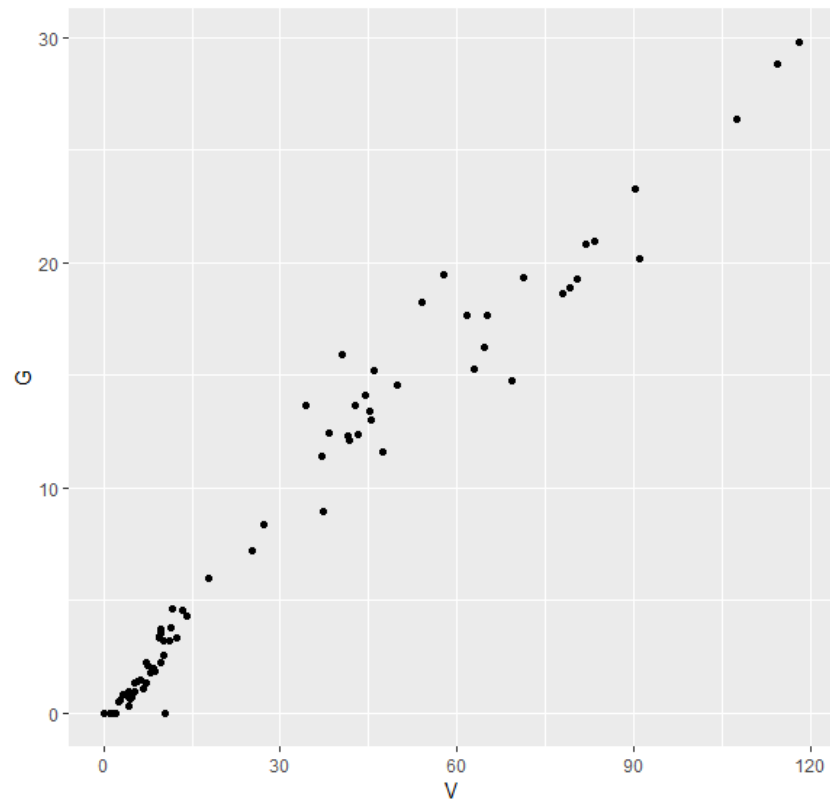
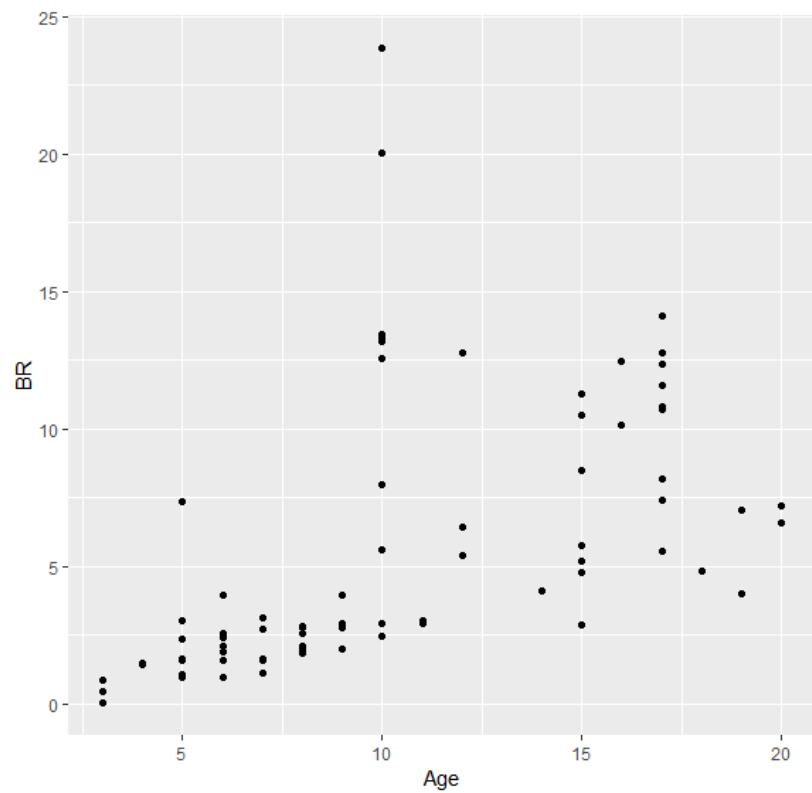




Korelacja, regresja, modele



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

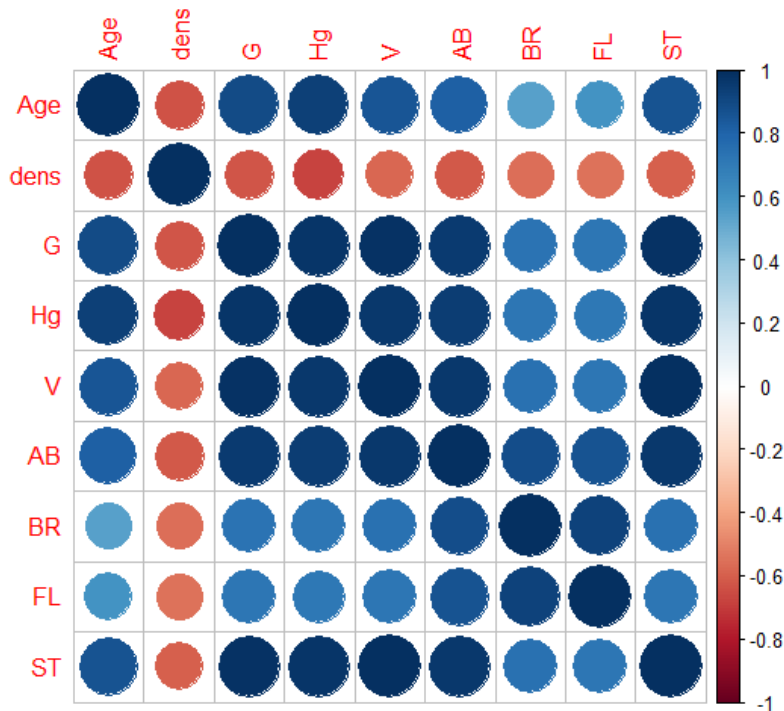
	AB	BR	FL	ST
Age	0.8144991	0.5485267	0.5914534	0.8621775
dens	-0.6186956	-0.5598806	-0.5479810	-0.5917399
G	0.9576488	0.7362057	0.7225715	0.9847925
Hg	0.9441822	0.7238736	0.7191310	0.9700117
V	0.9626557	0.7405488	0.7230836	0.9955775
AB	1.0000000	0.8822559	0.8623371	0.9669117
BR	0.8822559	1.0000000	0.9214561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

correlation matrix

`cor(dataset[,1:10])`

funkcja `cor()` dla
więcej niż dwóch
zmiennych


```
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= $10 \cdot \text{średnica} + 2$

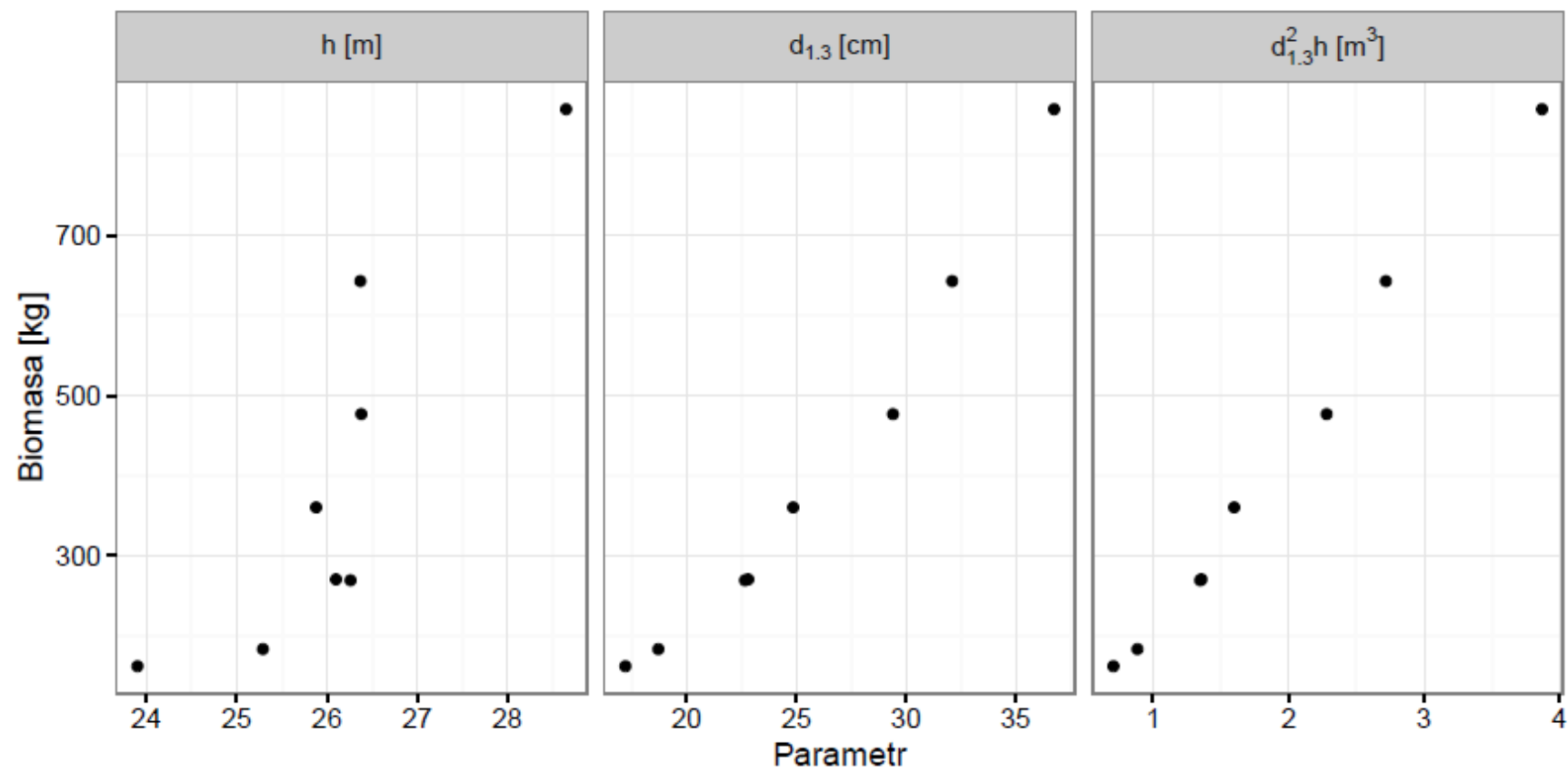
Regresja liniowa

Jak zmienia się masa drzewa wraz z przyrostem na grubość?

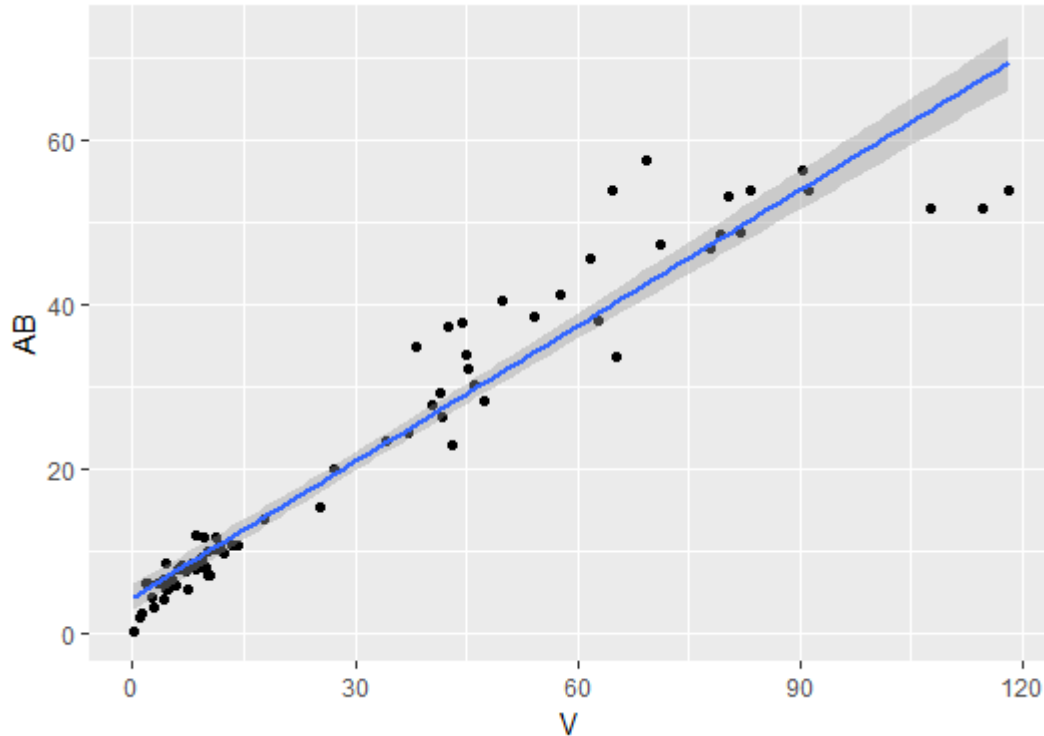
problem badawczy (poznanie tempa wzrostu)

problem aplikacyjny (możliwość estymacji)

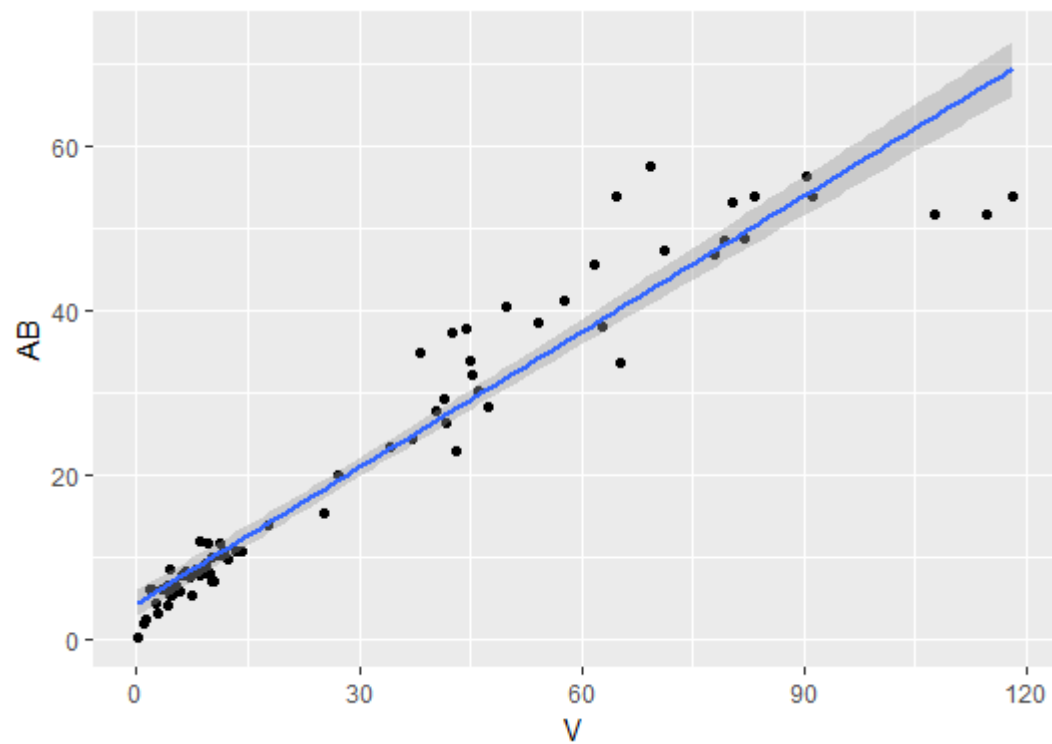
?



