

Korelacja i regresja

Współczynniki korelacji

r pearsona - najczęściej używany, parametryczny (zakłada rozkład normalny)

ρ Spearmana - nieparametryczny - korelacja rang

τ Kendalla

R^2 a r :

współczynnik determinacji - procent wyjaśnionej zmienności

```
sosny<-read.csv('sosna.csv',sep=';')
```

```
cor(sosny$AB,sosny$Age)
```

```
[1] 0.8144991
```

```
co to jest?
```

```
?cor
```

var, cov and cor compute the variance of x and the covariance or correlation of x and y if these are vectors. If x and y are matrices then the covariances (or correlations) between the columns of x and the columns of y are computed.

cov2cor scales a covariance matrix into the corresponding correlation matrix efficiently.

Usage

```
var(x, y = NULL, na.rm = FALSE, use)
```

```
cov(x, y = NULL, use = "everything",  
    method = c("pearson", "kendall", "spearman"))
```

```
cor(x, y = NULL, use = "everything",  
    method = c("pearson", "kendall", "spearman"))
```

x a numeric vector, matrix or data frame.

y NULL (default) or a vector, matrix or data frame with compatible dimensions to x. The default is equivalent to y = x (but more efficient).

na.rm logical. Should missing values be removed?

use an optional character string giving a method for computing covariances in the presence of missing values. This must be (an abbreviation of) one of the strings "everything", "all.obs", "**complete.obs**", "na.or.complete", or "**pairwise.complete.obs**".

method a character string indicating which correlation coefficient (or covariance) is to be computed. One of "pearson" (default), "kendall", or "spearman": can be abbreviated.

```
> cor(sosny$AB,sosny$Age, method = 'pearson')
```

```
[1] 0.8144991
```

```
> cor(sosny$AB,sosny$Age, method = 'spearman')
```

```
[1] 0.8771537
```

macierz korelacji

```
> cor(sosny[,c(5:13)])
```

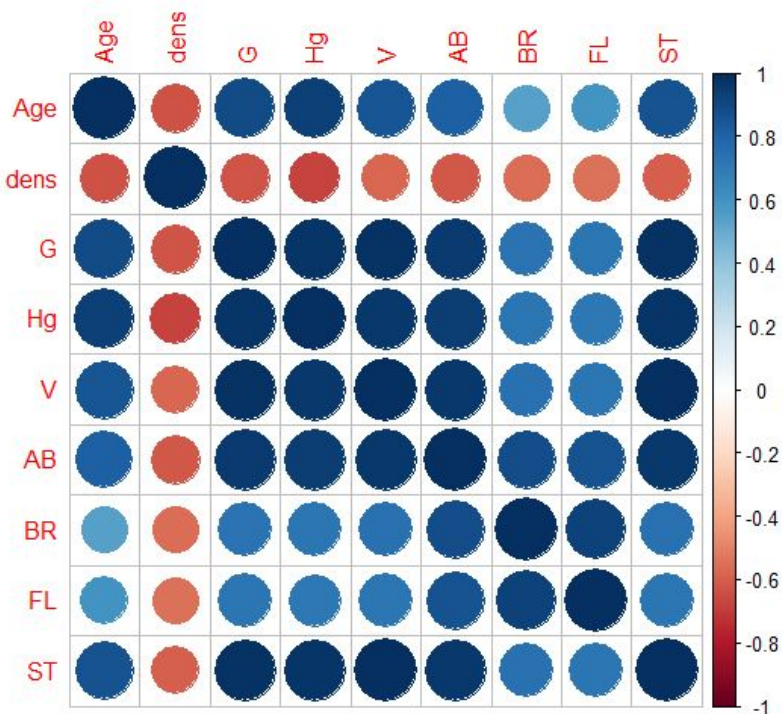
	Age	dens	G	Hg	V
Age	1.0000000	-0.6369490	0.8984927	0.9308940	0.8521970
dens	-0.6369490	1.0000000	-0.6222580	-0.6751222	-0.5784278
G	0.8984927	-0.6222580	1.0000000	0.9766615	0.9812217
Hg	0.9308940	-0.6751222	0.9766615	1.0000000	0.9636563
V	0.8521970	-0.5784278	0.9812217	0.9636563	1.0000000
AB	0.8144991	-0.6186956	0.9576488	0.9441822	0.9626557
BR	0.5485267	-0.5598806	0.7362057	0.7238736	0.7405488
FL	0.5914534	-0.5479810	0.7225715	0.7191310	0.7230836
ST	0.8621775	-0.5917399	0.9847925	0.9700117	0.9955775

	AB	BR	FL	ST
Age	0.8144991	0.5485267	0.5914534	0.8621775
dens	-0.6186956	-0.5598806	-0.5479810	-0.5917399
G	0.9576488	0.7362057	0.7225715	0.9847925
Hg	0.9441822	0.7238736	0.7191310	0.9700117
V	0.9626557	0.7405488	0.7230836	0.9955775
AB	1.0000000	0.8822559	0.8623371	0.9669117
BR	0.8822559	1.0000000	0.9214561	0.7470985
FL	0.8623371	0.9214561	1.0000000	0.7253943
ST	0.9669117	0.7470985	0.7253943	1.0000000

```
>
```



```
library(corrplot)
corrplot(cor(sosny[,6:13]))
```

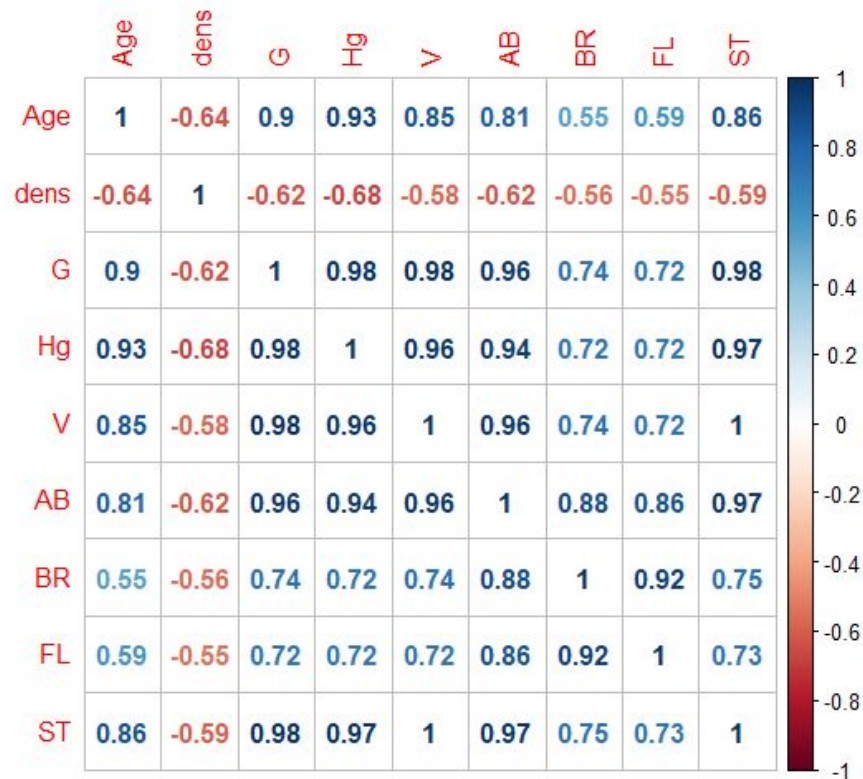


Logika składni:

`cor()` dla więcej niż 2 wektorów zwraca macierz korelacji

wrzucamy wynik `cor()` w funkcję `corrplot()`

```
corrplot(cor(sosny[,6:13]),method='num')
```



Jest wiele opcji prezentacji danych za pomocą tego pakietu:

<https://cran.r-project.org/web/packages/corrplot/vignettes/corrplot-intro.html>

Regresja

kompromis pomiędzy dwoma cechami

-przewidywanie (modelowanie) zmiennej zależnej (błąd modelu, np. RMSE)

-wyjaśnianie procesów (procent wyjaśnionej zmienności - R^2)

co tak naprawdę chcemy osiągnąć?

z czego możemy zrezygnować?

Korelacja a regresja

korelacja - miara współzależności

regresja - opis zależności

np. masa~średnicy $r=0,95$ masa= $10 \cdot \text{średnica} + 2$

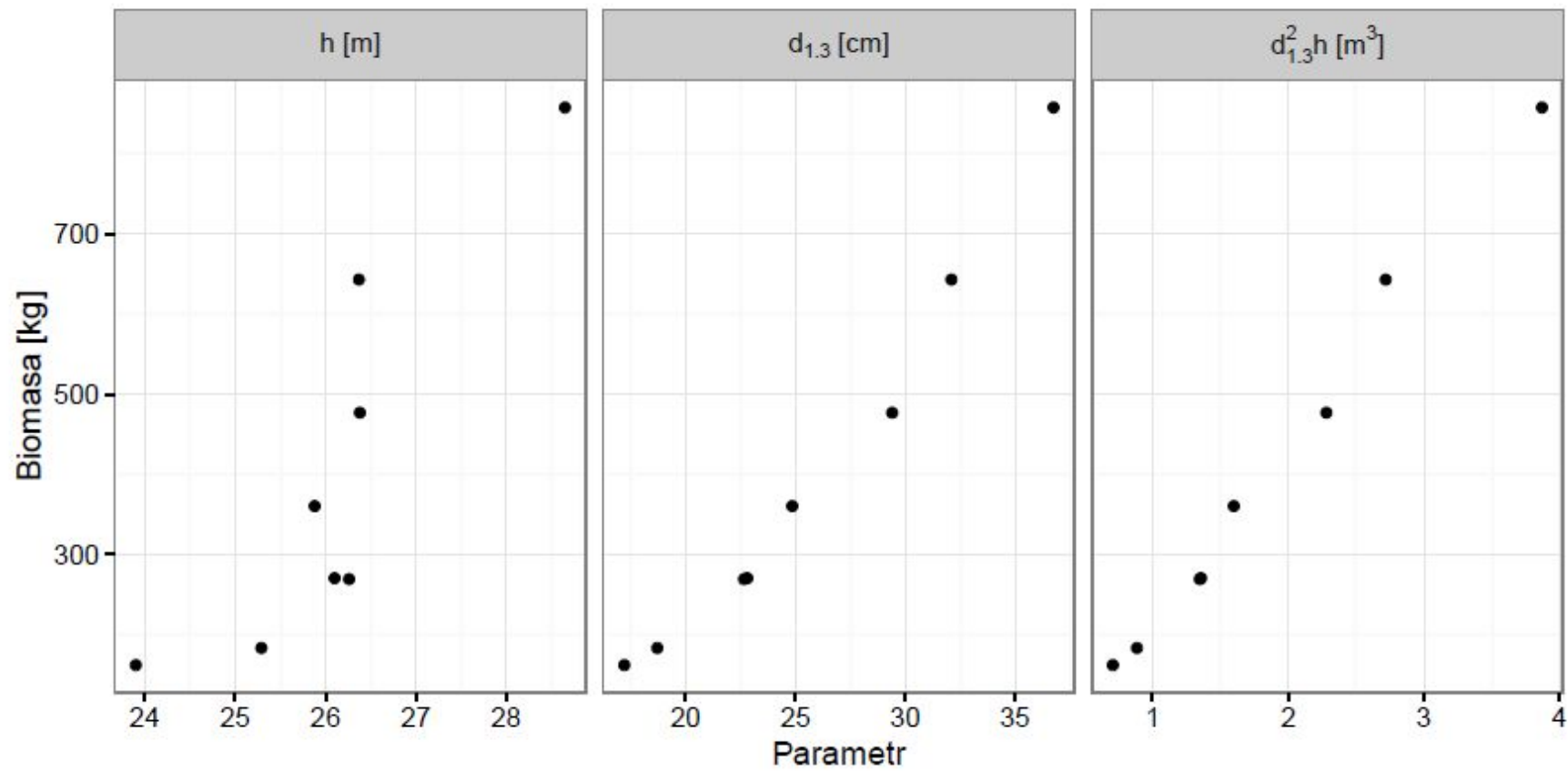
Regresja liniowa

Jak zmienia się masa drzewa wraz z przyrostem na grubość?

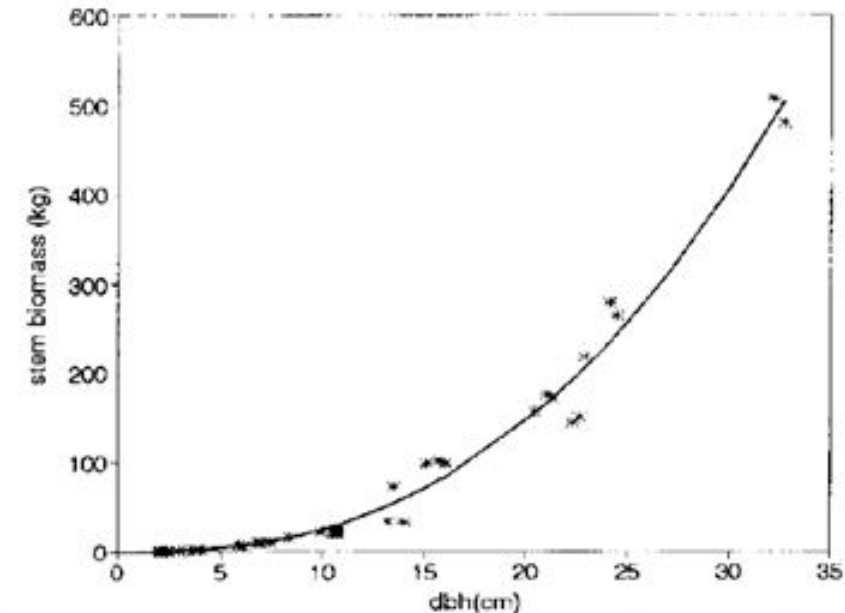
problem badawczy (poznanie tempa wzrostu)

problem aplikacyjny (możliwość estymacji)

?



Równania allometryczne



Ann. Sci. For. (1997) 54, 39-50
© Elsevier/INRA

39

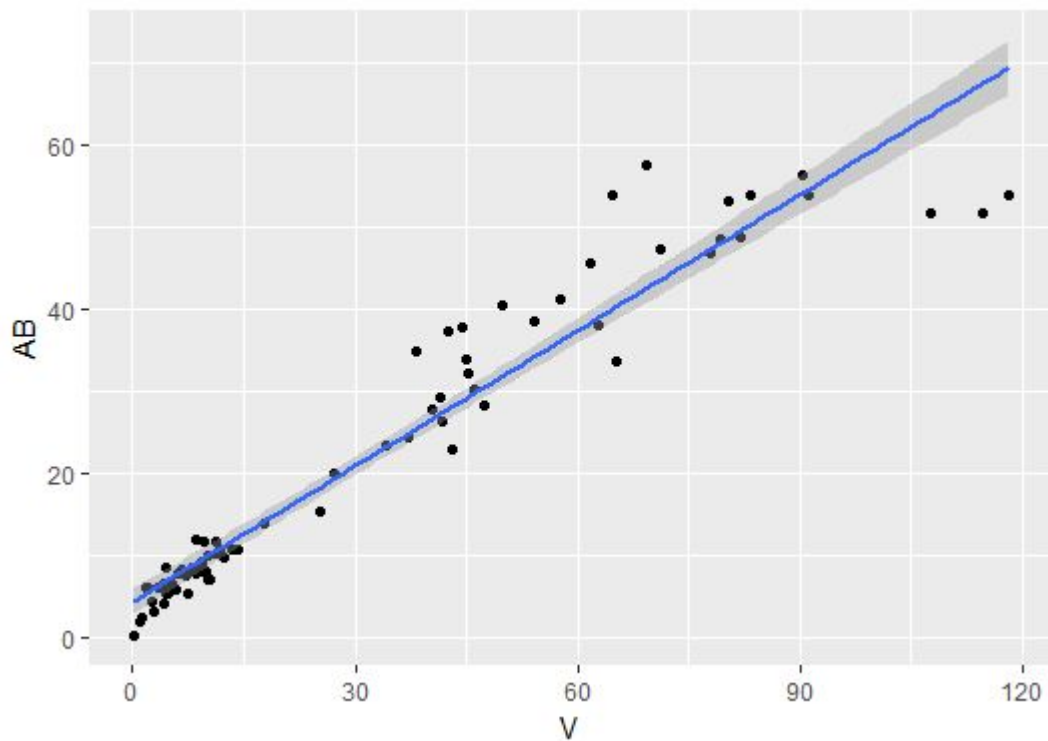
Original article

Allometric relationships for biomass and leaf area of
beech (*Fagus sylvatica* L.)

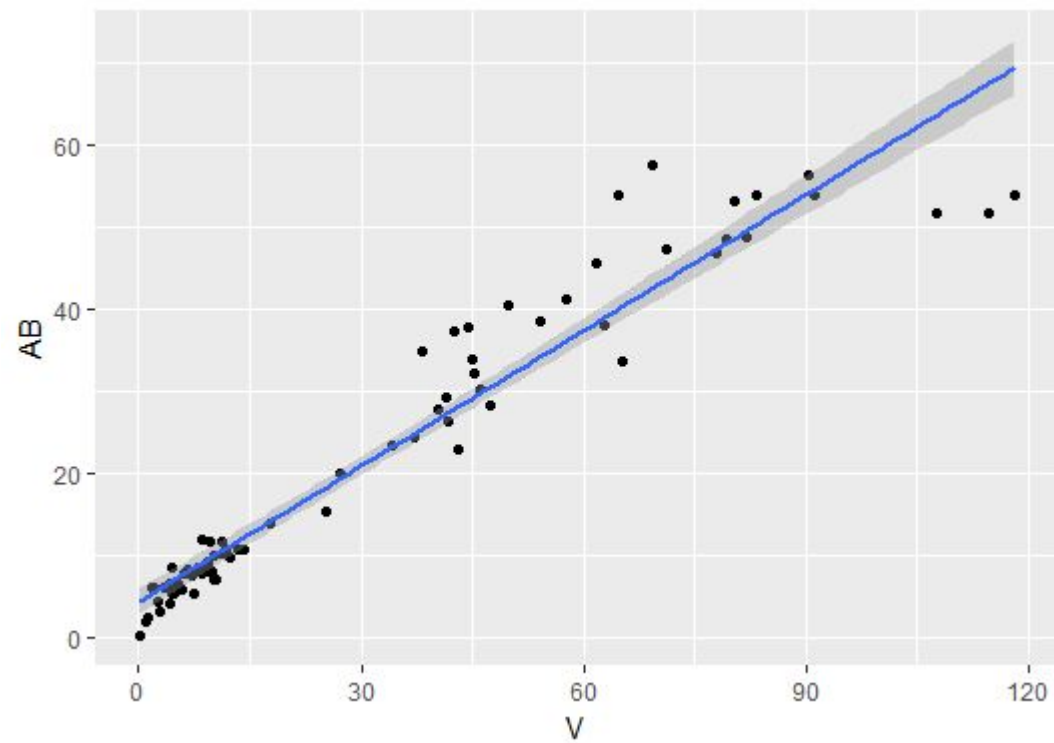
HH Bartelink



```
ggplot(sosny, aes(x=V,y=AB))+geom_point()+geom_smooth(method='lm')
```



jak to się stało?



