# Modele regresji i ich zastosowania
# Lista 1 i 2

Jan Solarz
243889

18 marca 2021

## Spis treści

**1 Zadania labolatoryjne** **1**
- 1.1 Zadanie 1 . . . 1
- 1.2 Zadanie 2 . . . 4
- 1.3 Zadanie 3 . . . 5
- 1.4 Zadanie 4 . . . 6
- 1.5 Zadanie 5 . . . 7
- 1.6 Zadanie 6 . . . 8
- 1.7 Zadanie 7 . . . 8
- 1.8 Zadanie 8 . . . 8
- 1.9 Zadanie 9 . . . 11
- 1.10 Zadanie 10 . . . 13

**2 Zadania teoretyczne** **17**

# 1 Zadania labolatoryjne

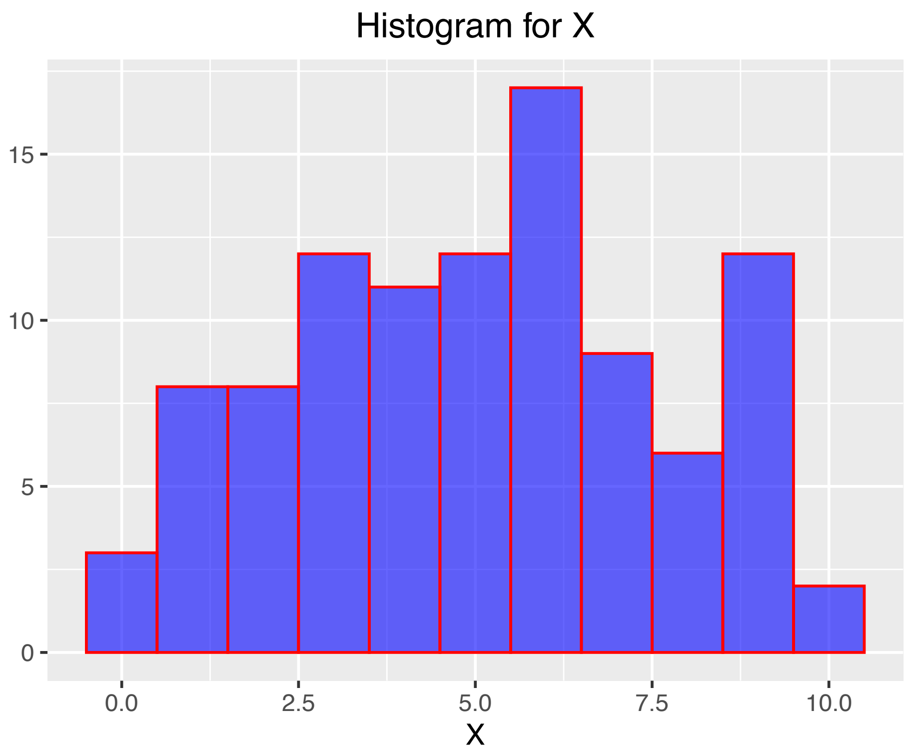
## 1.1 Zadanie 1

Analizy zaczynamy od zapoznania się z plikiem zawierającym zmienne $x$ i $y$ w stu rekordach.

```
lab1 <- read.delim2("~/Downloads/lab1.txt")
library(ggplot2)
attach(lab1)

my.summary <- function(x)
{
  wskazniki<- c(min(x),quantile(x,0.25), median(x), mean(x), quantile(x,0.75), max(x), c
  names(wskazniki) <- c("minimum", "Q1", "mediana", "średnia", "Q3", "maksimum","odch_st
  return(wskazniki)
}
sapply(lab1 , function(x) my.summary(as.numeric(x)) )
```

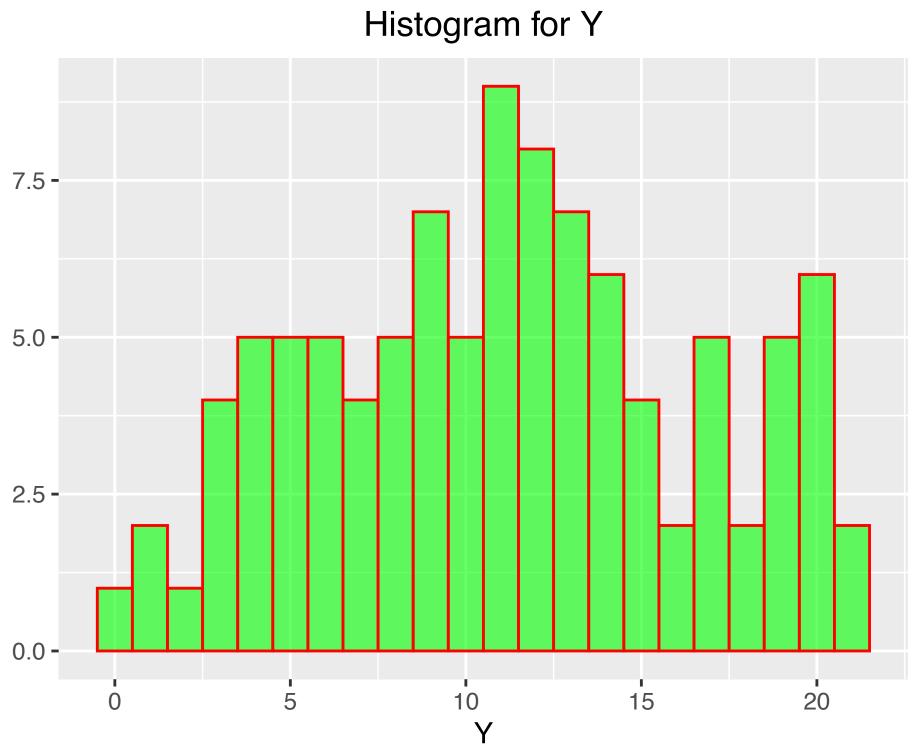
```
##                      x         y
## minimum      0.105200  0.231600
## Q1           3.016625  7.329675
## mediana      5.235750 11.408750
## średnia      5.020899 11.135093
## Q3           6.866625 14.966675
## maksimum     9.765800 21.009400
## odch_stand. 2.589414  5.367752
## wariancja    6.705065 28.812766
```

```
par(mfrow=c(1,2))
qplot(x,
      binwidth = 1,
      main = "Histogram for X",
      xlab = "X",
      fill=I("blue"),
      col=I("red"),
      alpha=I(.6))+
  theme(plot.title = element_text(hjust = 0.5))
```

