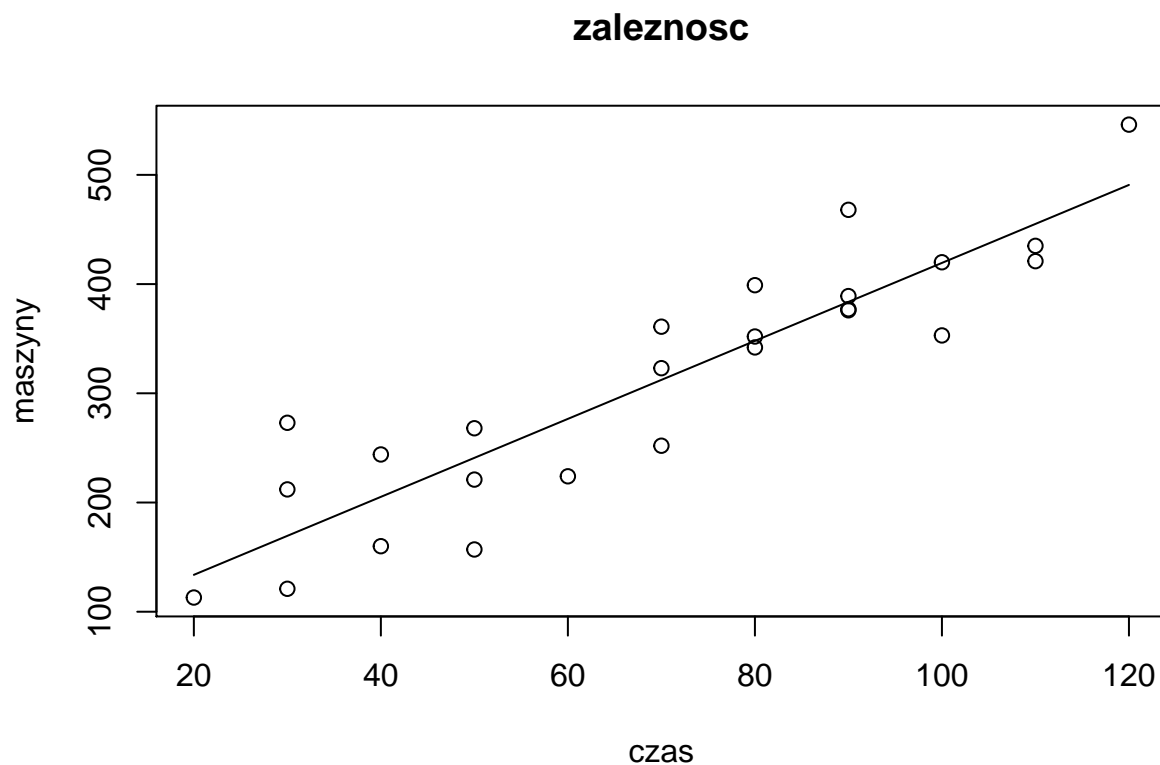


Lista 2

Aneta Przydrozna

Dane dotyczące ilości obsługiwanych maszyn w czasie
(w godzinach) na wykresie.



Ustalone zmienne Y-objaśniana (liczba kserokopiarek) i X-objaśniająca (liczba godzin) na wykresie układają się rosnąco, zatem ilość serwisowanych kopiarek zależy od ilości godzin. Widać również prostą regresji liniowej.