Model liniowy

$$y=a*x+b$$

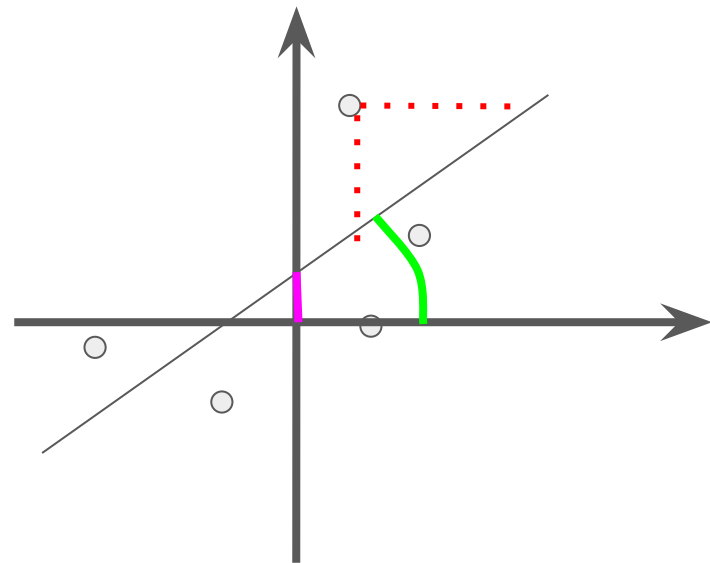
a - współczynnik kierunkowy, slope, regression coefficient, beta

nachylenie linii regresji (kąt)

b - wyraz wolny, intercept

punkt przecięcia z osią Y, położenie linii

dopasowanie - metodą najmniejszych kwadratów



Model liniowy

Y

zmienna zależna
odpowiedź (response)
zmienna modelowana

coś co chcemy
wymodelować

$$Y=ax+b$$

zapis matematyczny

$$Y\sim X$$

zapis w R
~ - tylda (pod Esc)

X

zmienna niezależna
predyktor

coś, co ma nam wyjaśniać Y

*ale nie parametr (parametr to a)

```
lm(V~AB,data=sosny)
```

Call:

```
lm(formula = AB ~ V, data = sosny)
```

Coefficients:

(Intercept)	V
4.3984	0.5498

$$AB = 4.3984 + 0.5498 \cdot V$$

summary(lm(AB~V,data=sosny))

```
> summary(lm(AB~V,data=sosny))
```

```
Call:
lm(formula = AB ~ V, data = sosny)
```

Residuals:

Min	1Q	Median	3Q	Max
-15.7506	-1.7161	-0.3876	1.1526	15.1380

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.39843	0.79045	5.564	3.88e-07 ***
V	0.54978	0.01785	30.794	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.915 on 75 degrees of freedom
Multiple R-squared: 0.9267, Adjusted R-squared: 0.9257
F-statistic: 948.3 on 1 and 75 DF, p-value: < 2.2e-16

```
> |
```

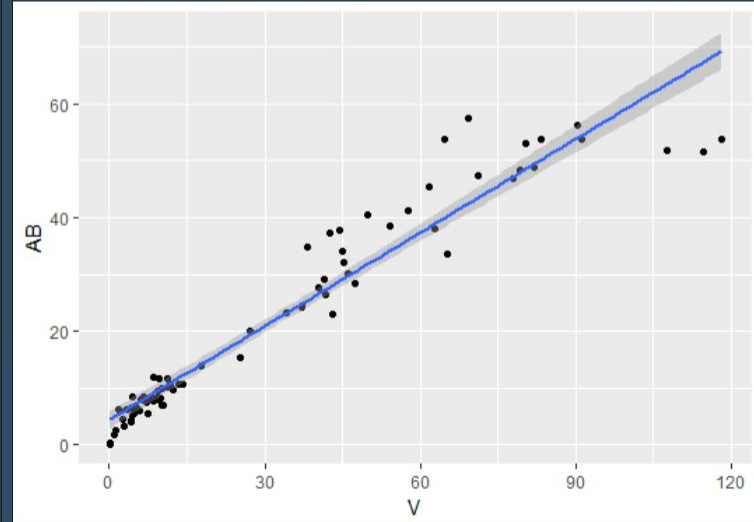
transy 1872 obs. of 2 variables

Functions

se function (x)

Files Plots Packages Help Viewer

Zoom Export Publish



Biologiczne znaczenie - effect size!

<https://www.amstat.org/asa/files/pdfs/P-ValueStatement.pdf>

<https://amstat.tandfonline.com/doi/pdf/10.1080/00031305.2016.1154108?needAccess=true>

Ronald L. Wasserstein, Allen L. Schirm & Nicole A. Lazar (2019) Moving to a World Beyond “ $p < 0.05$ ”, The American Statistician, 73:sup1, 1-19, DOI: 10.1080/00031305.2019.1583913

nieistotny biologicznie efekt (3%) - $p < 0.00001$ przy $n=300$

istotny efekt (800%) - $p > 0.05$ przy $n=3$

summary(lm(AB~V,data=sosny))

```
> summary(lm(AB~V,data=sosny))

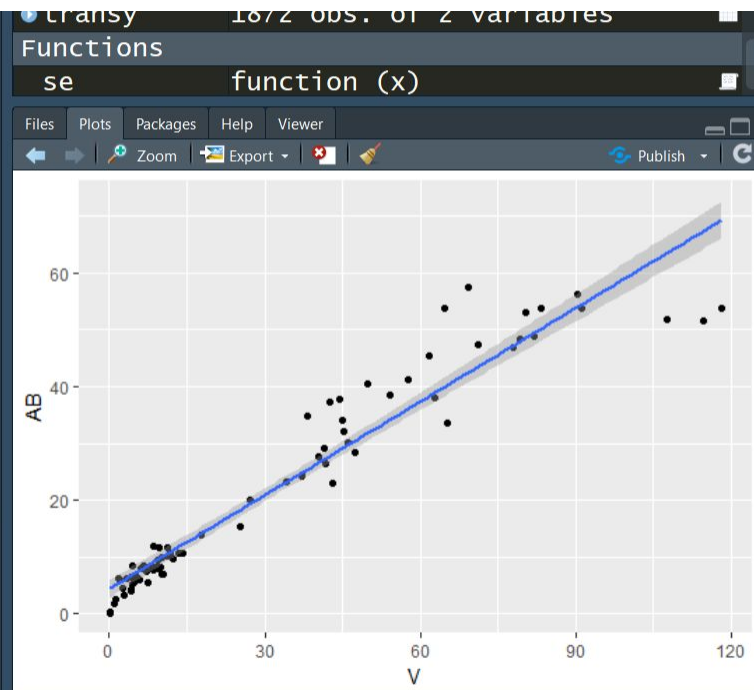
Call:
lm(formula = AB ~ V, data = sosny)

Residuals:
    Min       1Q   Median       3Q      Max
-15.7506  -1.7161  -0.3876   1.1526  15.1380

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.39843    0.79045   5.564 3.88e-07 ***
V            0.54978    0.01785  30.794 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.915 on 75 degrees of freedom
Multiple R-squared:  0.9267,    Adjusted R-squared:  0.9257
F-statistic: 948.3 on 1 and 75 DF,  p-value: < 2.2e-16

> |
```



summary(lm(AB~V,data=sosny))

```
> summary(lm(AB~V,data=sosny))
```

Call:
lm(formula = AB ~ V, data = sosny)

Residuals:

Min	1Q	Median	3Q	Max
-15.7506	-1.7161	-0.3876	1.1526	15.1380

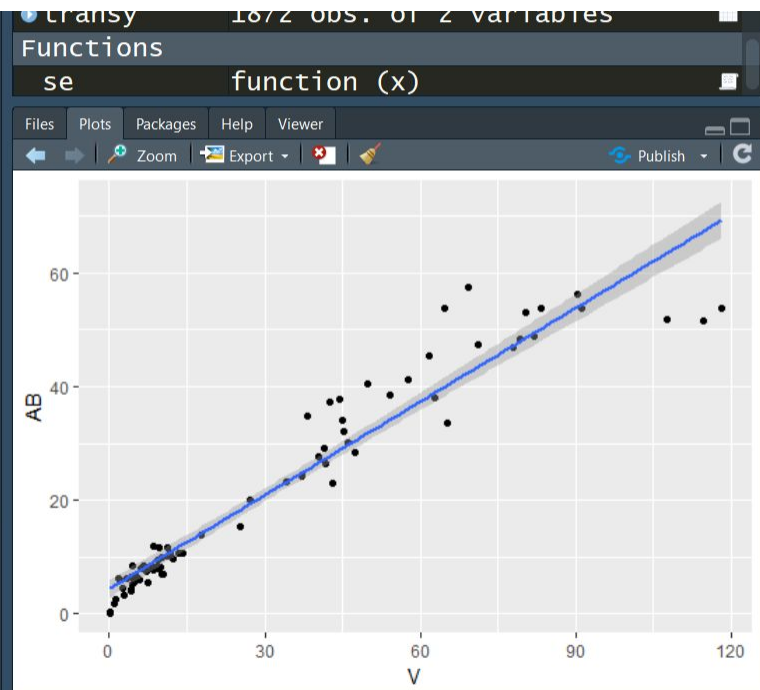
Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.39843	0.79045	5.564	3.88e-07 ***
V	0.54978	0.01785	30.794	< 2e-16 ***

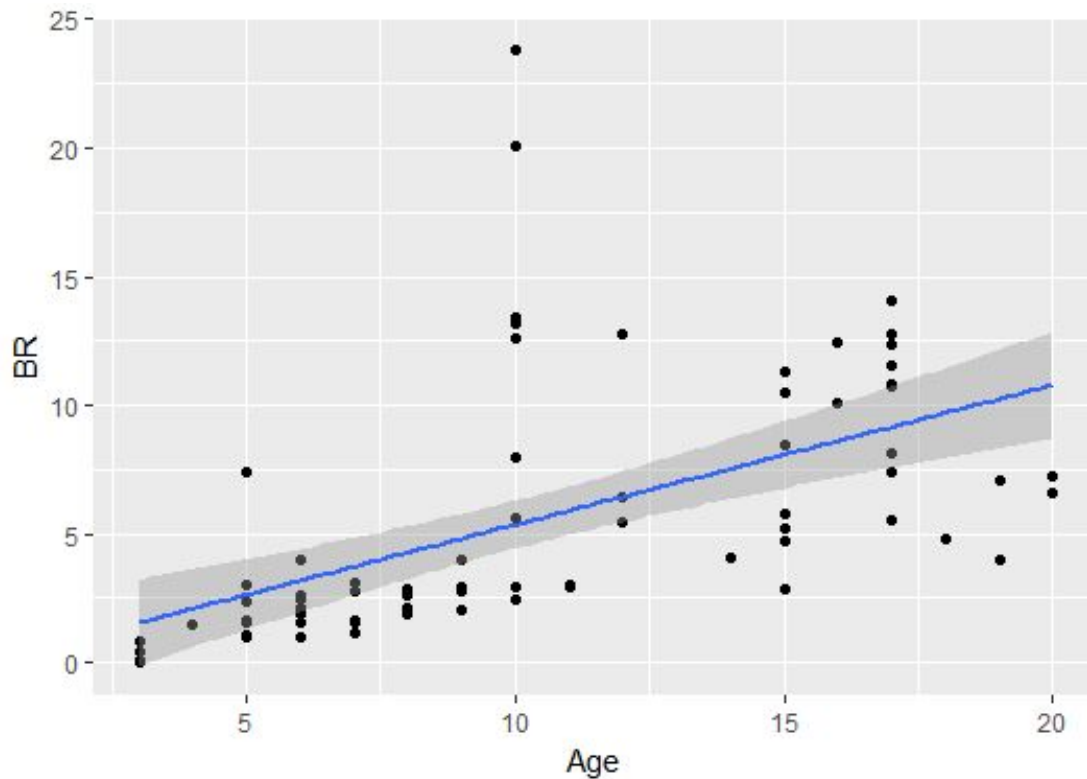
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.915 on 75 degrees of freedom
Multiple R-squared: 0.9267, Adjusted R-squared: 0.9257
F-statistic: 948.3 on 1 and 75 DF, p-value: < 2.2e-16

```
> |
```



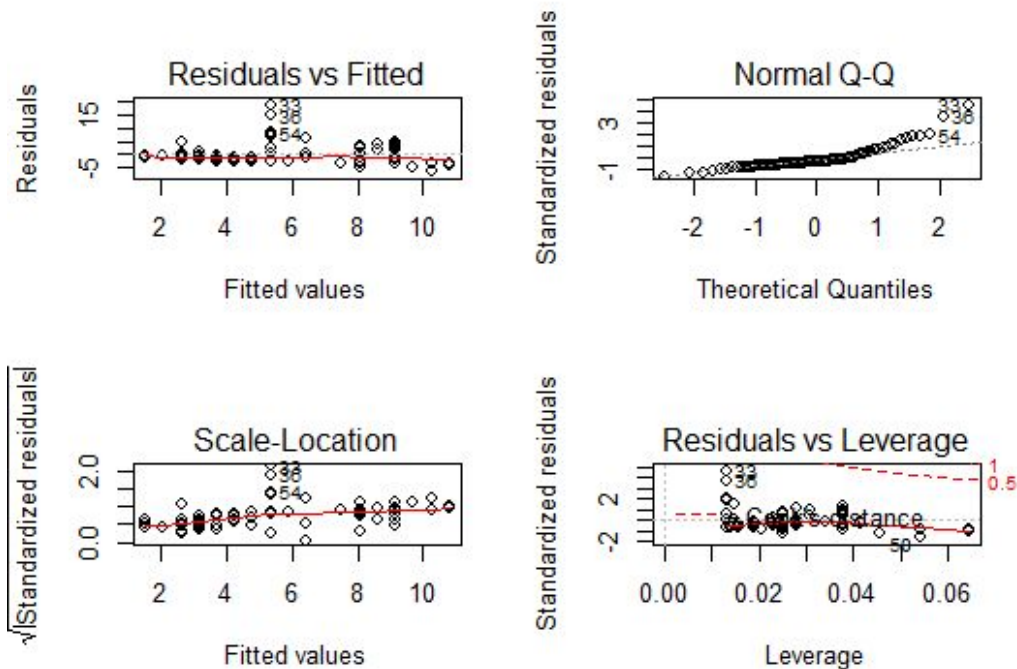
a tutaj? lepszy czy gorszy?



```
)  
> summary(lm(BR~Age,data=sosny))  
  
call:  
lm(formula = BR ~ Age, data = sosny)  
  
Residuals:  
    Min       1Q   Median       3Q      Max   
-6.2556 -2.2313 -1.0834  0.7993 18.5083  
  
Coefficients:  
            Estimate Std. Error t value Pr(>|t|)      
(Intercept) -0.08740     1.08768   -0.080    0.936      
Age          0.54361     0.09568    5.681 2.41e-07 ***  
---  
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 4.11 on 75 degrees of freedom  
Multiple R-squared:  0.3009,    Adjusted R-squared:  0.2916  
F-statistic: 32.28 on 1 and 75 DF,  p-value: 2.411e-07  
  
> |
```

