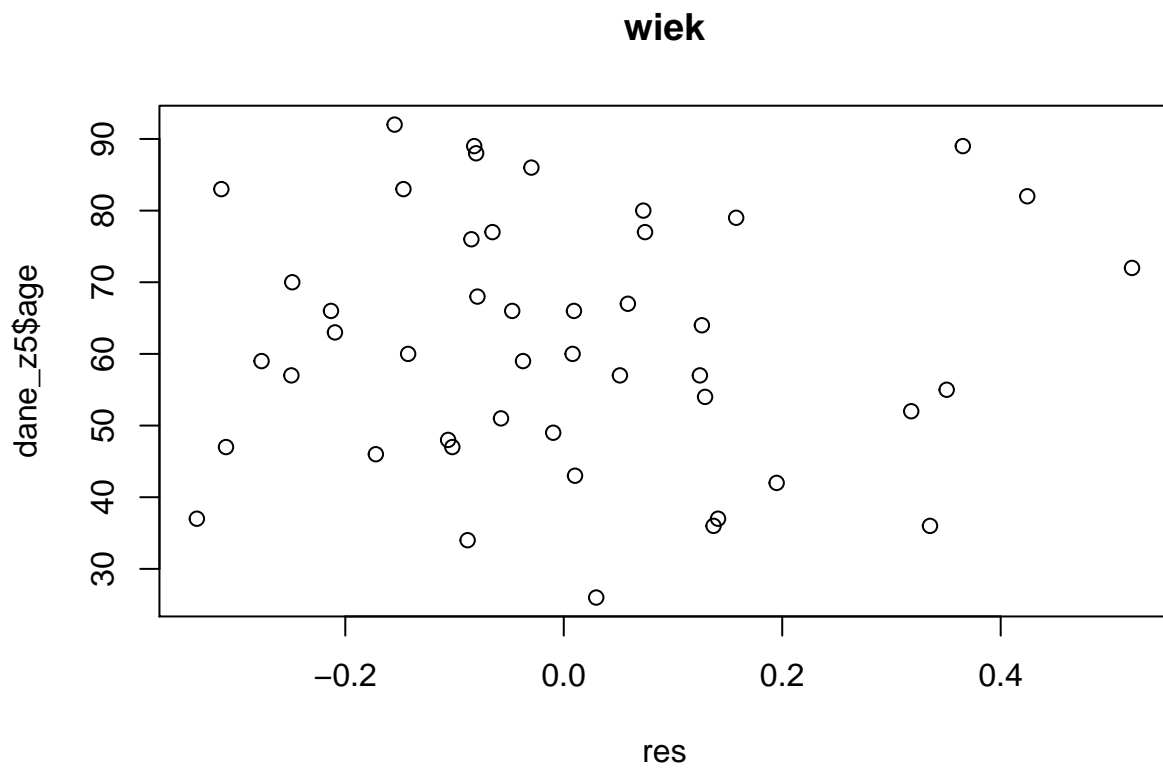
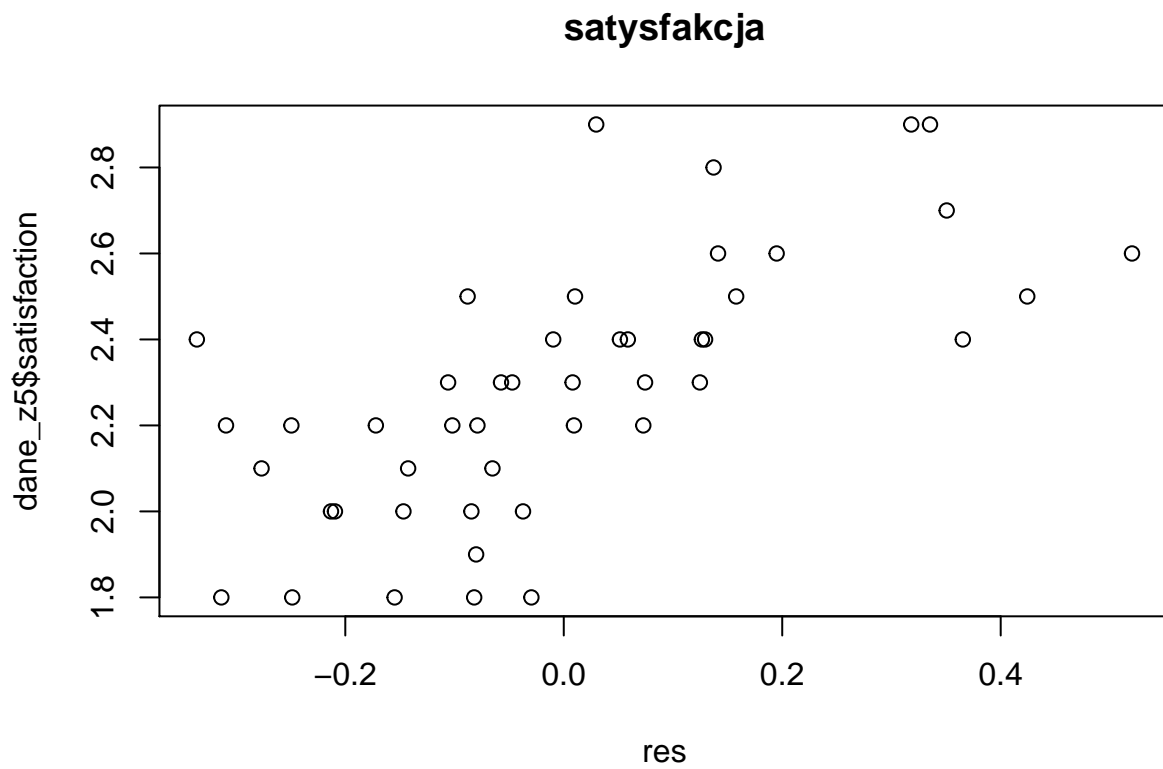
W zadaniu 5 analizujemy dane ‘CH06PR15’. Tworzymy regresje wieloraką i patrzymy na ‘summary’ i wyciągamy wnioski. Odrzucamy hipotezę zerową, ponieważ statystyka F jest duża, a p-wartość bardzo mała. Co najmniej jedna z bet jest różna od 0.

