

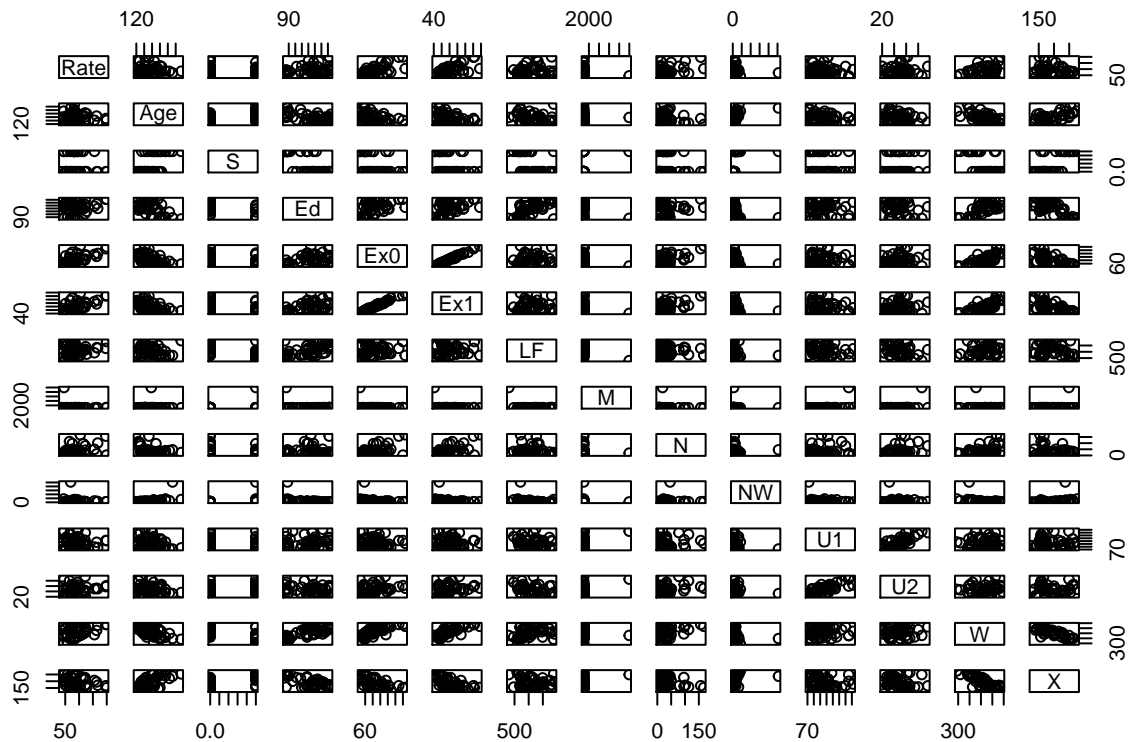
Raport 4

Romana Żmuda

20 05 2020

Zadanie 1

Wykres rozrzutu dla wszystkich par ze zbioru crime:



Zadanie 2

Macierz współczynników korelacji próbkowych Pearsona dla zmiennych z pliku crime:

```
cor(crime, method = "pearson")
```

```
##           Rate           Age           S           Ed           Ex0           Ex1
## Rate  1.00000000 -0.08951992 -0.09087073  0.32310629  0.68761660  0.66674511
## Age  -0.08951990  1.00000000  0.58435534 -0.53023964 -0.50696949 -0.51317336
## S    -0.09087073  0.584355338  1.00000000 -0.70274132 -0.37354720 -0.37616753
```

```

## Ed    0.32310629 -0.530239642 -0.70274132  1.00000000  0.48326144  0.49940958
## Ex0    0.68761660 -0.506969490 -0.37354720  0.48326144  1.00000000  0.99359772
## Ex1    0.66674511 -0.513173356 -0.37616753  0.49940958  0.99359772  1.00000000
## LF     0.18913745 -0.160948824 -0.50546948  0.56117795  0.12217106  0.10634960
## M     -0.16902237  0.004461682  0.19845402 -0.22557612 -0.19533933 -0.20932621
## N      0.33749589 -0.280637618 -0.04991832 -0.01722740  0.52608277  0.51378940
## NW    -0.06372231  0.457234083  0.39708823 -0.38365025 -0.18123271 -0.18655758
## U1    -0.05067083 -0.224380599 -0.17241931  0.01810345 -0.04277496 -0.05171199
## U2     0.17706371 -0.244843390  0.07169289 -0.21568155  0.18564957  0.16922422
## W      0.44149947 -0.670055056 -0.63694543  0.73599704  0.78724481  0.79426205
## X     -0.18585829  0.629962825  0.67496961 -0.69788198 -0.63228425 -0.65072345
##          LF          M          N          NW          U1          U2
## Rate  0.1891374 -0.169022371  0.33749589 -0.06372231 -0.05067083  0.17706371
## Age   -0.1609488  0.004461682 -0.28063762  0.45723408 -0.22438060 -0.24484339
## S     -0.5054695  0.198454020 -0.04991832  0.39708823 -0.17241931  0.07169289
## Ed    0.5611780 -0.225576121 -0.01722740 -0.38365025  0.01810345 -0.21568155
## Ex0    0.1221711 -0.195339325  0.52608277 -0.18123271 -0.04277496  0.18564957
## Ex1    0.1063496 -0.209326206  0.51378940 -0.18655758 -0.05171199  0.16922422
## LF     1.0000000 -0.288299113 -0.12367222 -0.22796997 -0.22939968 -0.42076249
## M     -0.2882991  1.000000000 -0.07853513 -0.04361084  0.33547258  0.33600928
## N     -0.1236722 -0.078535135  1.00000000  0.05942204 -0.03811995  0.27042159
## NW    -0.2279700 -0.043610839  0.05942204  1.00000000 -0.21818218 -0.10460448
## U1    -0.2293997  0.335472578 -0.03811995 -0.21818218  1.00000000  0.74592482
## U2    -0.4207625  0.336009279  0.27042159 -0.10460448  0.74592482  1.00000000
## W      0.2946323 -0.101849407  0.30826271 -0.34330695  0.04485720  0.09207166
## X     -0.1923744  0.213000458 -0.12466424  0.31281286 -0.11491193 -0.05777706
##          W          X
## Rate  0.44149947 -0.18585829
## Age   -0.67005506  0.62996283
## S     -0.63694543  0.67496961
## Ed    0.73599704 -0.69788198
## Ex0    0.78724481 -0.63228425
## Ex1    0.79426205 -0.65072345
## LF     0.29463231 -0.19237444
## M     -0.10184941  0.21300046
## N      0.30826271 -0.12466424
## NW    -0.34330695  0.31281286
## U1     0.04485720 -0.11491193
## U2     0.09207166 -0.05777706
## W      1.00000000 -0.84417172
## X     -0.84417172  1.00000000