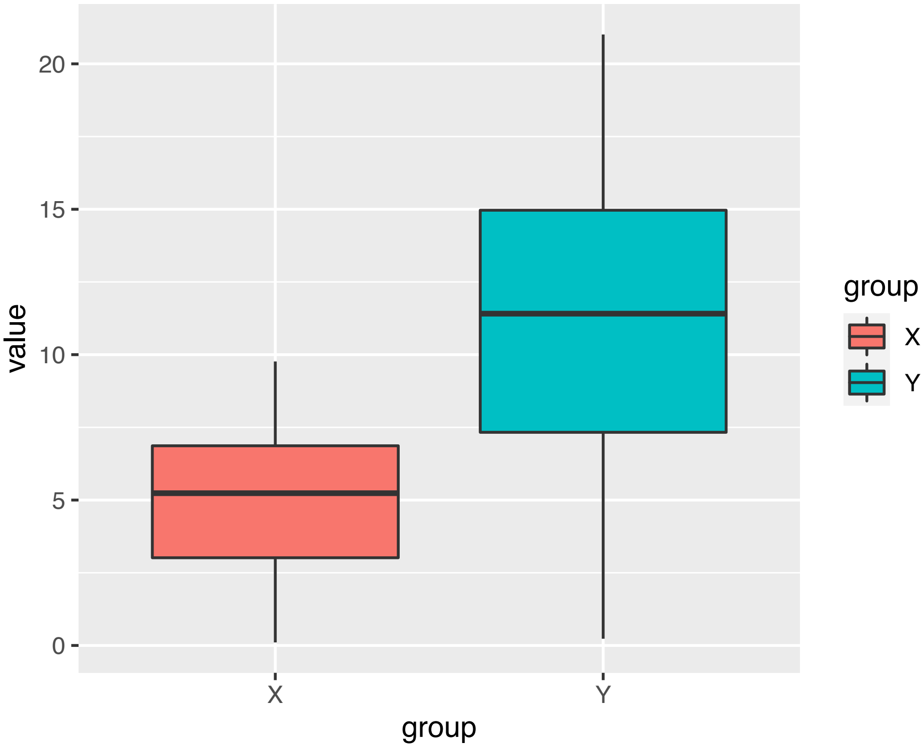



```
qplot(y,
      binwidth = 1,
      main = "Histogram for Y",
      xlab = "Y",
      fill=I("green"),
      col=I("red"),
```

```
      alpha=I(.6))+
  theme(plot.title = element_text(hjust = 0.5))
```




```
par(mfrow=c(1,1))
a = data.frame(group = "X", value = x)
b = data.frame(group = "Y", value = y)

plot.data = rbind(a,b)

ggplot(plot.data, aes(x=group, y=value, fill=group)) +
  geom_boxplot()
```

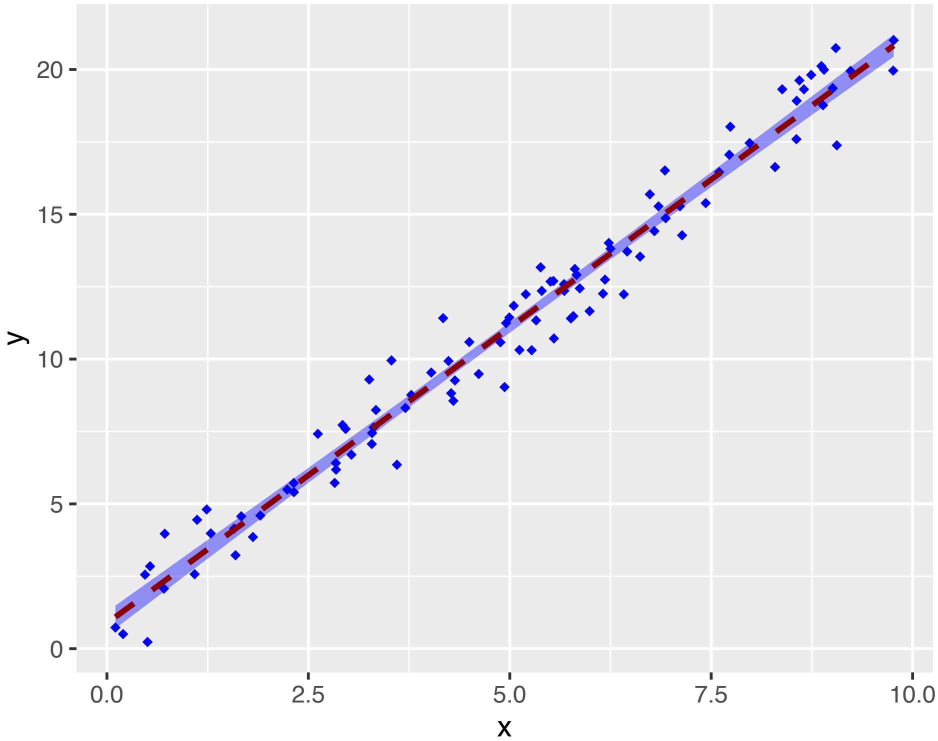



Rysunek 1: Boxplot dla x i y

Wskaźniki i wykorzystane wykresy pokazują że dane z $y$ są dużo bardziej zróżnicowane od $x$. Mamy rozstęp danych odpowiednio 10 i 21 jednostek. Histogramy i boxploty wskazują na wyraźne grupowanie się liczb w $x$ w mniejszą ilość klas niż w $y$.

