Porównywanie modeli

```
> g0 <- lm(species~1,data=gala.eaa)
> anova(g0)
Analysis of Variance Table

Response: species
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals 29 381081   13141
> summary(g0)

Call:
lm(formula = species ~ 1, data = gala.eaa)

Residuals:
   Min     1Q Median     3Q    Max
-83.23 -72.23 -43.23  10.77 358.77

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    85.23      20.93   4.072 0.000328 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 114.6 on 29 degrees of freedom
```

```
> g1 <- lm(species~elevation,data=gala.eaa)
> anova(g1)
Analysis of Variance Table

Response: species
          Df Sum Sq Mean Sq F value    Pr(>F)
elevation  1 207828  207828  33.587 3.177e-06 ***
Residuals 28 173254    6188
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> summary(g1)

Call:
lm(formula = species ~ elevation, data = gala.eaa)

Residuals:
     Min       1Q   Median       3Q      Max
-218.319  -30.721  -14.690    4.634  259.180

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 11.33511   19.20529   0.590     0.56
elevation    0.20079    0.03465   5.795 3.18e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 78.66 on 28 degrees of freedom
Multiple R-squared:  0.5454,    Adjusted R-squared:  0.5291
F-statistic: 33.59 on 1 and 28 DF,  p-value: 3.177e-06
```

```
> g2 <- lm(species~elevation+adjacent,data=gala.eaa)
> anova(g2)
Analysis of Variance Table

Response: species
          Df Sum Sq Mean Sq F value    Pr(>F)
elevation  1 207828  207828  56.112 4.662e-08 ***
adjacent   1  73251   73251  19.777 0.0001344 ***
Residuals 27 100003    3704
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> 207828+73251
[1] 281079

> 281079/2
[1] 140539.5

> 1-pf(140539/3704,2,27)
[1] 1.434361e-08


> summary(g2)

Call:
lm(formula = species ~ elevation + adjacent, data = gala.eaa)

Residuals:
    Min      1Q  Median      3Q     Max
-103.41  -34.33  -11.43   22.57  203.65

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.43287   15.02469   0.095 0.924727
elevation    0.27657    0.03176   8.707 2.53e-09 ***
adjacent    -0.06889    0.01549  -4.447 0.000134 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 60.86 on 27 degrees of freedom
Multiple R-squared:  0.7376,    Adjusted R-squared:  0.7181
F-statistic: 37.94 on 2 and 27 DF,  p-value: 1.434e-08




> anova(g1,g2)
Analysis of Variance Table

Model 1: species ~ elevation
Model 2: species ~ elevation + adjacent
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1     28 173254
2     27 100003  1     73251 19.777 0.0001344 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> 73251/3704
[1] 19.77619
```

test czy wsp przy elevation może być 0.3

```
> t <- (0.27657-0.3)/0.03176
> t
[1] -0.7377204

> 2*pt(t,27)
[1] 0.467048
```