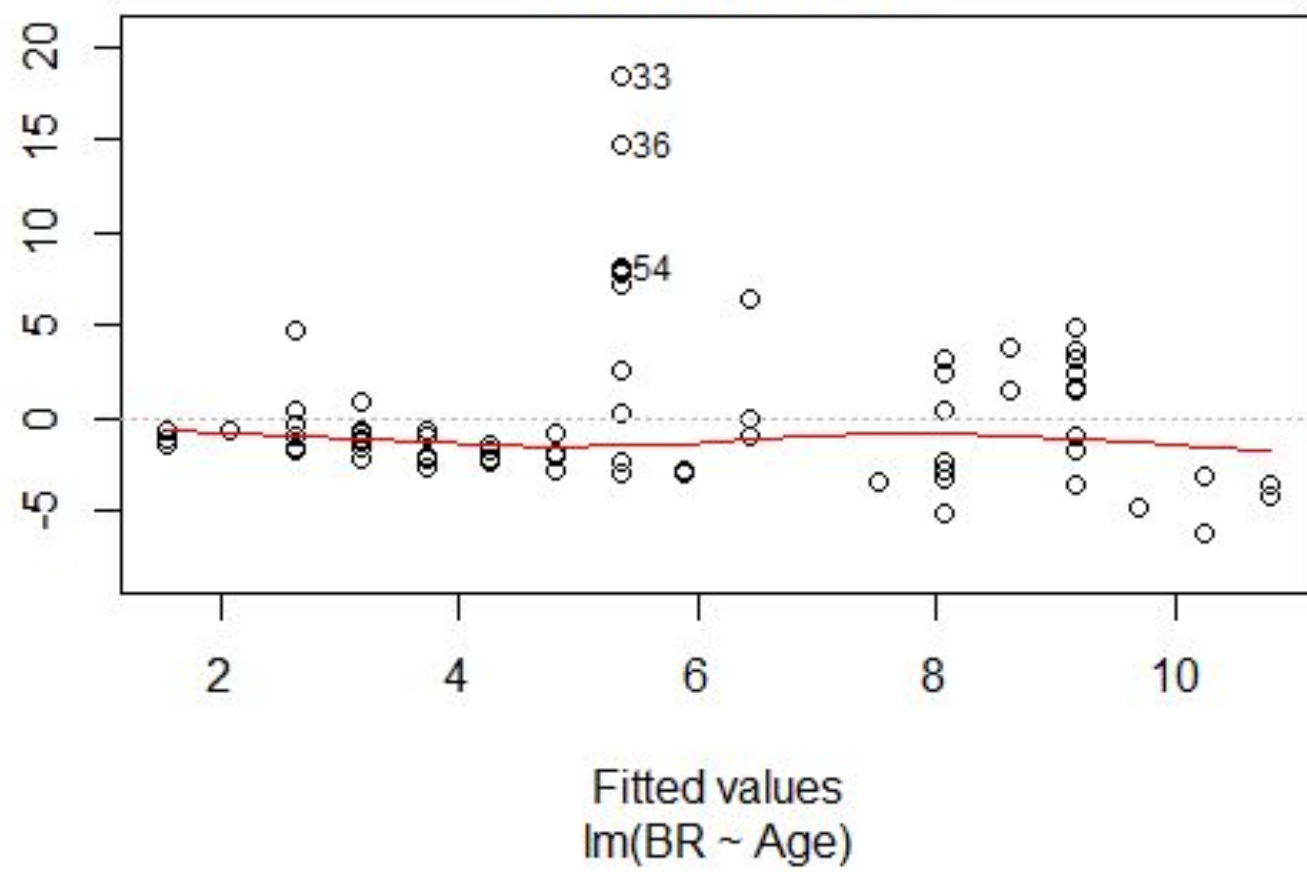
Diagnostyka modeli

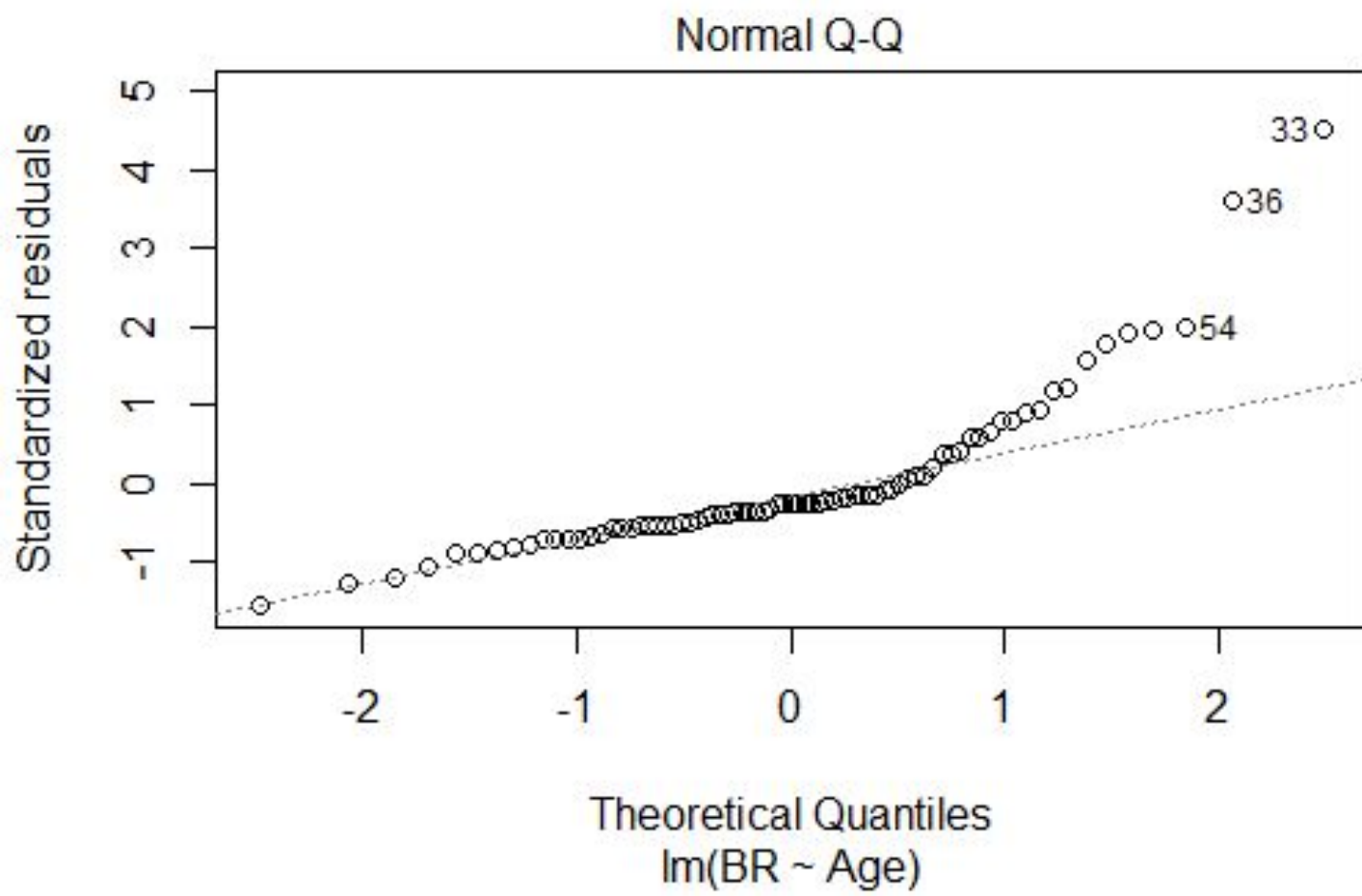
```
par(mfrow=c(2,2)) #podział wykresu na 4  
plot(lm(BR~Age,data=sosny))
```

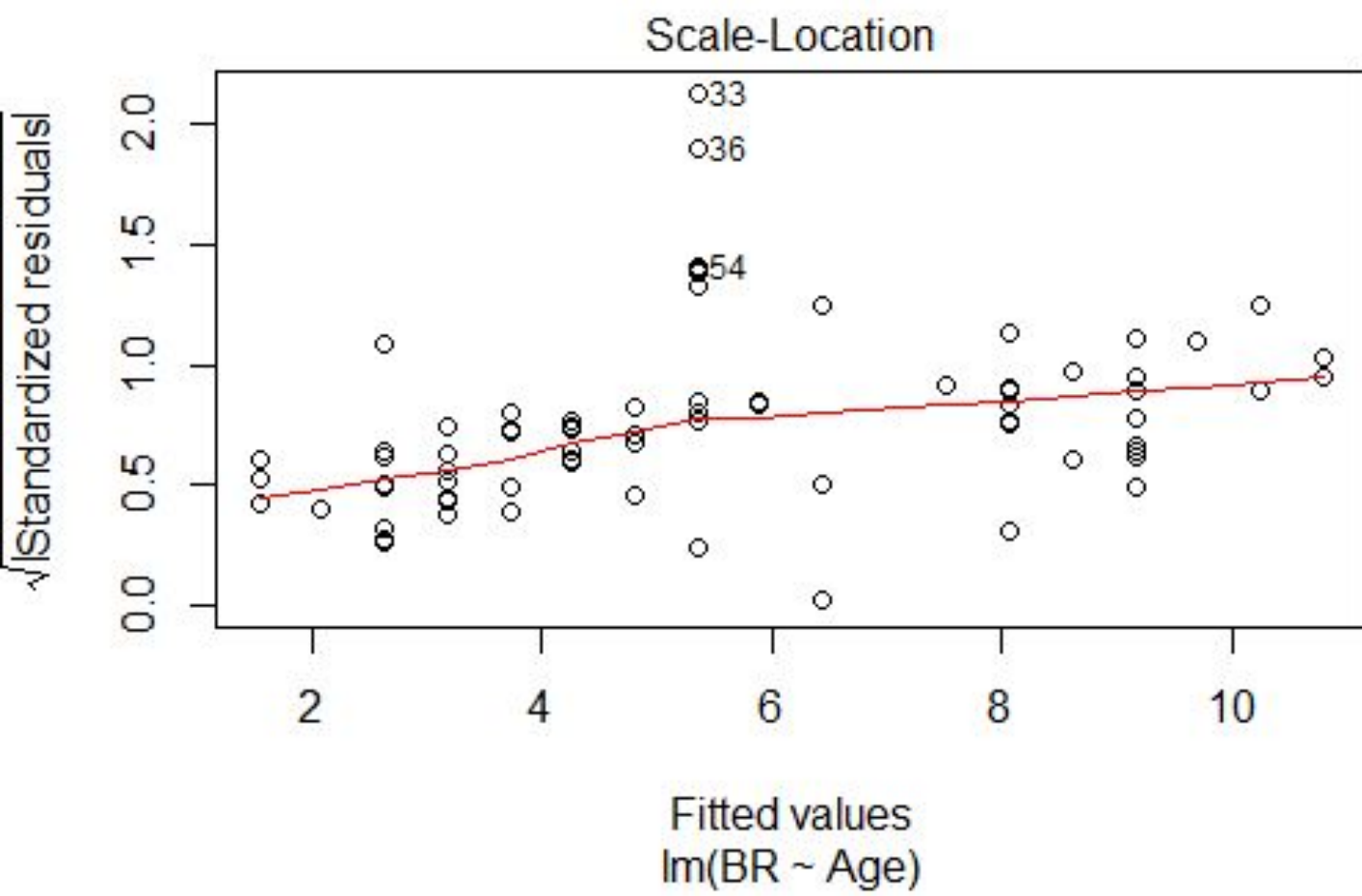


Residuals

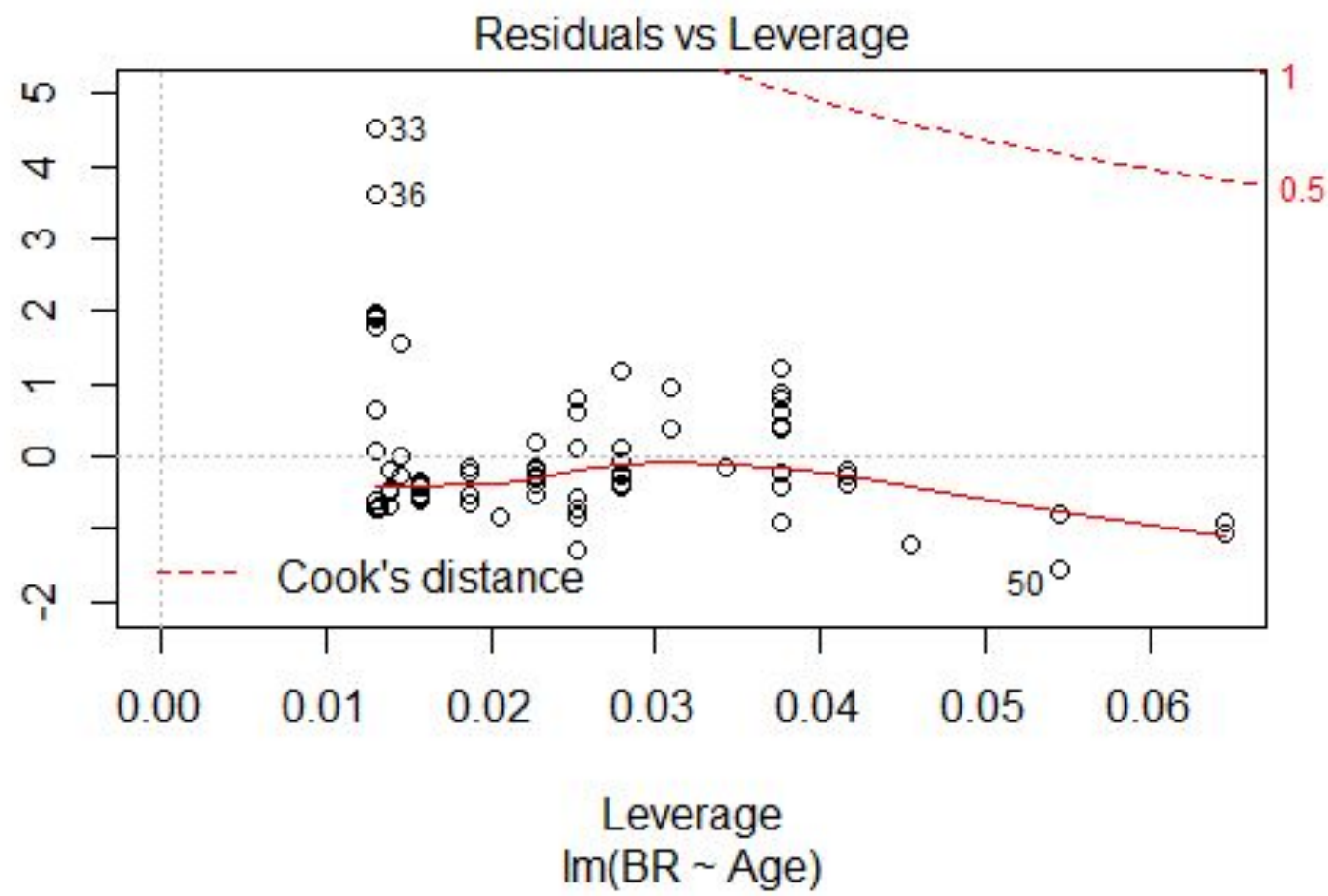
Residuals vs Fitted



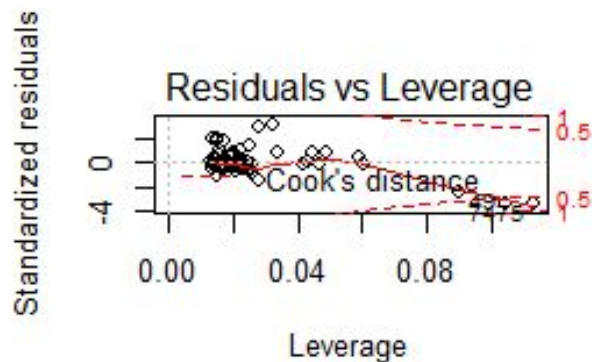
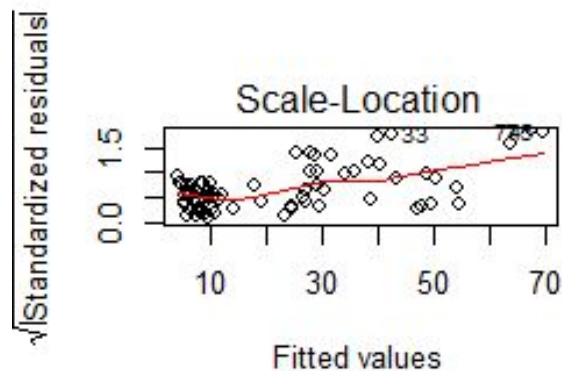
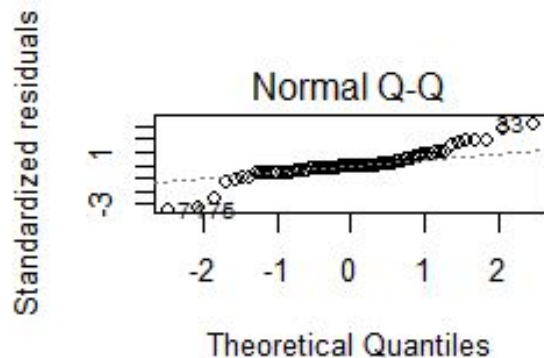
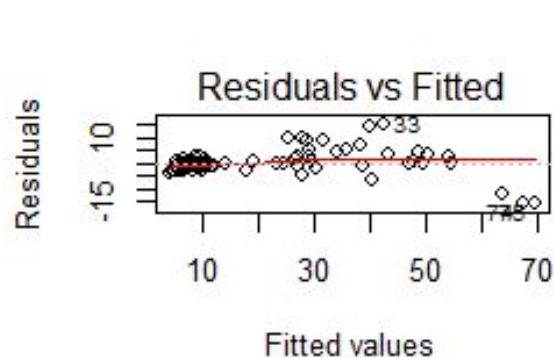




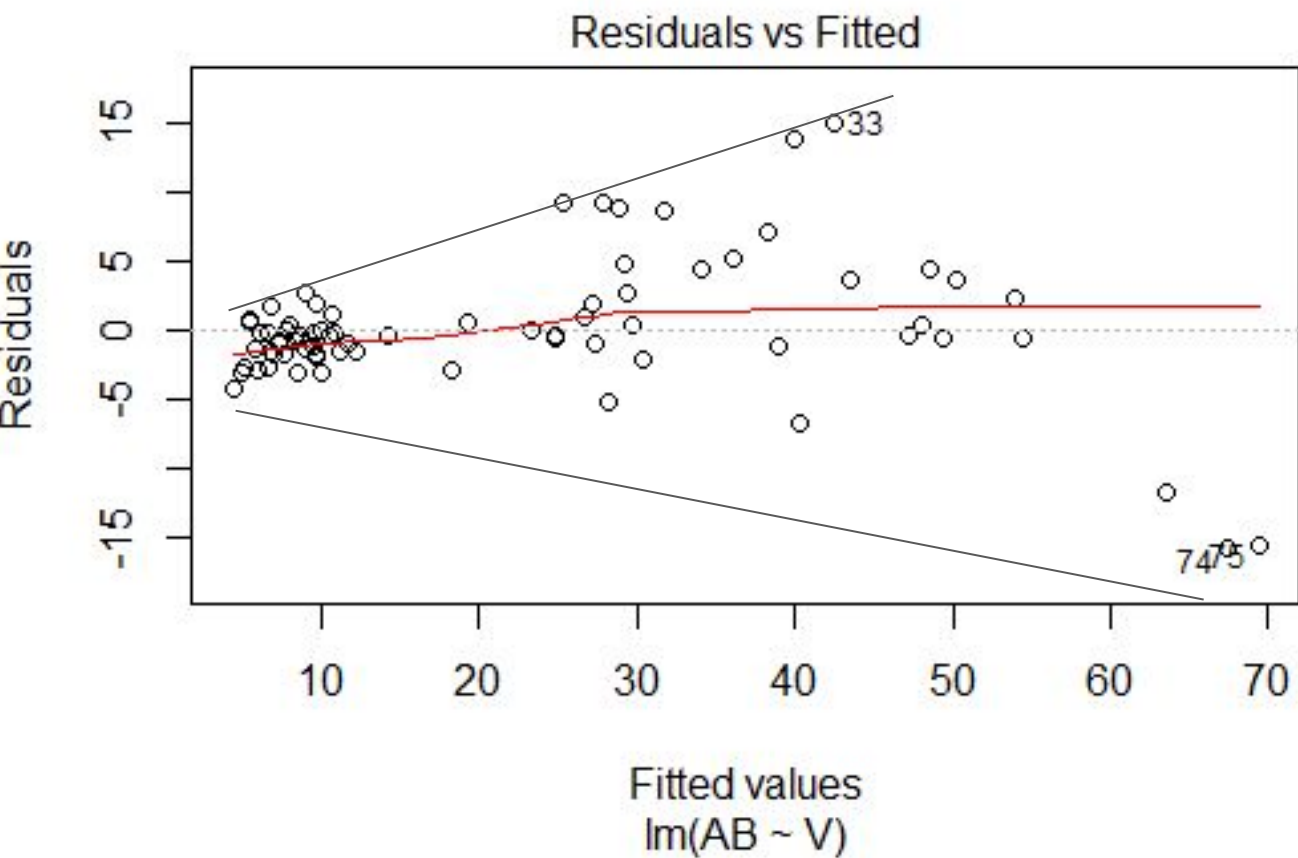
Standardized residuals

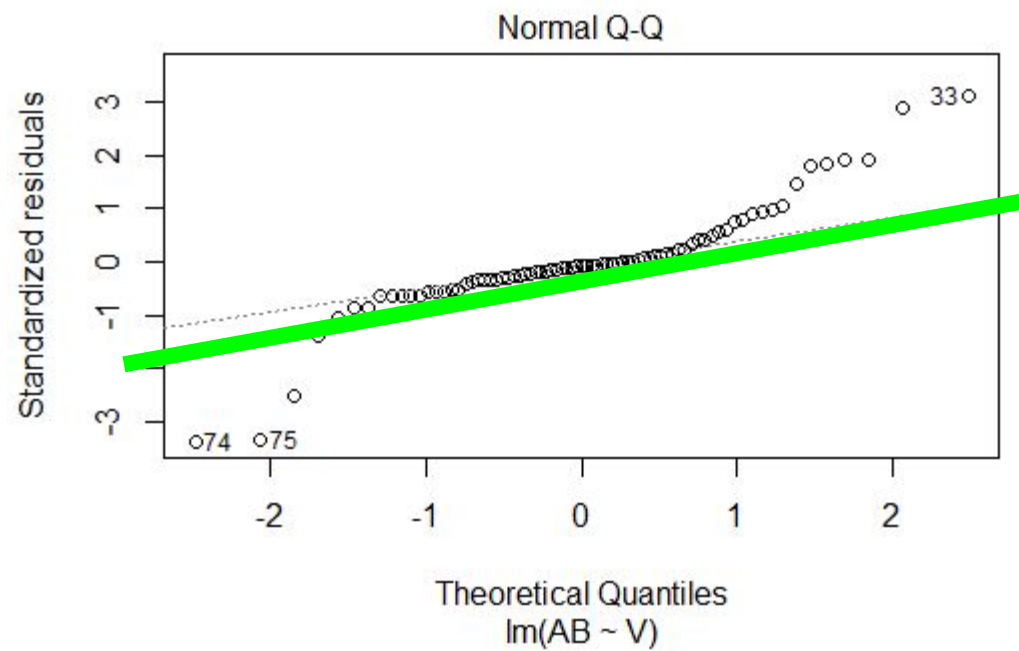


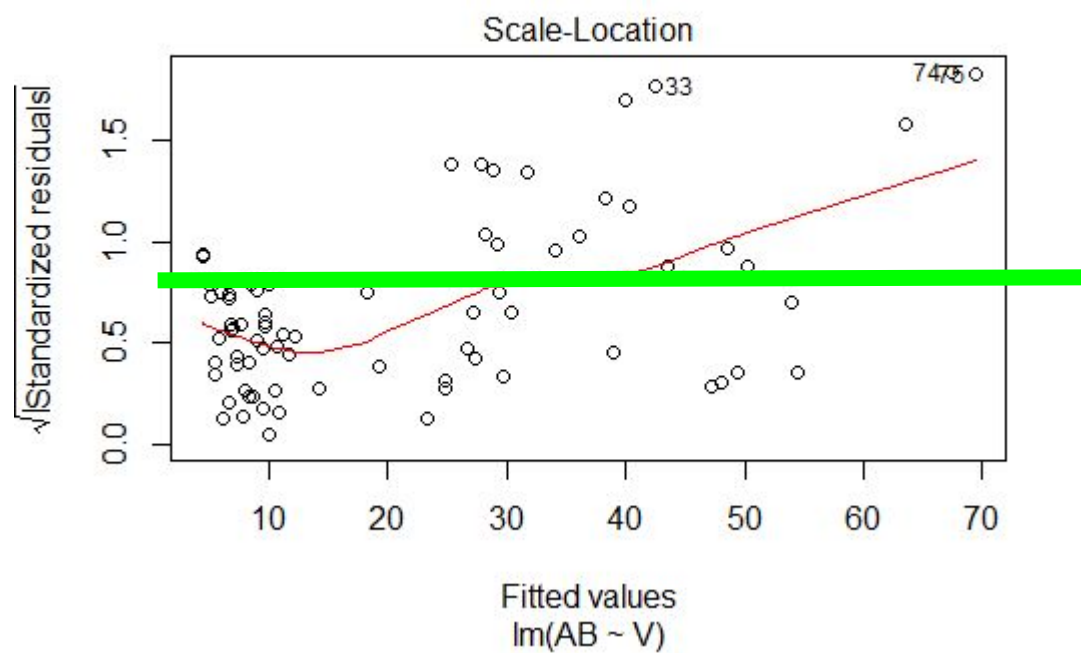
Jak wygląda dobry model? AB~V...