## 1.2 Zadanie 2

```
ggplot(lab1, aes(x = x, y = y)) +
  geom_point(shape=18, color="blue")+
  geom_smooth(method=lm,  linetype="dashed",
              color="darkred", fill="blue")

## `geom_smooth()` using formula 'y ~ x'
```




Rysunek 2: Wykres rozproszenia x i y

```r
cor(x,y)

## [1] 0.9853143
```

Korelacja między $x$ i $y$ jest wysoka i wynosi 0.985. Chmura punktów ma zdecydowanie w przybliżeniu liniowy charakter. W związku z tym możemy wykorzystać dany poniżej model regresji liniowej do opisania zależności jakie występują między $x$ i $y$

$$y = \beta_0 + \beta 1 \cdot x + \epsilon \tag{1}$$

## 1.3 Zadanie 3

```r
model <- lm( y ~  x, data=lab1)
names(model)

##  [1] "coefficients"  "residuals"     "effects"       "rank"
##  [5] "fitted.values" "assign"        "qr"            "df.residual"
##  [9] "xlevels"       "call"          "terms"         "model"

model.opis <- summary(model)
names(model.opis)

##  [1] "call"          "terms"         "residuals"     "coefficients"
##  [5] "aliased"       "sigma"         "df"            "r.squared"
##  [9] "adj.r.squared" "fstatistic"    "cov.unscaled"
```

```
model.opis

##
## Call:
## lm(formula = y ~ x, data = lab1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.00391 -0.68670  0.05062  0.55173  2.01107
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.87982    0.20178    4.36 3.21e-05 ***
## x            2.04252    0.03576   57.12  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9212 on 98 degrees of freedom
## Multiple R-squared:  0.9708,Adjusted R-squared:  0.9705
## F-statistic:  3263 on 1 and 98 DF,  p-value: < 2.2e-16

beta_0<-model.opis$coefficients[1]
beta_1<-model.opis$coefficients[2]


#beta_0 <- cor(y, x) *  sd(y) / sd(x)
#beta_1 <- mean(y) - beta1 * mean(x)
#betas<-rbind(c(beta0, beta1), coef(lm(y ~ x)))
```

Wartości estymatorów najmniejszych kwadratów parametrów dla $\beta_0$ i $\beta_1$ wynoszą $\hat{\beta}_0 = 0.8798193$, $\hat{\beta}_1 = 2.042517$

### 1.4 Zadanie 4

```
n=100
for (i in x) {
  for (j in y) {
    z1 = (y-beta_0-beta_1*x)^2
  }
}
sigma2 = sum(z1)/(n-2)
sigma2

## [1] 0.8486302
```

Znajdujemy wartość estymatora $\hat{\sigma^2}$ parametru $\sigma^2$, który jest równy 0.8486302. Korzystamy z prognozowanej przez model zmiennej objaśnianej $y$:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 \cdot x_i \tag{2}$$

## 1.5 Zadanie 5

```
attach(lab1)

## The following objects are masked from lab1 (pos = 3):
##
##     x, y

x<-x[1:100,]

## Error in x[1:100, ]: niepoprawna liczba wymiarów

mean(x)

## [1] 5.020899

for (i in x){
  x_1 = (x-mean(x))^2
}
SE2=sigma2/sum(x_1)
SE2

## [1] 0.00127844

T1 = beta_1/sqrt(SE2)
abs(T1) >= qt(p=0.99/2,df = n-2)*sqrt(SE2)

## [1] TRUE

model.opis

##
## Call:
## lm(formula = y ~ x, data = lab1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.00391 -0.68670  0.05062  0.55173  2.01107
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.87982    0.20178    4.36 3.21e-05 ***
## x            2.04252    0.03576   57.12  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9212 on 98 degrees of freedom
## Multiple R-squared:  0.9708,Adjusted R-squared:  0.9705
## F-statistic:  3263 on 1 and 98 DF,  p-value: < 2.2e-16
```

- Poprzez oszacowanie statystyki, która spełnia nierówność z treści zadania możemy odrzucić hipotezę zerową.
- p-wartość jest bardzo niska, mniejsza od 2.2e-16
- Przyjęty model regresji liniowej ma sens

## 1.6 Zadanie 6

```
beta_range = c(beta_1,beta_1) + c((qt(p=0.99/2,df = n-2)*sqrt(SE2)),-qt(p=0.99/2,df = n-
beta_range

## [1] 2.042068 2.042967
```

Przedział ufności dla $\beta_1$ na poziomie ufności 0.99 wynosi 2.042068, 2.042967. Jest on bardzo wąski.

## 1.7 Zadanie 7

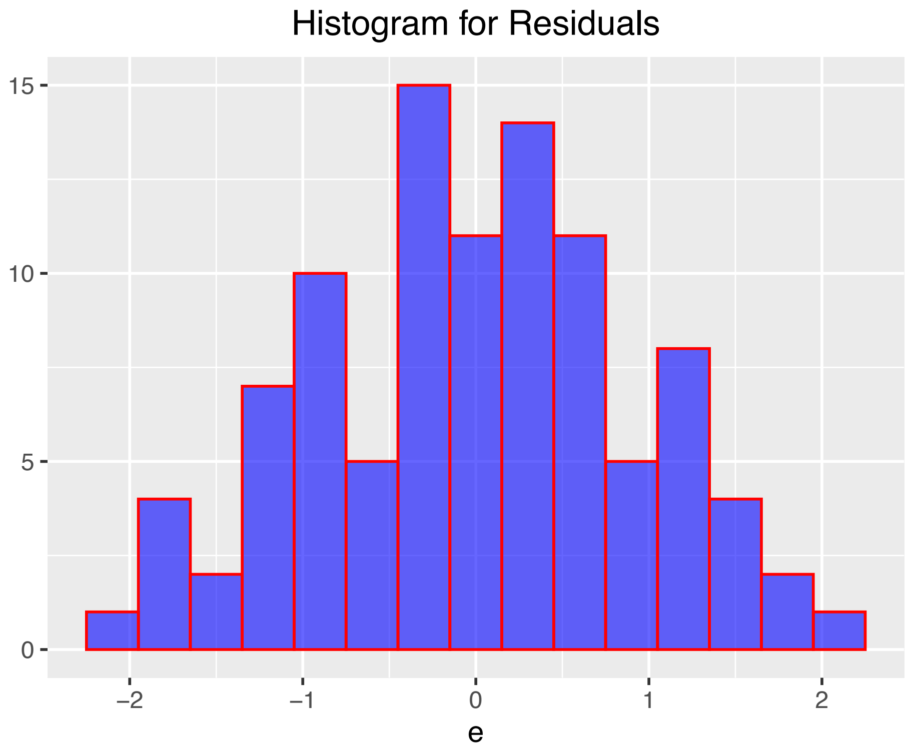
```
x0=1
Y_0= beta_0+beta_1*x0
Y_0

## [1] 2.922337

SEE2=sigma2*(1+1/n+(x0-mean(x))/sum(x_1))


beta_range2 = c(Y_0,Y_0) + c((qt(p=0.99/2,df = n-2)*sqrt(SEE2)),-qt(p=0.99/2,df = n-2)*s
beta_range2

## [1] 2.910738 2.933935
```

