Sprawozdanie 1

Krzysztof Kotlarz

14 03 2020

# Lista 1.

## Zadanie 2.

model.1 <- lm(dl.milk ~ sex + weight + ml.suppl + mat.weight + mat.height)  
summary(model.1)

##   
## Call:  
## lm(formula = dl.milk ~ sex + weight + ml.suppl + mat.weight +   
## mat.height)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.74201 -0.81173 -0.00926 0.78326 2.52646   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -11.681839 4.361561 -2.678 0.010363 \*   
## sexgirl -0.499532 0.312672 -1.598 0.117284   
## weight 1.349124 0.322450 4.184 0.000135 \*\*\*  
## ml.suppl -0.002233 0.001241 -1.799 0.078829 .   
## mat.weight 0.006212 0.023708 0.262 0.794535   
## mat.height 0.072278 0.030169 2.396 0.020906 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.075 on 44 degrees of freedom  
## Multiple R-squared: 0.5459, Adjusted R-squared: 0.4943   
## F-statistic: 10.58 on 5 and 44 DF, p-value: 1.03e-06

### (a)

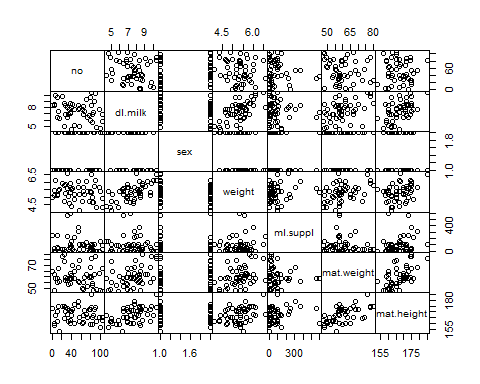
Zmienne istotne: *weight*, *mat.height*

* Przy zmiennej niezależnej *sex* == girl, *dl.milk* zmaleje o -11.68,
* przy wzorście *weight* o 1, *dl.milk* wzrośnie o 1.35,
* przy wzroście *ml.supply* o 1, *dl.milk* zmalaje o .0022,
* przy wzroście *mat.weight* o 1, *dl.milk* wzrośnie o .0062,
* przy wzroście *mat.weight* o 1, *dl.milk* wzrośnie o .072.
* punkt przecięcia osi -11.68

### (b)

Zmienna *sex* jako factor ponieważ ta zmienna przyjmuje tylko 2 poziomy, jest to zmienna jakościowa.

### (c)



Wykres przedstawia korelacje zmiennych “każda z każdą”. Zauważono korelacje pomiedzy zmiennymi:

* *weight* <-> *dl.milk*
* *mat.weight* <-> *mat.height*

Wykonano macierz korelacji w celu potwierdzenia powyższych zależności:

kfm.copy <- kfm  
kfm.copy$sex <- as.numeric(kfm.copy$sex)  
cor(kfm.copy)

## no dl.milk sex weight ml.suppl  
## no 1.00000000 -0.18068364 0.06193530 -0.1649280 -0.12978958  
## dl.milk -0.18068364 1.00000000 -0.29940126 0.6360448 -0.06351955  
## sex 0.06193530 -0.29940126 1.00000000 -0.2200106 -0.07136717  
## weight -0.16492803 0.63604482 -0.22001058 1.0000000 0.12838120  
## ml.suppl -0.12978958 -0.06351955 -0.07136717 0.1283812 1.00000000  
## mat.weight -0.13731160 0.43427002 -0.05303191 0.4081748 -0.07887363  
## mat.height -0.02356765 0.50504203 -0.11776734 0.3867571 0.18230263  
## mat.weight mat.height  
## no -0.13731160 -0.02356765  
## dl.milk 0.43427002 0.50504203  
## sex -0.05303191 -0.11776734  
## weight 0.40817476 0.38675714  
## ml.suppl -0.07887363 0.18230263  
## mat.weight 1.00000000 0.56473304  
## mat.height 0.56473304 1.00000000

Z wykresu zależności parami jak i macierzy korelacji wynikają następujące silnie skorelowane pary:

* *weight* <-> *dl.milk*
* *mat.weight* <-> *mat.height*

### (d)

Przy poziomie istotności = 5%

model.all <- lm(dl.milk ~ ., data = kfm)  
summary(model.all)