```
sosny<-read.csv('sosny.csv', sep=',' )  
ggplot(sosny, aes(x=V,y=AB))+geom_point()+geom_smooth(method='lm')
```



jak to się stało?



Model liniowy

$$y=a*x+b$$

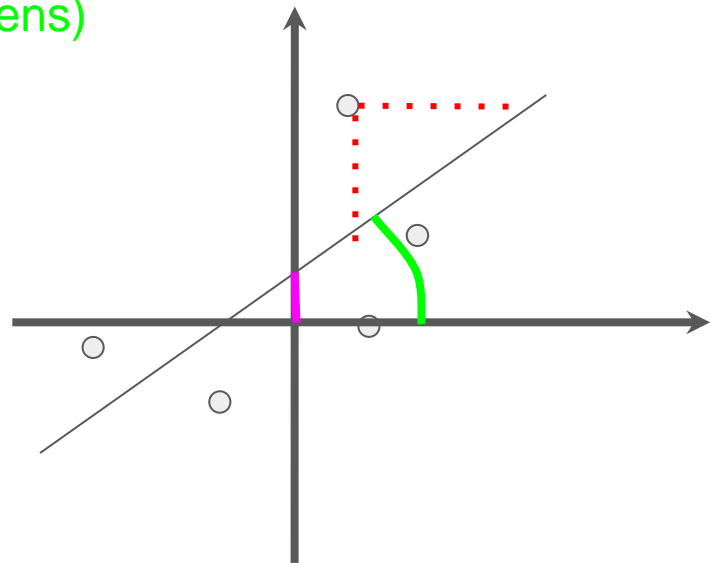
a - współczynnik kierunkowy, slope, regression coefficient, beta

nachylenie linii regresji (kąt – w zasadzie jego tangens)

b - wyraz wolny, intercept

punkt przecięcia z osią Y, położenie linii

dopasowanie - metodą najmniejszych kwadratów



Model liniowy

$$Y = ax + b$$

zapis matematyczny

$$Y \sim X$$

zapis w R

~ - tylda (pod Esc)

Y

zmienna zależna
odpowiedź (response)
zmienna modelowana

X

zmienna niezależna
predyktor

coś co chcemy
wymodelować

coś, co ma nam wyjaśniać Y

*ale nie parametr (parametr to a)

Założenie: rozkład normalny zmiennej zależnej
(lub zbliżony do normalnego)

```
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))

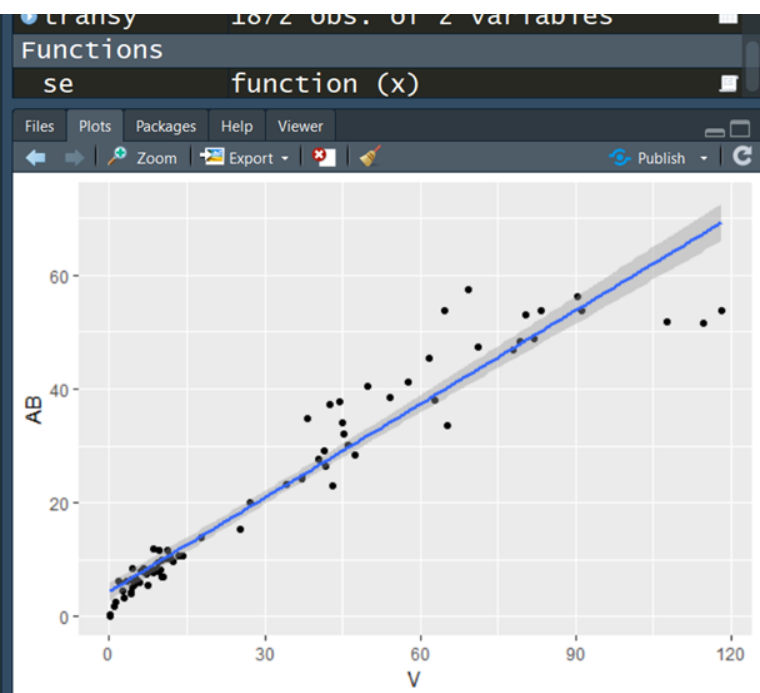
```
> summary(lm(AB~V,data=sosny))
```

```
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
```



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))

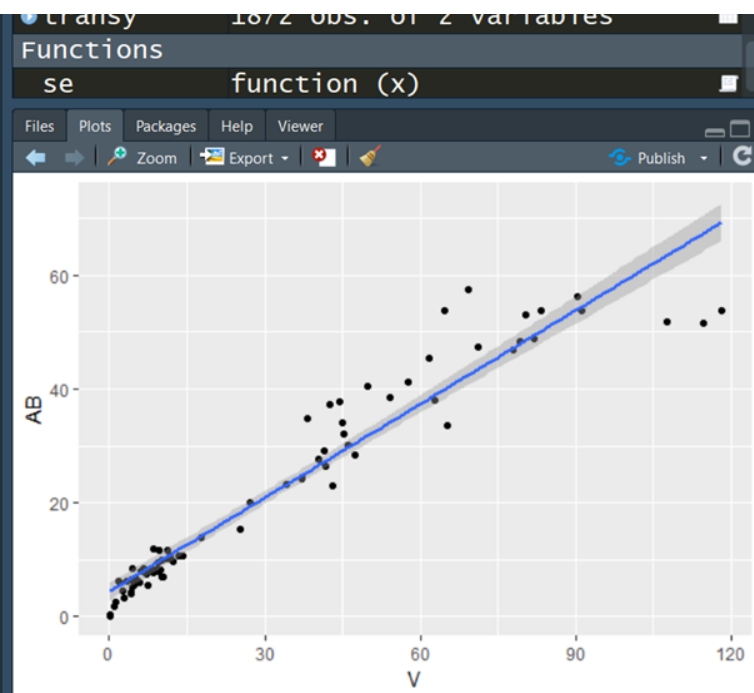
Call:
lm(formula = AB ~ V, data = sosny)

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

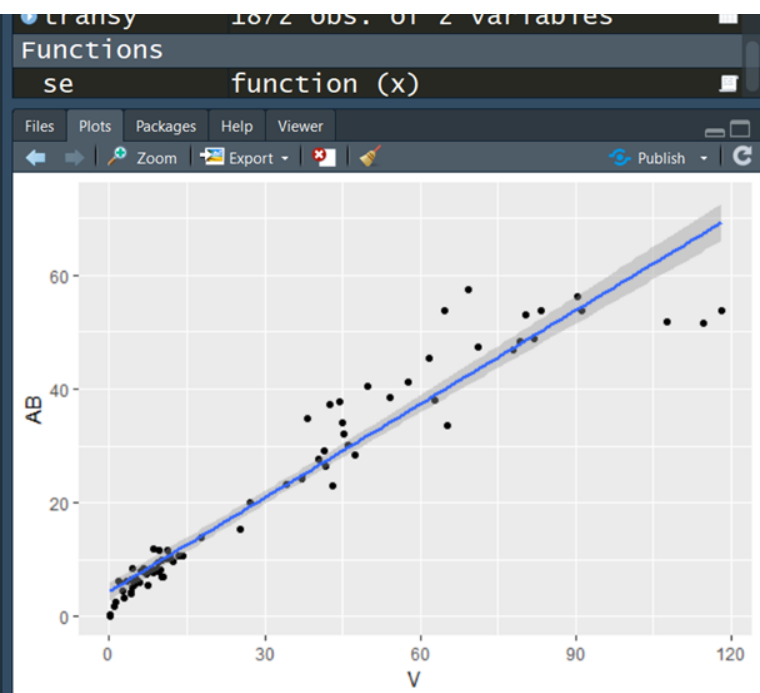
Coefficients:

	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 ***

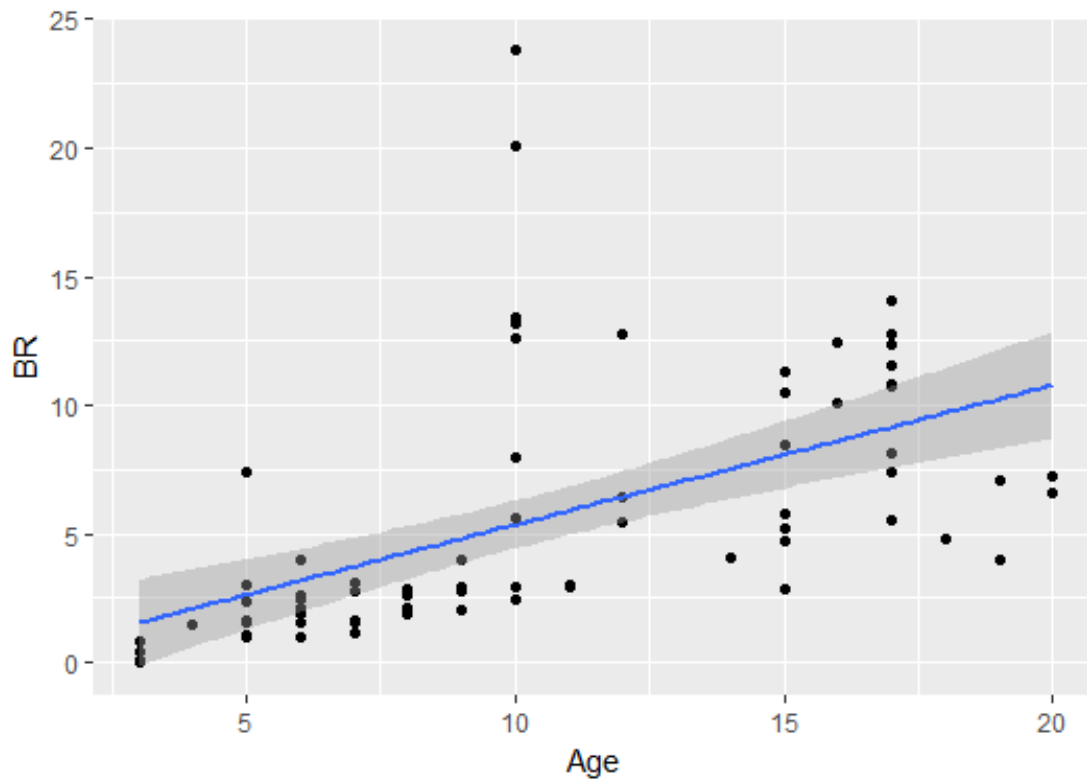
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

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
```

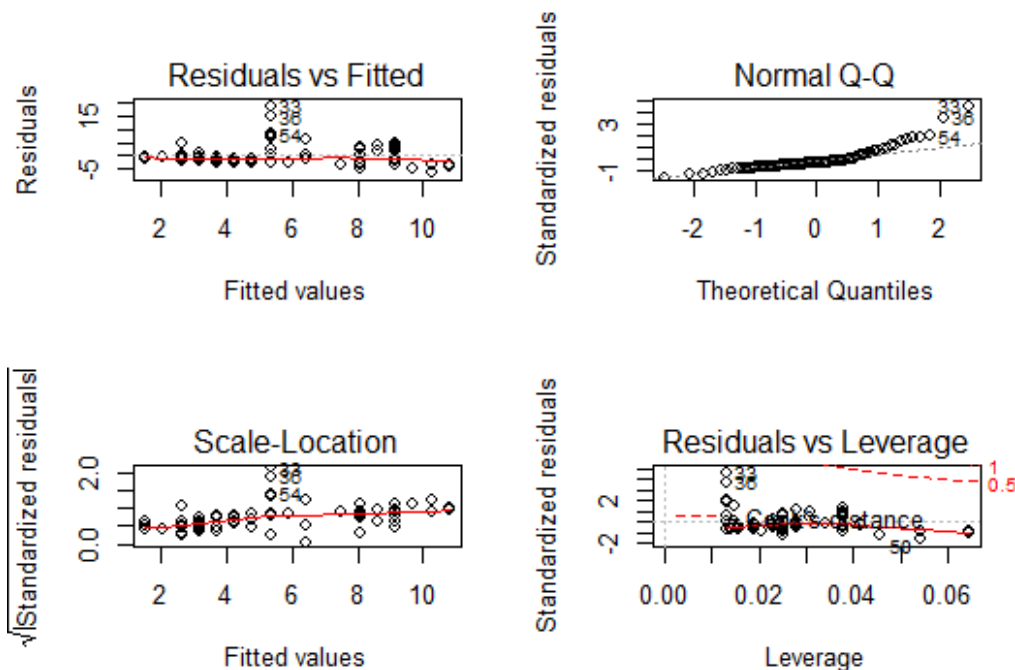
```
Multiple R-squared:  0.3009,    Adjusted R-squared:  0.2916
```

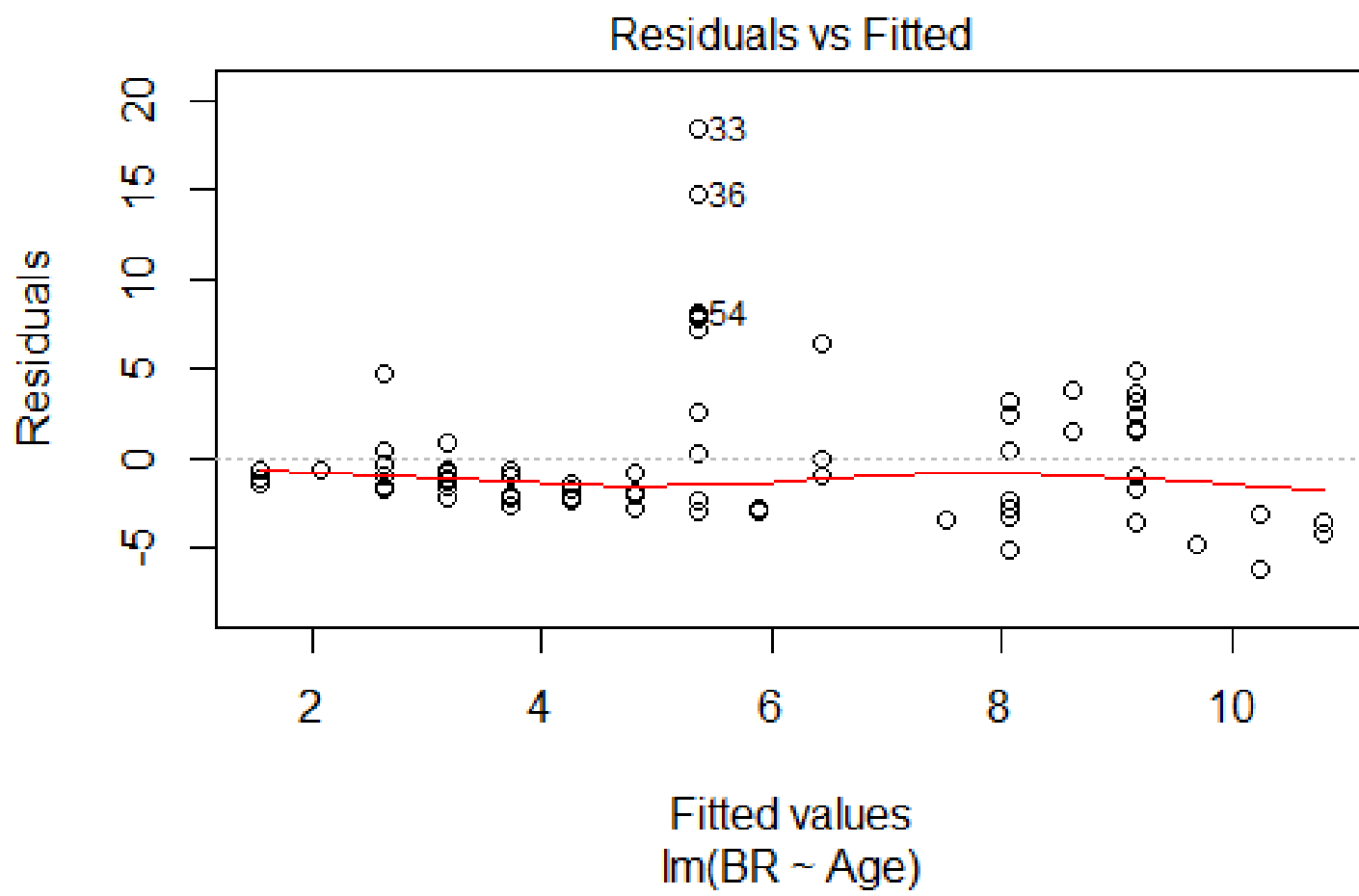
```
F-statistic: 32.28 on 1 and 75 DF,  p-value: 2.411e-07
```