- Prognozowana przez model wartość wynosi 2.922337
- Przedział ufności na poziomie ufności 0.99 wynosi 2.910738, 2.933935

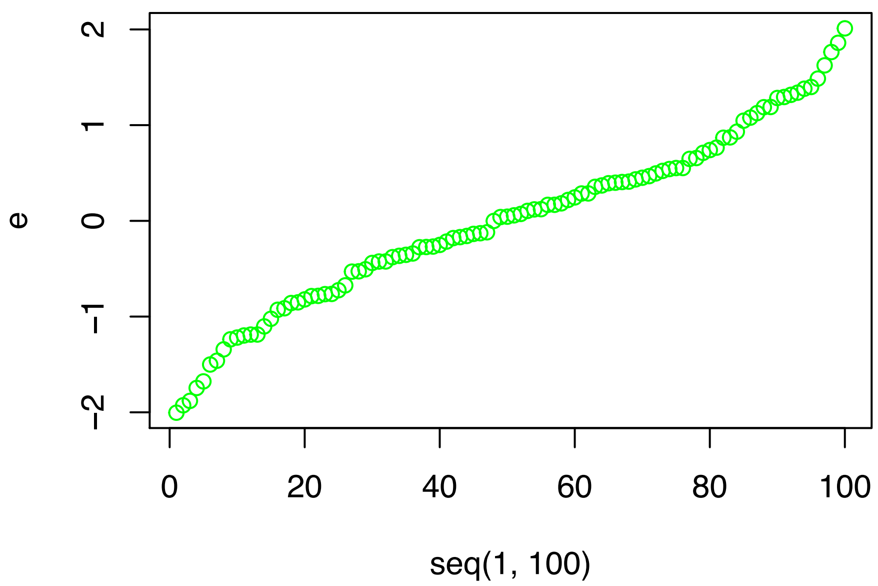
## 1.8 Zadanie 8

```
y_daszek= beta_0+beta_1*x

e=y-y_daszek

qplot(e,
      binwidth = 0.3,
      main = "Histogram for Residuals",
```

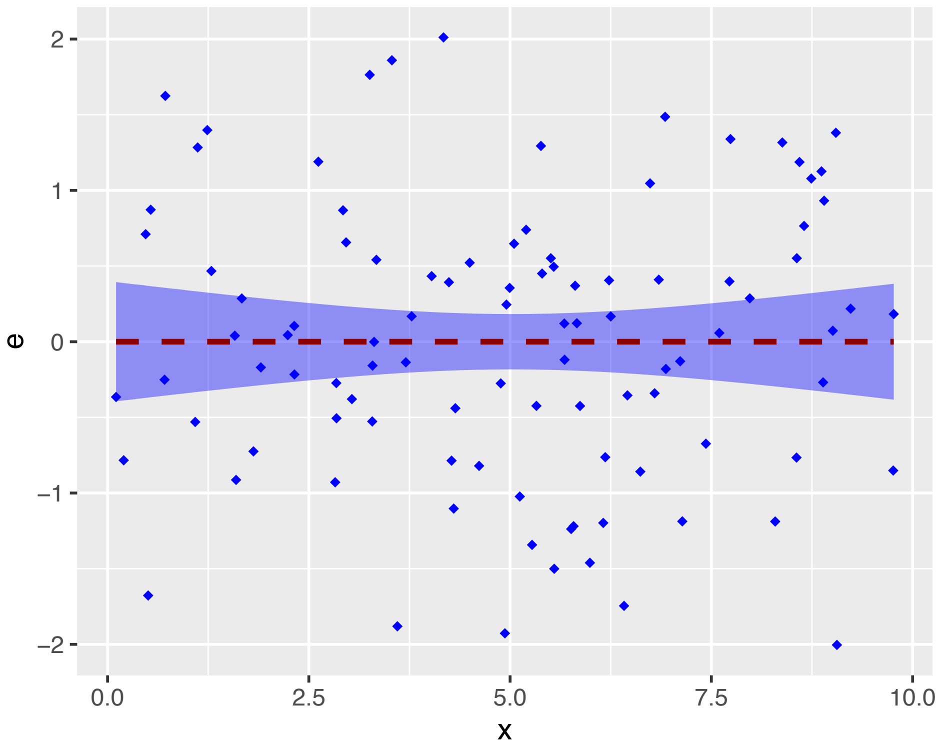
```
      xlab = "e",
      fill=I("blue"),
      col=I("red"),
      alpha=I(.6))+
  theme(plot.title = element_text(hjust = 0.5))
```