##   
## Call:  
## lm(formula = dl.milk ~ ., data = kfm)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.89286 -0.87720 0.06426 0.73663 2.28685   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -11.652909 4.357128 -2.674 0.010542 \*   
## no -0.005522 0.005286 -1.045 0.302010   
## sex2 -0.488757 0.312518 -1.564 0.125164   
## weight 1.311822 0.324088 4.048 0.000212 \*\*\*  
## ml.suppl -0.002432 0.001254 -1.939 0.059077 .   
## mat.weight 0.002453 0.023956 0.102 0.918925   
## mat.height 0.076445 0.030401 2.515 0.015739 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.074 on 43 degrees of freedom  
## Multiple R-squared: 0.5571, Adjusted R-squared: 0.4953   
## F-statistic: 9.015 on 6 and 43 DF, p-value: 2.189e-06

Usunięto nieistotną zmienna *mat.weight*

model.all\_1 <- lm(dl.milk ~ no + sex + weight + ml.suppl + mat.height, data = kfm)  
summary(model.all\_1)

##   
## Call:  
## lm(formula = dl.milk ~ no + sex + weight + ml.suppl + mat.height,   
## data = kfm)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.91656 -0.87688 0.05494 0.73957 2.28175   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -11.818720 3.999250 -2.955 0.005004 \*\*   
## no -0.005604 0.005167 -1.084 0.284057   
## sex2 -0.486724 0.308359 -1.578 0.121630   
## weight 1.320304 0.309779 4.262 0.000105 \*\*\*  
## ml.suppl -0.002466 0.001196 -2.061 0.045201 \*   
## mat.height 0.078083 0.025559 3.055 0.003815 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.062 on 44 degrees of freedom  
## Multiple R-squared: 0.557, Adjusted R-squared: 0.5067   
## F-statistic: 11.06 on 5 and 44 DF, p-value: 6.137e-07

Usunięto nieistotną zmienną *no*

model.all\_2 <- lm(dl.milk ~ sex + weight + ml.suppl + mat.height, data = kfm)  
summary(model.all\_2)

##   
## Call:  
## lm(formula = dl.milk ~ sex + weight + ml.suppl + mat.height,   
## data = kfm)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.77312 -0.81196 -0.00683 0.76988 2.52240   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -12.112571 3.997860 -3.030 0.00405 \*\*   
## sex2 -0.494675 0.308875 -1.602 0.11626   
## weight 1.372524 0.306612 4.476 5.14e-05 \*\*\*  
## ml.suppl -0.002313 0.001190 -1.943 0.05824 .   
## mat.height 0.076363 0.025560 2.988 0.00454 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.064 on 45 degrees of freedom  
## Multiple R-squared: 0.5452, Adjusted R-squared: 0.5047   
## F-statistic: 13.48 on 4 and 45 DF, p-value: 2.658e-07

Usunięto nieistotną zmienną *mat.height*

model.all\_3 <- lm(dl.milk ~ weight + ml.suppl + mat.height, data = kfm)  
summary(model.all\_3)

##   
## Call:  
## lm(formula = dl.milk ~ weight + ml.suppl + mat.height, data = kfm)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.06540 -0.74758 -0.02408 0.67488 2.79882   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -13.064926 4.020073 -3.250 0.00216 \*\*   
## weight 1.464781 0.306231 4.783 1.81e-05 \*\*\*  
## ml.suppl -0.002237 0.001209 -1.850 0.07074 .   
## mat.height 0.077600 0.025979 2.987 0.00451 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.082 on 46 degrees of freedom  
## Multiple R-squared: 0.5192, Adjusted R-squared: 0.4879   
## F-statistic: 16.56 on 3 and 46 DF, p-value: 1.953e-07

Usunięto nieistotną zmienną *sex*

model.all\_4 <- lm(dl.milk ~ weight + mat.height, data = kfm)  
summary(model.all\_4)