```
> |
```

Diagnostyka modeli

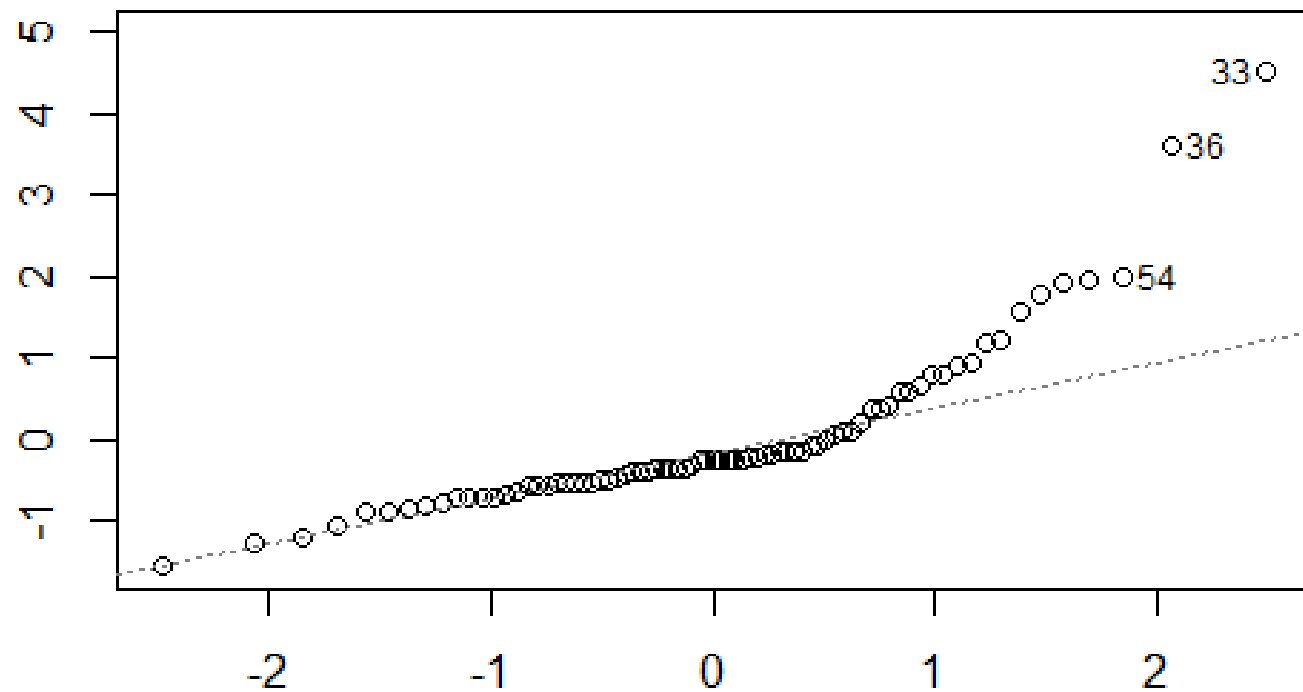
```
par(mfrow=c(2,2)) #podział wykresu na 4  
plot(lm(BR~Age,data=sosny))
```



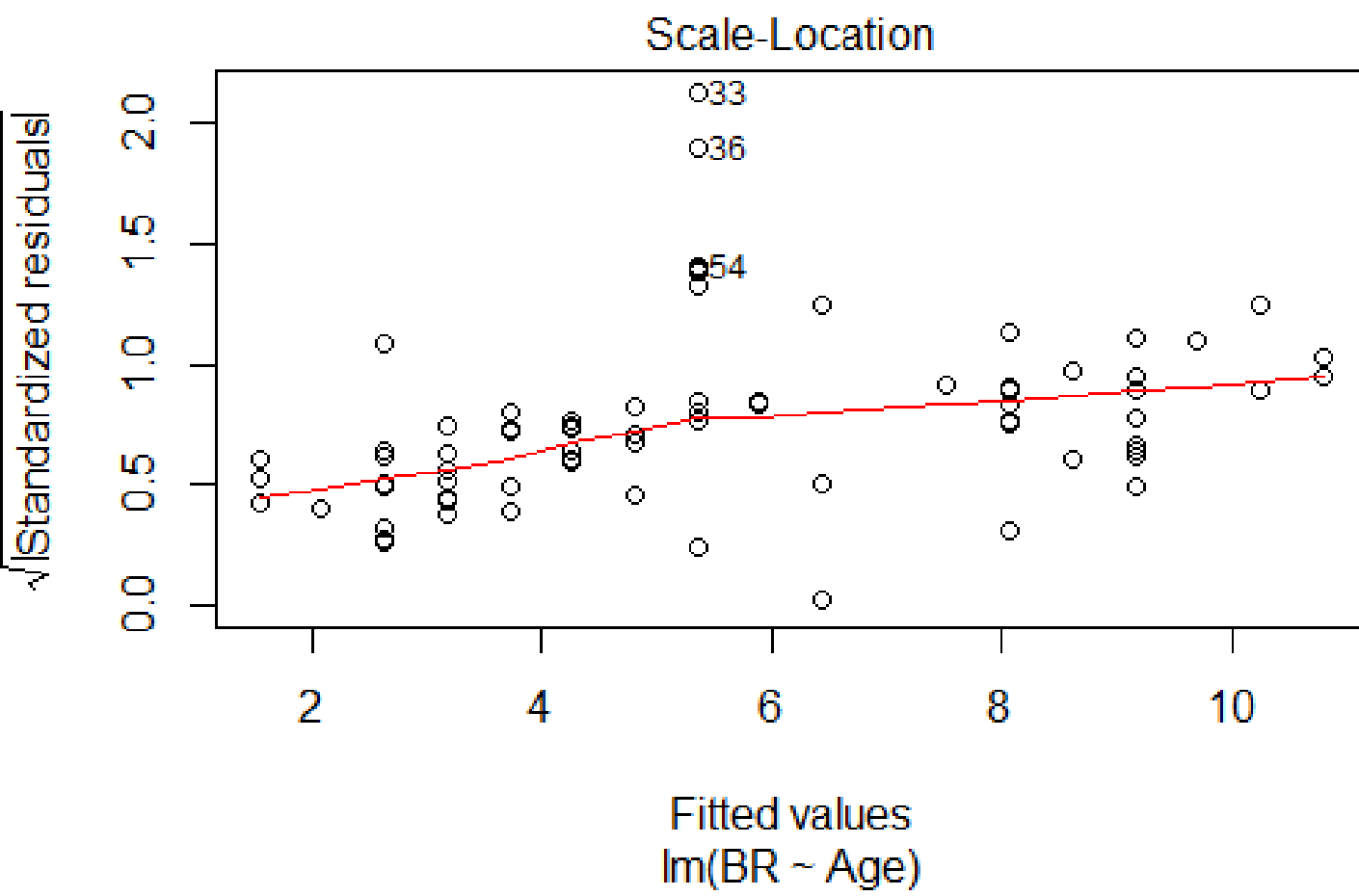


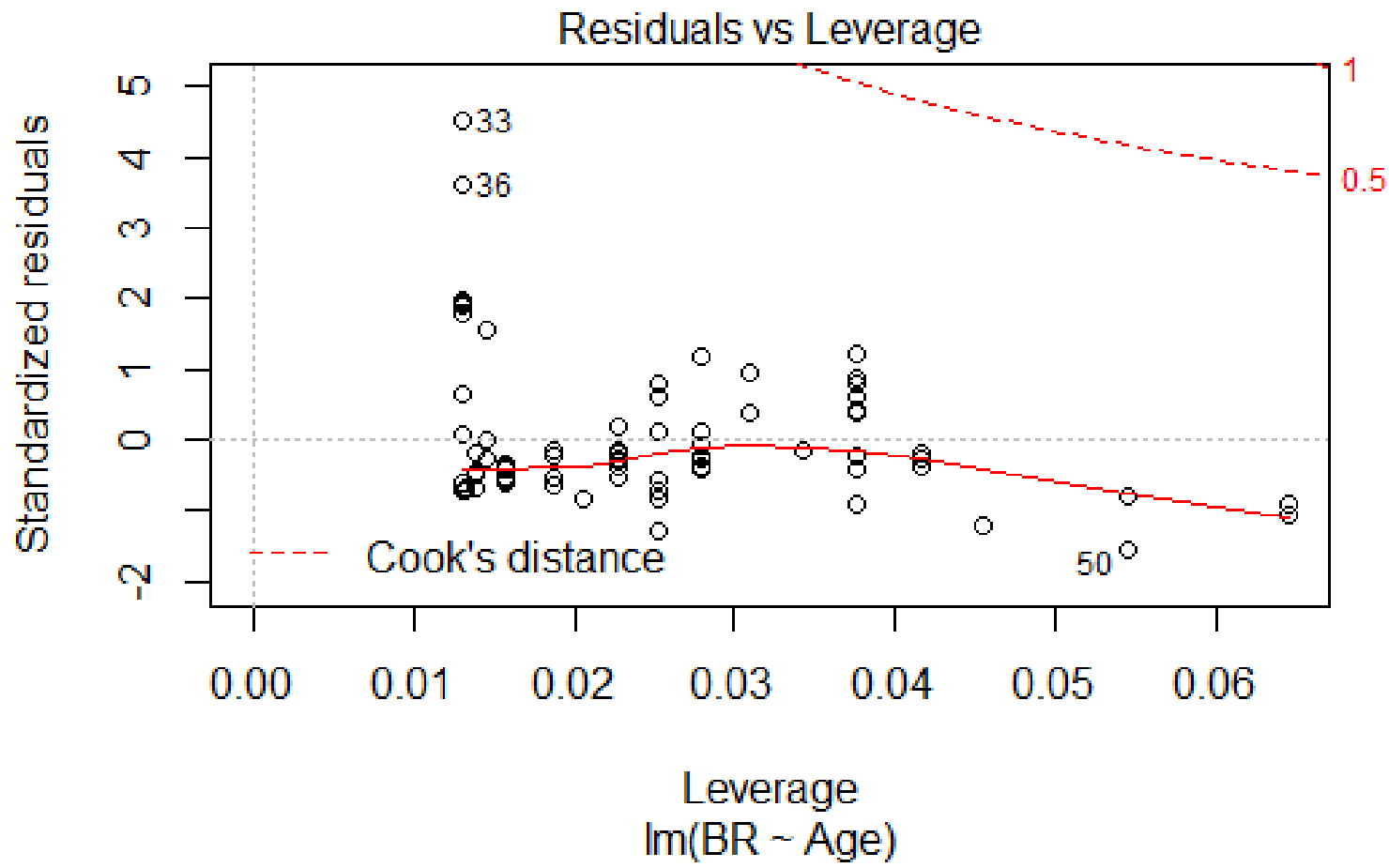
Normal Q-Q

Standardized residuals

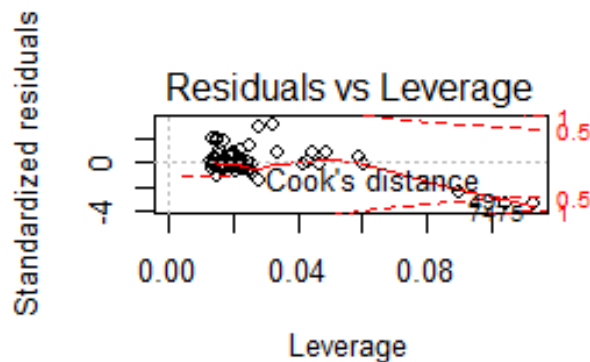
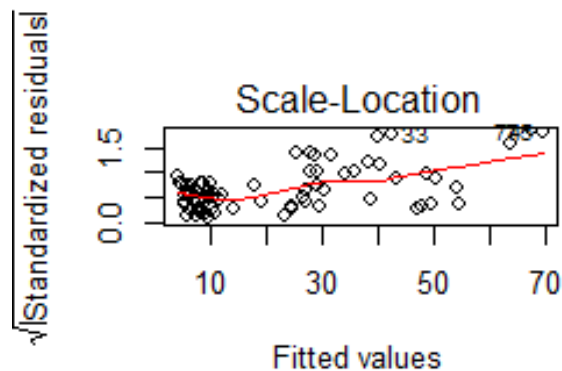
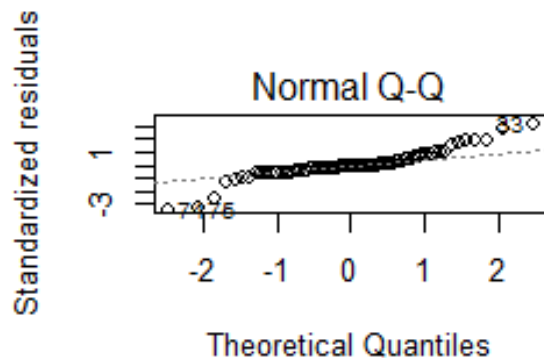
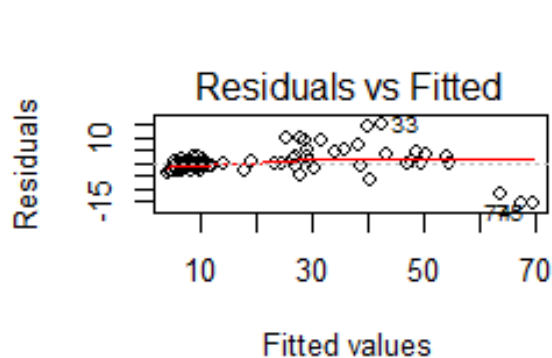


Theoretical Quantiles
lm(BR ~ Age)

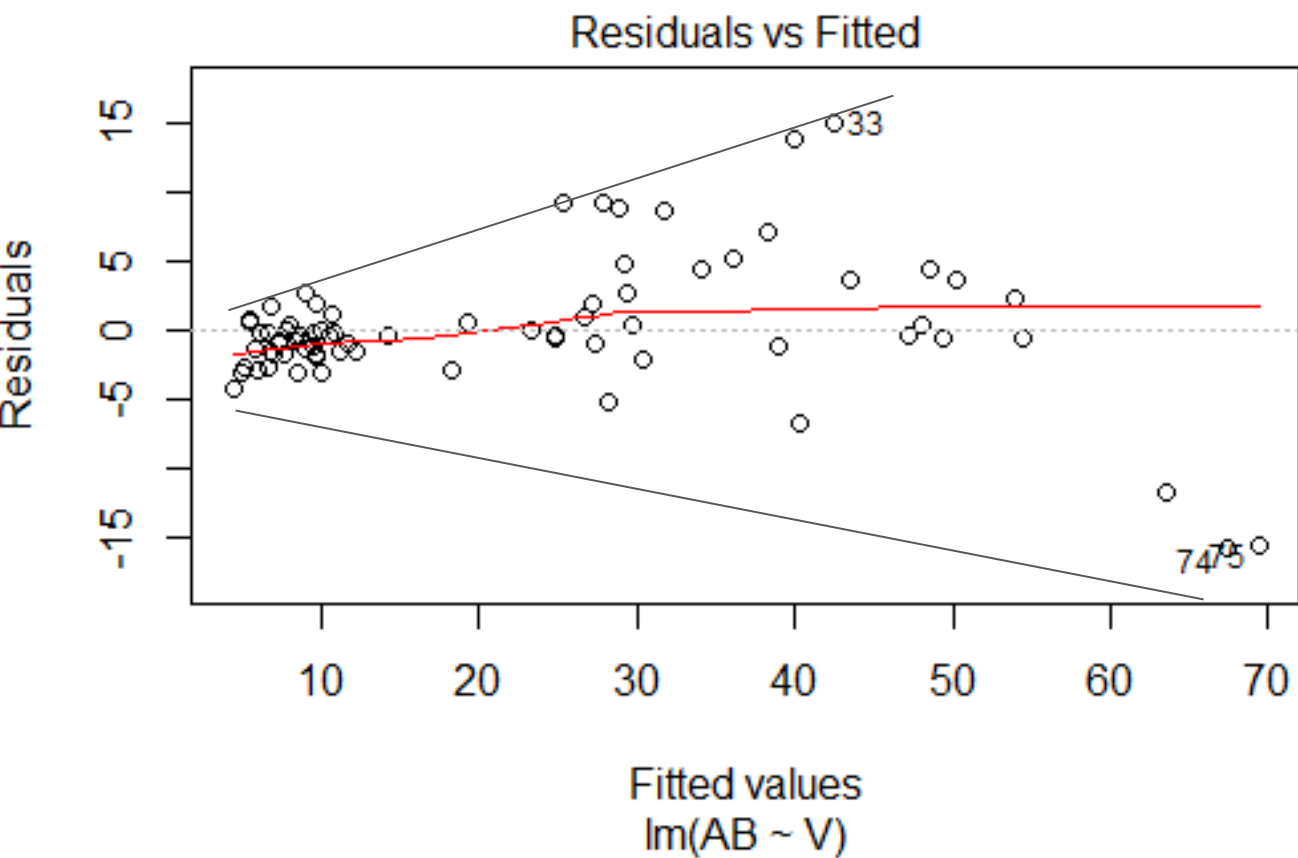


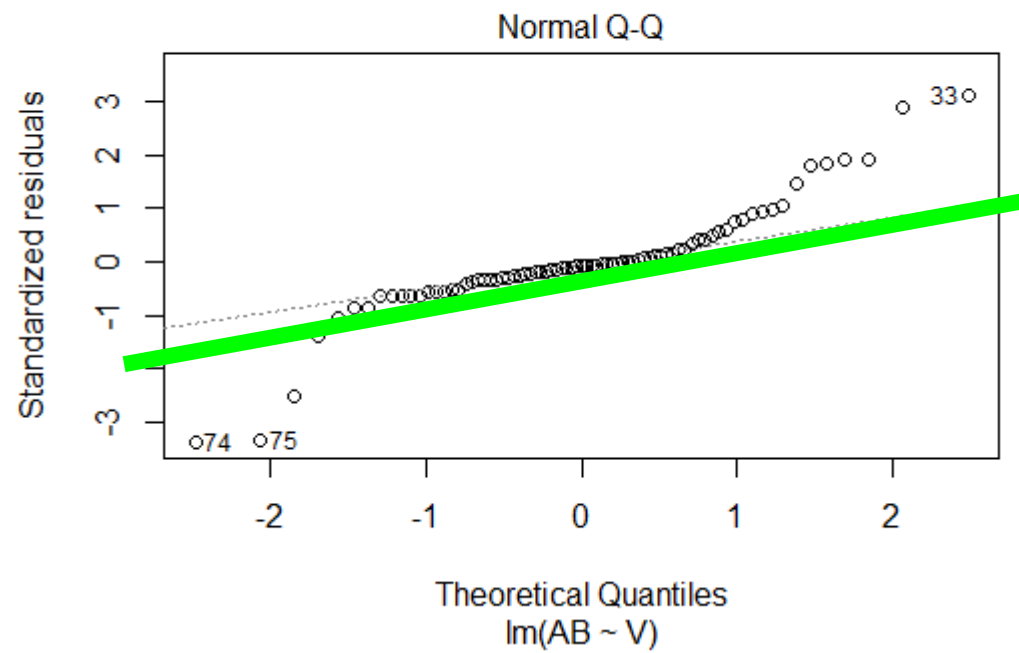


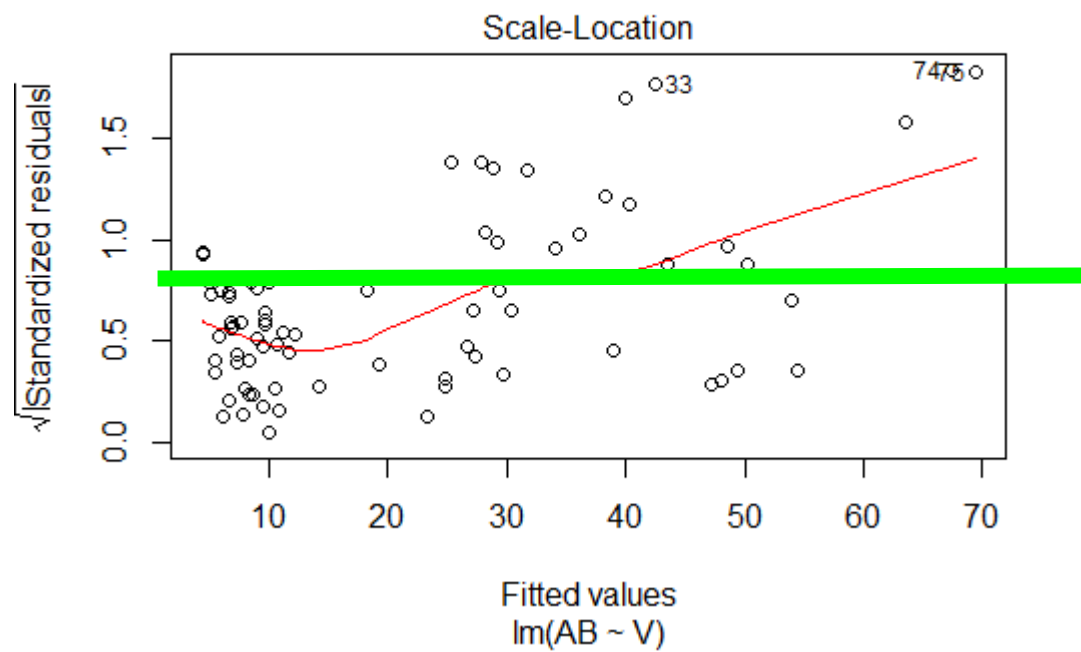
Jak wygląda dobry model? $AB \sim V \dots$

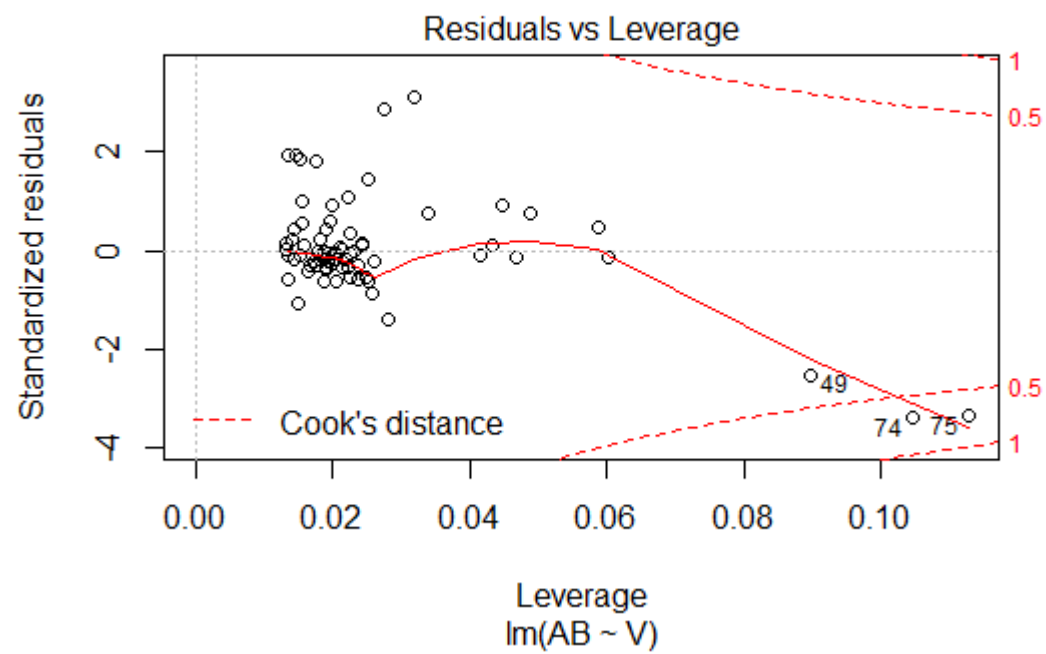


heteroskedastyczność!









Co się stało?

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

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

Inny sposób oceny jakości modeli

- pakiet DHARMa

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