```

Zadanie 3

- Na zmienną objaśnianą Rate największy wpływ mają zmienne informujące o wydatkach policyjnych, a więc zmienne Ex0 oraz Ex1
- Problem współliniowości to istnienie korelacji między zmiennymi, której wartość jest bliska 1, takie zjawisko występuje między Ex0 i Ex1

Zadanie 4

Poniżej konstruuje model regresji dla zmiennej objaśnianej Rate na całym zbiorze:

```
regresja<-lm(Rate~., data = crime)
```

Równanie regresji to podanie odpowiednik współczynników β oraz współczynnik wyrazu wolnego ϵ w naszym odpowiednio są to wartości w kolumnie Estimate, gdzie Intercept to wyraz wolny, a wartości poniżej to współczynniki β .

```
##
## Call:
## lm(formula = Rate ~ ., data = crime)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -38.873 -13.277  -0.191   8.858  57.838
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.146e+02  1.193e+02  -4.312 0.000138 ***
## Age          1.078e+00  4.377e-01   2.463 0.019164 *
## S            5.514e+00  1.329e+01   0.415 0.680945
## Ed           1.512e+00  6.708e-01   2.255 0.030915 *
## Ex0          1.877e+00  1.139e+00   1.648 0.108742
## Ex1         -8.789e-01  1.233e+00  -0.713 0.480862
## LF           5.758e-02  1.365e-01   0.422 0.675781
## M           -3.304e-03  3.277e-03  -1.008 0.320665
## N           -8.805e-02  1.254e-01  -0.702 0.487512
## NW          -2.112e-03  1.116e-02  -0.189 0.851005
## U1          -2.764e-01  3.913e-01  -0.706 0.484937
## U2           1.769e+00  8.995e-01   1.967 0.057630 .
## W           7.899e-02  1.062e-01   0.744 0.462435
## X           5.463e-01  2.067e-01   2.643 0.012480 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23.59 on 33 degrees of freedom
## Multiple R-squared:  0.733, Adjusted R-squared:  0.6279
## F-statistic:  6.97 on 13 and 33 DF, p-value: 3.197e-06
```

Wyznaczony współczynnik determinacji R^2 :

```
## [1] 0.7330418
```

Wyznaczony współczynnik determinacji $adj R^2$:

```
## [1] 0.6278765
```

Obliczenie prognozowanej przez ten model wartości współczynnika Rate, dla danych (150, 1, 90, 50, 60, 500, 950, 30, 300, 100, 40, 400, 200):

```
## [1] 36.24069
```

Zadanie 5

A)Poniżej konstruuje model regresji dla zmiennej objaśnianej Rate na zmiennych objaśniających $Ex1, X, Ed, Age, U2$:

```
regresja.a<-lm(Rate~Age+Ed+Ex1+U2+X, data = crime)
```