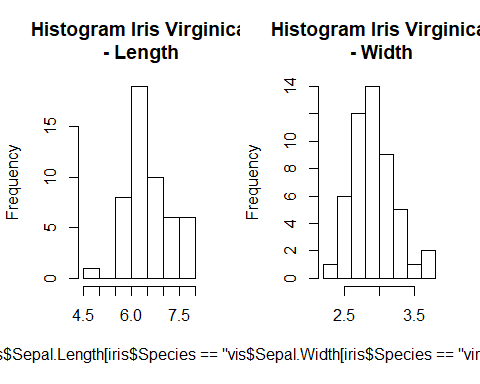
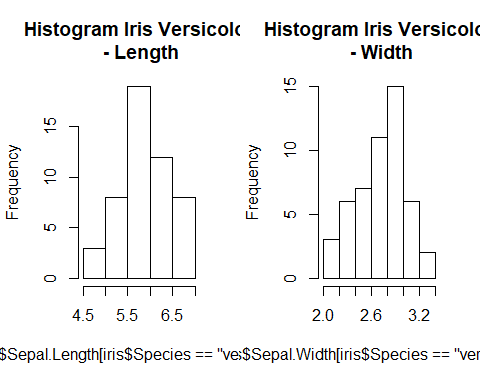
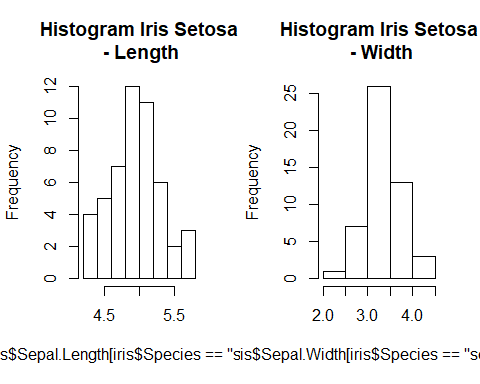
##   
## Call:  
## lm(formula = dl.milk ~ weight + mat.height, data = kfm)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.19598 -0.82149 0.01822 0.75582 2.83375   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -11.92014 4.07325 -2.926 0.00527 \*\*   
## weight 1.42862 0.31338 4.559 3.67e-05 \*\*\*  
## mat.height 0.07063 0.02636 2.680 0.01013 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.109 on 47 degrees of freedom  
## Multiple R-squared: 0.4835, Adjusted R-squared: 0.4615   
## F-statistic: 22 on 2 and 47 DF, p-value: 1.811e-07

Usunięto nieistotną zmienną *ml.supply*. Ostateczny model:

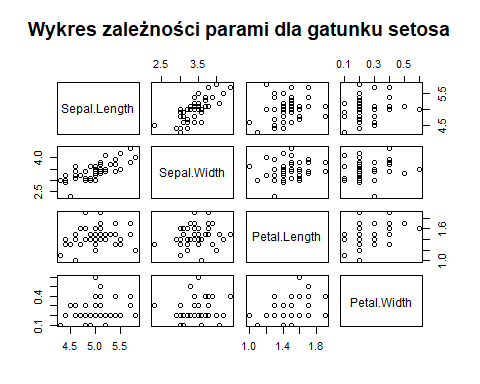
* Przy wzroście *weight* o jednostke, dl.milk wzrośnie o 1.43
* Przy wzroście *mat.weight* o jednostke, dl.milk wzrośnie o .071
* Punkt przecięcia osi OY -11.92

## Zadanie 2.

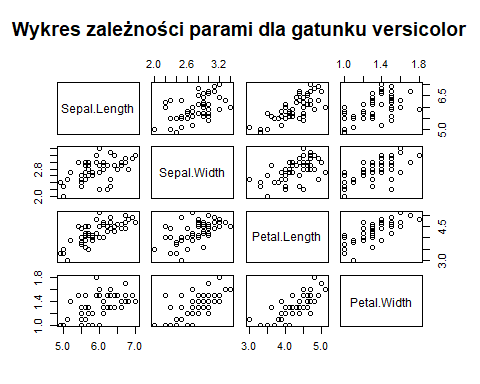
### (a)



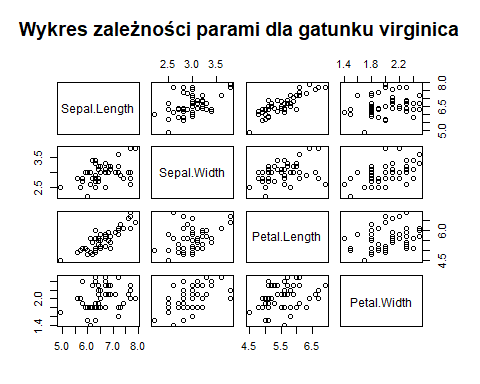
### (b)



Korelacja pomiędzy zmiennymi *Sepal.Length* <-> *Sepal.Length*



Korelacja pomiędzy zmiennymi *Sepal.Length* <-> *Petal.Length* oraz *Sepal.Width* <-> *Petal.Length*