```
library(DHARMa)
```

```
model.sr<-simulateResiduals(model)
```

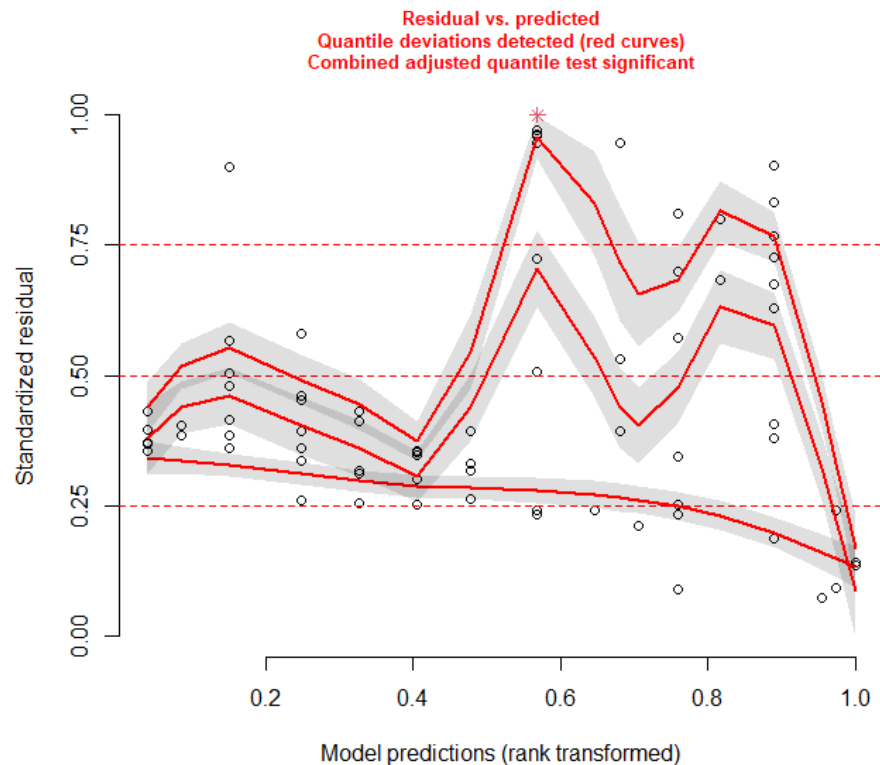
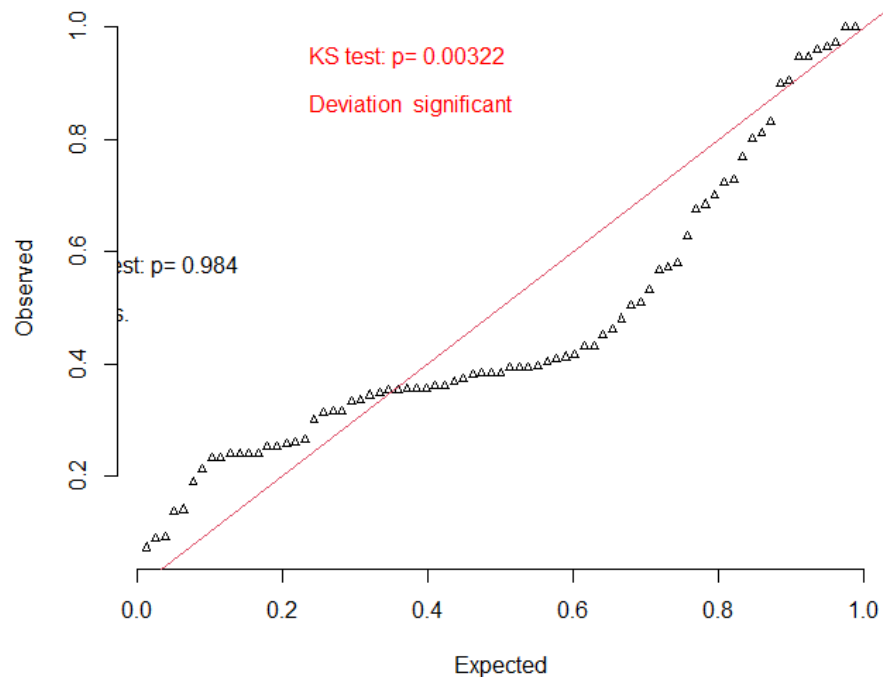
```
plot(model.sr)
```

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

Inny sposób oceny jakości modeli

DHARMA residual diagnostics

QQ plot residuals



Czy można tylko oceniać na podstawie R^2 ?

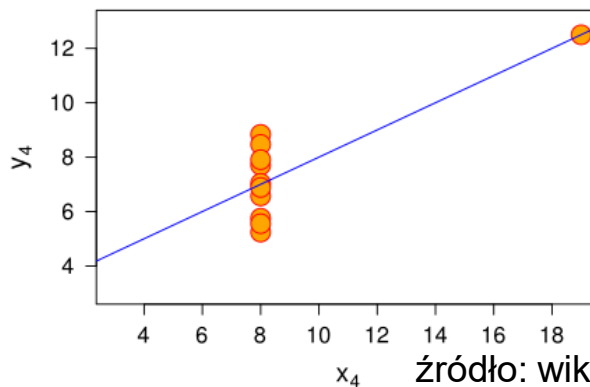
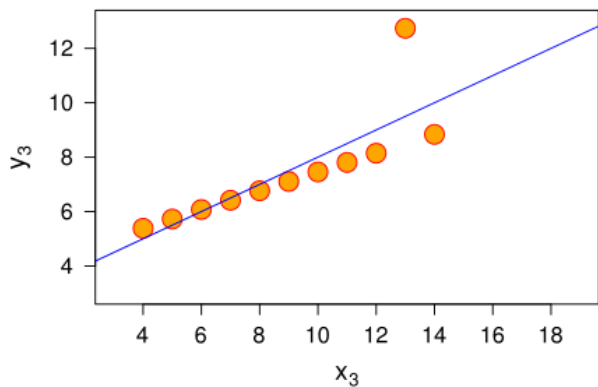
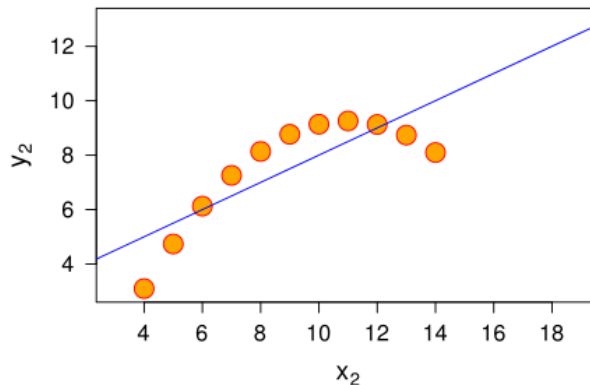
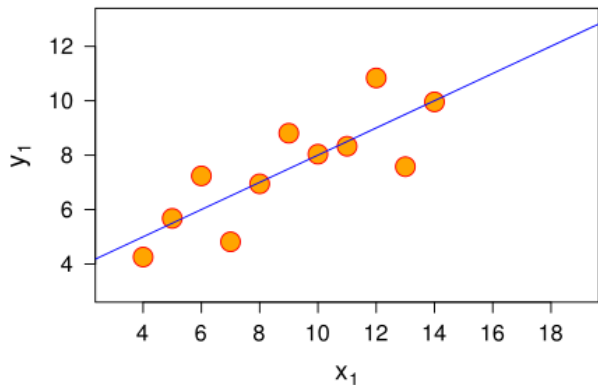
NIE!

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 modelu

inny typ rozkładu – uogólnienie LM na inne – GLM (jutro:)

Transformacje

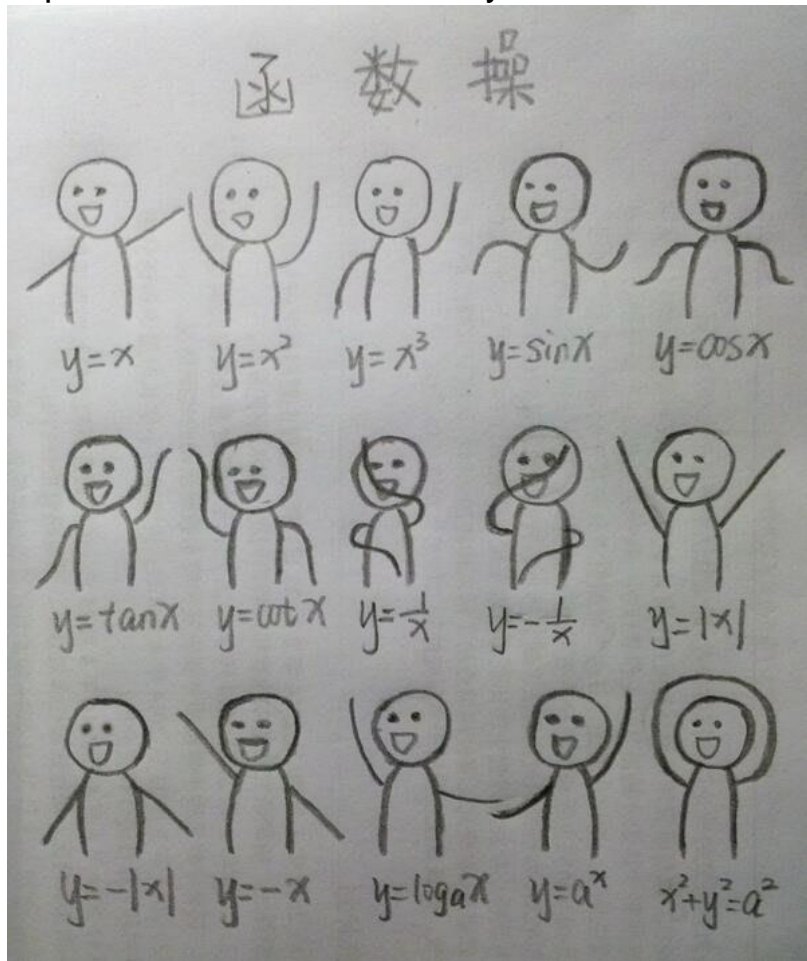
logarytm - $\log(x)$

pierwiastek - \sqrt{x}

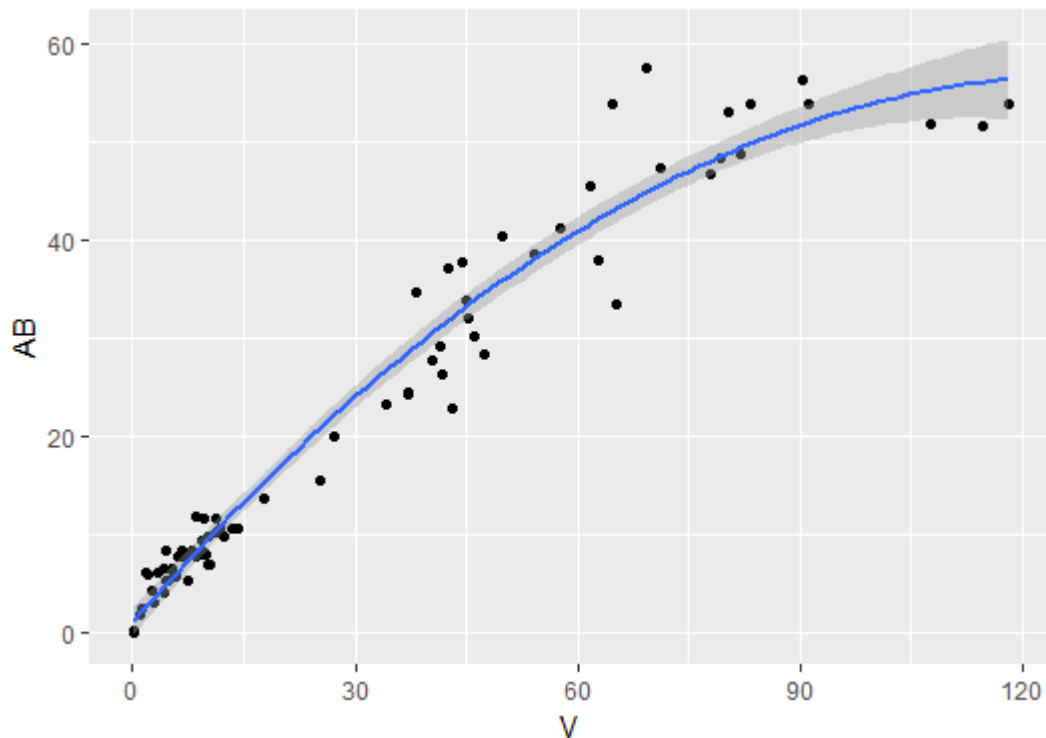
skalowanie i centrowanie $\text{scale}(x)$

transformacje potęgowe (Box-Cox, Yeo-Johnson)*

Inny typ modelu - funkcje



Równania kwadratowe $y=ax^2+bx+c$



Wielomian drugiego stopnia
Model kwadratowy
Parabola
polynomial

model:
`lm(y~poly(x,2), data)`

```
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	***

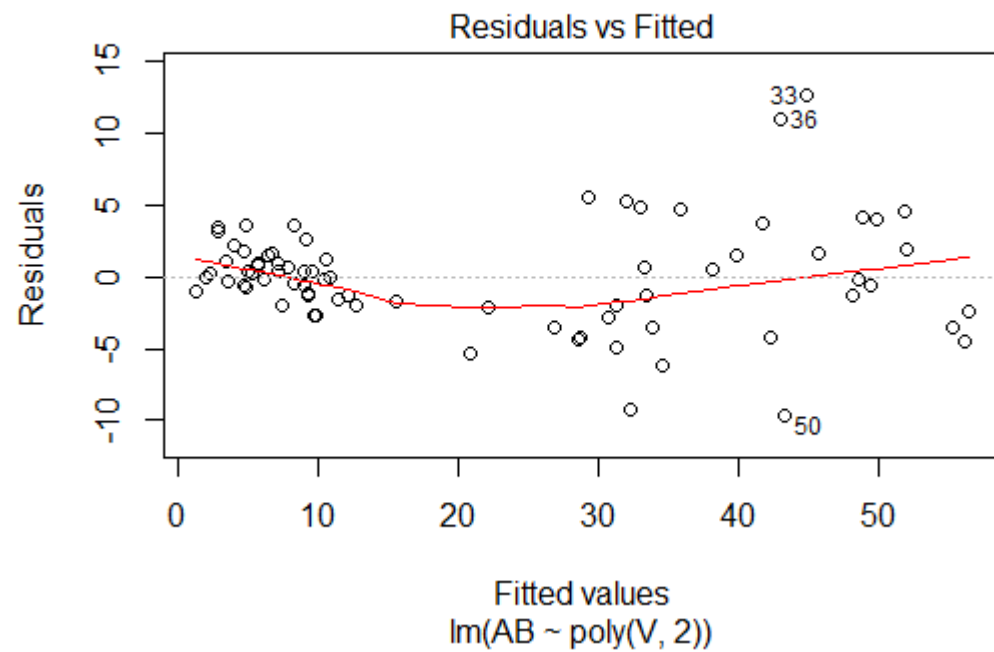
signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