Estymowany parametr $b_0 = 62.4$

Estymowany parametr $b_1 = 3.6$

Szacowane równanie regresji: $Y = 62.4 + 3.6 * X$

95% przedział ufności dla współczynnika nachylenia
wynosi (2.9 , 4.3)

```
summary(regresja)

##
## Call:
## lm(formula = Y ~ X, data = dane)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -83.876 -34.088  -5.982   38.826 103.528
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   62.366     26.177   2.382  0.0259 *
## X              3.570       0.347  10.290 4.45e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 48.82 on 23 degrees of freedom
## Multiple R-squared:  0.8215, Adjusted R-squared:  0.8138
## F-statistic: 105.9 on 1 and 23 DF,  p-value: 4.449e-10
```

Czy między zmiennymi nie ma zależności? O tym mówi hipoteza zerowa $H_0 : \beta_1 = 0$. Gdyby była to odrzucamy tę hipotezę i przyjmujemy alternatywę $H_A : \beta_1 \neq 0$, ale P-value jest bardzo małe więc przyjmujemy alternatywę. Statystyka R jest równa 0.8, zatem spory procent X wyjaśnia Y.

Szacowana średnia liczba naprawianych maszyn w czasie $X_0 = 11$

```
kwantyl=confint(regresja,level = 0.95)
n=length(dane$Y)
names(dane)<-c("X","Y")
regresja<-lm(Y~X,dane)
beta<-solve(t(X1)%*%X1)%*%t(X1)%*%Y1
Xo<-11
e<-regresja$residuals

prawy<-beta[1]+beta[2]*Xo+qt(1-0.05/2,n-2)*(sd(e)*sqrt((n-1)/(n-2)))*
  sqrt(((1/n)+(Xo-mean(dane$X))^2/(var(dane$X)*(n-1))))

lewy<-beta[1]+beta[2]*Xo-qt(1-0.05/2,n-2)*(sd(e)*sqrt((n-1)/(n-2)))*
  sqrt(((1/n)+(Xo-mean(dane$X))^2/(var(dane$X)*(n-1))))

przedzial<-c(lewy,prawy)
przedzial

## [1]  54.71898 148.55718
mean(przedzial)
```

```
## [1] 101.6381
```

Przedział ufności (na poziomie ufności $\alpha = 0.05$) Zatem średnia ilość kserokopiarek naprawionych w ciągu 11 godzin to 102.

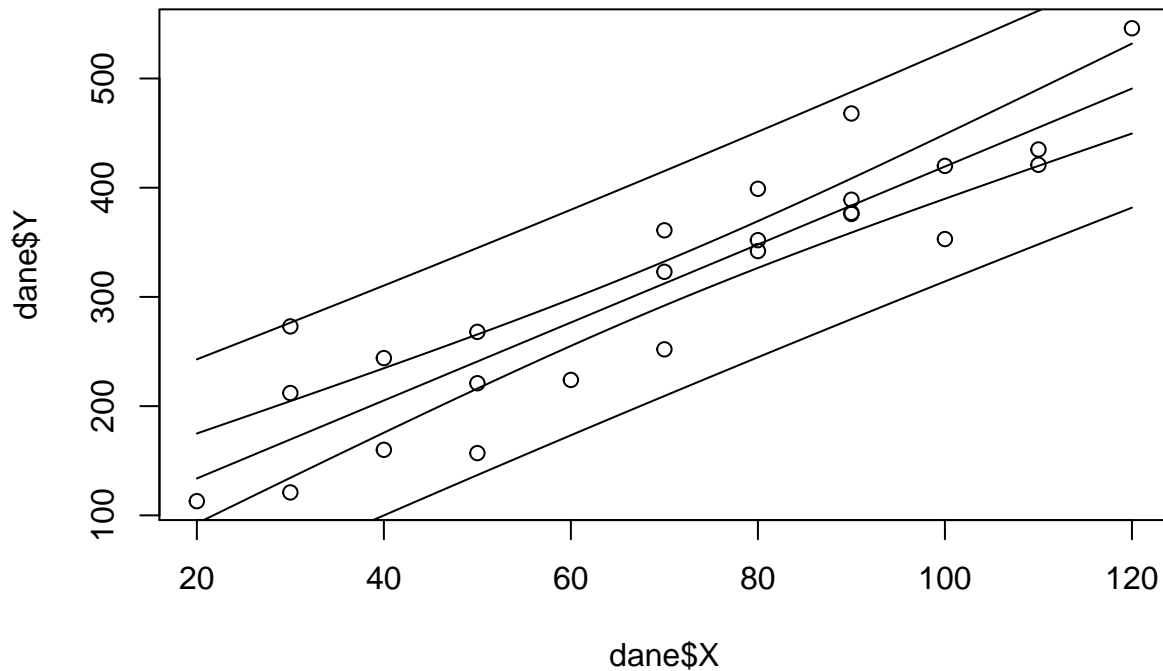
**Prognoza rzeczywistej ilości serwisowanych maszyn,
której można się spodziewać po 11 godzinach.**

```
prawy<-beta[1]+beta[2]*Xo+qt(1-0.05/2,n-2)*(sd(e)*sqrt((n-1)/(n-2)))*  
  sqrt(1+(1/n)+(Xo-mean(dane$X))^2/(var(dane$X)*(n-1)))  
  
lewy<-beta[1]+beta[2]*Xo-qt(1-0.05/2,n-2)*(sd(e)*sqrt((n-1)/(n-2)))*  
  sqrt(1+(1/n)+(Xo-mean(dane$X))^2/(var(dane$X)*(n-1)))  
przedzial<-c(lewy,prawy)  
przedzial
```

```
## [1] -9.726824 213.002985
```