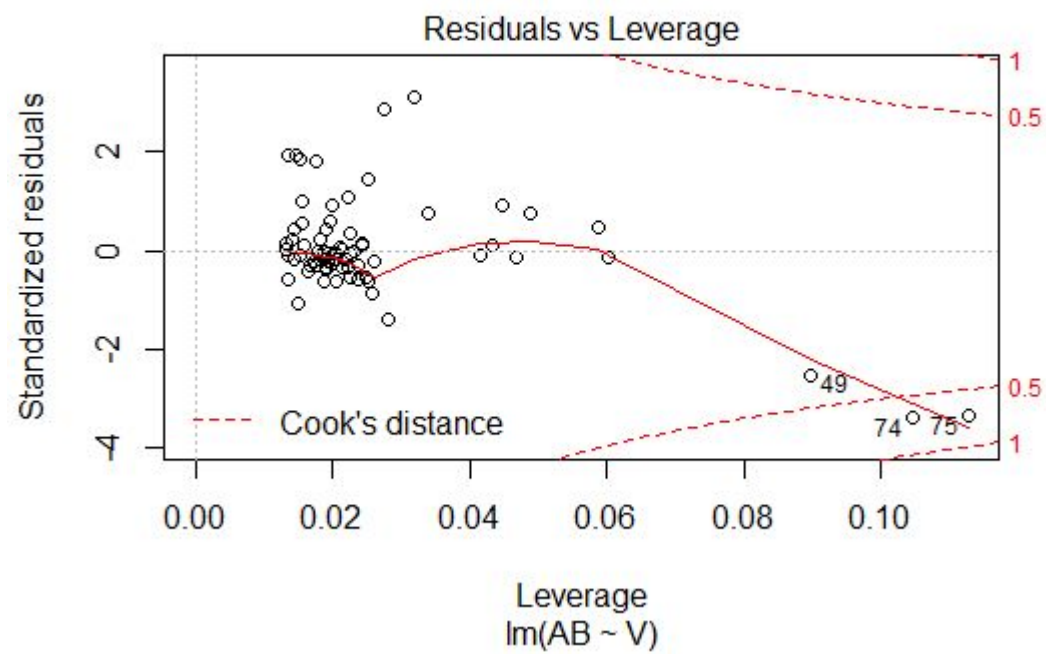


heteroskedastyczność!









Co się stało?

model gorszy wg R^2 ($BR \sim \text{age}$) okazał się mieć mniej problemów

model lepszy ($AB \sim V$) - heteroskedastyczność

Czy można tylko oceniać na podstawie R^2 ?

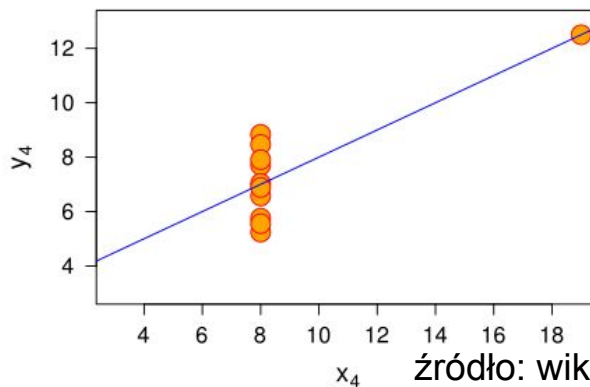
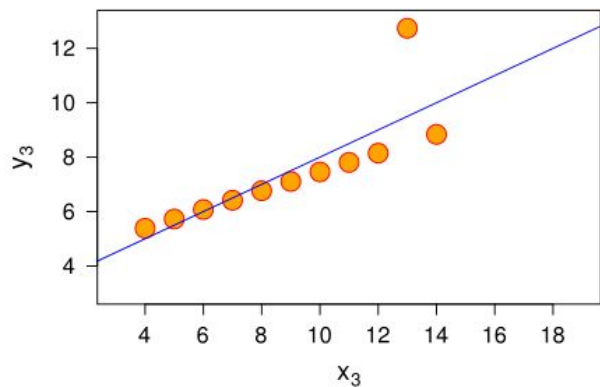
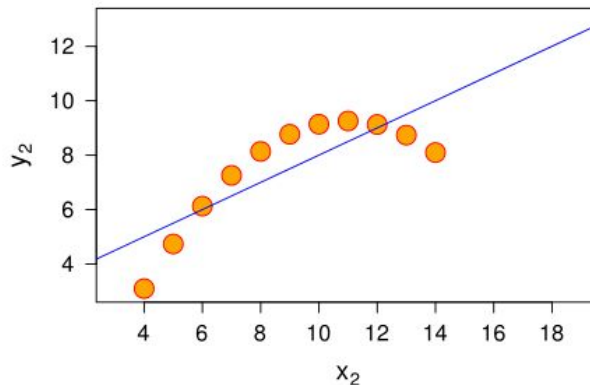
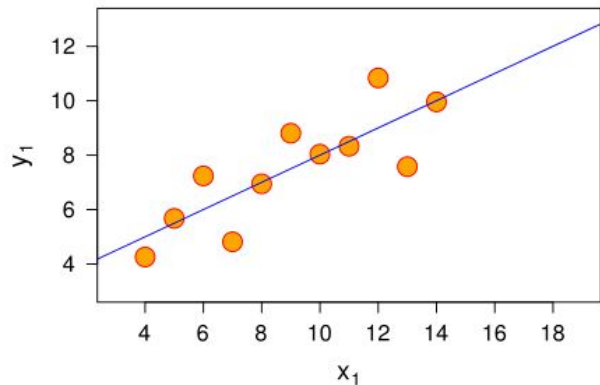
NIE!

R^2 nie uwzględnia kształtu rozkładu, heteroskedastyczności, outlierów...

kwadrat Anscombe'a

Kwadrat Ascombe'a

średnia $y=7,5$ średnia $x=9$
współczynnik $r^2=0,816$
równanie regresji:
 $y=3+0,5 \cdot x$



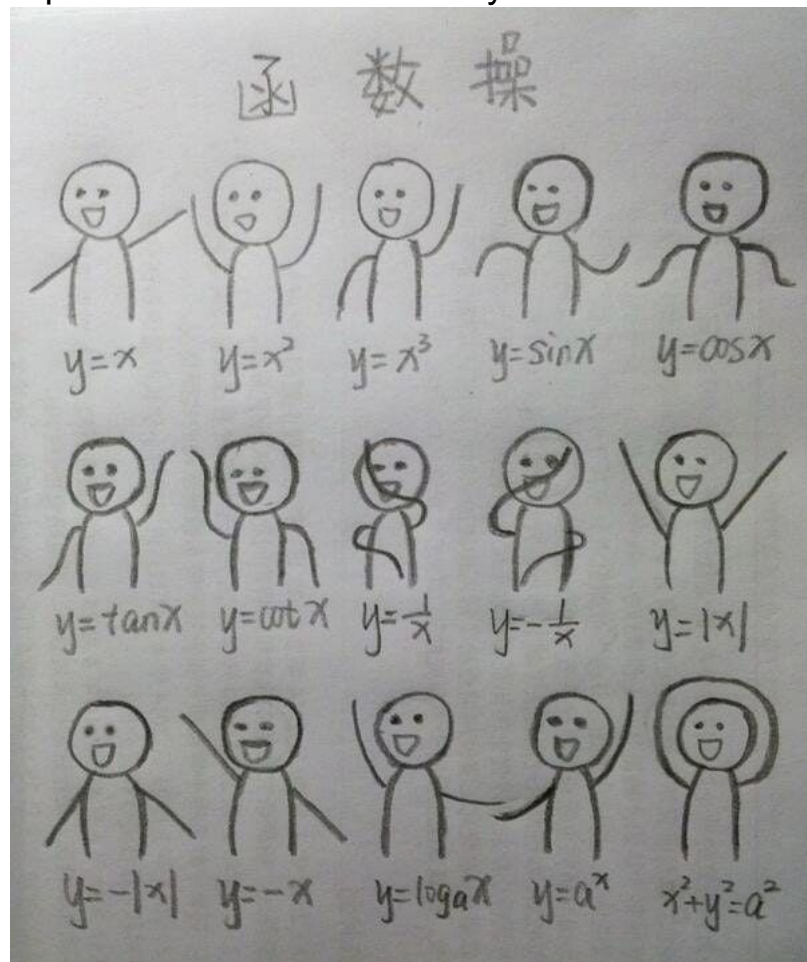
Co zrobić jak nie jest tak różowo?

transformacje (log, skalowanie...)

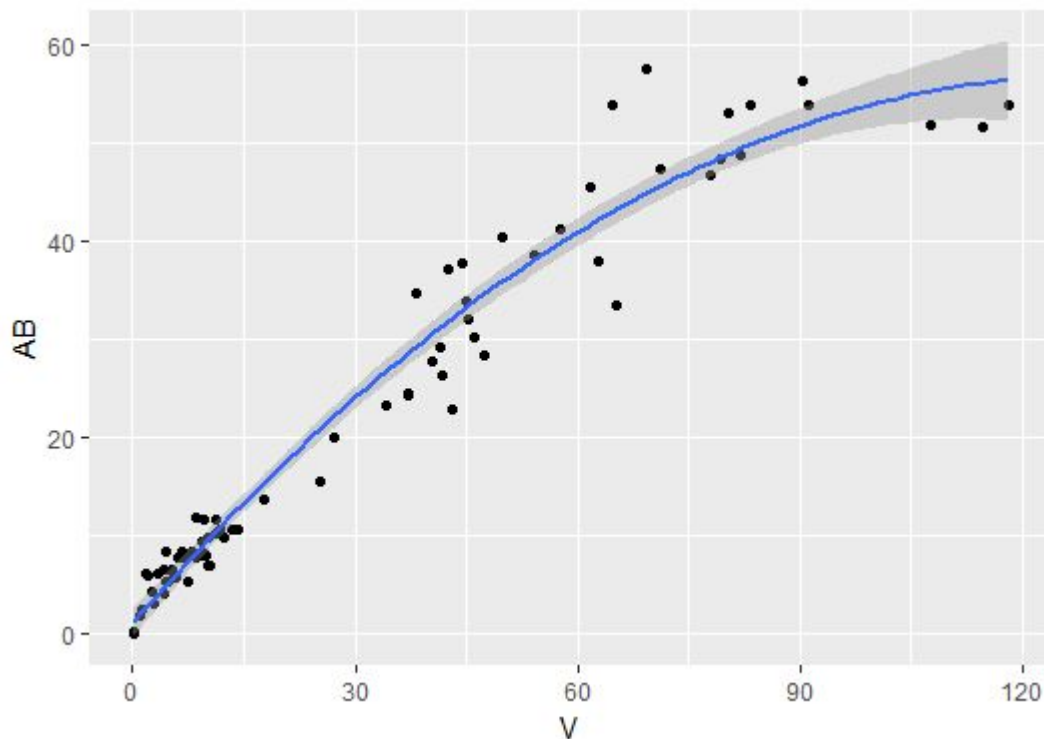
inny typ rozkładu

inny typ modelu

https://www.reddit.com/r/funny/comments/21h32s/dancing_math/



Parabole tańczą $y=ax^2+bx+c$



```
ggplot(sosny, aes(x=V,y=AB))+geom_point()+  
geom_smooth(method='lm',formula=y~poly(x,2))
```

```
> summary(lm(AB~poly(v,2),data=sosny))
```

```
Call:
lm(formula = AB ~ poly(v, 2), data = sosny)
```

```
Residuals:
```

	Min	1Q	Median	3Q	Max
	-9.6997	-1.9382	-0.1007	1.6296	12.7176

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	21.5752	0.4149	52.004	< 2e-16	***
poly(v, 2)1	151.3404	3.6405	41.571	< 2e-16	***
poly(v, 2)2	-28.8228	3.6405	-7.917	1.86e-11	***

```
---
```

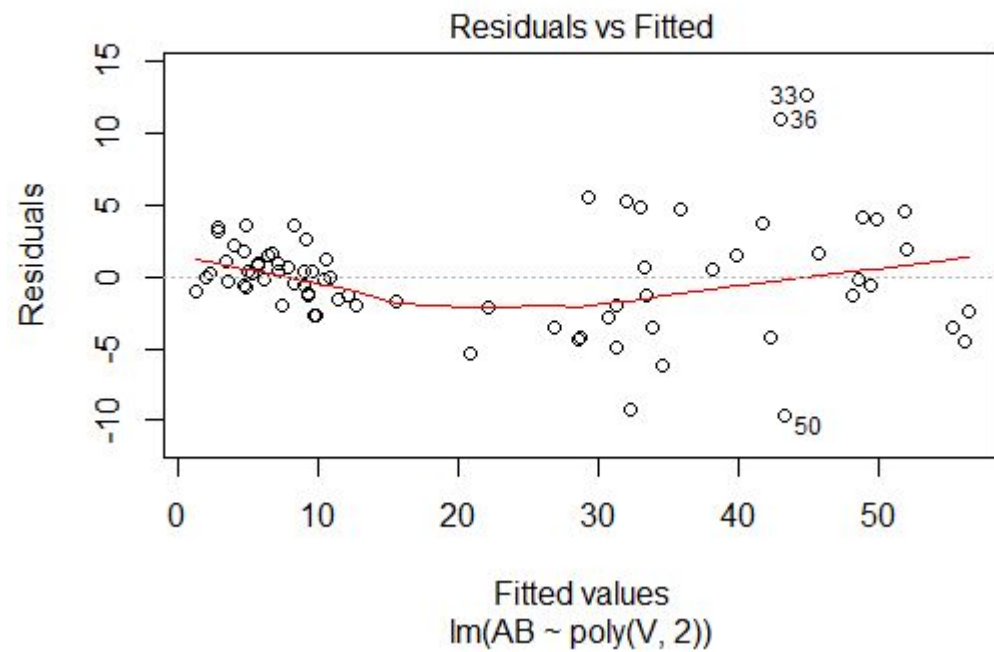
```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

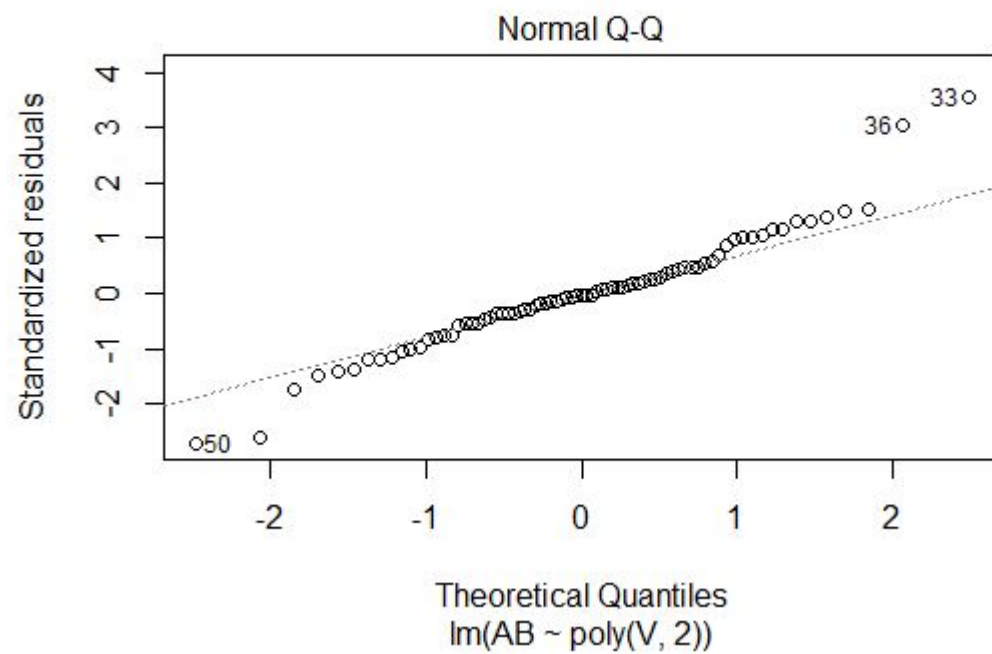
```
Residual standard error: 3.64 on 74 degrees of freedom
```

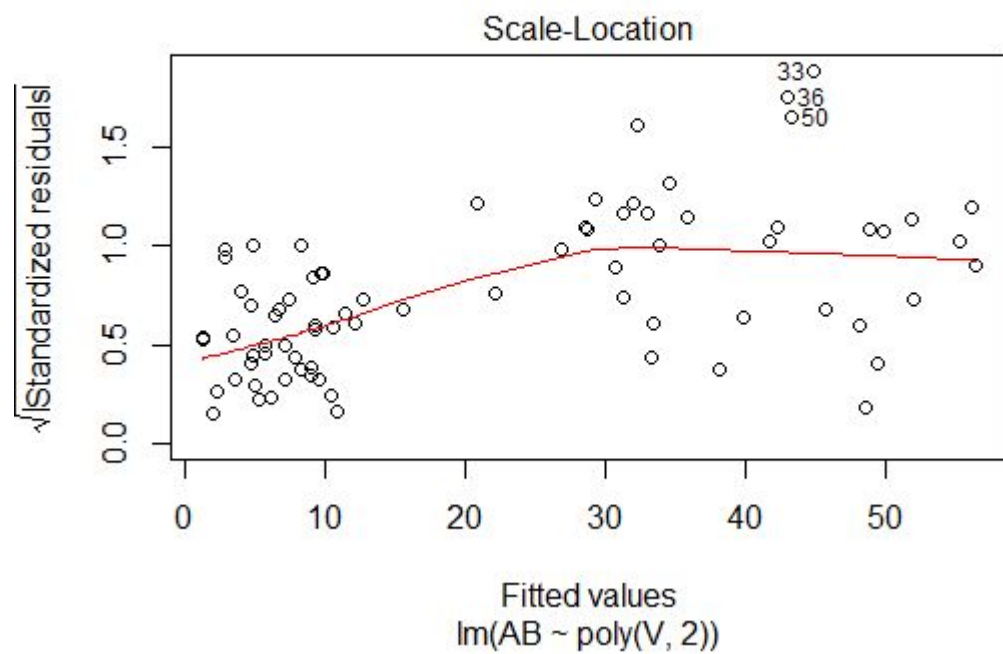
```
Multiple R-squared:  0.9603,    Adjusted R-squared:  0.9592
```

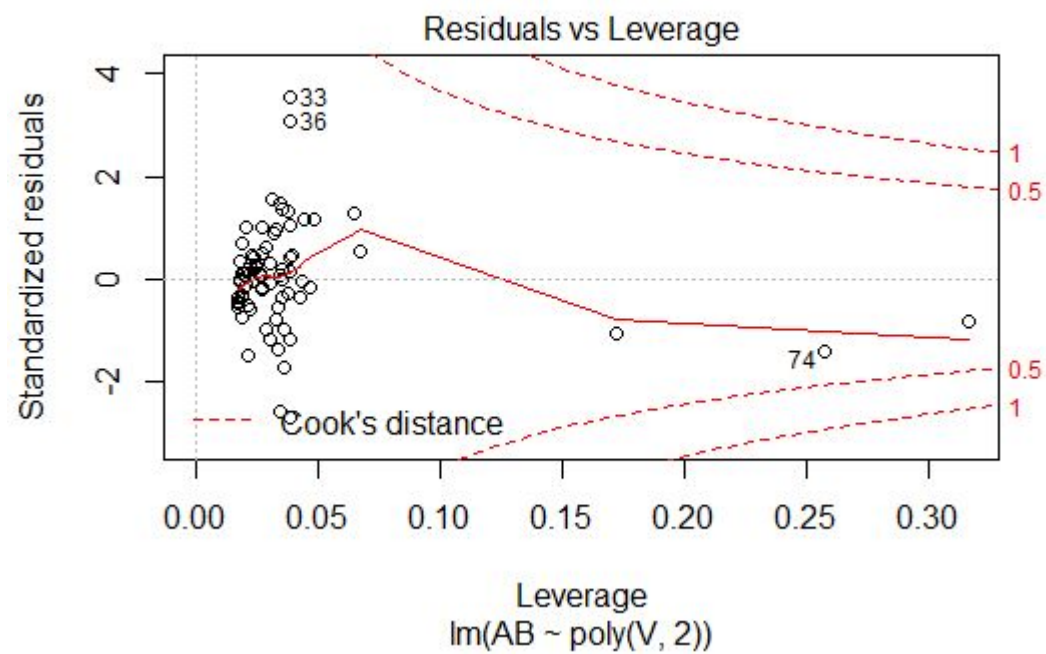
```
F-statistic: 895.4 on 2 and 74 DF,  p-value: < 2.2e-16
```