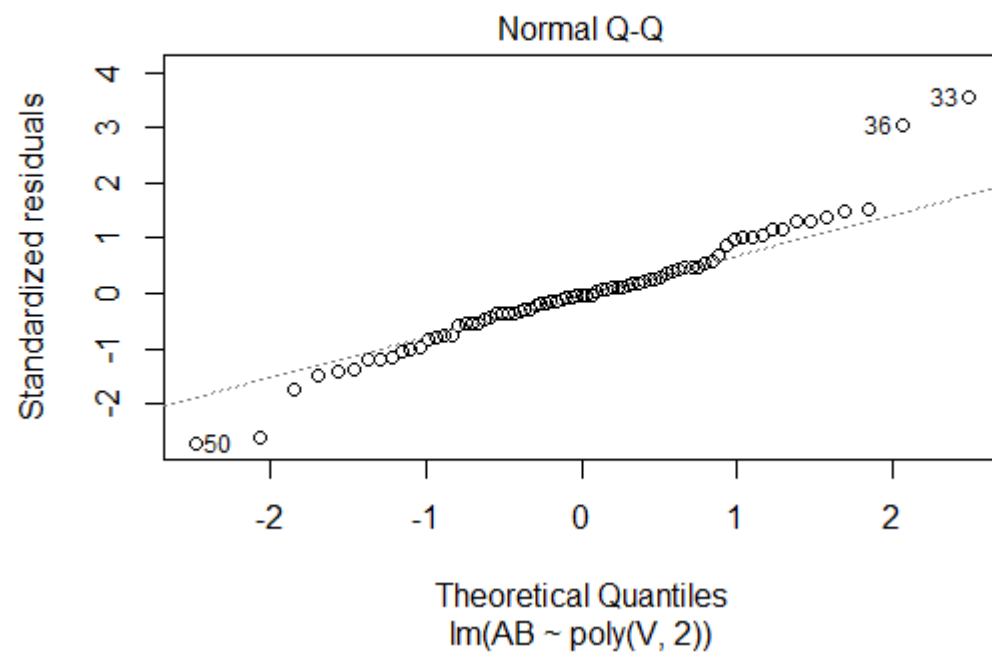
Residual standard error: 3.64 on 74 degrees of freedom

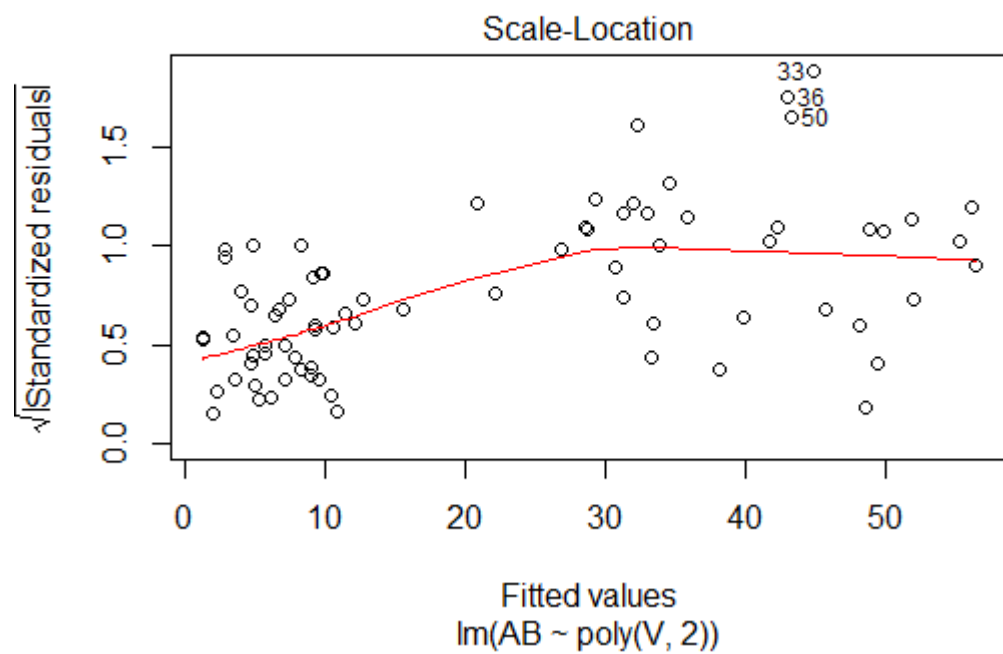
Multiple R-squared: 0.9603, Adjusted R-squared: 0.9592

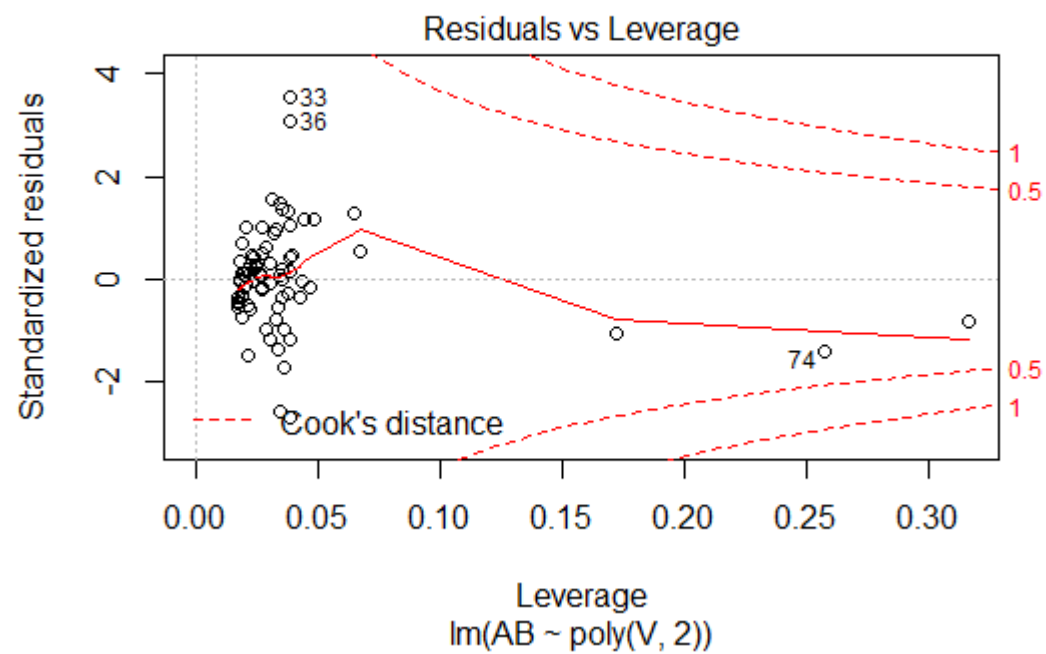
F-statistic: 895.4 on 2 and 74 DF, p-value: < 2.2e-16

```
> | 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

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

pierwiastek sumy kwadratów odchyleń / liczba obserwacji -

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

Nie mówmy dwa razy o tym samym

Rule of Thumb $VIF > 5 \Rightarrow$ 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)

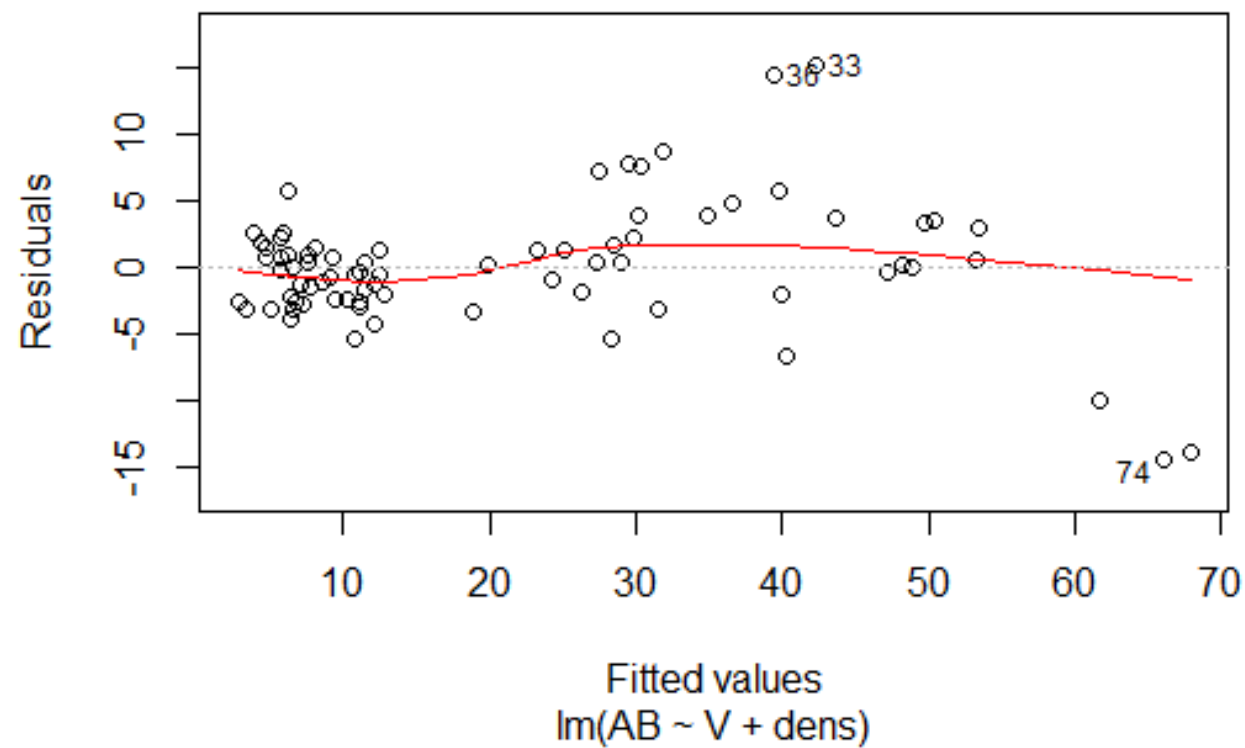
Residuals:
    Min       1Q   Median       3Q      Max
-14.537  -2.552  -0.056   1.745  15.193

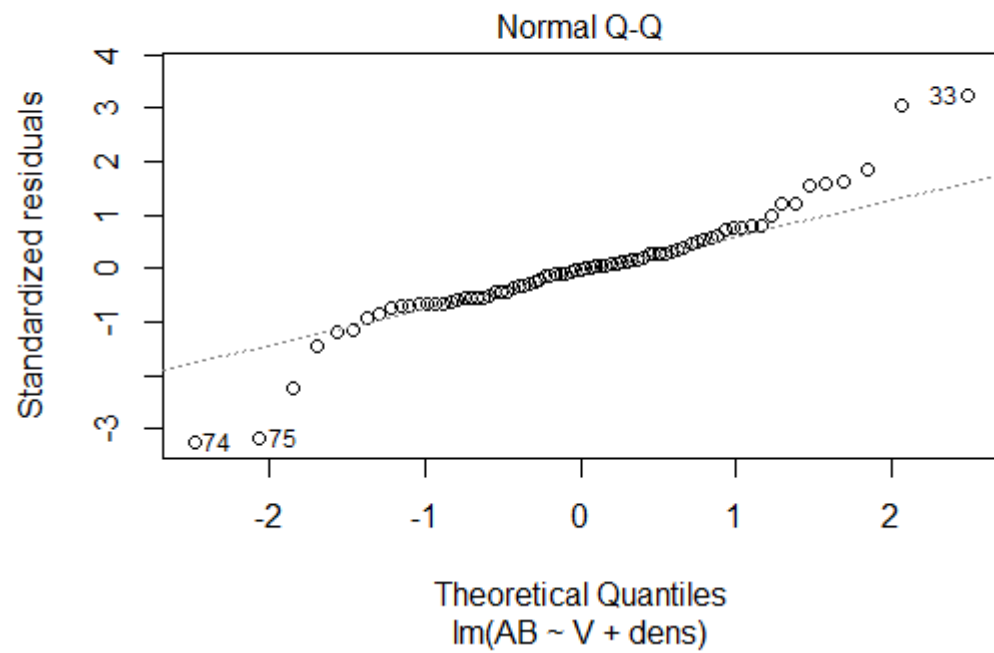
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  9.8163784   2.2893626   4.288 5.37e-05 ***
V             0.5190664   0.0211514  24.540 < 2e-16 ***
dens        -0.0006439   0.0002565  -2.510  0.0142 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

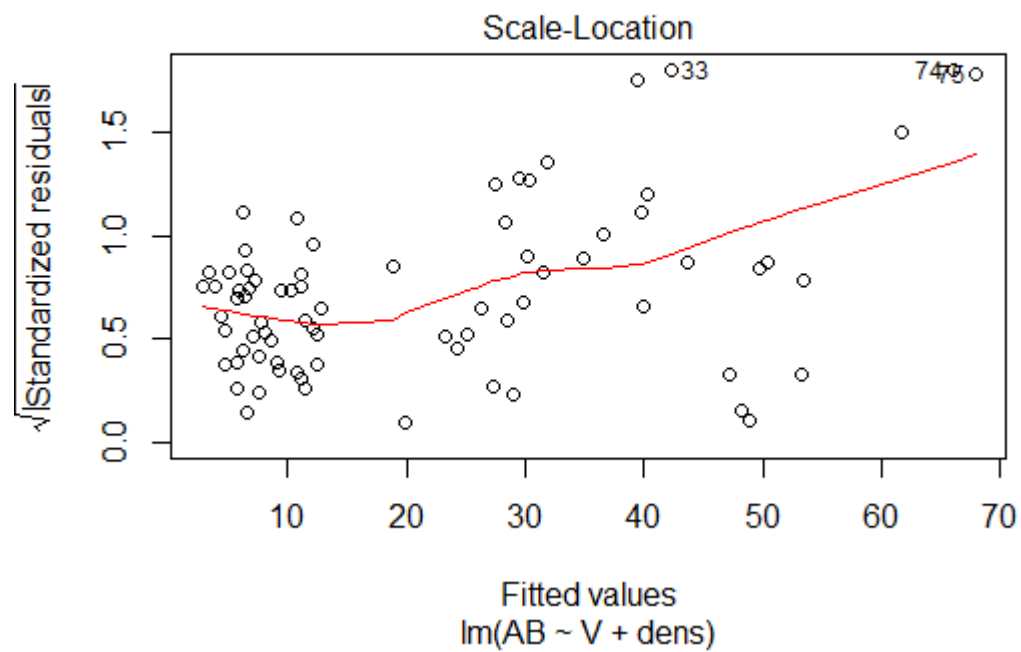
Residual standard error: 4.75 on 74 degrees of freedom
Multiple R-squared:  0.9325,    Adjusted R-squared:  0.9306
F-statistic: 510.8 on 2 and 74 DF,  p-value: < 2.2e-16

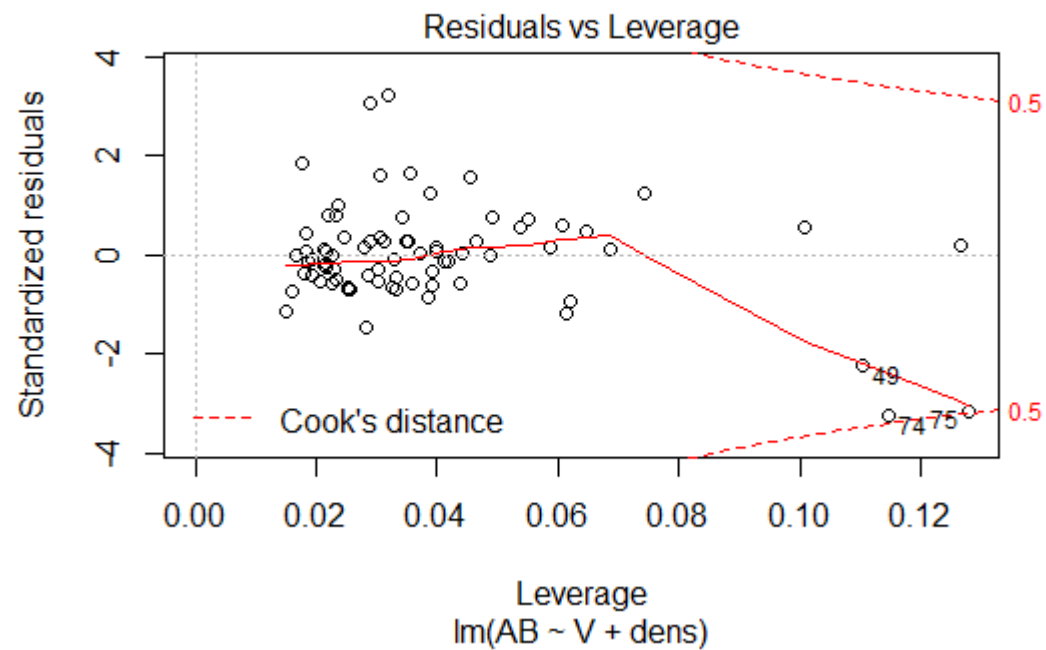
> AIC(model5)
[1] 463.3966
> vif(model5)
      V      dens
1.502807 1.502807
>
```

Residuals vs Fitted









Jak postępować z modelami?

Co trzeba sprawdzać?