Bazując na mniejszej liczbie danych popełniamy większy błąd. Prognoza rzeczywistej ilości serwisowanych maszyn jest mniej dokładna.

Powyższe przedziały naniesione na wykres.



Widać, że do szerszego przedziału (predykcji) wpadają wszystkie punkty. Przedział ufności nie obejmuje wszystkich.

Zakładamy, że $n = 40$, $\sigma^2 = 120$, $SSX = \sum (X_i - \bar{X})^2 = 1000$

Przy poziomie istotności $\alpha = 0.05$ i rzeczywistym nachyleniu $\beta_1 = 1$ szukamy moc odrzucenia hipotezy zerowej, że regresja nachylenie regresji wynosi 0.

```
n<-40
sigma2<-120
ssx<-1000
beta<-1
alpha<-0.05
stat1<-sigma2/ssx
tc<-qt(1-alpha/2,n-2)
stat2<-beta/sqrt(stat1)
p1<-pt(tc,n-2,stat2)
p2<-pt(-tc,n-2,stat2)
power<-1-p1+p2
power
```

```
## [1] 0.8032105
```

Mamy prawdopodobieństwo odrzucenia H_0 , gdy jest fałszywa. Wynosi ono 80%. Dla $\beta_1 \neq 0$ statystyka $t = \frac{b_1}{s(b_1)}$ ma rozkład Studenta(n-2). Mamy $|t| > t_c$, więc odrzucamy H_0 .

Funkcja $\beta_1 \in [-2, 2]$

```
## Warning in pt(tc, n - 2, delta): pełna precyzja może nie zostać osiągnięta
## w 'pnt{final}'
```

```
## Warning in pt(tc, n - 2, delta): pełna precyzja może nie zostać osiągnięta
## w 'pnt{final}'
```

```
## Warning in pt(tc, n - 2, delta): pełna precyzja może nie zostać osiągnięta
## w 'pnt{final}'
```

```
## Warning in pt(tc, n - 2, delta): pełna precyzja może nie zostać osiągnięta
## w 'pnt{final}'
```