Silna korelacja pomiędzy zmiennymi *Sepal.Length* <-> *Petal.Length*

### (c)

cor(iris[iris$Species=='setosa',c(1,3)])

## Sepal.Length Petal.Length  
## Sepal.Length 1.0000000 0.2671758  
## Petal.Length 0.2671758 1.0000000

Słaba korelacja *Sepal.Length* <-> *Petal.Length* (0.26)

cor(iris[iris$Species=='versicolor',c(1,3)])

## Sepal.Length Petal.Length  
## Sepal.Length 1.000000 0.754049  
## Petal.Length 0.754049 1.000000

Silna korelacja *Sepal.Length* <-> *Petal.Length*(0.75)

cor(iris[iris$Species=='virginica',c(1,3)])

## Sepal.Length Petal.Length  
## Sepal.Length 1.0000000 0.8642247  
## Petal.Length 0.8642247 1.0000000

Silna korelacja *Sepal.Length* <-> *Petal.Length* (0.86)

cor.test(iris$Sepal.Length[iris$Species == 'setosa'], iris$Petal.Length[iris$Species ==  
 'setosa'])

##   
## Pearson's product-moment correlation  
##   
## data: iris$Sepal.Length[iris$Species == "setosa"] and iris$Petal.Length[iris$Species == "setosa"]  
## t = 1.9209, df = 48, p-value = 0.0607  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.01206954 0.50776233  
## sample estimates:  
## cor   
## 0.2671758

Współczynnik korelacji *Sepal.Length* i *Petal.Lengt* o wartości 0.27 gatunku *setosa* nie jest istotny na poziomie ostotności = 5%

cor.test(iris$Sepal.Length[iris$Species == 'versicolor'], iris$Petal.Length[iris$Species ==  
 'versicolor'])

##   
## Pearson's product-moment correlation  
##   
## data: iris$Sepal.Length[iris$Species == "versicolor"] and iris$Petal.Length[iris$Species == "versicolor"]  
## t = 7.9538, df = 48, p-value = 2.586e-10  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.6020680 0.8532995  
## sample estimates:  
## cor   
## 0.754049

Zmienne *Sepal.Length* i *Petal.Length* gatunku *versicolor* są istotnie skorelowane ze współczynnikiem korelacji na poziomie 0.75 i wartością p 2.586e-10

cor.test(iris$Sepal.Length[iris$Species == 'virginica'], iris$Petal.Length[iris$Species ==  
 'virginica'])

##   
## Pearson's product-moment correlation  
##   
## data: iris$Sepal.Length[iris$Species == "virginica"] and iris$Petal.Length[iris$Species == "virginica"]  
## t = 11.901, df = 48, p-value = 6.298e-16  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.7714542 0.9210172  
## sample estimates:  
## cor   
## 0.8642247

Zmienne Sepal.Length i Petal.Lengt gatunku *virginica* są istotnie skorelowane ze współczynnikiem korelacji na poziomie 0.86 i wartością p 6.298e-16

cor(iris[iris$Species=='setosa',1:4])

## Sepal.Length Sepal.Width Petal.Length Petal.Width  
## Sepal.Length 1.0000000 0.7425467 0.2671758 0.2780984  
## Sepal.Width 0.7425467 1.0000000 0.1777000 0.2327520  
## Petal.Length 0.2671758 0.1777000 1.0000000 0.3316300  
## Petal.Width 0.2780984 0.2327520 0.3316300 1.0000000

Dla gatunku *setosa* najsilniej skorelowane są zmienne *Sepal.Width* <-> *Sepal.Width*

cor(iris[iris$Species=='versicolor',1:4])

## Sepal.Length Sepal.Width Petal.Length Petal.Width  
## Sepal.Length 1.0000000 0.5259107 0.7540490 0.5464611  
## Sepal.Width 0.5259107 1.0000000 0.5605221 0.6639987  
## Petal.Length 0.7540490 0.5605221 1.0000000 0.7866681  
## Petal.Width 0.5464611 0.6639987 0.7866681 1.0000000

Dla gatunku *versicolor* najsilniej skorelowane są zmienne *Petal.Length* <-> *Petal.Width*

cor(iris[iris$Species=='virginica',1:4])

## Sepal.Length Sepal.Width Petal.Length Petal.Width  
## Sepal.Length 1.0000000 0.4572278 0.8642247 0.2811077  
## Sepal.Width 0.4572278 1.0000000 0.4010446 0.5377280  
## Petal.Length 0.8642247 0.4010446 1.0000000 0.3221082  
## Petal.Width 0.2811077 0.5377280 0.3221082 1.0000000