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

AIC (pamiętamy o modelu zerowym)

wielkość efektu (estimate, wykresy)

rozrzut punktów (wykresy)

R^2

Selekcja modeli – funkcja step()

```
> step(lm(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:
```

```
lm(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)

Średnie brzegowe/ odpowiedzi brzegowe

- Model z więcej niż jedną zmienną trudniej pokazać
- Jak zmienia się wartość oczekiwana przy założeniu że pozostałe parametry są stałe?
- $\text{Masa} = 2.5 * \text{wysokość} + \text{typ1} * 0.2 + \text{typ2} * 0.5 + \text{typ3} * (-0.3)$
- Wiemy jak typ modyfikuje masę przy danej wysokości, ale jak pokazać średnie wartości z modelu?
- Co jak jest więcej zmiennych?

marginal means, emmeans, marginal response, partial dependence

Średnie brzegowe – marginal means – średnie z modelu

```
model<-lm(AB~V+type,data=sosny)
library(multcomp);library(emmeans)
```

```
summary(model)
```

Call:

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

Residuals:

Min	1Q	Median	3Q	Max
-15.360	-2.046	-0.242	1.509	16.415

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.94087	0.79884	4.933	4.93e-06 ***
V	0.53795	0.01845	29.155	< 2e-16 ***
typepostagric	5.03635	1.96117	2.568	0.0123 *
typepostind	1.42232	1.33221	1.068	0.2892

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

Residual standard error: 4.764 on 73 degrees of freedom

Multiple R-squared: 0.933, Adjusted R-squared: 0.9302

F-statistic: 338.7 on 3 and 73 DF, p-value: < 2.2e-16

```
emmeans(model, 'type')
```

type	emmean	SE	df	lower.CL	upper.CL
forest	20.7	0.69	73	19.4	22.1
postagric	25.8	1.82	73	22.2	29.4
postind	22.2	1.10	73	20.0	24.4

Confidence level used: 0.95

```
cld(emmeans(model, 'type'))
```

type	emmean	SE	df	lower.CL	upper.CL	.group
forest	20.7	0.69	73	19.4	22.1	1
postind	22.2	1.10	73	20.0	24.4	12
postagric	25.8	1.82	73	22.2	29.4	2

Confidence level used: 0.95

P value adjustment: tukey method for comparing a family of 3 estimates
significance level used: alpha = 0.05

NOTE: Compact letter displays can be misleading

because they show NON-findings rather than findings.

Consider using 'pairs()', 'pwpp()', or 'pwpm()' instead.

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:

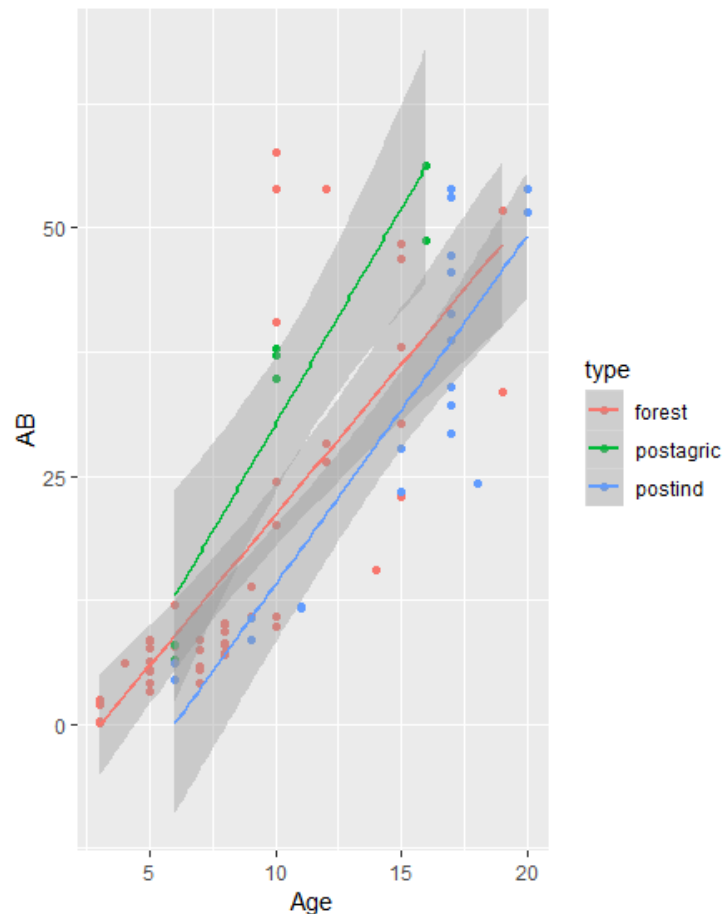
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-11.1761	2.7075	-4.128	9.61e-05 ***
Age	3.2661	0.2742	11.911	< 2e-16 ***
typepostagric	9.4329	4.0264	2.343	0.0219 *
typepostind	-6.2183	3.1153	-1.996	0.0497 *

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)
```

Odpowiedź brzegowa - ggeffects

```
mod11<-lm(AB~V+N+dens, sosny)
> summary(mod11)
```

Weźmy model z trzema zmiennymi

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

Residuals:

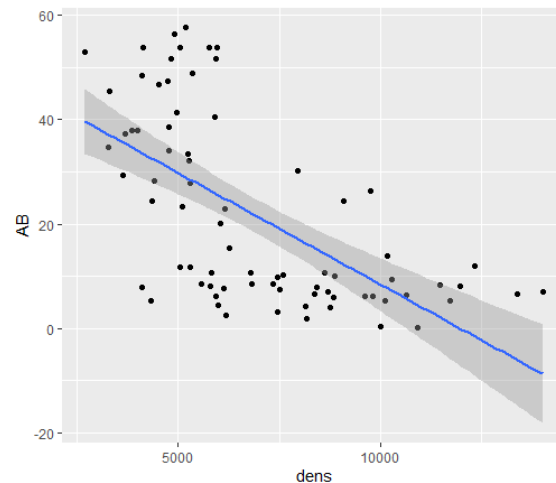
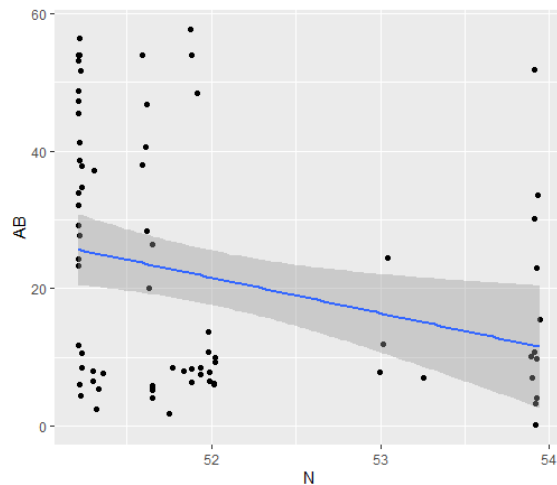
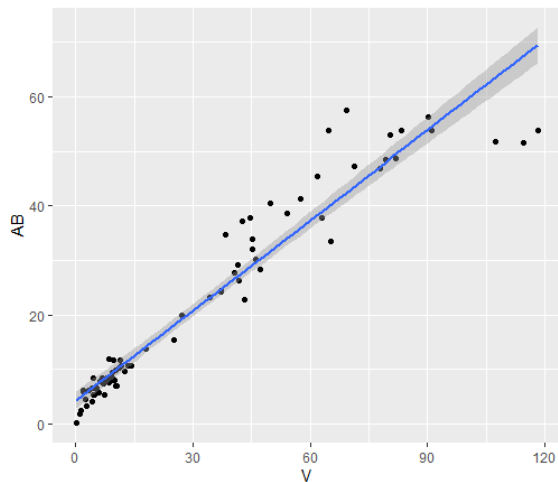
Min	1Q	Median	3Q	Max
-15.2934	-2.1345	-0.2878	1.3569	15.3180

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	83.6055072	29.5867642	2.826	0.00608 **
V	0.5192881	0.0204384	25.407	< 2e-16 ***
N	-1.4437916	0.5772877	-2.501	0.01463 *
dens	-0.0004569	0.0002589	-1.765	0.08175 .

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

Summary jedno, obrazki drugie



Dlaczego?

Każdy obrazek osobno, bez całościowego uwzględnienia modelu

Marginal response

- Odpowiedź brzegowa, zakładająca średni poziom pozostałych zmiennych
- $y = ax_1 + bx_2 + cx_3 + d$, podstawiamy pod x_1 poszczególne wartości
- pod x_2 i pod x_3 podstawiamy średnią wartość z zestawu danych
- mamy więc obraz jak wraz ze zmianą x_1 zmieni się y przy założeniu że pozostałe czynniki się nie zmieniają
- *ceteris paribus* – wszystkie bez jednego się nie zmieniają

```
ggpredict(mod11)
```

```
$V
```

```
# Predicted values of AB
```

```
V | Predicted | 95% CI
```

```
-----
```

```
0 | 5.75 | [3.90, 7.60]
```

```
15 | 13.54 | [12.13, 14.95]
```

```
30 | 21.33 | [20.20, 22.45]
```

```
45 | 29.12 | [27.98, 30.25]
```

```
60 | 36.91 | [35.49, 38.32]
```

```
75 | 44.69 | [42.83, 46.55]
```

```
90 | 52.48 | [50.11, 54.86]
```

```
120 | 68.06 | [64.58, 71.55]
```

```
Adjusted for:
```

```
* N = 52.01
```

```
• dens = 6055.00
```

```
$N
```

```
# Predicted values of AB
```

```
N | Predicted | 95% CI
```

```
-----
```

```
51.00 | 23.43 | [21.93, 24.93]
```

```
51.50 | 22.71 | [21.51, 23.90]
```

```
52.00 | 21.99 | [20.87, 23.10]
```

```
52.50 | 21.26 | [19.96, 22.57]
```

```
53.00 | 20.54 | [18.87, 22.21]
```

```
53.50 | 19.82 | [17.69, 21.95]
```

```
54.00 | 19.10 | [16.47, 21.72]
```

```
Adjusted for:
```

```
* V = 31.24
```

```
* dens = 6055.00
```

```
$dens
```

```
# Predicted values of AB
```

```
dens | Predicted | 95% CI
```

```
-----
```

```
2700 | 23.51 | [21.13, 25.88]
```

```
4100 | 22.87 | [21.10, 24.63]
```

```
5500 | 22.23 | [20.97, 23.48]
```

```
6900 | 21.59 | [20.56, 22.61]
```

```
8400 | 20.90 | [19.63, 22.17]
```

```
9800 | 20.26 | [18.48, 22.04]
```

```
11200 | 19.62 | [17.22, 22.02]
```

```
14000 | 18.34 | [14.61, 22.08]
```

```
Adjusted for:
```

```
* V = 31.24
```

```
* N = 52.01
```

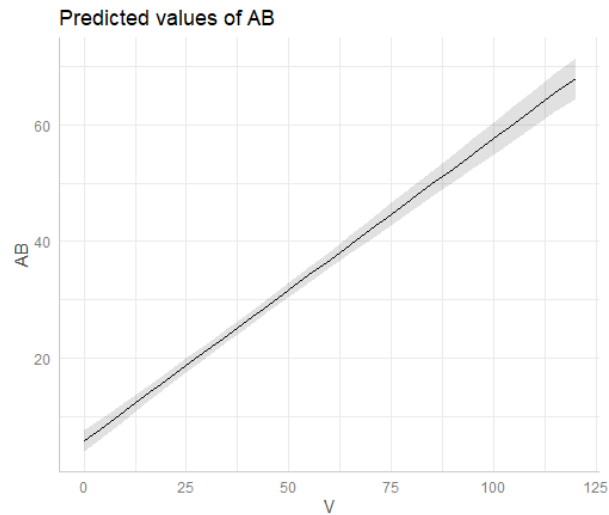
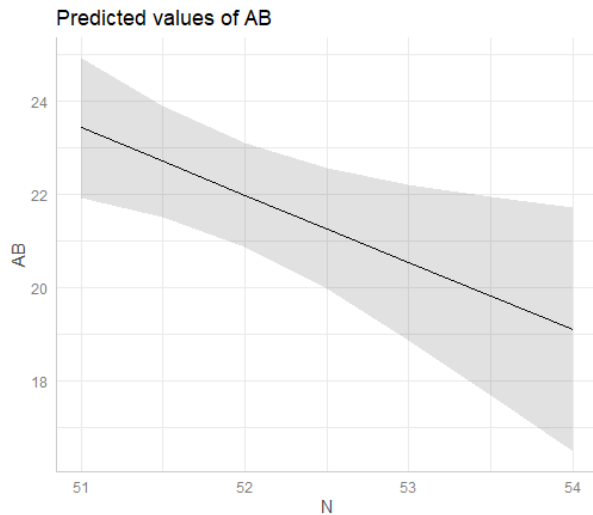
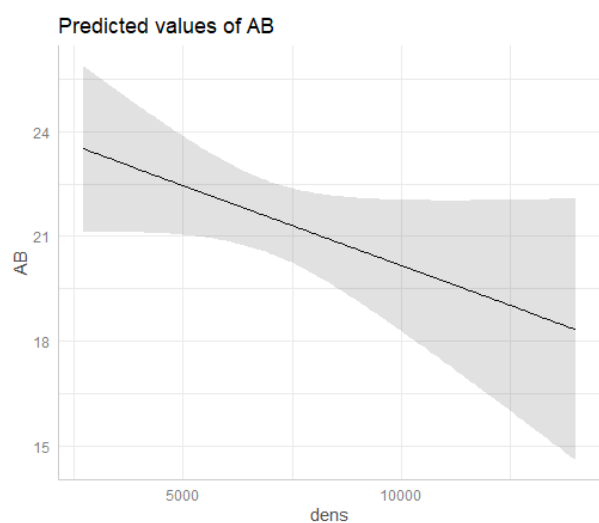
```
attr(,"class")
```

```
[1] "ggalleffects" "list"
```

```
attr(,"model.name")
```

```
[1] "mod11"
```

plot(ggpredict(mod1 1))



fajnie ale nie widać efektów – różne skale osi

Jak zrobić na jednym ggplotcie?

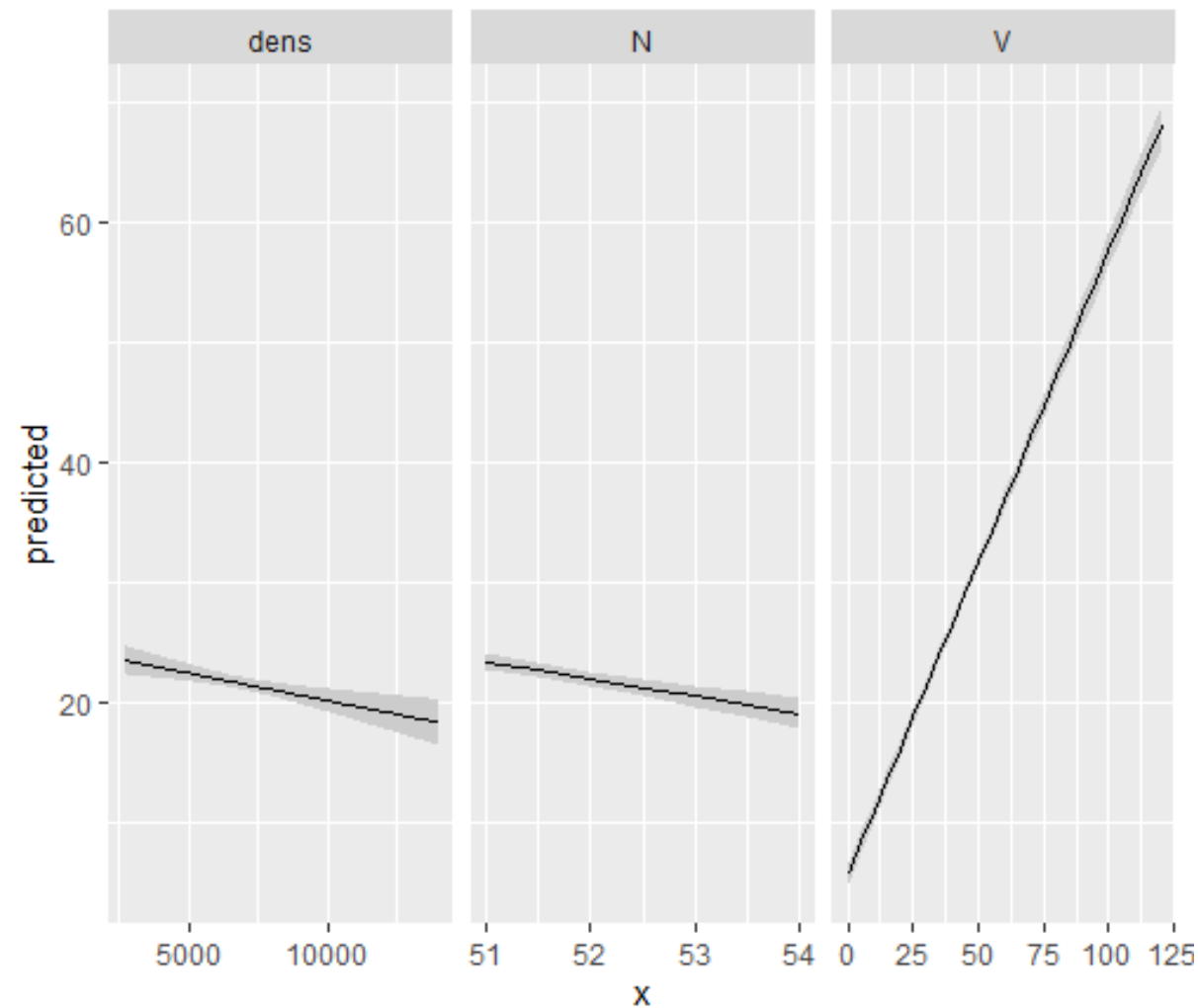
```
ggp1<-as.data.frame(ggpredict(mod11, 'V'))  
#stwórz ggpredict dla jednej zmiennej  
ggp1$group<-'V' #w kolumnie group wpisz nazwę zmiennej  
ggp2<-as.data.frame(ggpredict(mod11, 'N'))  
ggp2$group<-'N'  
ggp3<-as.data.frame(ggpredict(mod11, 'dens'))  
ggp3$group<-'dens'
```

```
ggp<-bind_rows(list(ggp1, ggp2, ggp3))  
#złącz listę tabel w jedno
```


Jak to wygląda?

```
head(ggp)
```

	x	predicted	std.error	conf.low	conf.high	group
1	0	5.748008	0.9440649	3.897675	7.598342	V
2	5	8.344449	0.8635288	6.651963	10.036934	V
3	10	10.940889	0.7880304	9.396378	12.485400	V
4	15	13.537330	0.7191579	12.127806	14.946853	V
5	20	16.133770	0.6589922	14.842169	17.425371	V
6	25	18.730210	0.6101146	17.534408	19.926013	V



```
ggplot(ggp, aes(x=x, y=predicted))  
+geom_ribbon(aes (ymin=predicted-  
std.error, ymax=predicted+std.error ),  
fill='gray80') +geom_line()  
+facet_wrap(~group, scales='free_x')
```

- widać effect size
- widać co wychodzi z modelu

Dodać rzeczywiste wartości?