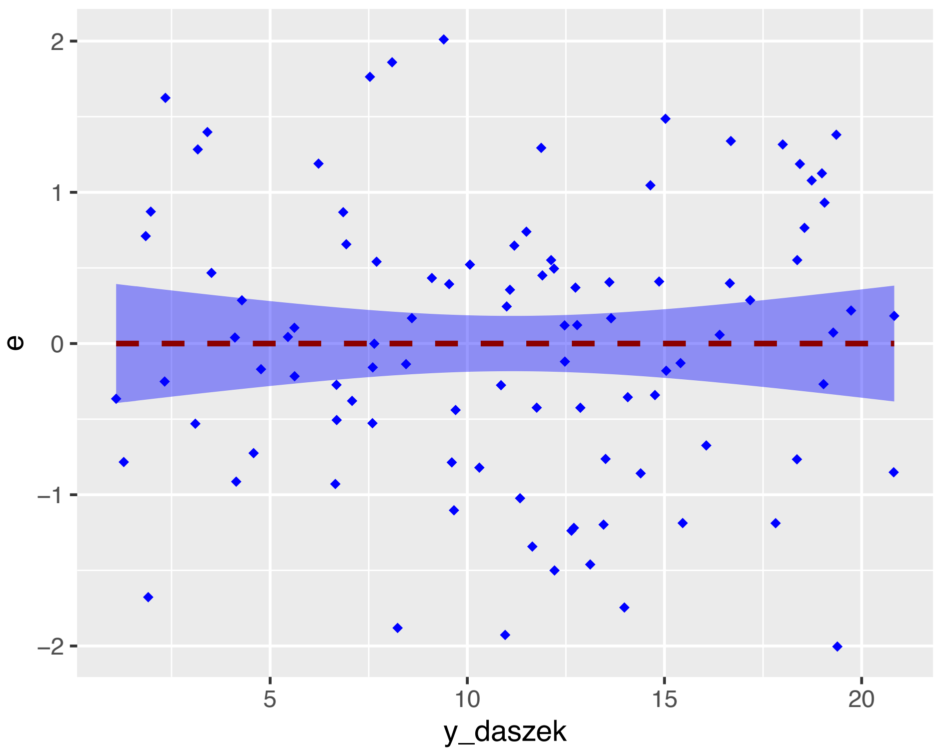

```
qqplot(seq(1,100),e, col="green")
```



```r
ggplot(lab1,aes(x = x, y = e)) +
  geom_point(shape=18, color="blue")+
  geom_smooth(method=lm,  linetype="dashed",
              color="darkred", fill="blue")

## `geom_smooth()` using formula 'y ~ x'
```




```r
data8<-cbind(y_daszek,e)

ggplot(lab1,aes(x = y_daszek, y = e)) +
  geom_point(shape=18, color="blue")+
  geom_smooth(method=lm,  linetype="dashed",
              color="darkred", fill="blue")

## `geom_smooth()` using formula 'y ~ x'
```



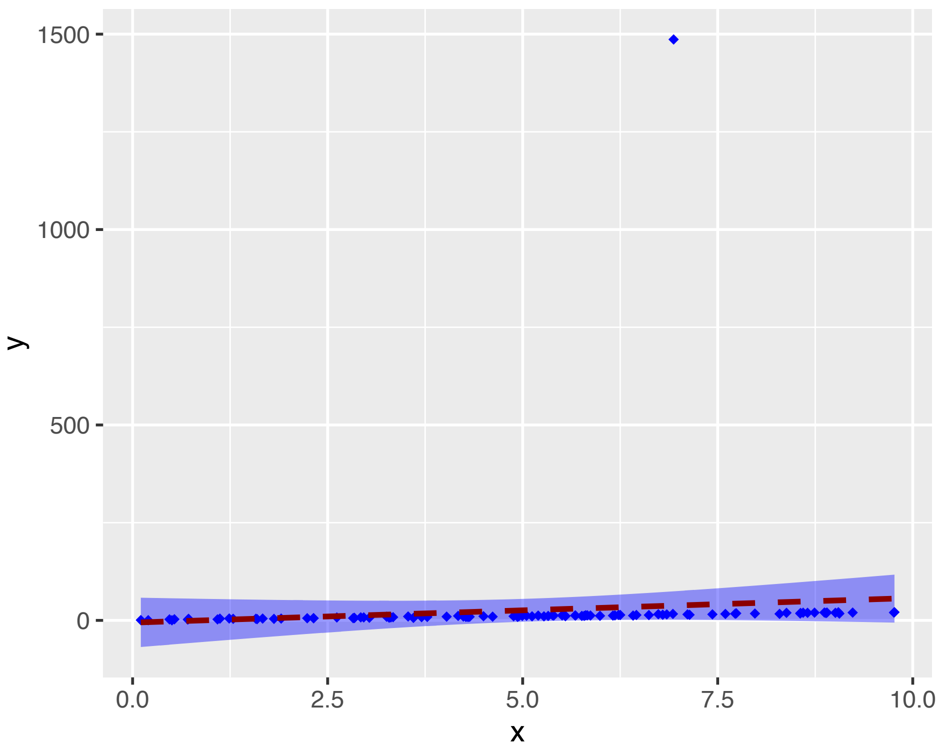

```
cor(e,y_daszek)

## [1] 9.100764e-16

cor(e,x)

## [1] 9.376956e-16
```

Nie ma korelacji między $x$ i e oraz między $y.daszek$ i $e$, korelacja jest równa prawie zeru, stąd model regresji liniowej poprawnie opisuje zależność między zmiennymi $x$ i $y$.

## 1.9 Zadanie 9

```
lab1[lab1 == 14.864] <- 1486.4
attach(lab1)

## The following objects are masked from lab1 (pos = 3):
##
##     x, y
## The following objects are masked from lab1 (pos = 4):
##
##     x, y

model <- lm( y ~  x, data=lab1)
names(model)

##  [1] "coefficients"  "residuals"     "effects"       "rank"
##  [5] "fitted.values" "assign"        "qr"            "df.residual"
##  [9] "xlevels"       "call"          "terms"         "model"
```

```r
model.opis <- summary(model)
names(model.opis)

##  [1] "call"          "terms"         "residuals"     "coefficients"
##  [5] "aliased"       "sigma"         "df"            "r.squared"
##  [9] "adj.r.squared" "fstatistic"    "cov.unscaled"

model.opis

##
## Call:
## lm(formula = y ~ x, data = lab1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
##  -35.67  -22.36  -15.41   -5.92 1448.52
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -5.706     32.302  -0.177    0.860
## x              6.285      5.724   1.098    0.275
##
## Residual standard error: 147.5 on 98 degrees of freedom
## Multiple R-squared:  0.01215,Adjusted R-squared:  0.002073
## F-statistic: 1.206 on 1 and 98 DF,  p-value: 0.2749

model.opis$coefficients[1]

## [1] -5.70636

model.opis$coefficients[2]

## [1] 6.285092

cor(x,y)

## [1] 0.1102418

ggplot(lab1, aes(x = x, y = y)) +
  geom_point(shape=18, color="blue")+
  geom_smooth(method=lm,  linetype="dashed",
             color="darkred", fill="blue")

## `geom_smooth()` using formula 'y ~ x'
```