Równanie regresji to podanie odpowiednik współczynników β oraz współczynnik wyrazu wolnego ϵ w naszym odpowiednio są to wartości w kolumnie Estimate, gdzie Intercept to wyraz wolny, a wartości poniżej to współczynniki β .

```
##
## Call:
## lm(formula = Rate ~ Age + Ed + Ex1 + U2 + X, data = crime)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -53.916 -13.952  -3.081   13.355   59.164
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -479.5483    100.3883  -4.777 2.29e-05 ***
## Age           1.1067     0.3819   2.898 0.006000 **
## Ed            1.6652     0.4815   3.458 0.001281 **
## Ex1           1.2689     0.1665   7.621 2.22e-09 ***
## U2            1.1204     0.4722   2.373 0.022426 *
## X             0.5253     0.1469   3.576 0.000913 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23.21 on 41 degrees of freedom
## Multiple R-squared:  0.679, Adjusted R-squared:  0.6399
## F-statistic: 17.35 on 5 and 41 DF,  p-value: 3.37e-09
```

Wyznaczony współczynnik determinacji R^2 :

```
## [1] 0.6790068
```

Wyznaczony współczynnik determinacji $adj R^2$:

```
## [1] 0.6398613
```

Obliczenie prognozowanej przez ten model wartości współczynnika Rate, dla danych (150, 90, 60, 40, 200):

```
## [1] 62.33506
```

B)Poniżej konstruuje model regresji dla zmiennej objaśnianej Rate na zmiennych objaśniających $EX0, LF, M, N, NW$:

```
regresja.b<-lm(Rate~Ex0+LF+M+N+NW, data = crime)
```

Równanie regresji to podanie odpowiednik współczynników β oraz współczynnik wyrazu wolnego ϵ w naszym odpowiednio są to wartości w kolumnie Estimate, gdzie Intercept to wyraz wolny, a wartości poniżej to współczynniki β .

```
##
## Call:
## lm(formula = Rate ~ Ex0 + LF + M + N + NW, data = crime)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -62.048 -16.267   2.798  18.395  53.936
##
```

```
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.443e+01  6.819e+01  -0.798    0.429
## Ex0          9.132e-01  1.808e-01   5.051 9.53e-06 ***
## LF           1.184e-01  1.170e-01   1.012   0.318
## M            1.813e-04  3.519e-03   0.052   0.959
## N           -2.187e-02  1.379e-01  -0.159   0.875
## NW           9.694e-03  1.230e-02   0.788   0.435
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 29.2 on 41 degrees of freedom
## Multiple R-squared:  0.4918, Adjusted R-squared:  0.4298
## F-statistic: 7.936 on 5 and 41 DF,  p-value: 2.658e-05
```

Wyznaczony współczynnik determinacji R^2 :

```
## [1] 0.4918136
```

Wyznaczony współczynnik determinacji $adj R^2$:

```
## [1] 0.4298397
```

Obliczenie prognozowanej przez ten model wartości współczynnika Rate,
dla danych (50, 500, 950, 30, 300):

```
## [1] 52.86884
```

Zadanie 6

Przyjrzyjmy się jeszcze raz wartością skorygowanych współczynników determinacji:

```
## [1] "Wartość adjR^2 dla zad 4 :"
```

```
## [1] 0.6278765
```

```
## [1] "Wartość adjR^2 dla zad 5a :"
```

```
## [1] 0.6398613
```

```
## [1] "Wartość adjR^2 dla zad 5b :"
```

```
## [1] 0.4298397
```

Najlepszym współczynnikiem $adj R^2$ jest ten najbliższy jedności, dlatego za najlepszy model uznajemy ten z zadania 5A. A jej model to:

```
##
## Call:
## lm(formula = Rate ~ Age + Ed + Ex1 + U2 + X, data = crime)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -53.916 -13.952  -3.081   13.355   59.164
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -479.5483    100.3883  -4.777 2.29e-05 ***
## Age           1.1067      0.3819   2.898 0.006000 **
## Ed            1.6652      0.4815   3.458 0.001281 **
```

```
## Ex1          1.2689      0.1665    7.621 2.22e-09 ***
## U2           1.1204      0.4722    2.373 0.022426 *
## X            0.5253      0.1469    3.576 0.000913 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23.21 on 41 degrees of freedom
## Multiple R-squared:  0.679, Adjusted R-squared:  0.6399
## F-statistic: 17.35 on 5 and 41 DF,  p-value: 3.37e-09
```

Zadanie 7

Przypomnijmy sobie wyliczenia dla zmiennych objaśniających z danych z zadania 4 dla modelu z zadania 4, 5a(Najlepszy), 5b:

```
## [1] "Wartość prognozowana dla zmiennych, zadanie 4 :"
```

```
## [1] 36.24069
```

```
## [1] "Wartość prognozowana dla zmiennych, zadanie 5A (Najlepszy) :"
```

```
## [1] 62.33506
```

```
## [1] "Wartość prognozowana dla zmiennych, zadanie 5B :"
```

```
## [1] 52.86884
```

Widać, że prognozy w znacznym stopniu się różnią. Mimo że wartości skorygowanych współczynników determinacji były podobne w zadaniu 4 oraz 5a to prognozy bardzo się różnią.