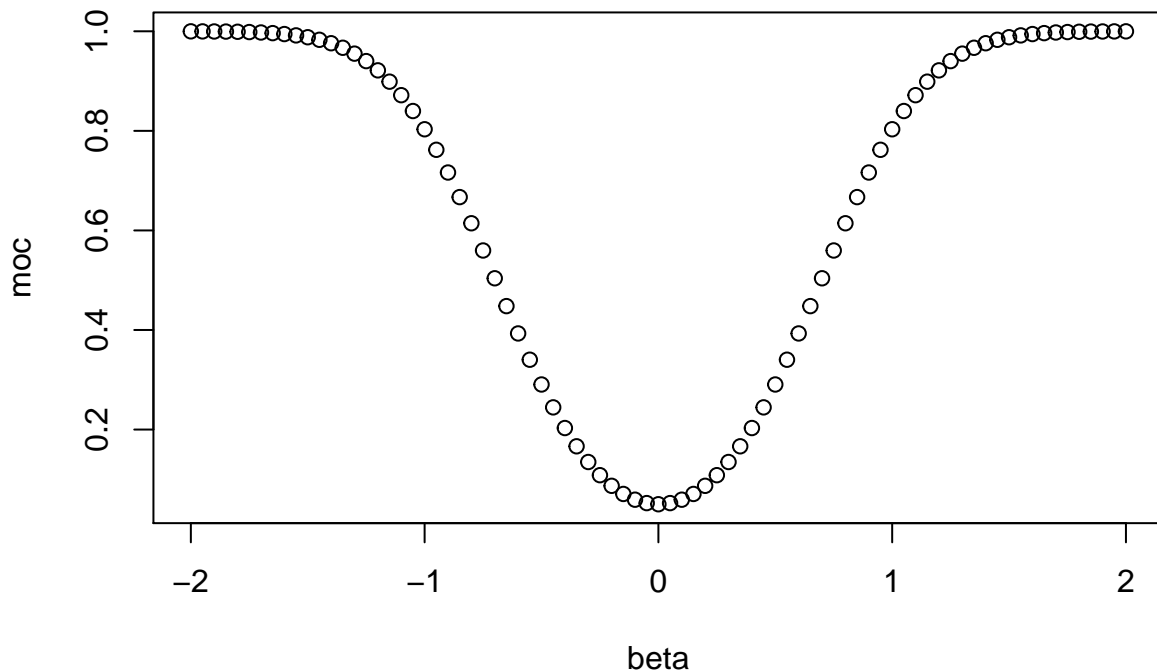
```
## Warning in pt(tc, n - 2, delta): pełna precyzja może nie zostać osiągnięta
## w 'pnt{final}'
```

```
## Warning in pt(tc, n - 2, delta): pełna precyzja może nie zostać osiągnięta
## w 'pnt{final}'
```

```
## Warning in pt(tc, n - 2, delta): pełna precyzja może nie zostać osiągnięta
## w 'pnt{final}'
```

```
## Warning in pt(tc, n - 2, delta): pełna precyzja może nie zostać osiągnięta
## w 'pnt{final}'
```

```
## Warning in pt(tc, n - 2, delta): pełna precyzja może nie zostać osiągnięta
## w 'pnt{final}'
```



Mamy wygenerowany wektor X dla rozkładu $N(0, \frac{1}{200}I)^T$ i 1000 wektorów

$$Y = 5 = \beta_1 X + \epsilon$$

$$\beta_1 = 0, \epsilon \sim N(0, 1)$$

Prawdopodobieństwo wpadnięcia 0 do przedziału, czyli przyjmowania H_0 wynosi 0.94.

$$\beta_1 = 0, \epsilon_1, \dots, \epsilon_{200} \sim \text{Exp}(\lambda = 1)$$

Moc testu powinna się wynosić więcej niż powyższego. Prawdopodobieństwo wyniosło 0.96.

$$\beta_1 = 1.5, \epsilon \sim N(0, 1)$$

Prawdopodobieństwo wpadnięcia 0 do przedziału wynosi 0.65

$$\beta_1 = 1.5, \epsilon_1, \dots, \epsilon_{200} \sim \text{Exp}(\lambda = 1)$$

Prawdopodobieństwo wpadnięcia 0 do przedziału wynosi 0.64

Widać, że niezerowa wartość β_1 przyczynia się do zmniejszonej częstotliwości przyjmowania hipotezy zerowej.