```
> | AB=-28,8228*V^2+151,3404*V+21,5752
```









Lepiej:)

Jak głęboko wchodzić?

Cel: predykcja/eksploracja?

Zastosowanie i oczekiwana dokładność

Przyzwolenie na błędy?

Ilość analizowanych danych...

inne miary jakości

AIC - kryterium informacyjne Akaikego

miara jakości dopasowania modelu

wartości można porównywać w ramach tej samej zmiennej objaśnianej

do czego może to służyć?

porównanie jakości modeli z modelem zerowym

model zerowy - intercept-only, $Y \sim 1$, czyli wstawiamy wszędzie średnią

porównanie jakości dwóch modeli między sobą

```
> AIC(lm(AB~1,data=sosny)) #model zerowy
```

```
[1] 666.9126
```

```
> AIC(lm(AB~V,data=sosny)) #model liniowy
```

```
[1] 467.6902
```

```
> AIC(lm(AB~poly(V,2),data=sosny)) #model kwadratowy
```

```
[1] 422.4431
```

jak prezentować wyniki?

Table 7
Relationships between tree stand characteristics (predictors) and BCEFs for particular biomass components [Mg m^{-3}].

Biomass component	Predictor	Model type (Eq. no.)	a	SE	b	SE	c	SE	RMSE	R ²	AIC	AIC ₀
AB	A	(12)	0.61691	0.06542	5.33448	1.03961	0.35147	0.05317	< 0.0001	0.709	44.473	134.220
	H _g	(12)	0.62412	0.05746	3.26428	0.35316	0.87235	0.10597	< 0.0001	0.729	39.084	–
	N	(11)	0.00376	0.00537	0.63451	0.15950	–	–	0.0119	0.186	120.599	–
	V	(12)	0.67209	0.04559	2.09454	0.15847	0.22621	0.02754	< 0.0001	0.764	28.410	–
BR	A	(12)	0.07917	0.07343	0.99264	0.22166	0.18209	0.06920	< 0.0001	0.438	–42.693	–2.843
	H _g	(12)	0.10830	0.06310	0.69410	0.13060	0.45800	0.18110	< 0.0001	0.378	–34.930	–
	N	(13)	0.46840	0.07462	–1080.0	422.1	–	–	< 0.0001	0.081	–7.283	–
	V	(12)	0.16099	0.03062	0.55175	0.08779	0.14350	0.03963	< 0.0001	0.428	–41.339	–
FL	A	(12)	0.09562	0.05773	5.67080	1.45618	0.42193	0.07180	< 0.0001	0.637	42.380	115.452
	H _g	(12)	0.08925	0.05133	3.09855	0.38049	1.01347	0.13006	< 0.0001	0.682	31.089	–
	N	(13)	0.91070	0.15630	–3192.17130	884.35630	–	–	< 0.0001	0.150	32.470	–
	V	(12)	0.12295	0.04423	1.86903	0.16629	0.26766	0.03813	< 0.0001	0.700	27.958	–
ST	A	(12)	0.37645	0.01009	6.30763	9.27127	1.11956	0.48019	< 0.0001	0.316	–174.461	–149.642
	H _g	(12)	0.38054	0.00548	3.65437	1.24671	5.30087	0.92841	< 0.0001	0.736	–246.734	–
	N	(11)	0.14574	0.08838	0.11308	0.06883	–	–	< 0.0001	0.035	–150.317	–
	V	(13)	0.37748	0.00526	0.03547	0.00240	–	–	< 0.0001	0.747	–252.212	–

*RMSE - pierwiastek średniego błędu kwadratowego

jakie są odchyły

$\sqrt{\text{sum}(\text{zmienna} - \text{predict}(\text{model}))^2 / \text{length}(\text{zmienna})}$

pierwiastek sumy kwadratów odchyleń / liczba obserwacji -

One function to rule the all - predict

Modele są po to, aby je wykorzystywać

```
> model<-lm(AB~poly(V,2),data=sosny)
```

```
> nowedane<-data.frame(V=c(1,2,3,10))
```

```
> predict(model,nowedane)
```

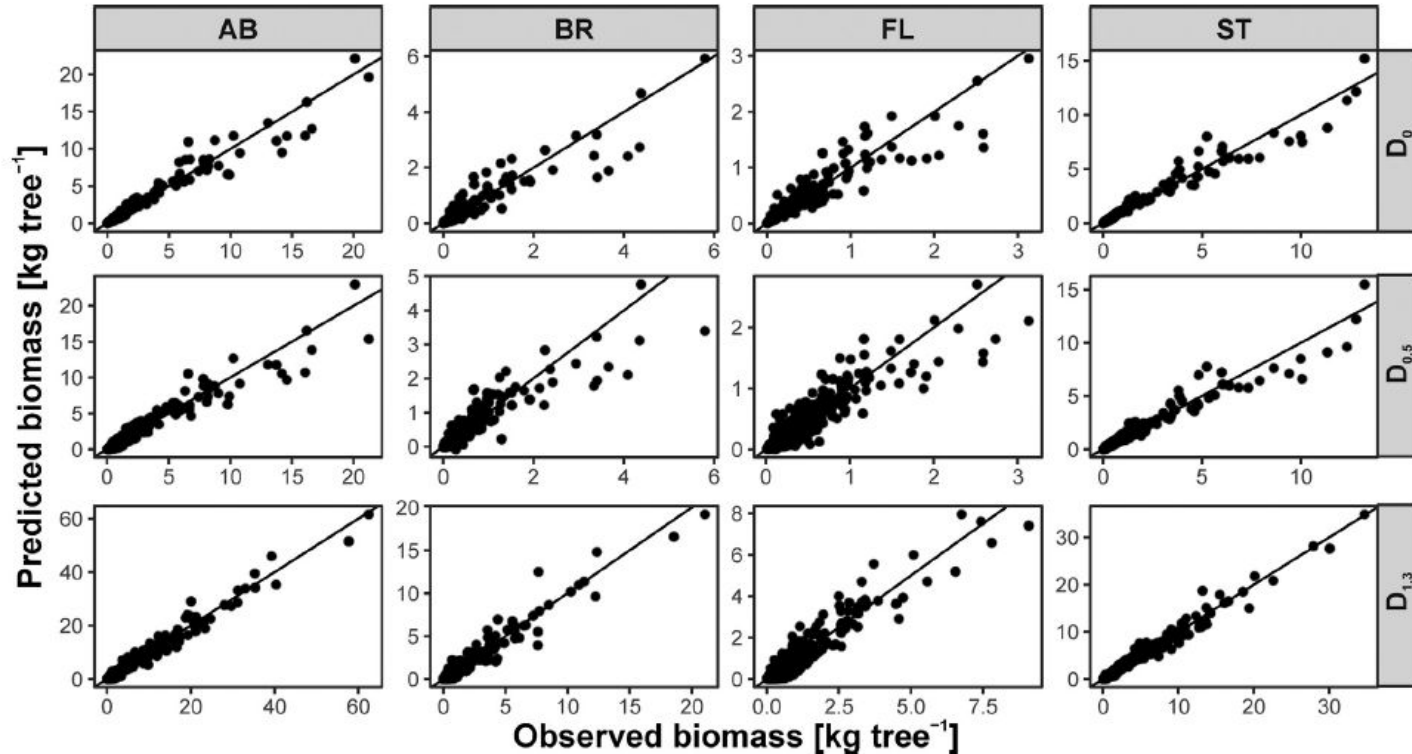
1	2	3	4
---	---	---	---

2.028546	2.882362	3.729461	9.471079
----------	----------	----------	----------

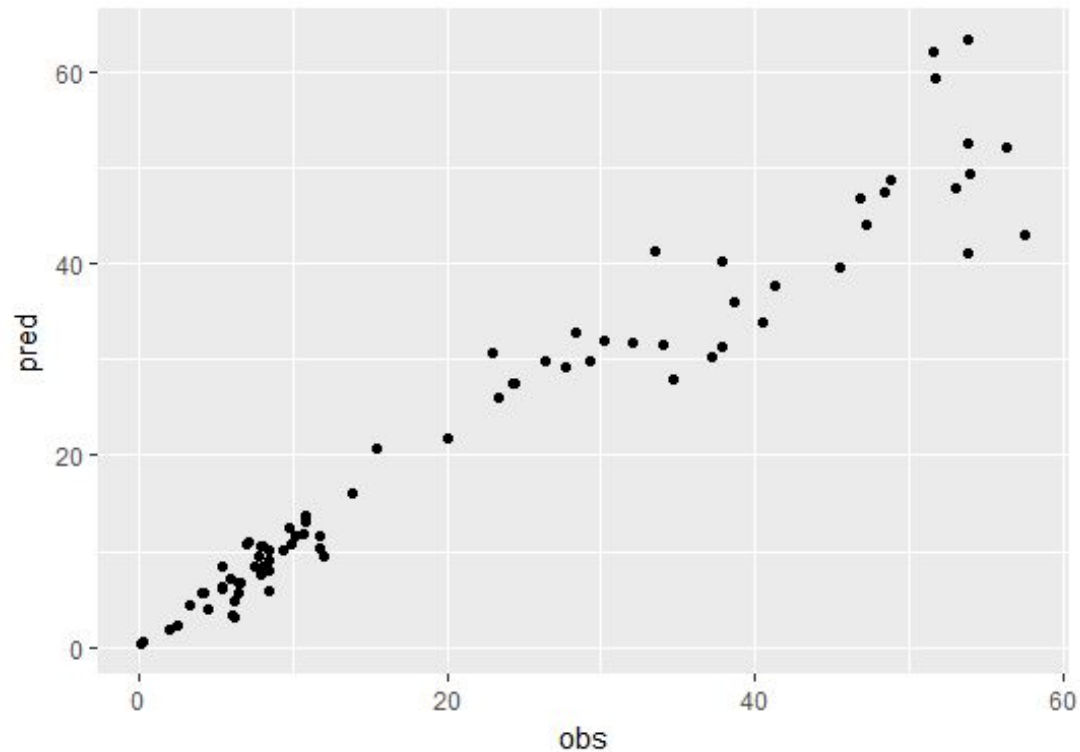
Inny sposób oceny - predicted vs. observed

A.M. Jagodziński et al.

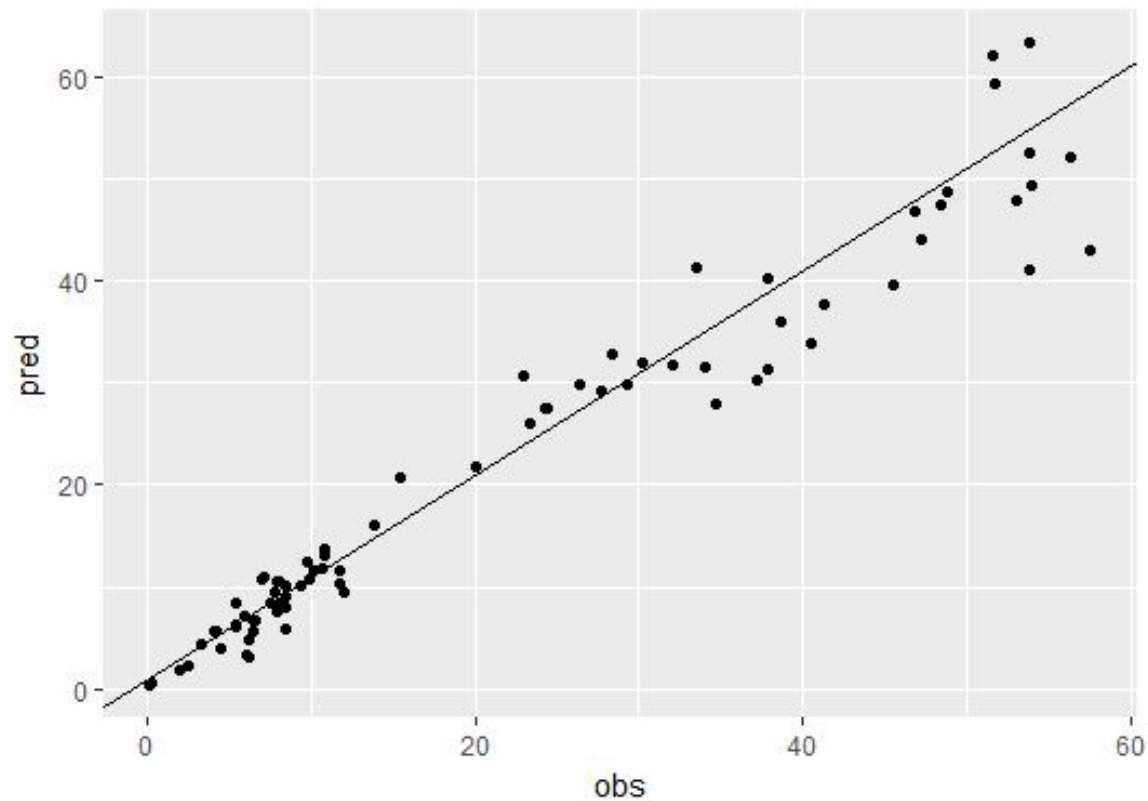
Forest Ecology and Management 409 (2018) 74–83



```
data.frame(obs=sosny$AB, pred=predict(model,sosny$V))%>%  
ggplot(aes(x=obs,y=pred))+geom_point()
```




```
data.frame(obs=sosny$AB, pred=predict(model,sosny$V))%>%  
ggplot(aes(x=obs,y=pred))+geom_point()+geom_abline(intercept=1)
```



Przyroda nie znosi prostoty

proste zależności są rzadko spotykane w przyrodzie

często chcemy zbadać wpływ kilku cech

po udanym wypadzie w teren mamy aż nadto danych

czy możemy wrzucić je wszystkie na raz i zobaczyć co wyjdzie?

Poprawianie jakości modelu

```
Console R Markdown
E:/Nauka/stat_narz/R/BSS/bssR/

> model0<-lm(AB~1,data=sosny)
> model1<-lm(AB~V,data=sosny)
> model2<-lm(AB~V+Hg,data=sosny)
> summary(model2)

call:
lm(formula = AB ~ V + Hg, data = sosny)

Residuals:
    Min       1Q   Median       3Q      Max
-14.5941  -2.0876  -0.5072   1.8221  16.3335

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.4508     1.6538   0.877   0.3832
V              0.4224     0.0655   6.449 1.03e-08 ***
Hg            1.8020     0.8932   2.017  0.0473 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.817 on 74 degrees of freedom
Multiple R-squared:  0.9305,    Adjusted R-squared:  0.9286
F-statistic: 495.6 on 2 and 74 DF,  p-value: < 2.2e-16
```

urościło nieznacznie, co z AIC?