- trzeba zmeltować tabelkę z danymi

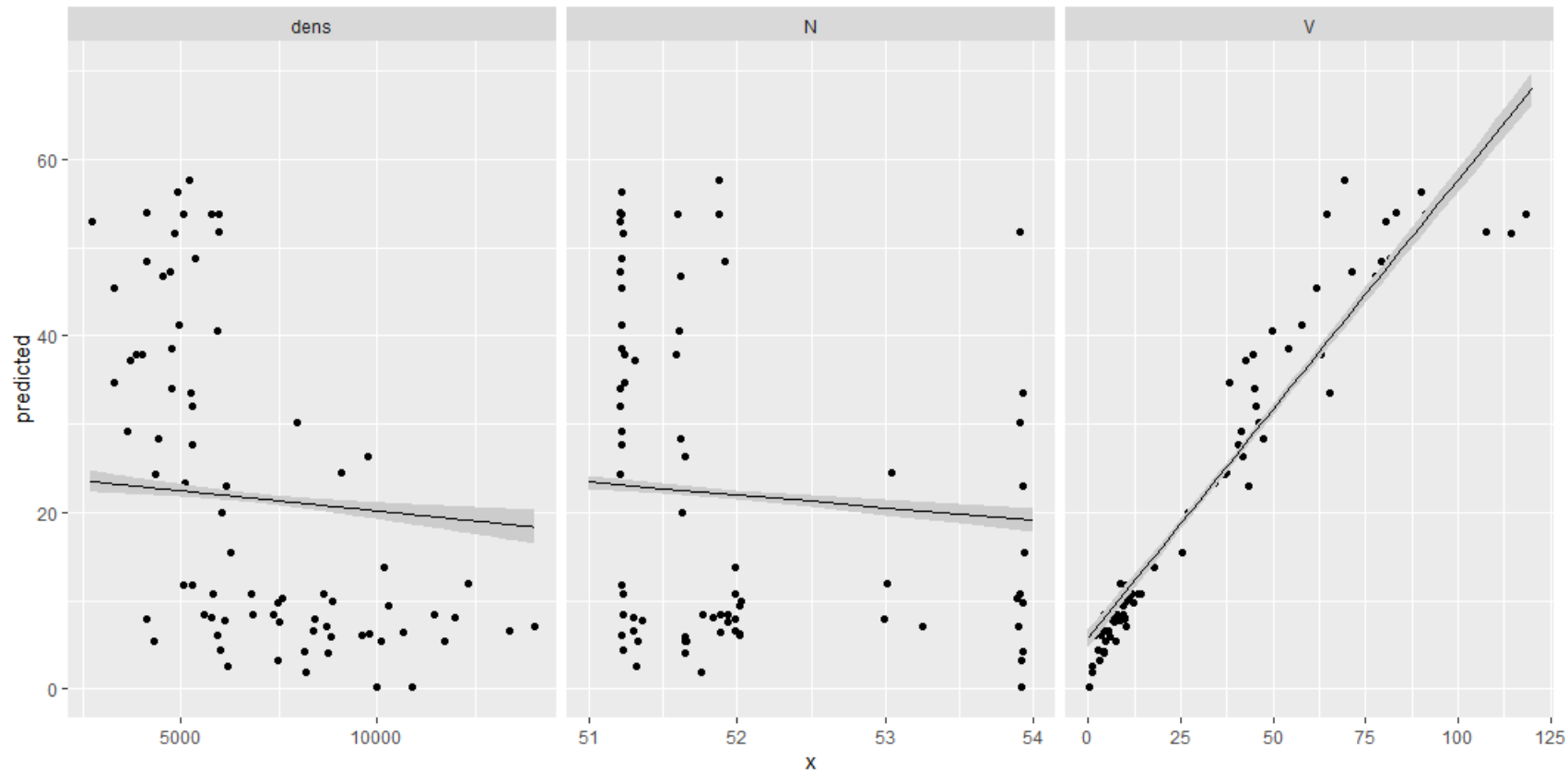
```
gg.points<-reshape2::melt(sosny[,c(2,6,9,10)], id.vars='AB')
```

```
> head(gg.points)
```

	AB	variable	value
1	2.4730	N	51.3192
2	0.2778	N	53.9168
3	0.1738	N	53.9161
4	1.8986	N	51.7522
5	1.8986	N	51.7523
6	6.2242	N	52.0138

```
colnames(gg.points)<-c('predicted','group','x')
```

```
#zmiana nazw kolumn na odpowiadające w tabeli ggp
```



```
ggplot(ggp, aes(x=x, y=predicted)) +geom_point(data=gg.points)
+geom_ribbon(aes(ymin=predicted-std.error ,ymax=predicted+std.error ),fill='gray80')
+geom_line()+facet_wrap(~group, scales='free_x')
```

#każdy geom_ może mieć swoją własną tabelę wejściową, definiowaną argumentem data=nazwa.tabeli

Outliery

obserwacje odstające

wyjątki

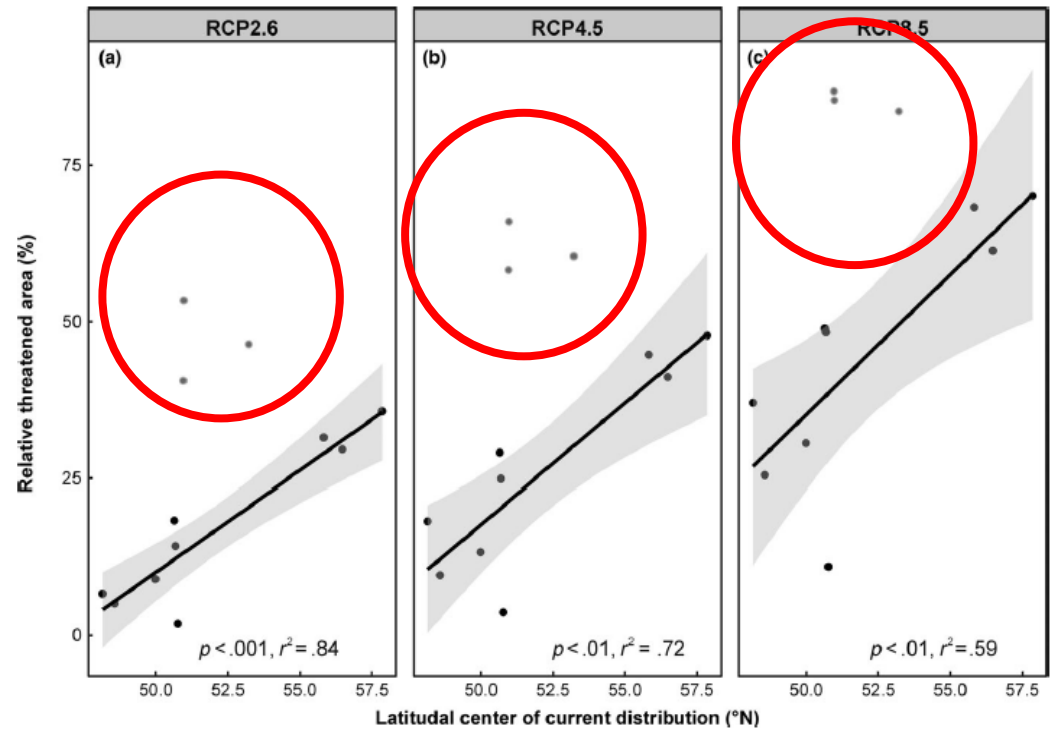


FIGURE 7 The relationship between threatened proportion of current distribution and latitudinal center of distribution for each climate change scenario: optimistic (RCP2.6; panel a), moderate (RCP4.5; panel b), and pessimistic (RCP8.5; panel c) calculated for averaged GCMs outputs. Three species outlying from the trend—*L. decidua*, *P. menziesii*, and *Q. rubra* were removed from the regression analysis and are represented by gray dots. Gray areas around the regression lines represent standard errors (SE) of each model

Received: 20 July 2017 | Accepted: 30 August 2017
DOI: 10.1111/gcb.13925

PRIMARY RESEARCH ARTICLE

WILEY | Globa

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

Marcin K. Dyderski^{1,2} | Sonia Paż³ | Lee E. Frelich⁴ | Andrzej M. Jagodziński^{1,2}

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:

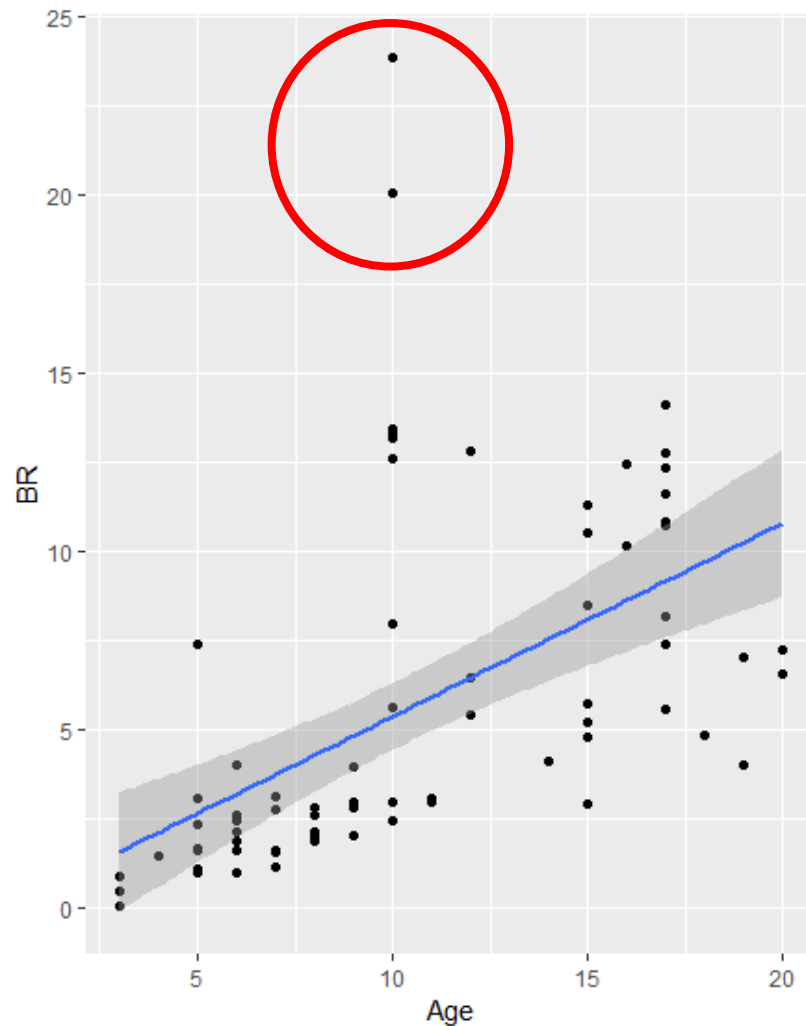
Min	1Q	Median	3Q	Max
-6.2556	-2.2313	-1.0834	0.7993	18.5083

Coefficients:

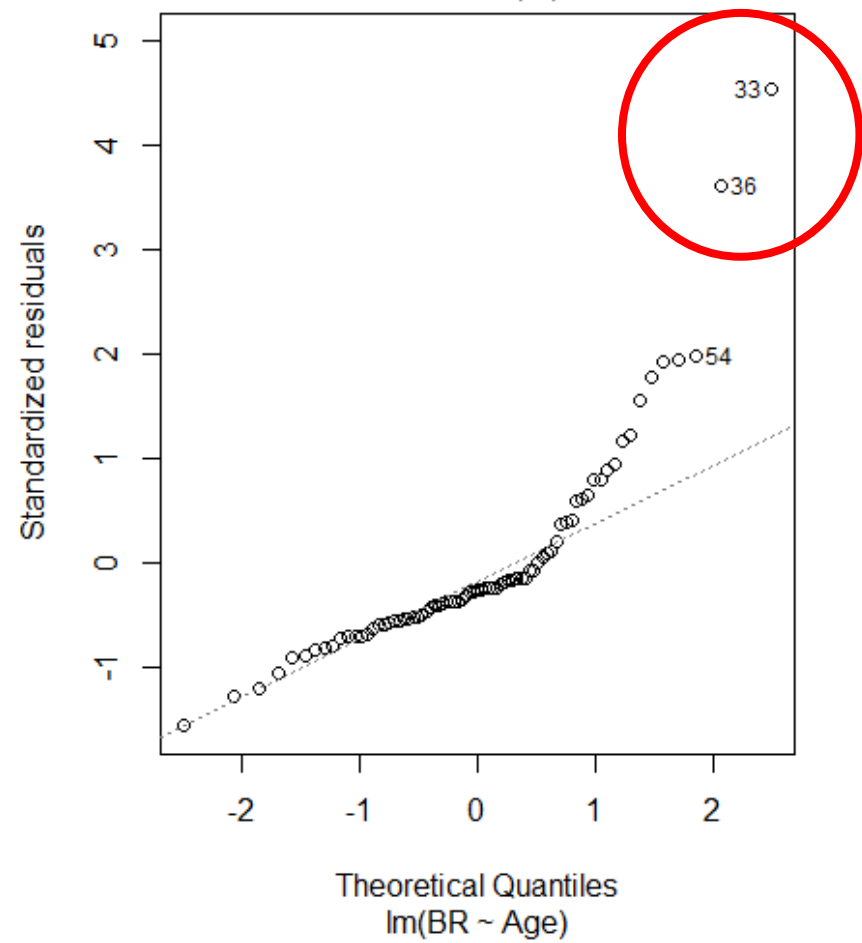
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.08740	1.08768	-0.080	0.936
Age	0.54361	0.09568	5.681	2.41e-07 ***

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

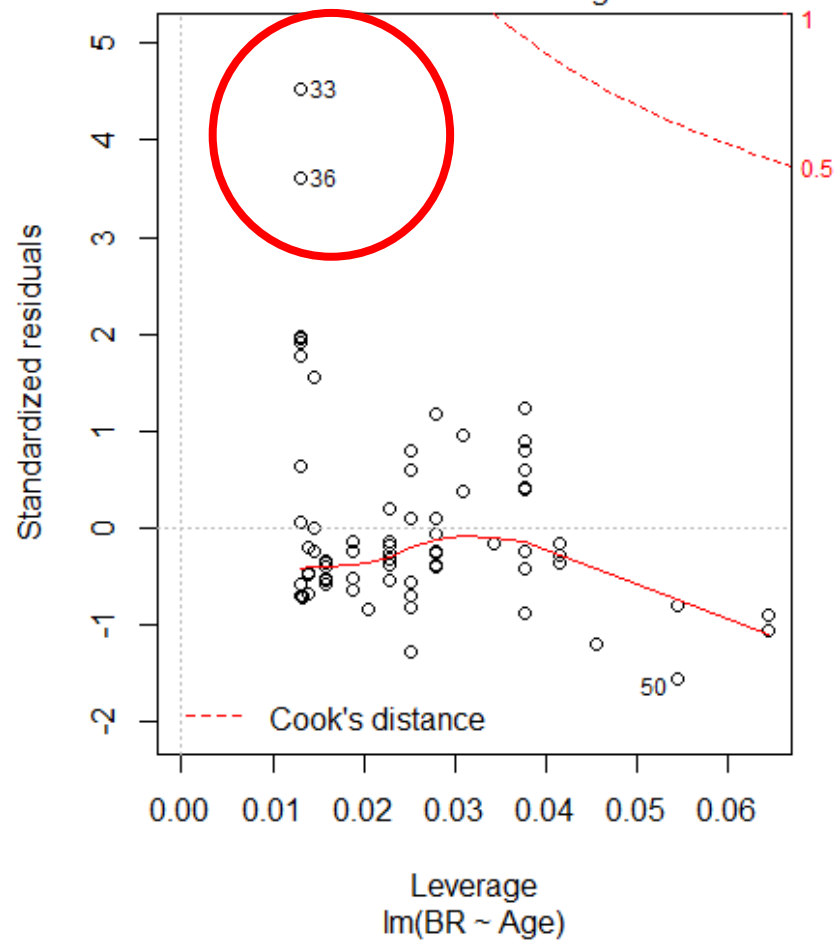
Residual standard error: 4.11 on 75 degrees of freedom
Multiple R-squared: 0.3009, Adjusted R-squared: 0.2916
F-statistic: 32.28 on 1 and 75 DF, p-value: 2.411e-07