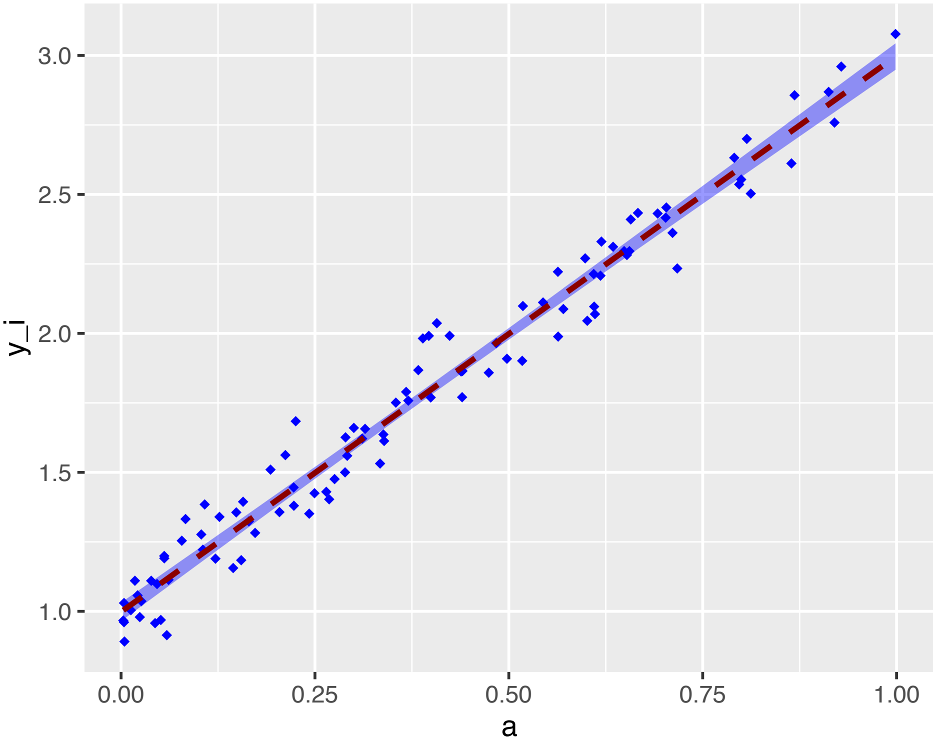



Po modyfikacji jednej zmiennej korelacja estymatory się diametralnie zmieniły, zarówno korelacja. Obserwacja jest odsająca (odbiega od trendu) oraz wpływowa (wpływa na wartości estymatorów i korelacji)

## 1.10 Zadanie 10

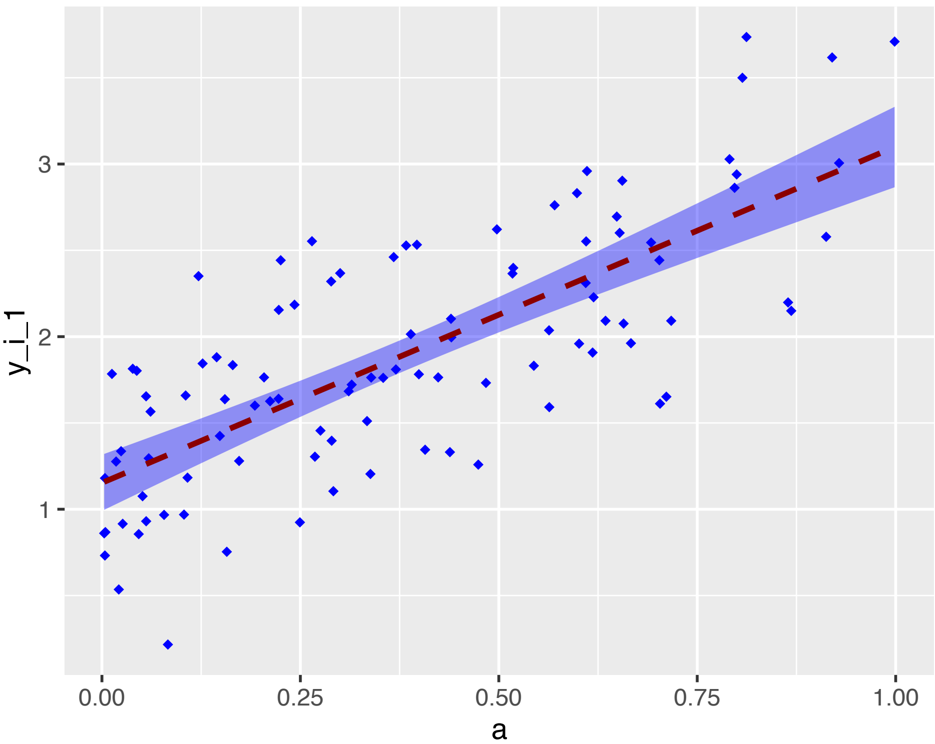
```
alpha=0.1
beta_1=1
beta_2=2
a<-runif(100,0,1)
b<-rnorm(100,0,0.1)
y_i=beta_1+beta_2*a+b
y_i

##   [1] 1.1991891 0.9575195 0.9792601 1.4030683 1.0042272 1.3236233 2.1113730
##   [8] 1.0982938 2.2958221 1.1840066 1.5596475 1.1894853 2.5536544 1.6839744
##  [15] 1.6129790 1.1907880 2.4314429 1.9911667 1.9013027 0.8912648 2.2960319
##  [22] 1.3845578 1.5316918 1.8678440 0.9608431 1.7894532 3.0771578 2.2216984
##  [29] 0.9688864 2.0694693 1.4300036 2.4164225 1.3570466 2.4333502 1.3942233
##  [36] 2.0876357 2.2700162 1.8631924 2.2336966 2.5029511 1.5000343 1.7694704
##  [43] 0.9142646 1.2540368 2.0963761 1.4249189 1.4757915 2.5360531 1.2765705
##  [50] 1.9915369 1.0364221 1.8647850 2.2136018 1.3559727 1.2822449 1.3398075
##  [57] 1.7509016 2.2077012 1.6363954 1.3319987 1.1132891 2.0367808 2.3116341
##  [64] 2.9600560 1.0573411 1.7580967 2.8568143 2.3617091 1.6599640 1.6259621
##  [71] 1.1558383 1.9650424 2.6318513 1.4458857 1.5097794 1.8585716 1.3799444
##  [78] 1.7704001 2.3302458 2.0453914 2.4530366 2.6115577 1.3513294 1.5621545
##  [85] 1.9087573 2.6998868 2.0984862 1.1099518 1.6565288 2.2820573 2.7584640
```

```
##  [92] 1.9885583 1.9819144 1.0299330 1.2217359 2.8689305 1.1100520 1.6213341
##  [99] 2.4100528 0.9668315

#b

ggplot(lab1, aes(x = a, y = y_i)) +
  geom_point(shape=18, color="blue")+
  geom_smooth(method=lm,  linetype="dashed",
              color="darkred", fill="blue")

## `geom_smooth()` using formula 'y ~ x'
```