```
> AIC(model1,model2,model0)
```

	df	AIC
--	----	-----

model1	3	467.6902
--------	---	----------

<u>model2</u>	<u>4</u>	<u>465.5678</u>
---------------	----------	-----------------

model0	2	666.9126
--------	---	----------

```
model3<-lm(AB~V+Hg+dens,data=sosny)
```

```
AIC(model3)
```

```
[1] 464.6941
```

```
model4<-lm(AB~V+Hg+E,data=sosny)
```

```
AIC(model4)
```

```
[1] 466.0577
```

VIF - variance inflation factor

Rule of Thumb $VIF > 10 \Rightarrow$ problem, ale...

<https://statisticalhorizons.com/multicollinearity>

<https://pdfs.semanticscholar.org/ed1f/4466a0982f3e8de202de01ecceb473d11893.pdf>

z czego wynika VIF? czy ma to biologiczne znaczenia dla badanej cechy?

library(car)

```
> vif(model1)
```

```
Error in vif.default(model1) : model contains fewer than 2 terms
```

```
> vif(model2)
```

V	Hg
---	----

14.01216	14.01216
----------	----------

```
> vif(model3)
```

V	Hg	E
---	----	---

15.573167	16.907578	1.359008
-----------	-----------	----------

```
> vif(model4)
```

V	Hg	E
---	----	---

15.573167	16.907578	1.359008
-----------	-----------	----------

Dlaczego tak?

$$V=G*H*f$$

f - wskaźnik kształtu którego nie znamy

V - miąższość

G - pole powierzchni przekroju pierśnicowego

H - wysokość

cechy są ze sobą silnie związane


```
model5<-lm(AB~V+dens,data=sosny)
```

```
Call:
lm(formula = AB ~ V + dens, data = sosny)

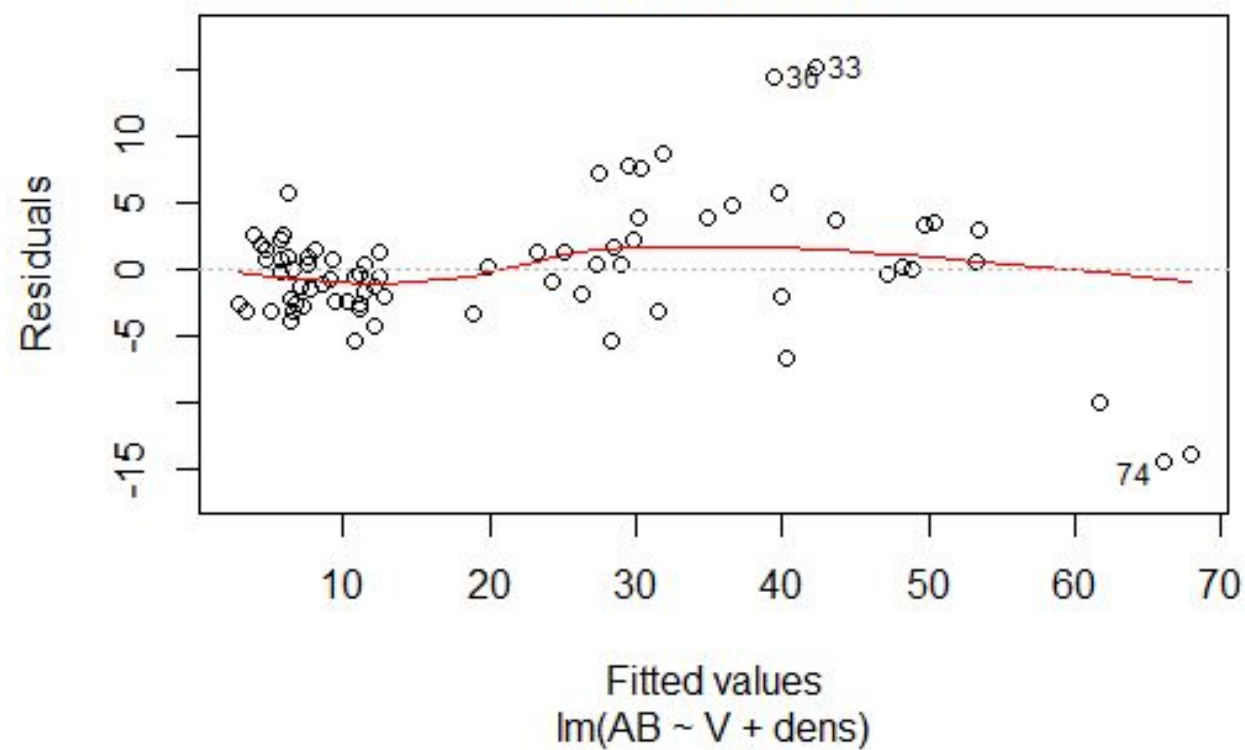
Residuals:
    Min       1Q   Median       3Q      Max
-14.537  -2.552  -0.056   1.745  15.193

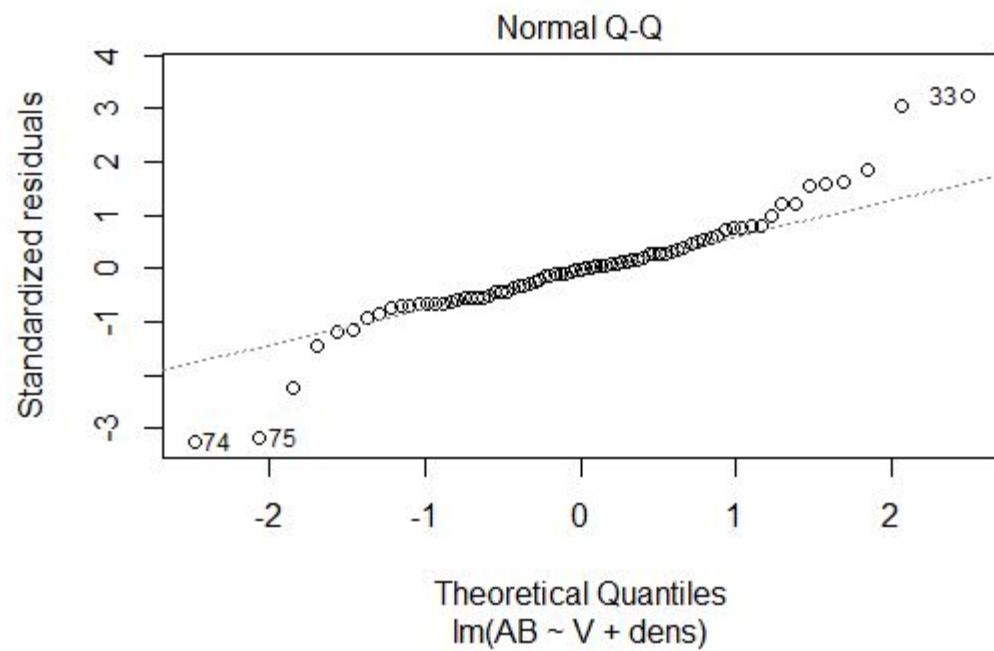
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  9.8163784   2.2893626   4.288 5.37e-05 ***
V             0.5190664   0.0211514  24.540 < 2e-16 ***
dens        -0.0006439   0.0002565  -2.510  0.0142 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

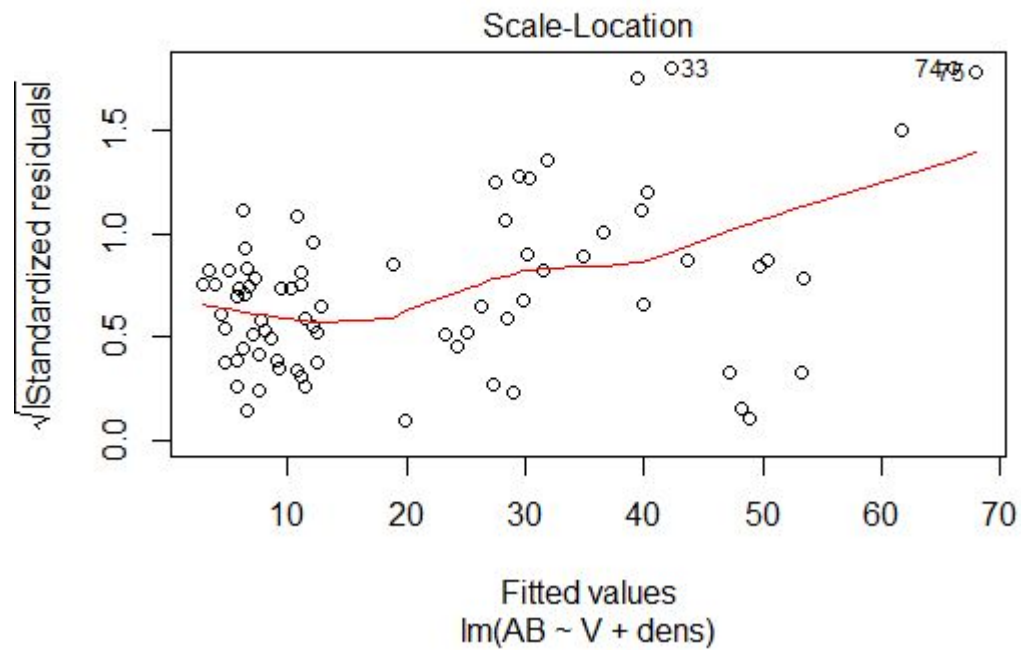
Residual standard error: 4.75 on 74 degrees of freedom
Multiple R-squared:  0.9325,    Adjusted R-squared:  0.9306
F-statistic: 510.8 on 2 and 74 DF,  p-value: < 2.2e-16

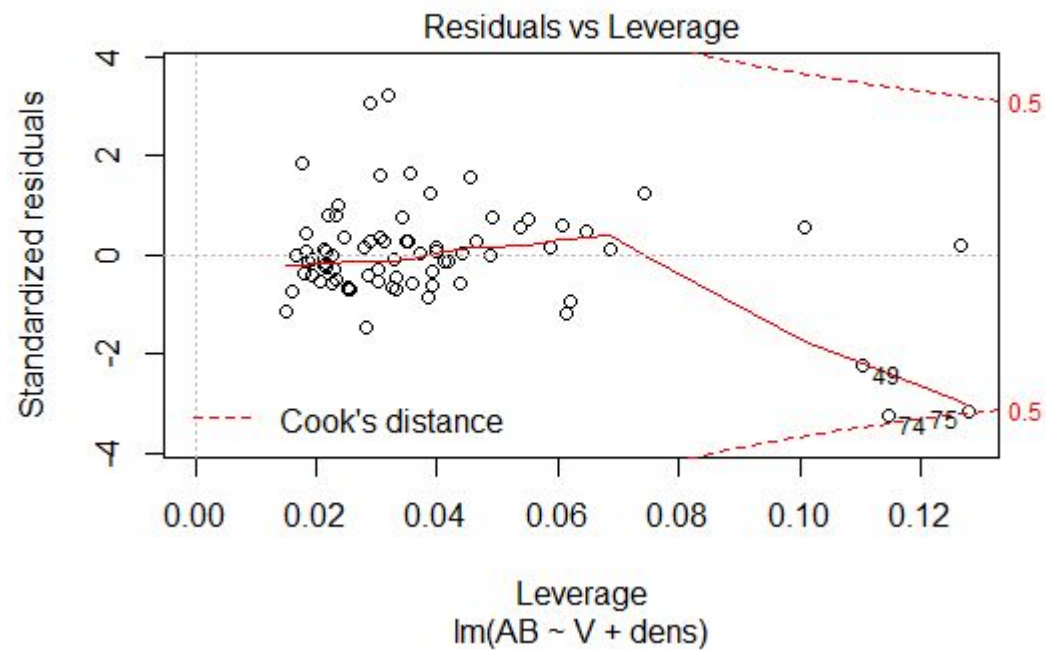
> AIC(model5)
[1] 463.3966
> vif(model5)
      V      dens
1.502807 1.502807
>
```

Residuals vs Fitted









i tak dalej...

Co trzeba sprawdzać?

VIFy (chyba że się zna korelacje między cechami i unika współliniowości)

AIC (pamiętamy o modelu zerowym)

R^2

wykresy diagnostyczne... - chyba że nie zależy nam na dokładności

With four parameters I can fit an elephant, and with five I can make him wiggle his trunk.

Drawing an elephant with four complex parameters

Jürgen Mayer
Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstr. 108, 01307 Dresden,
Germany

Khaled Khairy
European Molecular Biology Laboratory, Meyerhofstraße, 1, 69117 Heidelberg, Germany

Jonathan Howard
Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstr. 108, 01307 Dresden,
Germany

(Received 20 August 2008; accepted 5 October 2009)

We define four complex numbers representing the parameters needed to specify an elephantine shape. The real and imaginary parts of these complex numbers are the coefficients of a Fourier coordinate expansion, a powerful tool for reducing the data required to define shapes. © 2010 American Association of Physics Teachers
[DOI: 10.1119/1.3254017]

A turning point in Freeman Dyson's life occurred during a meeting in the Spring of 1953 when Enrico Fermi criticized the complexity of Dyson's model by quoting Johnny von Neumann: "With four parameters I can fit an elephant, and with five I can make him wiggle his trunk." Since then it has become a well-known saying among physicists, but nobody has successfully implemented it.

To parametrize an elephant, we note that its perimeter can be described as a set of points $(x(t), y(t))$, where t is a parameter that can be interpreted as the elapsed time while going along the path of the contour. If the speed is uniform, t becomes the arc length. We expand x and y separately² as a Fourier series

$$x(t) = \sum_{k=0}^{\infty} (A_k^x \cos(kt) + B_k^x \sin(kt)), \quad (1)$$

$$y(t) = \sum_{k=0}^{\infty} (A_k^y \cos(kt) + B_k^y \sin(kt)), \quad (2)$$

where A_k^x , B_k^x , A_k^y , and B_k^y are the expansion coefficients. The lower indices k apply to the k th term in the expansion, and the upper indices denote the x or y expansion, respectively.

Using this expansion of the x and y coordinates, we can analyze shapes by tracing the boundary and calculating the coefficients in the expansions (using standard methods from Fourier analysis). By truncating the expansion, the shape is smoothed. Truncation leads to a huge reduction in the information necessary to express a certain shape compared to a pixelated image, for example. Székely *et al.*³ used this approach to segment magnetic resonance imaging data. A similar approach was used to analyze the shapes of red blood cells,⁴ with a spherical harmonics expansion serving as a 3D generalization of the Fourier coordinate expansion.

The coefficients represent the best fit to the given shape in the following sense. The $k=0$ component corresponds to the center of mass of the perimeter. The $k=1$ component corresponds to the best fit ellipse. The higher order components

trace out elliptical corrections analogous to Ptolemy's epicycles.⁵ Visualization of the corresponding ellipses can be found at Ref. 6.

We now use this tool to fit an elephant with four parameters. Wei⁷ tried this task in 1975 using a least-squares Fourier sine series but required about 30 terms. By analyzing the picture in Fig. 1(a) and eliminating components with amplitudes less than 10% of the maximum amplitude, we obtained an approximate spectrum. The remaining amplitudes were

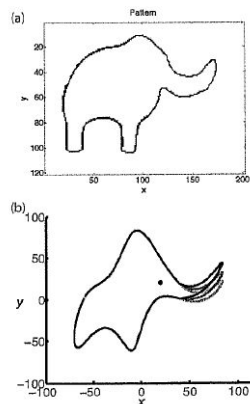


Fig. 1. (a) Outline of an elephant. (b) Three snapshots of the wagging trunk.

~John vonNeumann

niebezpieczne narzędzie

```
> step(1m(AB~V+Hg+G+dens+Age,sosny))
```

```
Start:  AIC=232.44
```

```
AB ~ V + Hg + G + dens + Age
```

	Df	Sum of Sq	RSS	AIC
- dens	1	21.067	1369.4	231.63
<none>			1348.3	232.44
- V	1	57.497	1405.8	233.65
- G	1	109.161	1457.5	236.43
- Hg	1	123.181	1471.5	237.17
- Age	1	272.744	1621.1	244.62

```
Step:  AIC=231.63
```

```
AB ~ V + Hg + G + Age
```

	Df	Sum of Sq	RSS	AIC
<none>			1369.4	231.63
- V	1	41.015	1410.4	231.90
- G	1	118.331	1487.7	236.01
- Hg	1	216.220	1585.6	240.92
- Age	1	311.616	1681.0	245.42

```
Call:
```

```
1m(formula = AB ~ V + Hg + G + Age, data = sosny)
```

```
Coefficients:
```

```
(Intercept)
```

```
V
```

```
Hg
```

```
G
```

```
Age
```

dlaczego niebezpieczne?

pozwała szybko i bez refleksy
wybrać model o najmniejszym
AIC, bez sprawdzania
pozostałych parametrów.

Pomocne, ale zawsze należy
obejrzeć model

Nie chroni przed
interkorelacją, outlierami,
brakiem logiki

Inna opcja - Multi-Model Inference - library(MuMIn)

uśrednianie modeli i wybór modelu w oparciu o AIC lub wagi Akaikego

tworzymy model globalny - ze wszystkimi zmiennymi które mogą wejść do modelu

funkcja dredge() buduje podzbiór wszystkich możliwych kombinacji zmiennych

oblicza AIC i wagi Akaikego

```
library(MuMIn)
global.model<-lm(AB~V+G+Hg+dens+Age,
data=sosny,na.action = na.fail)
```

#na.action=na.fail jest kluczowe!