Dla gatunku *virginica* najsilniej skorelowane są zmienne *Petal.Length* <-> *Sepal.Length*

### (d)

iris.setosa <- iris[iris$Species=='setosa',1:4]

attach(iris.setosa)  
model.all <- lm(Petal.Length ~ Sepal.Length + Sepal.Width + Petal.Width, data = iris.setosa)  
summary(model.all)

##   
## Call:  
## lm(formula = Petal.Length ~ Sepal.Length + Sepal.Width + Petal.Width,   
## data = iris.setosa)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.38868 -0.07905 0.00632 0.10095 0.48238   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.86547 0.34331 2.521 0.0152 \*  
## Sepal.Length 0.11606 0.10162 1.142 0.2594   
## Sepal.Width -0.02865 0.09334 -0.307 0.7602   
## Petal.Width 0.46253 0.23410 1.976 0.0542 .  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1657 on 46 degrees of freedom  
## Multiple R-squared: 0.1449, Adjusted R-squared: 0.08914   
## F-statistic: 2.598 on 3 and 46 DF, p-value: 0.06356

model.1 <- lm(Petal.Length ~ Sepal.Length + Petal.Width, data = iris.setosa)  
summary(model.1)

##   
## Call:  
## lm(formula = Petal.Length ~ Sepal.Length + Petal.Width, data = iris.setosa)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.40293 -0.08425 0.00073 0.09455 0.47839   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.88128 0.33614 2.622 0.0117 \*  
## Sepal.Length 0.09342 0.06925 1.349 0.1838   
## Petal.Width 0.45959 0.23164 1.984 0.0531 .  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1641 on 47 degrees of freedom  
## Multiple R-squared: 0.1432, Adjusted R-squared: 0.1067   
## F-statistic: 3.926 on 2 and 47 DF, p-value: 0.0265

model.2 <- lm(Petal.Length ~ Petal.Width, data = iris.setosa)  
summary(model.2)

##   
## Call:  
## lm(formula = Petal.Length ~ Petal.Width, data = iris.setosa)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.43686 -0.09151 -0.03686 0.09018 0.46314   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.32756 0.05996 22.141 <2e-16 \*\*\*  
## Petal.Width 0.54649 0.22439 2.435 0.0186 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1655 on 48 degrees of freedom  
## Multiple R-squared: 0.11, Adjusted R-squared: 0.09144   
## F-statistic: 5.931 on 1 and 48 DF, p-value: 0.01864

* Przy wzroście wartości Petal.Width o 1, zmienna Petal.Length wzrośnie o 0.54
* Punkt przecięcia osi OY 1.33

# Lista 2.

## Zadanie 3.

model.1 <- lm(stack.loss ~ Water.Temp + Acid.Conc.)  
summary(model.1)

##   
## Call:  
## lm(formula = stack.loss ~ Water.Temp + Acid.Conc.)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -7.8197 -2.8600 -0.6656 3.1235 8.3295   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -51.2361 18.5374 -2.764 0.0128 \*   
## Water.Temp 2.7320 0.3949 6.919 1.82e-06 \*\*\*  
## Acid.Conc. 0.1290 0.2329 0.554 0.5866   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.137 on 18 degrees of freedom  
## Multiple R-squared: 0.7704, Adjusted R-squared: 0.7449   
## F-statistic: 30.2 on 2 and 18 DF, p-value: 1.772e-06

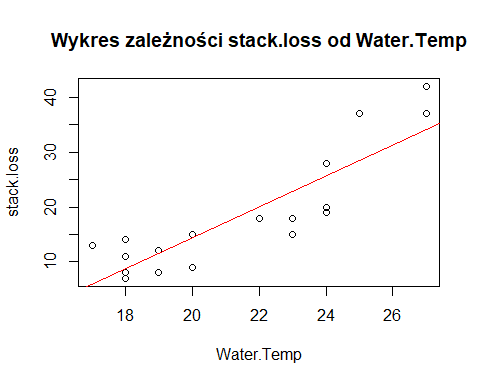
W modelu powinna się znaleźć jedynie zmienna *Water.Temp*, zmienna *Acid.Conc.* jest nieisttona na poziomie a 5%

model.2 <- lm(stack.loss ~ Water.Temp)  
summary(model.2)