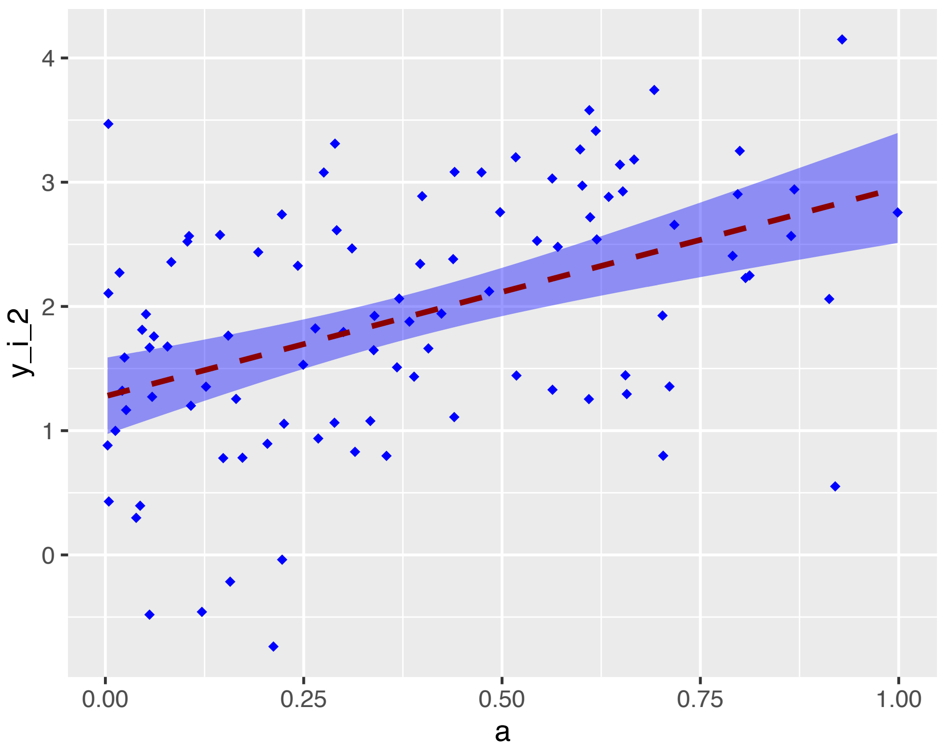



```
#c

model <- lm( y_i ~  a)
model.opis <- summary(model)
model.opis$coefficients[1]

## [1] 0.9981244

model.opis$coefficients[2]

## [1] 2.000613

model.opis$r.squared

## [1] 0.9698184
```

```
#wartości parametrów są bardzo przybliżone


#d

b1<-rnorm(100,0,0.5)
b2<-rnorm(100,0,1)
y_i_1=beta_1+beta_2*a+b1
y_i_2=beta_1+beta_2*a+b2



ggplot(lab1, aes(x = a, y = y_i_1)) +
  geom_point(shape=18, color="blue")+
  geom_smooth(method=lm,  linetype="dashed",
              color="darkred", fill="blue")

## `geom_smooth()` using formula 'y ~ x'
```




```
ggplot(lab1, aes(x = a, y = y_i_2)) +
  geom_point(shape=18, color="blue")+
  geom_smooth(method=lm,  linetype="dashed",
              color="darkred", fill="blue")

## `geom_smooth()` using formula 'y ~ x'
```




```
model <- lm( y_i_1 ~  a)
model.opis <- summary(model)
model.opis$coefficients[1]

## [1] 1.152889

model.opis$coefficients[2]

## [1] 1.948239

model.opis$r.squared

## [1] 0.5595151

model <- lm( y_i_2 ~  a)
model.opis <- summary(model)
model.opis$coefficients[1]

## [1] 1.277037

model.opis$coefficients[2]

## [1] 1.678628

model.opis$r.squared

## [1] 0.2074275
```

- Tak, chmura ma charakter liniowy

- Estymatory najmniejszych kwadratów wynoszą odpowiednio 1.035504 i 1.950445, są więc zbliżone do $\beta_0$=1 i $\beta_1$=2
- Zdecydowanie gdy rośnie sigma precyzja estymatorów maleje.
- 0.9729966 0.4788748 0.2437685 - wzrost R2 wraz ze wzrostem sigmy.

# 2 Zadania teoretyczne

Zadania teoretyczne

2.1

$$S(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - \beta_0 + \beta_1 x_i)^2$$

$$(\hat{\beta}_0, \hat{\beta}_1) = \arg\min_{(\beta_0,\beta_1)} \left[ (y_i - \beta_0 - \beta_1 x_i)^2 \right]$$

rozw. problemu minimalizacyjnego

Liczymy pochodne cząstkowe $S(\beta_0, \beta_1)$ w $(\hat{\beta}_0, \hat{\beta}_1)$, będzie się zerują. ( $S(\beta_0,\beta_1)$ jest wypukła - wartość najm. osiąga w $(\hat{\beta}_0, \hat{\beta}_1)$.