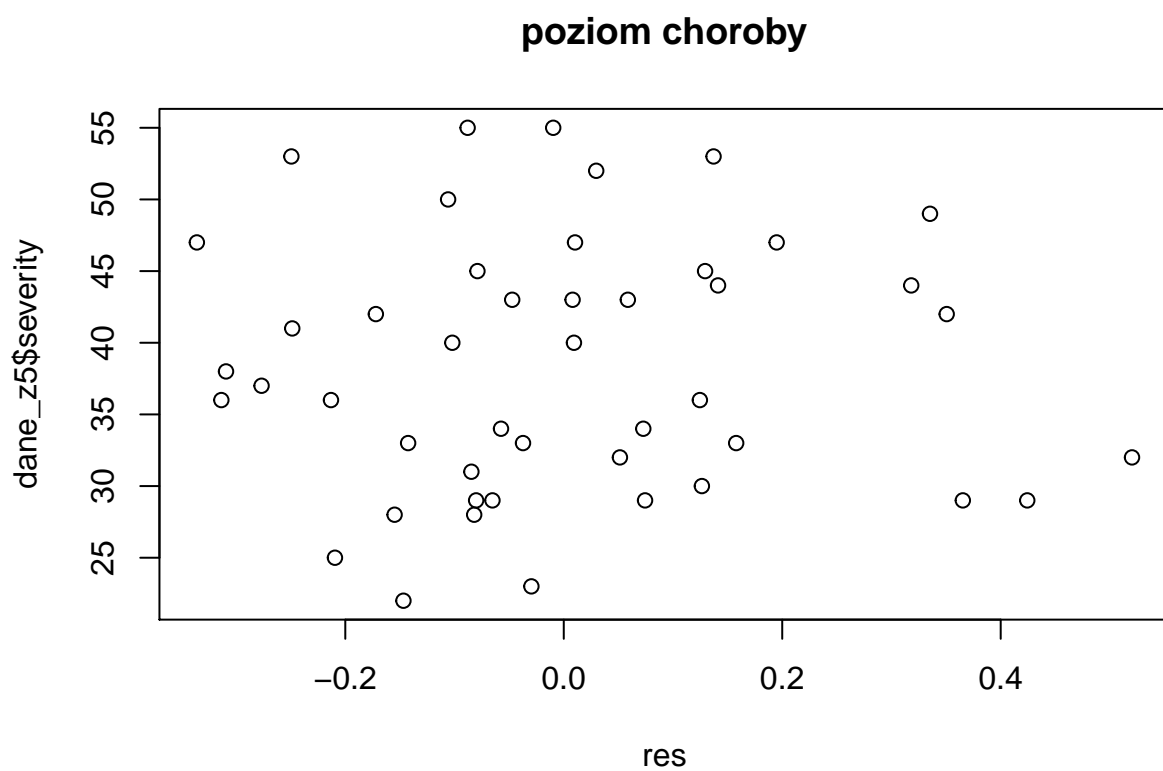
```
##           2.5 %           97.5 %
## age      -0.01209411  0.0003730895
## severity -0.00974994  0.0136060385
## anxiety   0.01146717  0.0488283055

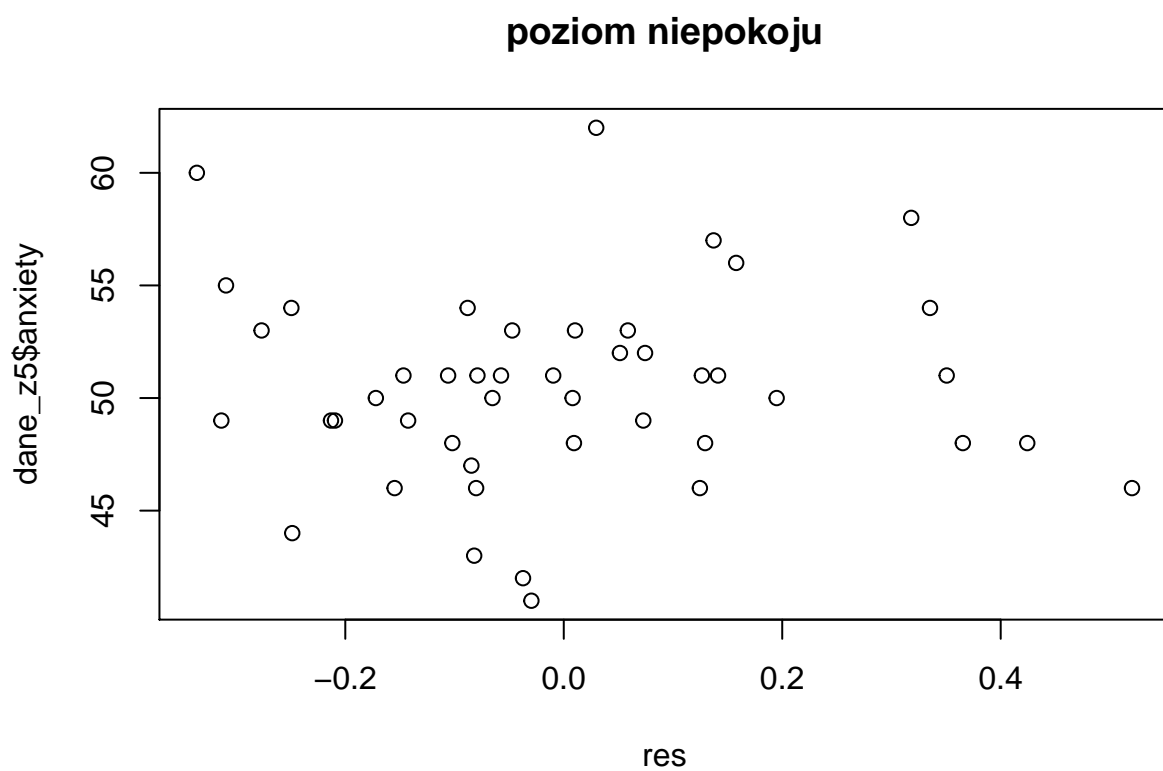
##           Estimate Std. Error   t value    Pr(>|t|)
## age      -0.005860509  0.003088873 -1.8972967  0.064678127
## severity  0.001928049  0.005786678  0.3331876  0.740650314
## anxiety   0.030147737  0.009256597  3.2568922  0.002232272
```