##   
## Call:  
## lm(formula = stack.loss ~ Water.Temp)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -7.8904 -3.6206 0.3794 2.8398 8.4747   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -41.9109 7.6056 -5.511 2.58e-05 \*\*\*  
## Water.Temp 2.8174 0.3567 7.898 2.03e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.043 on 19 degrees of freedom  
## Multiple R-squared: 0.7665, Adjusted R-squared: 0.7542   
## F-statistic: 62.37 on 1 and 19 DF, p-value: 2.028e-07

Po usunięciu nieostotnej zmiennej, wyraz wolny zyskał wyższą istotność. Współczynnik dopasowania R2 = 0.75

### (ii)



Brak potencjanych punktów o dużej dzwigni.

### (v, vi)

reszty.stud <- rstudent(model.2)  
outlier <- reszty.stud[abs(reszty.stud) > 2]

Brak obserwacji odstających wyznaczonych za pomocą reszt studentyzowanych dla rstudent > 2.

### (vii)

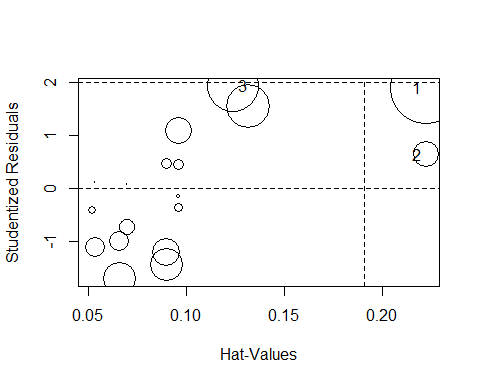
outlierTest(model.2)

## No Studentized residuals with Bonferroni p < 0.05  
## Largest |rstudent|:  
## rstudent unadjusted p-value Bonferroni p  
## 3 1.917911 0.071129 NA

outlierTest nie również nie wykrywa obserwacji odstających na poziomie 5%

### (viii)

influencePlot(model.2)



## StudRes Hat CookD  
## 1 1.8759536 0.2221163 0.44361619  
## 2 0.6282651 0.2221163 0.05820793  
## 3 1.9179109 0.1239276 0.22802446

influencePlot wskazuje dla CoocD = n/21 = 0.19: obserwacje 1 i 3

### (x)

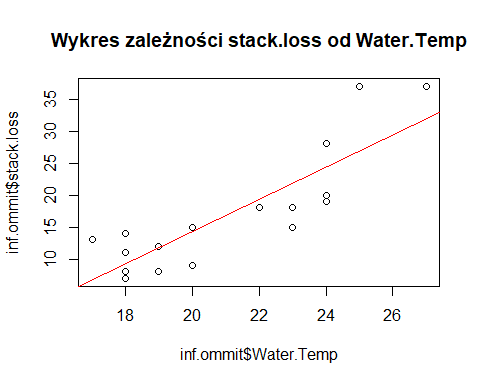
## Influence measures of  
## lm(formula = stack.loss ~ Water.Temp) :  
##   
## dfb.1\_ dfb.Wt.T dffit cov.r cook.d hat inf  
## 1 -0.8120 0.88850 1.0024 1.002 0.443616 0.2221 \*  
## 2 -0.2719 0.29756 0.3357 1.372 0.058208 0.2221 \*  
## 3 -0.4954 0.56604 0.7213 0.877 0.228024 0.1239   
## 4 -0.0840 0.10049 0.1466 1.195 0.011205 0.0898   
## 5 0.0135 -0.02713 -0.0964 1.153 0.004859 0.0517   
## 6 0.1057 -0.13991 -0.2663 1.070 0.035440 0.0658   
## 7 0.2579 -0.30853 -0.4500 0.986 0.095958 0.0898   
## 8 0.2161 -0.25850 -0.3771 1.050 0.069480 0.0898   
## 9 0.1787 -0.23659 -0.4503 0.886 0.092259 0.0658   
## 10 0.2842 -0.25073 0.3540 1.084 0.062041 0.0956   
## 11 0.2842 -0.25073 0.3540 1.084 0.062041 0.0956   
## 12 0.5280 -0.48072 0.6018 1.000 0.168744 0.1316   
## 13 0.1170 -0.10323 0.1457 1.205 0.011087 0.0956   
## 14 0.0140 -0.01167 0.0208 1.197 0.000228 0.0696   
## 15 -0.0426 0.03756 -0.0530 1.228 0.001482 0.0956   
## 16 -0.0959 0.08458 -0.1194 1.214 0.007469 0.0956   
## 17 -0.1359 0.11299 -0.2011 1.129 0.020720 0.0696   
## 18 -0.1359 0.11299 -0.2011 1.129 0.020720 0.0696   
## 19 -0.1241 0.08886 -0.2656 1.030 0.034813 0.0536   
## 20 0.0124 -0.00888 0.0265 1.176 0.000372 0.0536   
## 21 0.0124 -0.00888 0.0265 1.176 0.000372 0.0536