```
dred<-dredge(global.model)
dred
```

mamy selekcje modeli i ich współczynniki,
AICc - AIC dla małych prób
delta - różnica w AICc w stosunku do najlepszego modelu

weights - wagi Akaikiego
można je interpretować luźno jako
“prawdopodobieństwo że dany model jest najlepszym modelem pod względem AIC w zbiorze modeli-kandydatów”

źródło interpretacji:

<https://link.springer.com/article/10.3758/BF03206482>

Global model call: lm(formula = AB ~ V + G + Hg + dens + Age, data = sosny, na.action = na.fail)

Model selection table

	(Intrc)	Age	dens	G	Hg	V	df	logLik	AICc	delta	weight
14	4.644	-1.57100		1.4680	5.3670		5	-221.209	453.3	0.00	0.319
30	4.582	-1.32400		1.0300	4.5360	0.1418	6	-220.073	453.3	0.08	0.306
32	7.889	-1.26000	-0.0002998	0.9929	3.8280	0.1785	7	-219.476	454.6	1.31	0.165
16	6.102	-1.57100	-0.0001315	1.5010	5.1510		6	-221.084	455.4	2.10	0.111
26	2.274	-1.08700			5.3560	0.3158	5	-223.264	457.4	4.11	0.041
28	6.320	-1.02100	-0.0003578		4.4750	0.3522	6	-222.474	458.1	4.88	0.028
24	13.380	-0.70170	-0.0006240	1.1820		0.3108	6	-222.842	458.9	5.62	0.019
22	7.419	-0.59970		1.3700		0.2789	5	-225.718	462.3	9.02	0.004
19	9.816		-0.0006439			0.5191	4	-227.698	464.0	10.69	0.002
23	8.397		-0.0005050	0.5347		0.3889	5	-226.571	464.0	10.72	0.001
20	12.590	-0.29060	-0.0007630			0.5520	5	-226.816	464.5	11.21	0.001
21	4.039		0.7695		0.3529	4	-228.354	465.3	12.00	0.001	
27	7.277		-0.0005090		0.8580	0.4649	5	-227.347	465.5	12.28	0.001
25	1.451			1.8020	0.4224	4	-228.784	466.1	12.86	0.001	
31	8.219		-0.0004971	0.5214	0.0720	0.3876	6	-226.569	466.3	13.07	0.000
29	2.575		0.5440	0.9591	0.3428	5	-227.967	466.8	13.51	0.000	
8	13.640	-0.97780	-0.0004926	2.5260			5	-228.529	467.9	14.64	0.000
17	4.398				0.5498	3	-230.845	468.0	14.75	0.000	
6	8.811	-0.87260		2.5670		4	-230.136	468.8	15.56	0.000	
18	4.878	-0.07853			0.5602	4	-230.779	470.1	16.85	0.000	
10	-2.010	-1.76700		10.8500		4	-232.947	474.4	21.18	0.000	
12	-3.863	-1.76100	0.0001866		10.9800	5	-232.750	476.3	23.08	0.000	
5	3.828		2.0980		3	-235.591	477.5	24.25	0.000		
13	1.549		1.6860	1.5000		4	-234.788	478.1	24.87	0.000	
7	6.040		-0.0002576	2.0470		4	-235.195	478.9	25.68	0.000	
15	3.014		-0.0001322	1.7190	1.2830	5	-234.699	480.2	26.98	0.000	
9	-6.688			7.3530	3	-245.955	498.2	44.97	0.000		
11	-9.035		0.0002385		7.5340	4	-245.725	500.0	46.74	0.000	
4	3.070	2.58900	-0.0011640			4	-287.564	583.7	130.42	0.000	
2	-9.012	2.98100				3	-289.534	585.4	132.13	0.000	
3	51.240		-0.0042850			3	-312.879	632.1	178.82	0.000	
1	21.580					2	-331.456	667.1	213.81	0.000	

Zmienne kategoryczne - porównanie między grupami

Model liniowy ze zmienną jakościową (tekstem)

czy drzewostany sosnowe różnią się AB pomiędzy typem?

AB~type

summary(lm(AB~type,data=sosny))

są gwiazdki, ale jak to zinterpretować?

AB=16.455+ jeśli postagric 16.329
jeśli postind 13.999

jeśli forest? jako referencja (0)

Call:

lm(formula = AB ~ type, data = sosny)

Residuals:

Min	1Q	Median	3Q	Max
-26.233	-10.566	-6.249	10.815	41.138

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	16.455	2.379	6.915	1.41e-09 ***
typepostagric	16.329	6.790	2.405	0.01868 *
typepostind	13.999	4.452	3.145	0.00239 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 16.83 on 74 degrees of freedom

Multiple R-squared: 0.1524, Adjusted R-squared: 0.1295

F-statistic: 6.653 on 2 and 74 DF, p-value: 0.002202

ANOVA a model liniowy

model liniowy - daje nam informacje o wpływie współczynników i postać modelu

poziom referencyjny (pierwszy poziom zmiennej grupującej) - różnice w stosunku do niego

analiza wariancji - sprawdza czy są różnice między którąkolwiek z par poziomów zmiennej grupującej

```
an<-aov(lm(AB~type,data=sosny)) #przypisujemy do obiektu, notacja jak w modelu  
summary(an)
```

```
          Df Sum Sq Mean Sq F value Pr(>F)  
type        2  3767  1883.4   6.653 0.0022 **  
Residuals   74 20949   283.1
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

są różnice* - tylko pomiędzy którym z wariantów?

wiadomo tylko że jest różnica, aby sprawdzić co się różni - test *post hoc* (test *posteriori*)

test *post hoc* wykonujemy tylko jeśli wyszły istotne różnice]

*ale pieprzyć p-values (informacyjnie)

Co po ANOVIE?

Analiza wariancji wykazała różnice istotne na poziomie $p < 0.05$

Mamy trzy warianty - to znaczy że który się różni od którego?

testów *post hoc* jest wiele, np.:

test HSD Tukeya (najczęściej stosowany, konserwatywny)

test LSD Fishera (mniej konserwatywny)

test Duncana (mało konserwatywny)

Dwie wersje testu Tukeya

```
> library(agricolae)
> HSD.test(an, 'type', console = T)
```

Study: an ~ "type"

HSD Test for AB

Mean Square Error: 283.089

dziwactwa ze składnią:

an - obiekt typu aov

'type' - nazwa zmiennej

grupującej

console=T - bo inaczej nie

pokaże wyniku:(

Groups according to probability of means differences and alpha level(0.05)

Treatments with the same letter are not significantly different.

type, means

	AB	std	r	Min	Max
forest	16.45475	16.34306	50	0.1738	57.5928
postagric	32.78376	18.97096	7	6.5507	56.3193
postind	30.45343	17.32281	20	4.4608	53.9193

Alpha: 0.05 ; DF Error: 74

Critical Value of Studentized Range: 3.382467

AB groups

postagric	32.78376	a
postind	30.45343	a
forest	16.45475	b

Ok, ale skąd się wzięły literki?

powiązania w parach - pojedyncze testy par
zobaczmy jak to wygląda w innym pakiecie:

```
library(multcomp)  
summary(glht(an,mcp(type='Tukey')))
```

```
Simultaneous Tests for General Linear Hypotheses  
Multiple Comparisons of Means: Tukey Contrasts  
Fit: aov(formula = lm(AB ~ type, data = sosny))  
Linear Hypotheses:
```

	Estimate	Std. Error	t value	Pr(> t)
postagric - forest == 0	16.329	6.790	2.405	0.04607 *
postind - forest == 0	13.999	4.452	3.145	0.00628 **
postind - postagric == 0	-2.330	7.389	-0.315	0.94509

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
(Adjusted p values reported -- single-step method)
```

pakiet multcomp pozwala na więcej
test Tukeya można zastosować także
dla bardziej złożonych modeli

składnia:

summary - żeby dał wyniki
glht - funkcja która potrzebuje
modelu i nazwy zmiennej

an - nasz obiekt z anową

mcp() - funkcja do porównań
wielokrotnych, dajemy jej dwa
argumenty:
nazwę zmiennej
typ testu ('Tukey')

Literki!

forest i postagric różnią się - nie mogą mieć tych samych

forest i postind - tak samo

postind i postagric - nie różnią się - muszą mieć taką samą

Szybszy sposób - trzeba glht() wrzucić w cld() #trochę liczy

```
cld(glht(an,mcp(type='Tukey')))
```

```
forest postagric postind
```

```
"a"      "b"      "b"
```

Termin badań - interpretacja

Termin badań - numeric czy factor?

ANOVA - wpływ terminu na wynik-> co to znaczy?

LM - wpływ terminu na wynik -? co to znaczy?

ANOVA - w przynajmniej jednej parze kombinacji terminów są różnice

LM - jest trend

Co jest ważniejsze dla biologa?

Kilka uwag o nazewnictwie

One-way ANOVA - jednoczynnikowa analiza wariancji

Jeden “X”, jeden “Y”

Two-way ANOVA, dwuczynnikowa analiza wariancji, multivariate ANOVA, ANOVA
- co najmniej dwa “X” i jeden “Y”

MANOVA - multidimensional lub mulivariate ANOVA :

co najmniej jeden “X” i co najmniej dwa “Y”

Dwuczynnikowa ANOVA

```
> an2<-aov(lm(AB~type+Soil.type,data=sosny))
```

```
> summary(an2)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
type	2	3767	1883.4	7.161	0.00152 **
Soil.type	7	3326	475.1	1.806	0.10040
Residuals	67	17623	263.0		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Model z interakcją

znaki w formule: + addytywność (wspólne oddziaływanie)

: interakcja *addytywność i interakcja

```
> an2<-aov(lm(AB~type:Soil.type,data=sosny))
> summary(an2)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
type:Soil.type	10	7125	712.5	2.673	0.0083 **
Residuals	66	17590	266.5		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
> an2<-aov(lm(AB~type*Soil.type,data=sosny))
> summary(an2)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
type	2	3767	1883.4	7.067	0.00166 **
Soil.type	7	3326	475.1	1.783	0.10554
type:Soil.type	1	32	32.2	0.121	0.72917
Residuals	66	17590	266.5		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Co znaczy interakcja?

Różne różnice pomiędzy grupami jednego czynnika w grupach drugiego czynnika

np. gatunki iglaste w ramach jednej grupy będą miały większy przyrost biomasy niż w ramach drugiej grupy

Czy sprawdzać wszystkie możliwe interakcje (full-factorial)?

Nie, bo testujemy coś dla czego nie ma uzasadnienia?

Tak, bo może za kilkanaście lat ktoś będzie miał uzasadnienie?

...?

ostrożnie z interakcjami, mogą wyjść efekty uboczne;)

MANOVA

Chcemy sprawdzić oddziaływanie na zbiór cech,

np. jak typ lasu wpływa na parametry wzrostowe sosen:

```
> mm<-manova(cbind(AB,V,Hg,G)~type,data=sosny)
> summary(mm)
```

	Df	Pillai approx	F num	Df den	Df	Pr(>F)
type	2	0.44808	5.197	8	144	1.028e-05 ***
Residuals	74					

```
---
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

nie ma testów post hoc, generalnie rzadko używana - wtedy gdy trzeba wykazać jednoczesny i spójny wpływ jednej lub kilku cech na zestaw kilku cech

Analiza kowariancji - ANCOVA

znając dane wiemy że mamy dużo młodych drzewostanów wśród przemysłowych i porolnych. Nie wiemy ile mają lat a to może chyba mieć jakieś znaczenie;)

W tym celu sprawdzimy wpływ typu lasu na biomasę biorąc pod uwagę wiek

ANCOVA pomoże ocenić efekty zmiennych liczbowych i kategorycznych

zmienna liczbową - np. wiek

zmienna kategoryczna - np. typ lasu


```
summary(lm(AB~Age+type,data=sosny))
```

Call:

```
lm(formula = AB ~ Age + type, data = sosny)
```

Residuals:

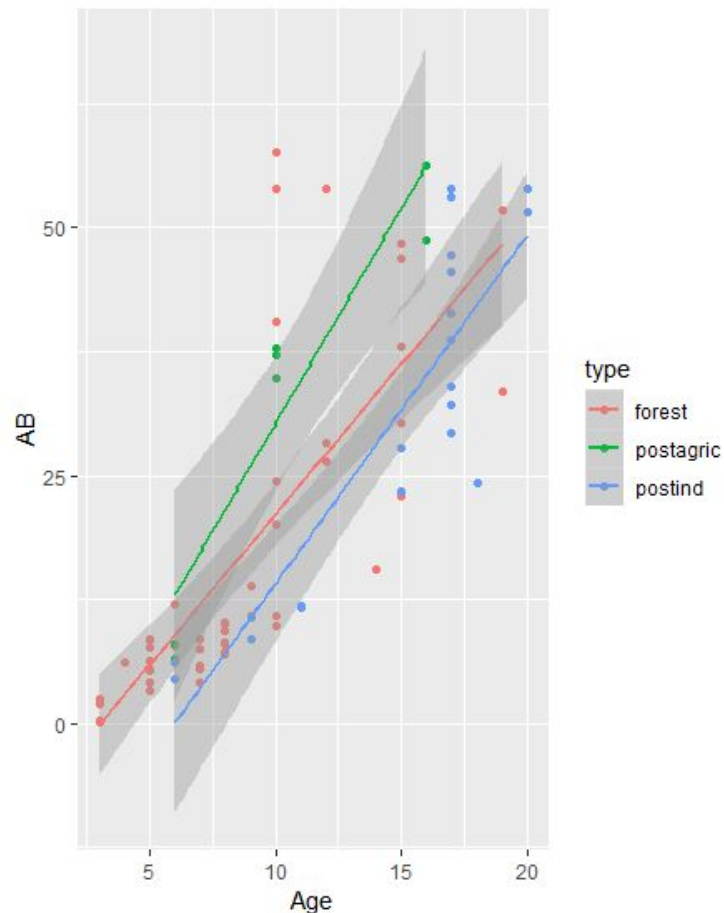
Min	1Q	Median	3Q	Max
-19.081	-6.287	-0.533	3.676	36.108

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-11.1761	2.7075	-4.128	9.61e-05 ***
Age	3.2661	0.2742	11.911	< 2e-16 ***
typepostagric	9.4329	4.0264	2.343	0.0219 *
typepostind	-6.2183	3.1153	-1.996	0.0497 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.874 on 73 degrees of freedom
Multiple R-squared: 0.7121, Adjusted R-squared: 0.7002
F-statistic: 60.17 on 3 and 73 DF, p-value: < 2.2e-16



```
ggplot(sosny, aes(x=Age,y=AB,col=type))  
+geom_point()+geom_smooth(method='lm')
```

Analiza kowariancji dla modelu

```
> an3<-aov(lm(AB~Age+type,data=sosny))
> summary(an3)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Age	1	16396	16396	168.189	< 2e-16 ***
type	2	1202	601	6.167	0.00335 **
Residuals	73	7117	97		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
cld(glht(an3, mcp(type='Tukey')))
```

	forest	postagric	postind
	"ab"	"a"	"b"

```
summary(glht(an3, mcp(type='Tukey')))
```

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Fit: aov(formula = lm(AB ~ Age + type, data = sosny))


Linear Hypotheses:

	Estimate	Std. Error	t value	Pr(> t)
postagric - forest == 0	9.433	4.026	2.343	0.05466 .
postind - forest == 0	-6.218	3.115	-1.996	0.11727
postind - postagric == 0	-15.651	4.478	-3.495	0.00224 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)

Outliery

How much does climate change threaten European forest tree species distributions?

Marcin K. Dyderski^{1,2} | Sonia Paž³ | Lee E. Frelich⁴ | Andrzej M. Jagodziński^{1,2} 

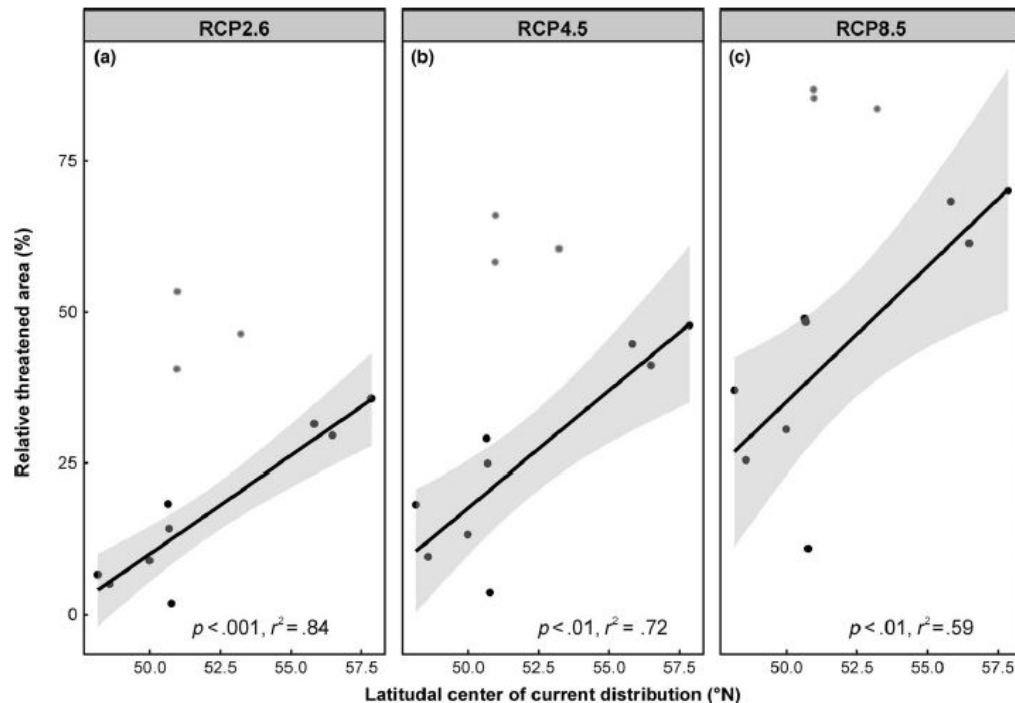


FIGURE 7 The relationship between threatened proportion of current distribution and latitudinal center of distribution for each climate

Formalna podstawa?

Outlier wynikający z błędu pomiaru

Outlier wynikający ze struktury danych (filogeneza, biologia, układ doświadczenia)

Outlier będący poprawnym pomiarem

Usunąć? Zostawić? Co jeśli Recenzent się uprze przy konieczności uzasadnienia?

Statystyki liczbowe

For inspection of outliers we checked Cook's distances and leverage values using *hat* statistics implemented in the `base::influence()` function. To identify outliers we used Bonferroni *p*-values for Studentized residuals t-tests using the `car::outlierTest()` function. After visual interpretation of diagnostic plots both for calcareous and granite grasslands we adopted 0.15 as threshold leverage values, as most of the observations had relatively low leverage values, ranging from 0.07 in calcareous and from 0.05 to 0.15 in granite grasslands. After that, we excluded two outliers: plot no. 71 (elevation 1039 m a.s.l.) from calcareous and plot no. 57 (elevation 2123 m a.s.l.) from granite grasslands due to high leverage values (0.20 and 0.24, respectively) and due to the `car::outlierTest()` $p < 0.05$ in both cases.