$$\frac{\partial S(\hat{\beta}_0, \hat{\beta}_1)}{\partial \hat{\beta}_0} = \frac{\partial \left(\sum_{i=1}^{n} (y_i^2 - 2 y_i \hat{\beta}_0 - 2 y_i x_i \hat{\beta}_1 + \hat{\beta}_0^2 + 2\hat{\beta}_0 \hat{\beta}_1 x_i + \hat{\beta}_1^2 x_i^2)\right)}{\partial \hat{\beta}_0}$$

$$= -2 \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$\frac{\partial S(\hat{\beta}_0, \hat{\beta}_1)}{\partial \hat{\beta}_1} = -2 \sum_{i=1}^{n} (y_i x_i - \beta_0 x_i - \hat{\beta}_1 x_i^2) =$$

$$= -2 \sum_{i=1}^{n} (y_i \cancel{x_i} - \beta_0 - \hat{\beta}_1 x_i) x_i = 0$$

Mamy

$$\begin{cases} -2 \sum_{i=1}^{n} y_i + 2n\hat{\beta}_0 + 2\hat{\beta}_1 \sum_{i=1}^{n} x_i = 0 \\ -2 \sum_{i=1}^{n} x_i y_i + 2 \hat{\beta}_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = 0 \end{cases} \qquad \sum_{i=1}^{n} x_i = \bar{x} \quad \sum_{i=1}^{n} y_i = \bar{y}$$

$$\begin{cases} \hat{\beta}_0 = \dfrac{\sum_{i=1}^{n} y_i - \beta_1 \sum_{i=1}^{n} x_i}{n} = \bar{y} - \hat{\beta}_1 \bar{x} \\ \sum_{i=1}^{n} (x_i y_i - x_i \bar{y} + \hat{\beta}_1 x_i \bar{x} - \hat{\beta}_1 x_i^2) = 0 \end{cases}$$

$$\Rightarrow \begin{cases} \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \\ \hat{\beta}_1 = \dfrac{\sum_{i=1}^{n} (x_i y_i - x_i \bar{y})}{\sum_{i=1}^{n} (x_i^2 - x_i \bar{x})} \end{cases}$$

Fakt

$$\sum_{i=1}^{n} (\bar{x}^2 - x_i \bar{x}) = 0 \quad *$$

$$\sum_{i=1}^{n} (\bar{x}\bar{y} - y_i \bar{x}) = 0 \quad **$$

Przekształcamy $\hat{\beta}_1$

$$\hat{\beta}_1 = \frac{\sum (x_i y_i - x_i \bar{y} + \overbrace{\bar{x}\bar{y} - y_i \bar{x}}^{**})}{\sum (x_i^2 - x_i \bar{x} + \underbrace{\bar{x}^2 - x_i \bar{x}}_{*})} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

2.2

Cel: $\hat{\beta}_1 = r \cdot \frac{s_y}{s_x}$, gdzie $S_x$ - wariancja próbkowa $x$

$S_y$ - kowariancja próbkowa $y$

$r$ - współczynnik Pearsona

$$S_y = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2}$$

$$S_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

korzystamy z końcowej formy $\hat{\beta}_1$ z zad. 1

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \underbrace{\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2}\cdot\sqrt{\sum(x_i-\bar{x})^2}}\cdot\frac{\sqrt{\sum(y_i-\bar{y})^2}}{\sqrt{\sum(y_i-\bar{y})^2}}}_{\to r} =$$

$$= r \cdot \frac{\sqrt{\frac{1}{n}\sum(y_i-\bar{y})}}{\sqrt{\frac{1}{n}\sum(x_i-\bar{x})}} = r \cdot \frac{S_y}{S_x}$$

Z. 3

- współczynnik dla $(x_1, e_1), \ldots, (x_n, e_n)$

Jeśli model regresji liniowej poprawnie opisuje zależność między $x$ i $y$ to $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ $\quad i \in \mathbb{N}$

resydua: $e_i = y_i - \hat{y}_i$

współczynnik korelacji $r_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}}$

Liczymy współczynnik korelacji

Mianownik na potrzeby obliczeń zastępujemy $T_1$

Mamy

$$r = \frac{\sum (x_i - \bar{x})(e_i - \bar{e})}{T_1} = \frac{\sum (x_i - \bar{x})(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i - \bar{e})}{T_1} =$$

$$= \frac{\sum x_i y_i - \hat{\beta}_0 \sum x_i - \hat{\beta}_1 \sum x_i^2 - \bar{x} \sum y_i + \hat{\beta}_0 \bar{x} \cdot n + \hat{\beta}_1 \bar{x} \sum x_i}{T_1} =$$

$$= \frac{\hat{\beta}_0(\underbrace{\bar{x} \cdot n - \sum x_i}_{\bar{x} \cdot n}) + \hat{\beta}_1 \overbrace{\sum (\bar{x}^2 - \bar{x} \cdot x_i)}^{0} + \sum (x_i y_i - \bar{x} y_i)}{T_1} =$$

$$\bar{e} = \frac{\sum e_i}{n} = \frac{\sum (y_i - \hat{y}_i)}{n} = \frac{\sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)}{n} = \frac{-n\hat{\beta}_0}{n} + \bar{y} - \hat{\beta}_1 \bar{x} = -\hat{\beta}_0 + \hat{\beta}_0 = 0$$

$$= \frac{\sum \left( \frac{x_i y_i}{\cancel{T_1}} - \frac{\bar{x} y_i}{\cancel{T_1}} \right)}{T_1} = \cancel{\frac{\sum x_i \bar{y} - \bar{x} \bar{y}}{T_1}} \cancel{=0} = 0$$

$$= \frac{\sum (x_i y_i - \bar{x} y_i) + \sum (\bar{x}\bar{y} - \bar{y} x_i)}{T_1}$$

Na wykresach punkty powinne być chaotycznie rozrzucone. Brak korelacji.

$T_1 \quad \sum (x_i y_i - \ldots$

2.3

- współczynnik dla $(\hat{y}_1, e_1), \dots, (\hat{y}_n, e_n)$

$$r = \frac{\sum(\hat{y}_i - \bar{\hat{y}})(e_i - \bar{e})}{T_2} =$$

$$\boxed{\begin{aligned} &\bar{e} = 0 \\ &\bar{\hat{y}} = \frac{\sum(\hat{\beta}_0 + \hat{\beta}_1 x_i)}{n} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} \end{aligned}}$$

$$= \frac{\sum(\hat{\beta}_0 + \hat{\beta}_1 x_i - \cancel{\beta_0} \hat{\beta}_0 - \beta_1 \bar{x})(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)}{T_2}$$

$$= \frac{\sum \hat{\beta}_1 (x_i - \bar{x})(y_i - \beta_0 - \hat{\beta}_1 x_i)}{T_2}$$ dalej analogicznie jak dla $(x_n, e_n)$ $= 0$