Influence measures wskazuje obserwacje 1 oraz 2 jako wpływowe. Po porównaniu wyników z influencePlot wybrano oberwacje 1 jako wpływową.

### (xi)

inf.ommit <- stackloss[-c(1),]  
model.3 <- lm(inf.ommit$stack.loss~inf.ommit$Water.Temp)

plot(inf.ommit$stack.loss~inf.ommit$Water.Temp, main = 'Wykres zależności stack.loss od Water.Temp')  
abline(model.3, col= 'red')



summary(model.3)

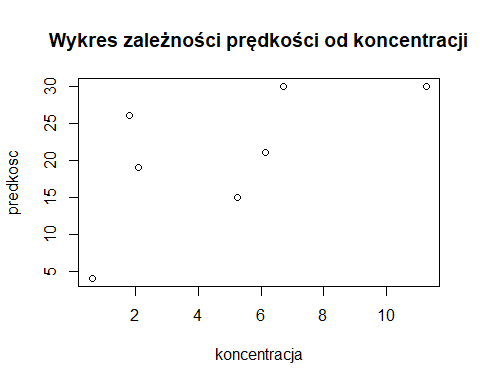
##   
## Call:  
## lm(formula = inf.ommit$stack.loss ~ inf.ommit$Water.Temp)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6.8431 -3.7843 -0.5049 3.9167 10.1176   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -36.1078 7.7873 -4.637 0.000205 \*\*\*  
## inf.ommit$Water.Temp 2.5196 0.3709 6.793 2.32e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.738 on 18 degrees of freedom  
## Multiple R-squared: 0.7194, Adjusted R-squared: 0.7038   
## F-statistic: 46.15 on 1 and 18 DF, p-value: 2.315e-06

Ostateczny wspołczynnik dopasowania modelu po usunięciu obserwacji odstających R2 = 0.7

## Zadanie 4.

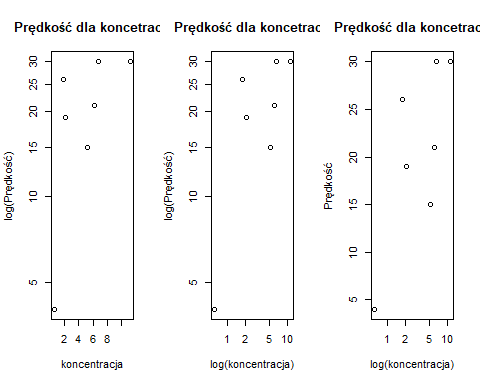
dane <-  
 read.table('http://theta.edu.pl/wp-content/uploads/2018/03/DANE\_predkosc\_reakcji.txt')  
  
dane <- rename(dane, predkosc = V1, koncentracja = V2)

### (i)

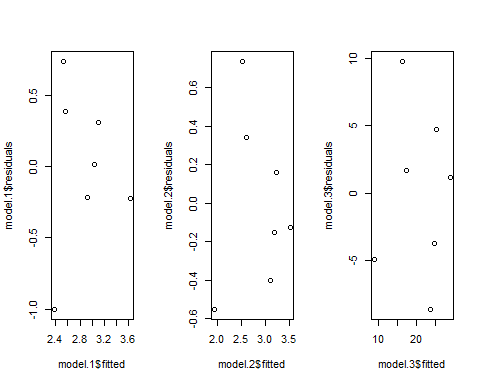


Przedstawiona zależność nie ma charakteru liniowego

### (ii)



Żadna z powyższych transformacji nie prowadzi do uzyskania liniowego charakteru. Wykres Prędkość~log(koncentracja), najbardziej zbliżony do charakteru liniowego.