Źródło: Czortek et al. 2018. Cessation of livestock grazing and windthrow drive a shift in plant species composition in the Western Tatra Mts. *Tuexenia* 38: 177–196. https://www.zobodat.at/pdf/Tuexenia_NS_38_0177-0196.pdf

Rozważmy taki model

```
> model<-lm(BR~Age, data=sosny)  
> summary(model)
```

Call:
lm(formula = BR ~ Age, data = sosny)

Residuals:

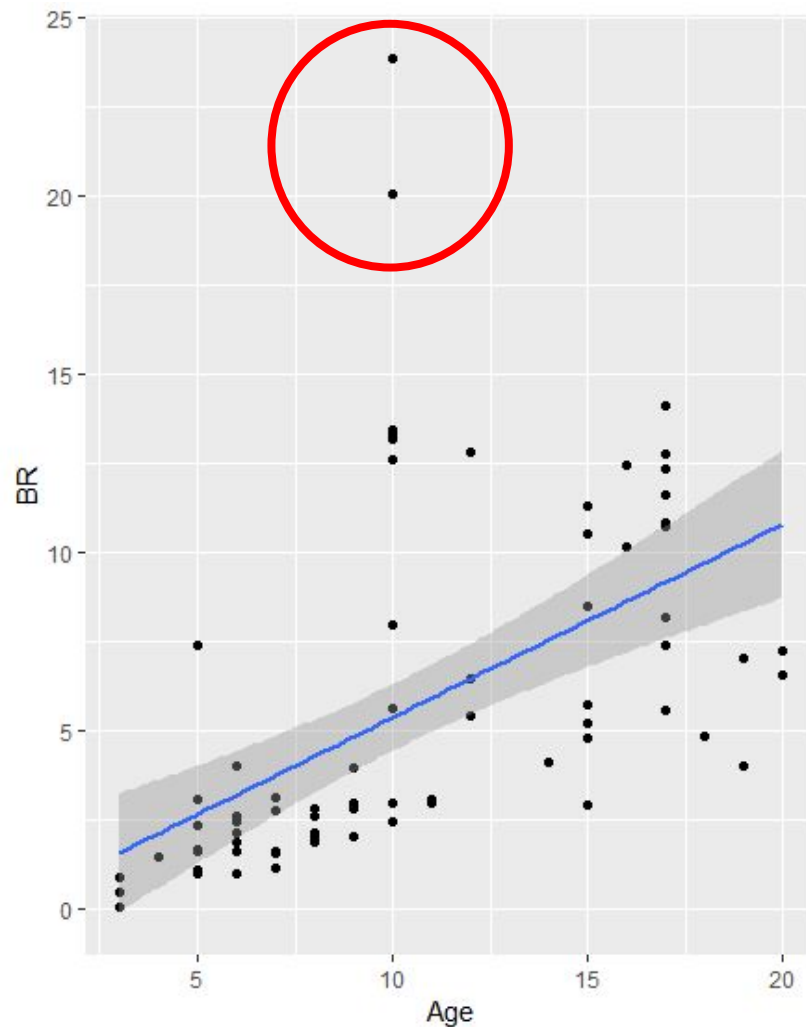
Min	1Q	Median	3Q	Max
-6.2556	-2.2313	-1.0834	0.7993	18.5083

Coefficients:

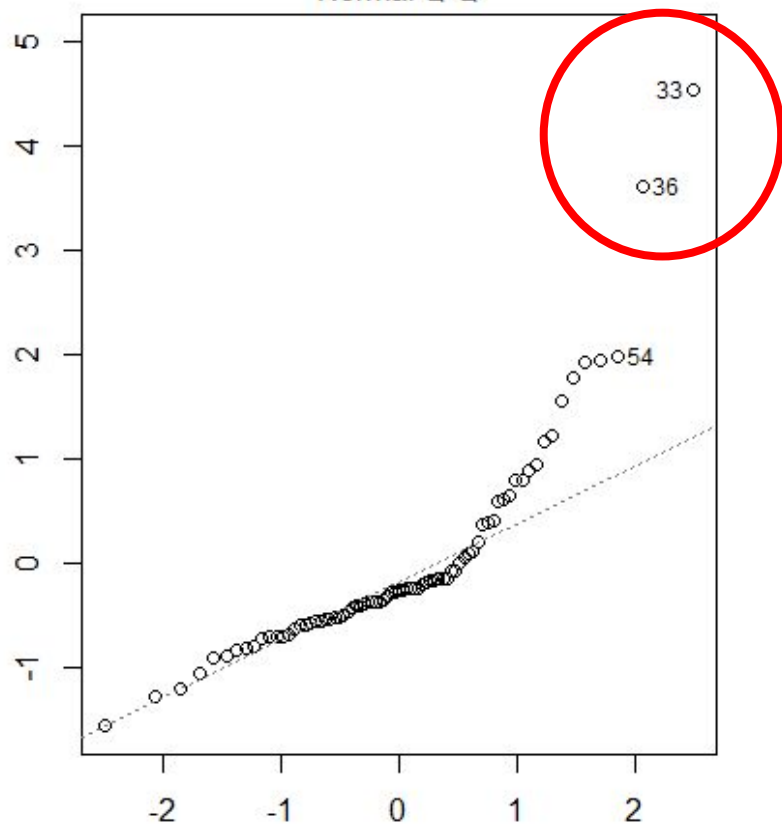
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.08740	1.08768	-0.080	0.936
Age	0.54361	0.09568	5.681	2.41e-07 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.11 on 75 degrees of freedom
Multiple R-squared: 0.3009, Adjusted R-squared: 0.2916
F-statistic: 32.28 on 1 and 75 DF, p-value: 2.411e-07

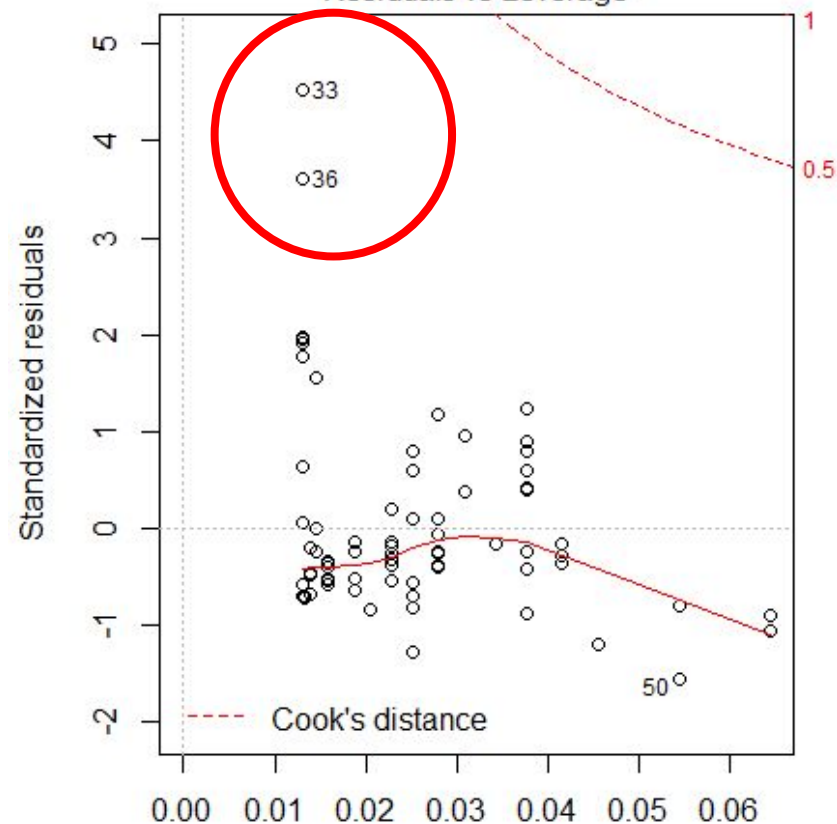


Normal Q-Q



Theoretical Quantiles
 $\ln(\text{BR} \sim \text{Age})$

Residuals vs Leverage



Leverage
 $\ln(\text{BR} \sim \text{Age})$

```
as.data.frame(influence(model))
```

	hat	coefficients..Intercept.	coefficients.Age	sigma	wt.res
1	0.04155579	-3.933644e-02	2.900932e-03	4.136538	-0.706535320
2	0.04155579	-8.295232e-02	6.117460e-03	4.133604	-1.489935320
3	0.04155579	-8.381529e-02	6.181101e-03	4.133525	-1.505435320
4	0.04155579	-6.202961e-02	4.574479e-03	4.135273	-1.114135320
5	0.04155579	-6.202961e-02	4.574479e-03	4.135273	-1.114135320
6	0.03422738	-3.269323e-02	2.320786e-03	4.136651	-0.660547628
7	0.03422738	-3.136678e-02	2.226626e-03	4.136709	-0.633747628
8	0.02798310	-1.267398e-02	8.554936e-04	4.137246	-0.291659937
9	0.02798310	1.797893e-02	-1.213578e-03	4.137101	0.413740063
....					
30	0.01575502	-5.745441e-02	2.754013e-03	4.129255	-2.212896862
31	0.01384724	-1.722269e-02	5.882398e-04	4.136193	-0.849509170
32	0.01384724	-3.814717e-02	1.302914e-03	4.131521	-1.881609170
33	0.01302358	2.706275e-01	-2.640268e-03	3.525303	18.508278522
34	0.01302358	3.444622e-03	-3.360607e-05	4.137297	0.235578522
35	0.01302358	1.143815e-01	-1.115917e-03	4.034866	7.822578522
36	0.01302358	2.149178e-01	-2.096759e-03	3.762979	14.698278522
37	0.01302358	-4.253562e-02	4.149817e-04	4.123362	-2.909021478
38	0.01302358	3.816302e-02	-3.723222e-04	4.126102	2.609978522
39	0.01302358	-3.526558e-02	3.440544e-04	4.127752	-2.411821478
40	0.01462865	2.132902e-02	6.081039e-03	4.069968	6.352053905
41	0.01462865	-3.423447e-03	-9.760465e-04	4.135665	-1.019546095
42	0.01462865	5.553506e-06	1.583340e-06	4.137388	0.001653905
43	0.02057022	2.738037e-02	-7.104097e-03	4.117704	-3.431870712
44	0.02516719	3.927169e-02	-7.544299e-03	4.123642	-2.862183020
45	0.02516719	-3.319243e-02	6.376441e-03	4.127573	2.419116980
46	0.02516719	-5.488588e-03	1.054387e-03	4.137120	0.400016980

zwraca nam wartości statystyki hat,
wartości współczynników przy odrzuceniu
obserwacji, wartość SD po odrzuceniu
obserwacji oraz ważone reszty

można coś zobaczyć, ale nie zawsze;)

mówi o wpływie na współczynniki modelu
(coefficients)


```
library(car)
```

```
outlierTest(model)
```

```
      rstudent unadjusted p-value Bonferonni p
```

```
33 5.284650      1.2239e-06  0.00009424
```

```
36 3.931709      1.8820e-04  0.01449200
```

test t-studenta na istotność średniego przesunięcia od linii regresji - daje p-value surowe oraz po poprawce Bonferonniego

Wskazuje które obserwacje można usunąć

```
model<-lm(BR~Age, data=sosny)
summary(model)
```

Call:
lm(formula = BR ~ Age, data = sosny)

Residuals:

Min	1Q	Median	3Q	Max
-6.2556	-2.2313	-1.0834	0.7993	18.5083

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.08740	1.08768	-0.080	0.936
Age	0.54361	0.09568	5.681	2.41e-07 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.11 on 75 degrees of freedom
Multiple R-squared: 0.3009, Adjusted R-squared: 0.2916
F-statistic: 32.28 on 1 and 75 DF, p-value: 2.411e-07

```
model2<-lm(BR~Age, data=sosny[-c(33,36),])
summary(model2)
```

Call:
lm(formula = BR ~ Age, data = sosny[-c(33, 36),])

Residuals:

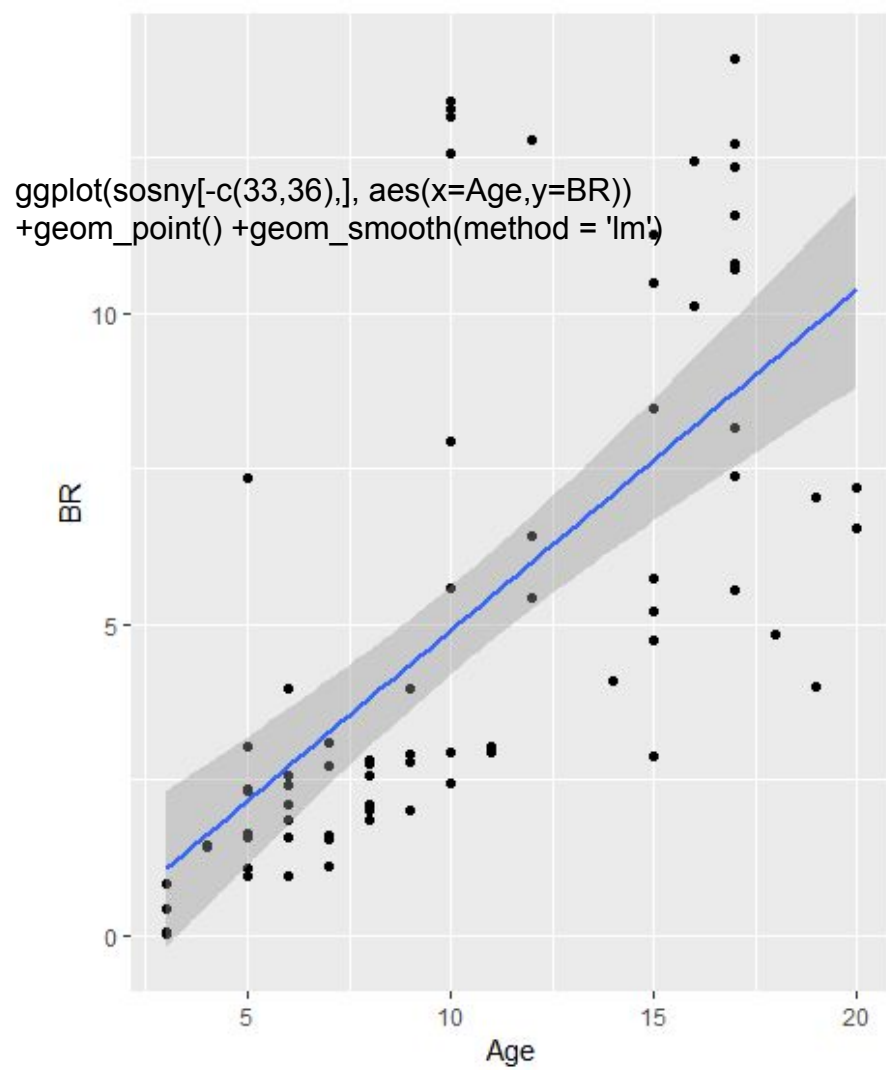
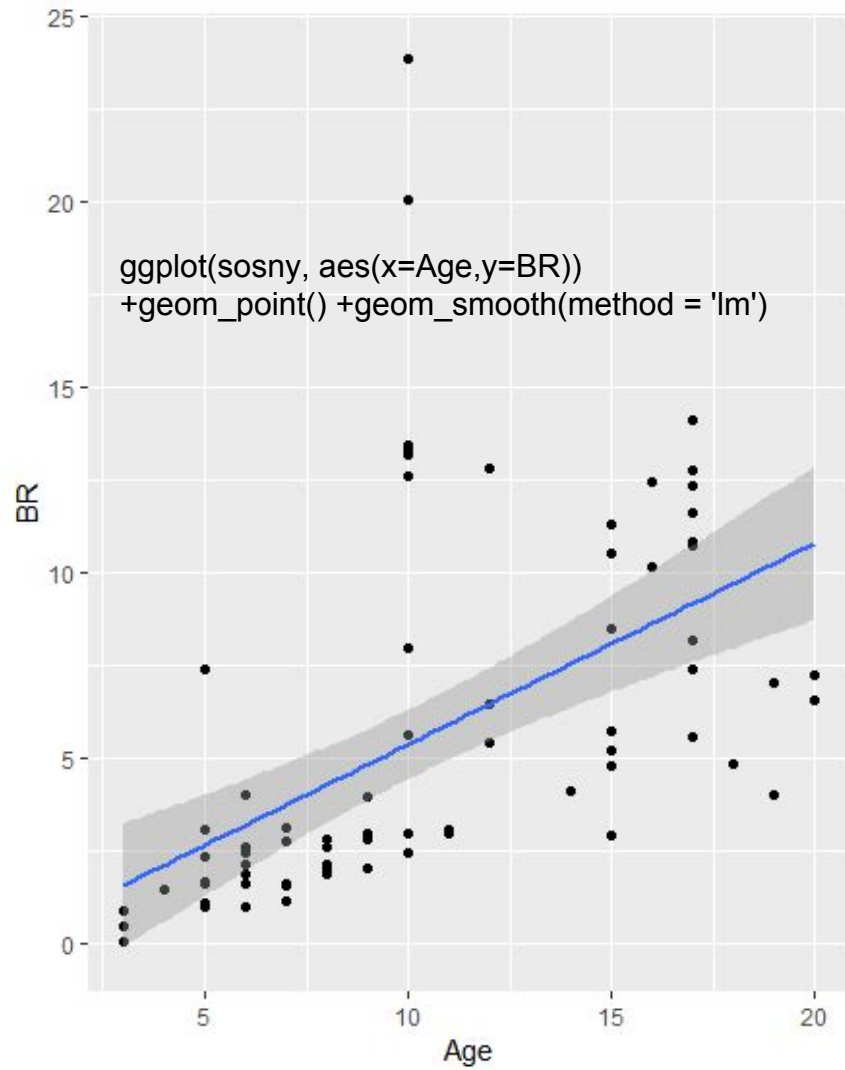
Min	1Q	Median	3Q	Max
-5.8548	-1.7888	-0.6365	0.8509	8.5087

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.57944	0.81816	-0.708	0.481
Age	0.54841	0.07176	7.643	6.6e-11 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.082 on 73 degrees of freedom
Multiple R-squared: 0.4445, Adjusted R-squared: 0.4369
F-statistic: 58.41 on 1 and 73 DF, p-value: 6.598e-11



Podsumowanie

obrazki, obrazki, obrazki!

-pomogą dobrać narzędzie i typ rozkładu

biologiczne znaczenie (effect size) > rozkład błędów > AIC > R^2

p-value - pomocniczo, **nie podajemy że coś jest istotne statystycznie**

nie $p < 0.05$, tylko dokładna wartość

za dużo predyktorów nie można - dwóch strażników - VIF i AIC

correlation does not imply causation!

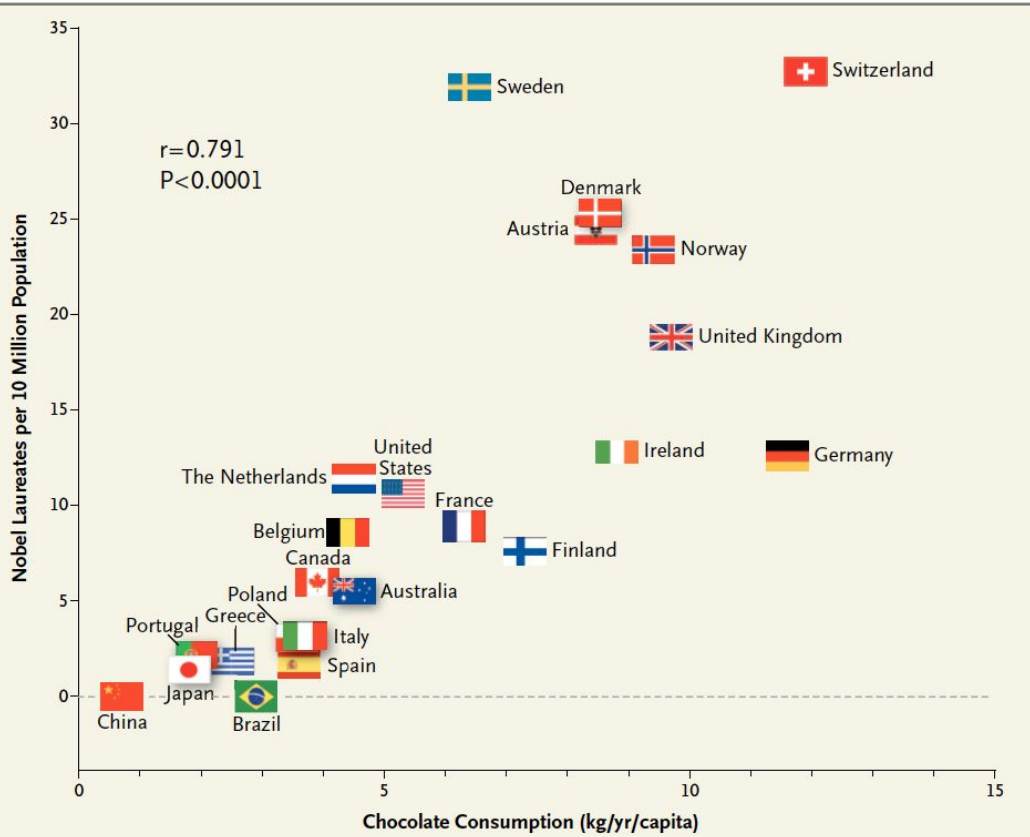


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

<https://blogs.scientificamerican.com/the-curious-wavefunction/chocolate-consumption-and-nobel-prizes-a-bizarre-juxtaposition-if-there-ever-was-one/>