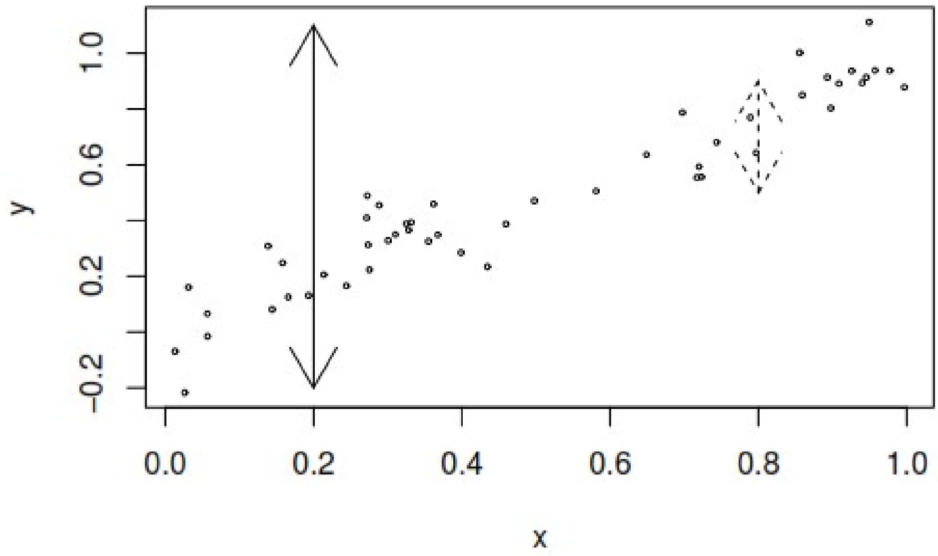


Figure 2.2: Variation in the response y when x is known is denoted by dotted arrows while variation in y when x is unknown is shown with the solid arrows

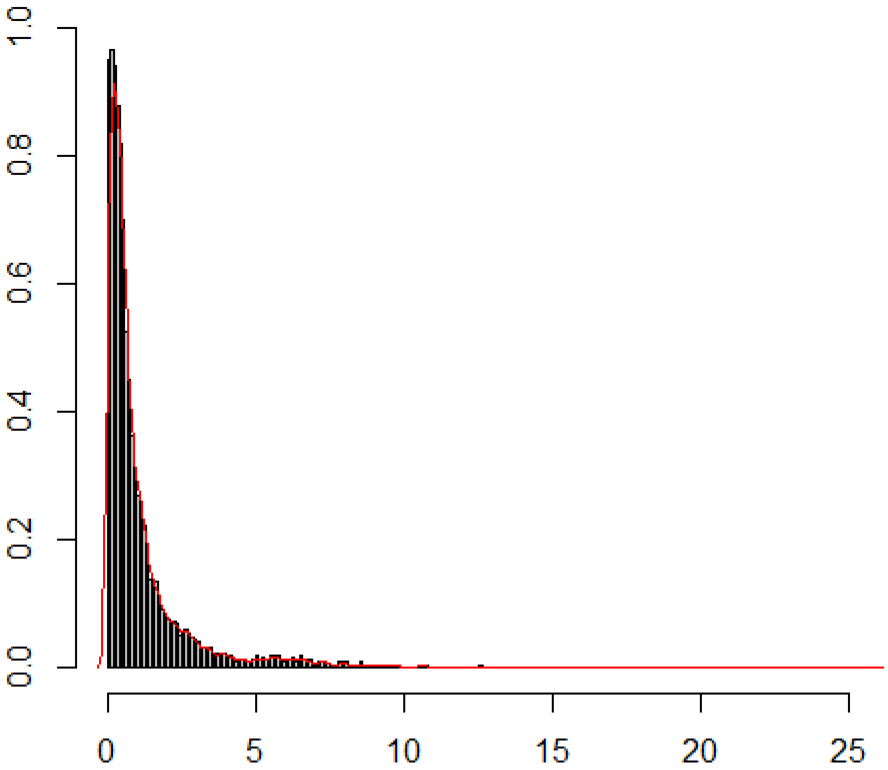
Permutacyjna metoda oceny istotności regresji

(bez założenia normalności)

```
f <-numeric(10000)
for (i in 1:10000){
  g <- lm(sample(gala$Species)~elevation+adjacent,data=gala.eaa)
  f[i] <- summary(g)$fstatistic[1]
}
length(f>37.94)/10000
quantile(f,probs = c(0.95,0.99,0.999))
histdens(f)
```

```
> length(f>37.94)/10000
[1] 1
```

```
> quantile(f,probs = c(0.95,0.99,0.999))
      95%       99%     99.9%
```




```
 4.691446  8.742880 14.156021
```