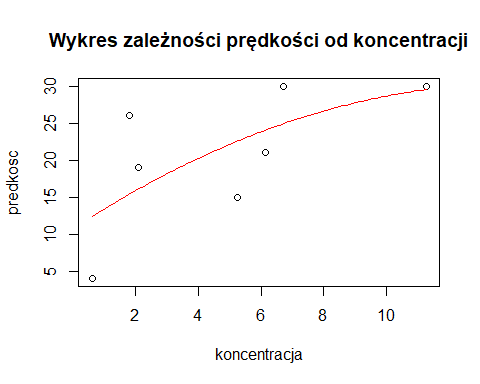
Losowy charakter reszt dla modelu 1, 2 i 3. Dla modelu 2 reszty skupiają sie wokół 0, ten model sugeruje dobre dopasowanie.

### (iii)

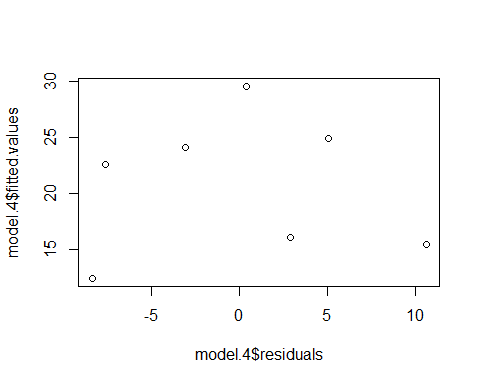
model.4 <- lm(predkosc ~ koncentracja + I(koncentracja ^ 2))  
summary(model.4)

##   
## Call:  
## lm(formula = predkosc ~ koncentracja + I(koncentracja^2))  
##   
## Residuals:  
## 1 2 3 4 5 6 7   
## -7.5878 5.0674 0.4279 -8.3738 10.6267 -3.0900 2.9296   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 10.69371 7.90800 1.352 0.248  
## koncentracja 2.77768 3.27419 0.848 0.444  
## I(koncentracja^2) -0.09782 0.27230 -0.359 0.738  
##   
## Residual standard error: 8.435 on 4 degrees of freedom  
## Multiple R-squared: 0.4478, Adjusted R-squared: 0.1717   
## F-statistic: 1.622 on 2 and 4 DF, p-value: 0.3049

### (iv)



Wszystkie wspołczynniki są nieistotne statystycznie ,współczynnik dopasowania R2 = 0.17



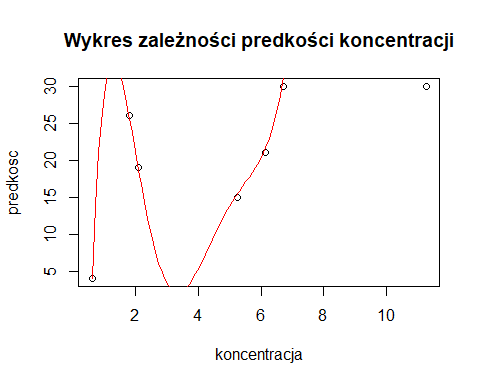
Charakter reszt jest losowy, lecz nie skupiają sie wokół zera. Model nie jest odpowiedni do opisu danych.

### (v)

model.5 <- lm(  
 predkosc ~ koncentracja + I(koncentracja ^ 2) + I(koncentracja ^ 3) + I(koncentracja ^ 4) + I(koncentracja ^ 5) + I(koncentracja ^ 6)  
)  
summary(model.5)

##   
## Call:  
## lm(formula = predkosc ~ koncentracja + I(koncentracja^2) + I(koncentracja^3) +   
## I(koncentracja^4) + I(koncentracja^5) + I(koncentracja^6))  
##   
## Residuals:  
## ALL 7 residuals are 0: no residual degrees of freedom!  
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -120.76838 NA NA NA  
## koncentracja 325.45467 NA NA NA  
## I(koncentracja^2) -246.63164 NA NA NA  
## I(koncentracja^3) 83.77377 NA NA NA  
## I(koncentracja^4) -14.25726 NA NA NA  
## I(koncentracja^5) 1.18499 NA NA NA  
## I(koncentracja^6) -0.03785 NA NA NA  
##   
## Residual standard error: NaN on 0 degrees of freedom  
## Multiple R-squared: 1, Adjusted R-squared: NaN   
## F-statistic: NaN on 6 and 0 DF, p-value: NA

### (vi)



Wspołczynnik dopasowania = 1. Nie jest to model odpowiedni, gdyż dopasowana krzywa jest dopasowana do wszystkich punktów zbioru, nie ma ona wartości predykcyjnej dla przyszłych obserwacji.