W zadaniu 6 patrzymy na odpowiednie wartości w tabelach i odczytujemy T-statystyki oraz p-wartości dla każdej ze zmiennych. Liczba stopni swobody wynosi 42, ponieważ obserwacji jest 46 a zmiennych objaśniających jest 3 oraz Intercept. Zatem $42 = 46 - 4$.

Zadanie 7

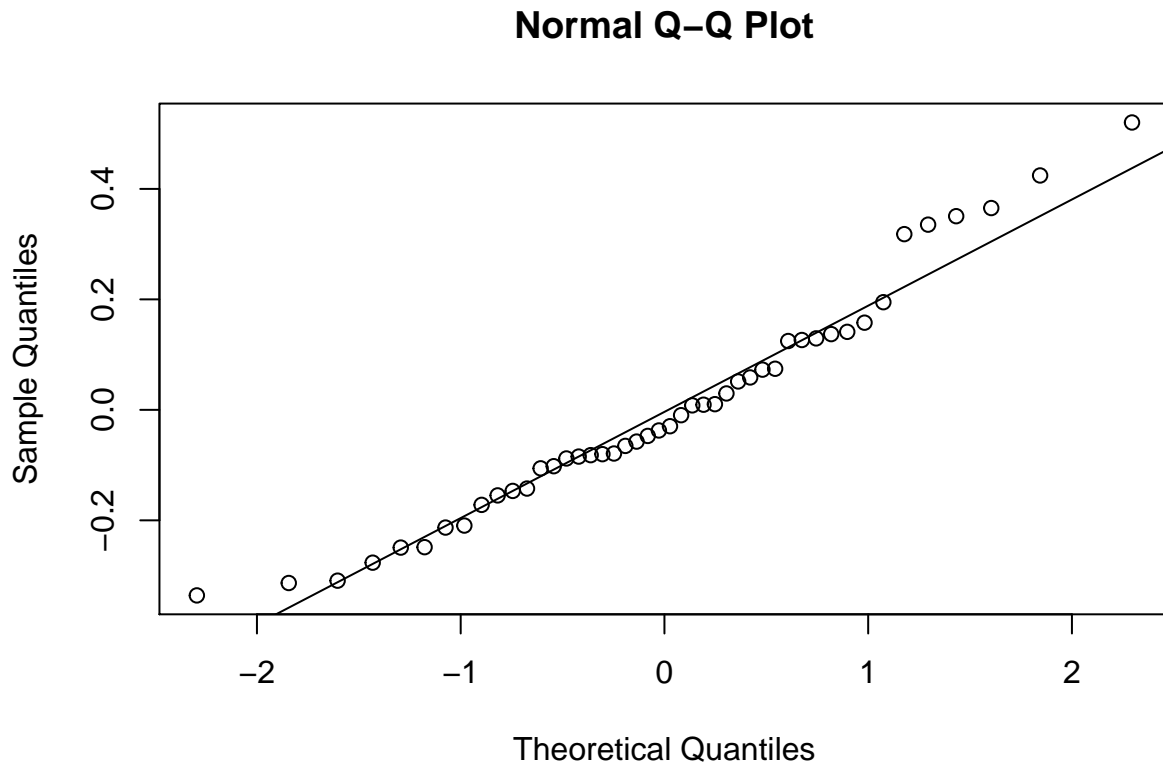






W zadaniu 7 rysujemy wykresy residuów względem Y oraz względem każdej zmiennej z osobna. Dla e vs Y wykres ma strukturę liniową, dla pozostałych nie ma struktury.

Zadanie 8



```
##  
## Shapiro-Wilk normality test  
##  
## data: regres_z5$residuals  
## W = 0.96286, p-value = 0.1481
```

W zadaniu 8 porównujemy dwie metody sprawdzające czy residua pochodzą z rozkładu normalnego. Z testu Shapiro-Wilka wynika, że na poziomie istotności 0.05 nie odrzucamy hipotezy zerowej ($p\text{-wartosc} = 0.1481$) i residua pochodzą z rozkładu normalnego, co zgadza się z wykresem qnorm.