Niech

$$\varepsilon \sim N(0, \sigma^2 I)$$

$$y \sim N(X\beta, \sigma^2 I)$$

$$\hat{\beta} = (X^T X)^{-1} X^T y \sim N(\beta, (X^T X)^{-1} \sigma^2)$$

Confidence Intervals for β

We start with the simultaneous regions. Some results from multivariate analysis show that

dla

$$\frac{(\hat{\beta}-\beta)^T X^T X(\hat{\beta}-\beta)}{\sigma^2} \sim \chi^2_p$$

and

$$\frac{(n-p)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-p}$$

and these two quantities are independent. Hence

$$\frac{(\hat{\beta}-\beta)^T X^T X(\hat{\beta}-\beta)}{p\hat{\sigma}^2} \sim \frac{\chi^2_p/p}{\chi^2_{n-p}/(n-p)} \equiv F_{p,n-p}$$

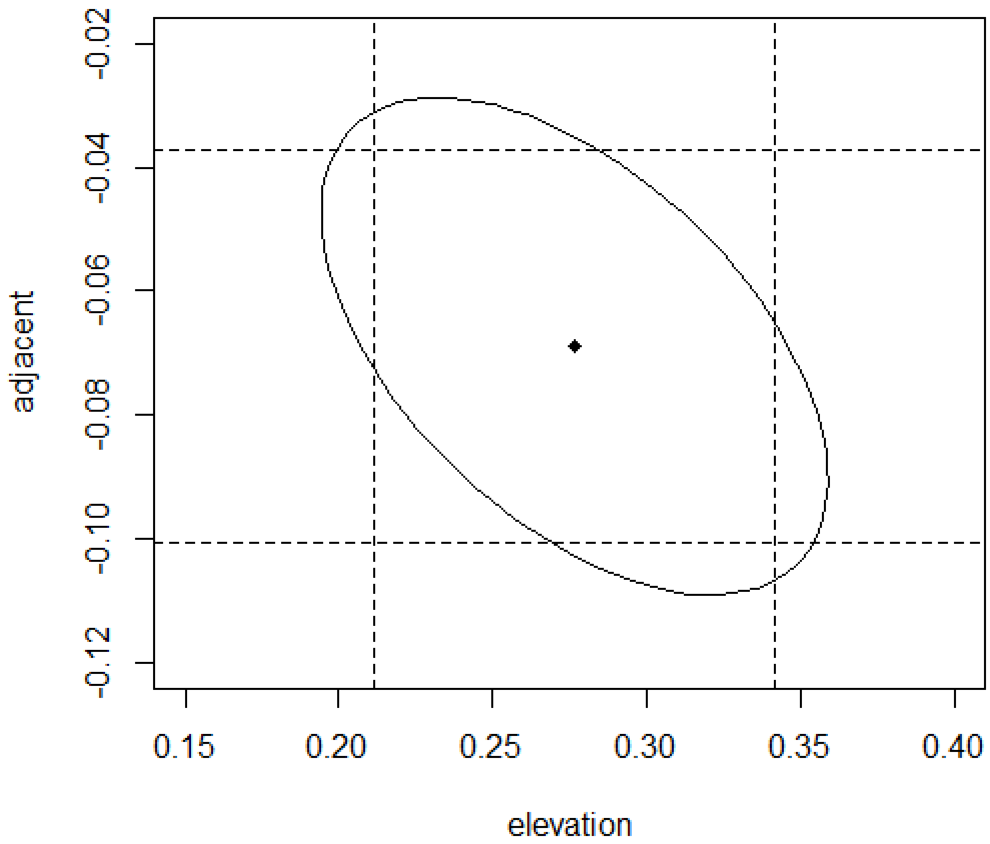
So to form a $100(1-\alpha)$ % confidence region for β, take β such that

$$(\hat{\beta}-\beta)^T X^T X(\hat{\beta}-\beta) \leq p\hat{\sigma}^2 F^{(\alpha)}_{p,n-p}$$

pojedynczych parametrów

$$\hat{\beta}_i \pm t^{(\alpha/2)}_{n-p}\hat{\sigma}\sqrt{(X^TX)^{-1}_{ii}}$$

```
> g2 <- lm(species~elevation+adjacent,data=gala.eaa)
> summary(g2)

Call:
lm(formula = species ~ elevation + adjacent, data = gala.eaa)

Residuals:
    Min      1Q  Median      3Q     Max 
-103.41  -34.33  -11.43   22.57  203.65 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.43287   15.02469   0.095 0.924727    
elevation    0.27657    0.03176   8.707 2.53e-09 ***
adjacent    -0.06889    0.01549  -4.447 0.000134 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1 

Residual standard error: 60.86 on 27 degrees of freedom
Multiple R-squared:  0.7376,    Adjusted R-squared:  0.7181 
F-statistic: 37.94 on 2 and 27 DF,  p-value: 1.434e-08 

> qt(0.975,27)

[1] 2.051831
> c(0.27657-2.051831*0.03176,0.27657+2.051831*0.03176)#elevation
[1] 0.2114038 0.3417362
> c( -0.06889-2.051831*0.01549, -0.06889+2.051831*0.01549)#adjacent
[1] -0.10067286 -0.03710714

> plot(ellipse(g2,c(2,3)),type="l",xlim=c(0.15,0.4),ylim=c(-0.12,-0.02))

> points(g2$coef[2],g2$coef[3],pch=18)

> abline(v=c(0.27657-2.051831*0.03176,0.27657+2.051831*0.03176),lty=2)
> abline(h=c( -0.06889-2.051831*0.01549, -0.06889+2.051831*0.01549),lty=2)
```

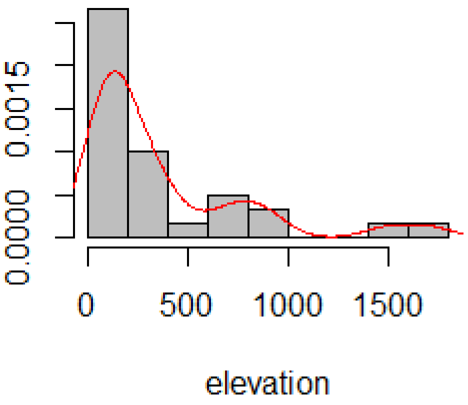


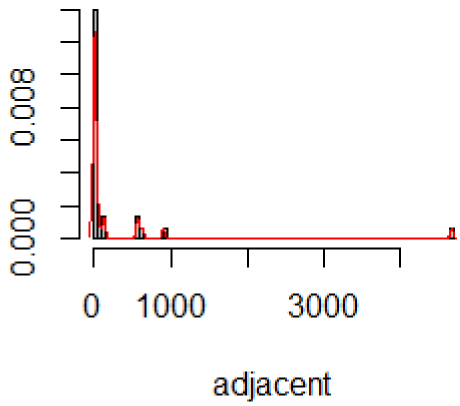
Dlaczego przecinają elipsę ufności?

```
> cor(gala.eaa$elevation,gala.eaa$adjacent)
[1] 0.5364578
```

im wyższa wartość elevation tym wyższa wartość adjacent a więc współczynniki są ujemnie skorelowane

Przedziały ufności dla prognoz






```
> x0 <- c(1,200,100)
> y0 <- sum(x0*g2$coef)
> y0
[1] 49.85798
```

dla wartości oczekiwanej:

```
> x <- cbind(1,gala.eaa[,3:4])
> x <- as.matrix(x)
> xtxi <- solve(t(x) %*% x)
> bm <- sqrt(x0 %*% xtxi %*% x0) *2.051831 * 60.86
> bm
         [,1]
[1,] 24.61241
> c(y0-bm,y0+bm)
[1] 25.24557 74.47039
```

dla indywidualnej wartości

```
> bm <- sqrt(1+x0 %*% xtxi %*% x0) *2.051831 * 60.86
> bm
         [,1]
[1,] 127.2768
> c(y0-bm,y0+bm)
[1] -77.41886 177.13483
```

INACZEJ

```
> predict(g2,data.frame(elevation=200,adjacent=100),se=T)
$fit
       1
49.85798

$se.fit
[1] 11.99514

$df
```

```
[1] 27

$residual.scale
[1] 60.85898


> sqrt(x0 %*% xtxi %*% x0) * 60.85898
         [,1]
[1,] 11.99514
```