Normal Q-Q



Residuals vs Leverage



```
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
 (`wt_residuals`)

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	0.08768	-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:

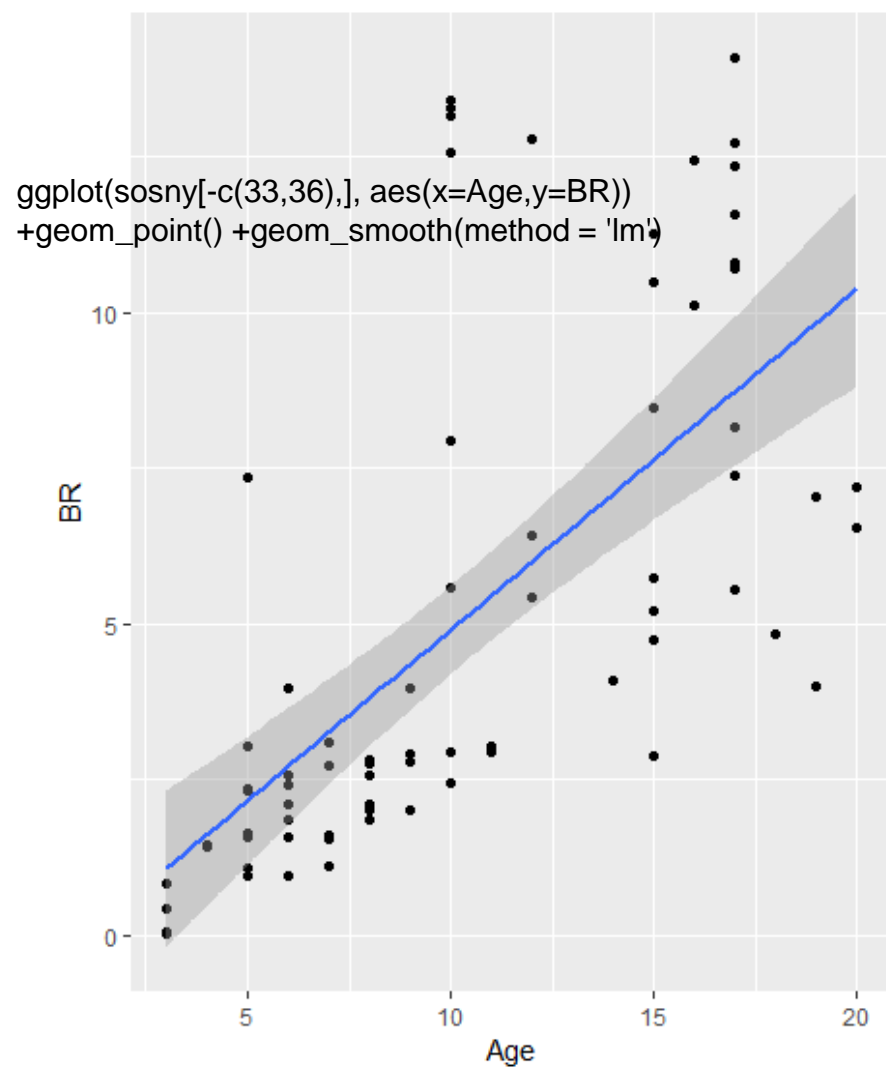
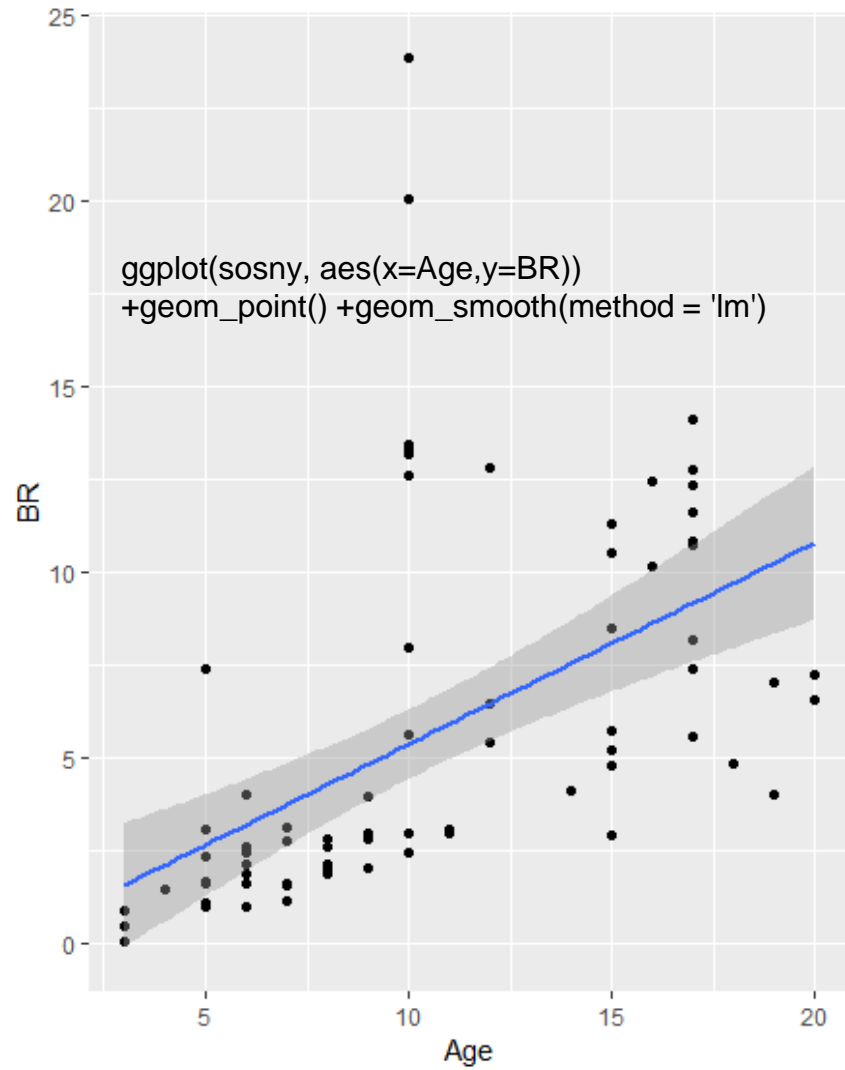
Min	1Q	Median	3Q	Max
-5.8548	-1.7888	-0.6365	0.8509	8.5087

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



Modele liniowe z efektami losowymi i stałymi (LMM)

a.k.a. modele mieszane, modele z efektem losowym, powtarzalne pomiary,

linear mixed-effect models/ generalized linear mixed-effects model,

Po co nam modele?

- przybliżenie procesu
- estymacja – szacowanie wartości oczekiwanej
- uwzględnianie dodatkowych czynników – co by było gdyby?
- zgodność z założeniami – niezależność obserwacji

efekt stały a efekt losowy

stały - związany z działaniem czynnika

losowy - związany z elementami które powinny być niezależne, a mogą mieć wpływ, np. wariant doświadczenia, powtórzenie, termin, itp.

Przykład

3 terminy badań na **6** blokach po **10** poletek **4** gatunków drzew - światło i odczyn

światło, odczyn, gatunek drzewa - efekty stałe

termin badań, poletko, blok - efekty losowe

Po co efekty losowe

niezależność obserwacji - możliwość wnioskowania i ekstrapolacji

zależność czasowa, przestrzenna, filogenetyczna, osobnicza...

uwypuklenie trendu z uwzględnieniem kontekstu

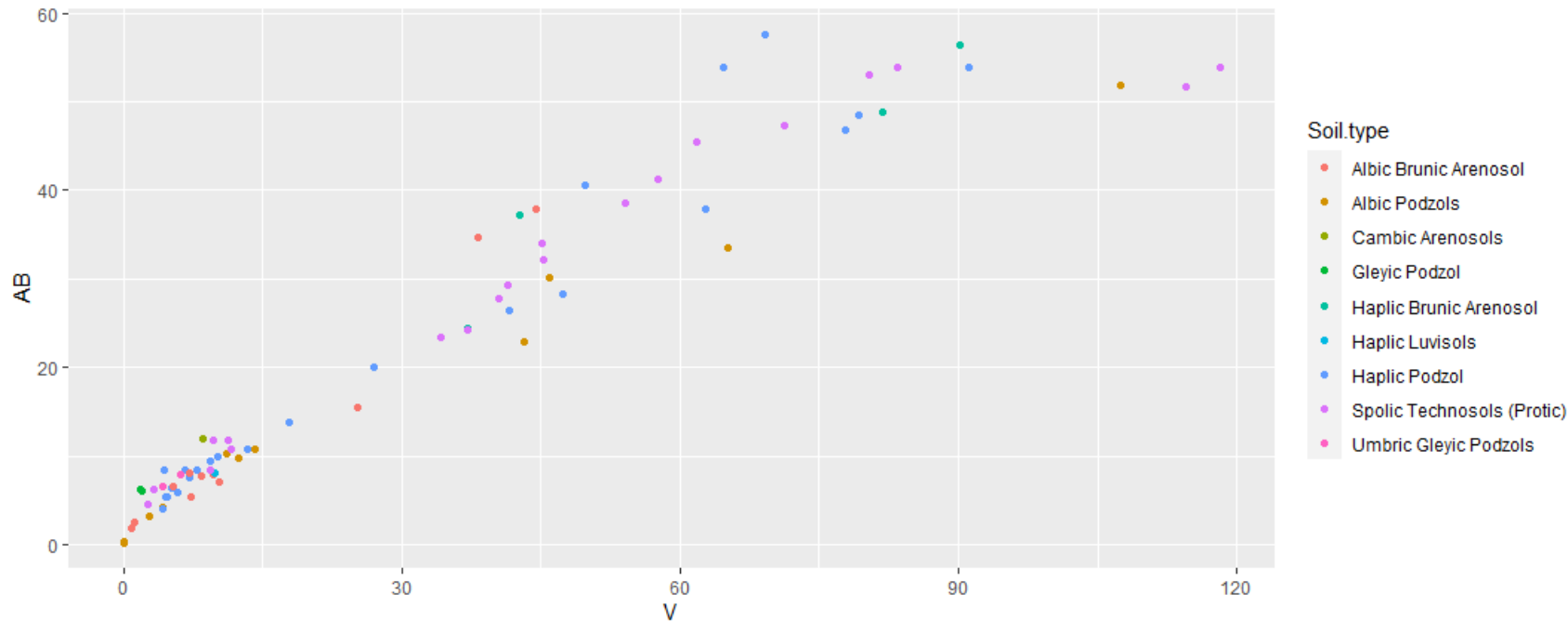
ocena zmienności na różnych poziomach hierarchii

“a co jeśli jakieś poletko jest odwiedzane przez psy i koty - to też może mieć wpływ na zbiorowiska mszaków epifitycznych przez dopływ biogenów i wydrapywanie?” - plot-specific effects

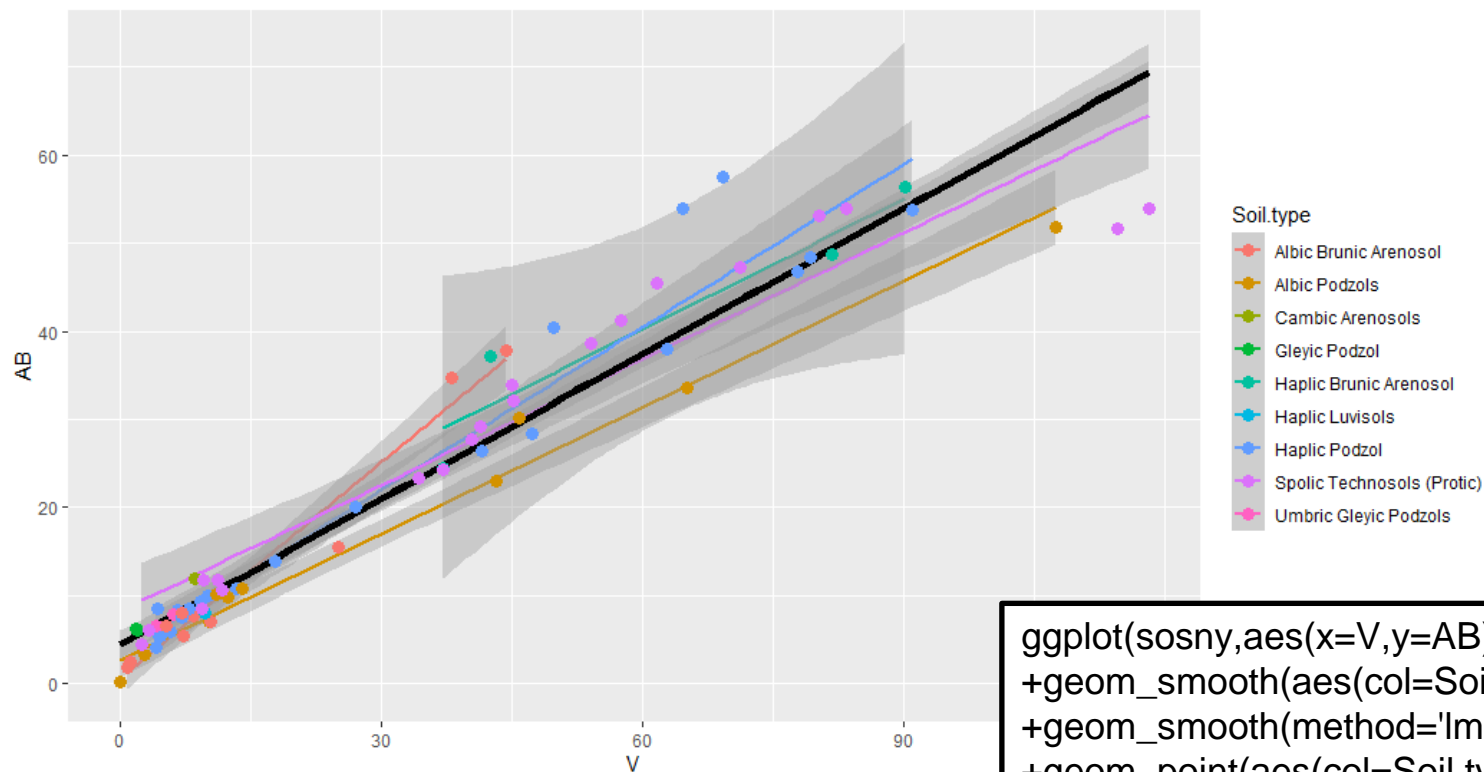
Przykłady

- przestrzenna: po 4 poletka na 10 transektach – nr transektu jako RE
- czasowa: 20 niezależnych poletek w 3 terminach – termin jako RE
- filogenetyczna: 100 gatunków z 40 rodzin – gatunek:rodzina jako RE
- osobnicza: kleszcze z 20 saren, pięć części ciała na sarnę – ID sarny jako RE

Sosny – różne wzorce dla różnych typów gleb



Jak wygląda linia trendu dla całości i dla typów?



```
ggplot(sosny,aes(x=V,y=AB))  
+geom_smooth(aes(col=Soil.type),method='lm')  
+geom_smooth(method='lm',lwd=2,col='black')  
+geom_point(aes(col=Soil.type),size=3)
```

model<-lm(AB~V, sosny)

- Model liniowy
- $AB = a \cdot V + b$

summary(model)

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

mod.mix<-lmer(AB~V+(1|Soil.type), sosny)

- Model liniowy z efektem losowym i stałym
- $AB = a \cdot V + (b + u_1)$, gdzie u_1 dla każdego poziomu Soil.type

summary(mod.mix)

Linear mixed model fit by REML. t-tests use Satterthwaite's method

[lmerModLmerTest]

Formula: AB ~ V + (1 | Soil.type)

Data: sosny

REML criterion at convergence: 466

Scaled residuals:

	Min	1Q	Median	3Q	Max
	-3.2871	-0.3629	-0.1629	0.2790	3.0615

Random effects:

Groups	Name	Variance	Std.Dev.
Soil.type	(Intercept)	1.755	1.325
	Residual	22.777	4.772

Number of obs: 77, groups: Soil.type, 9

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(> t)
(Intercept)	4.47608	0.94963	11.51329	4.713	0.000562 ***
V	0.54579	0.01812	73.36032	30.120	< 2e-16 ***

Prosty przykład na początek

```
hotspots<-read.csv('datasety/hotspots.csv',sep=';')  
mod<-lm(plants~mammals,data=hotspots)  
summary(mod)
```

Call:

```
lm(formula = plants ~ mammals, data = hotspots)
```

Residuals:

Min	1Q	Median	3Q	Max
-631.69	-192.69	-82.69	133.72	1294.67

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	48.746	28.716	1.698	0.0903.
mammals	31.824	3.114	10.219	<2e-16 ***

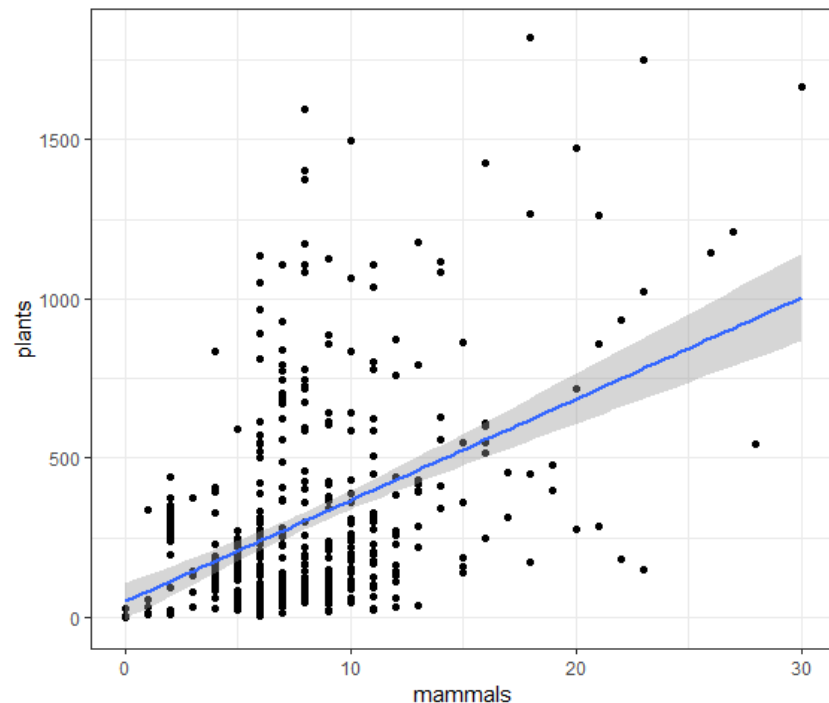
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 294.8 on 447 degrees of freedom

(160 observations deleted due to missingness)

Multiple R-squared: 0.1894, Adjusted R-squared: 0.1876

F-statistic: 104.4 on 1 and 447 DF, p-value: < 2.2e-16



```
ggplot(hotspots, aes(x=mammals,y=plants))+geom_point()  
+theme_bw()+geom_smooth(method='lm')
```

```
mod<-lm(plants~mammals+mainl,data=hotspots)
```

Call:
lm(formula = plants ~ mammals + mainl, data = hotspots)

Residuals:

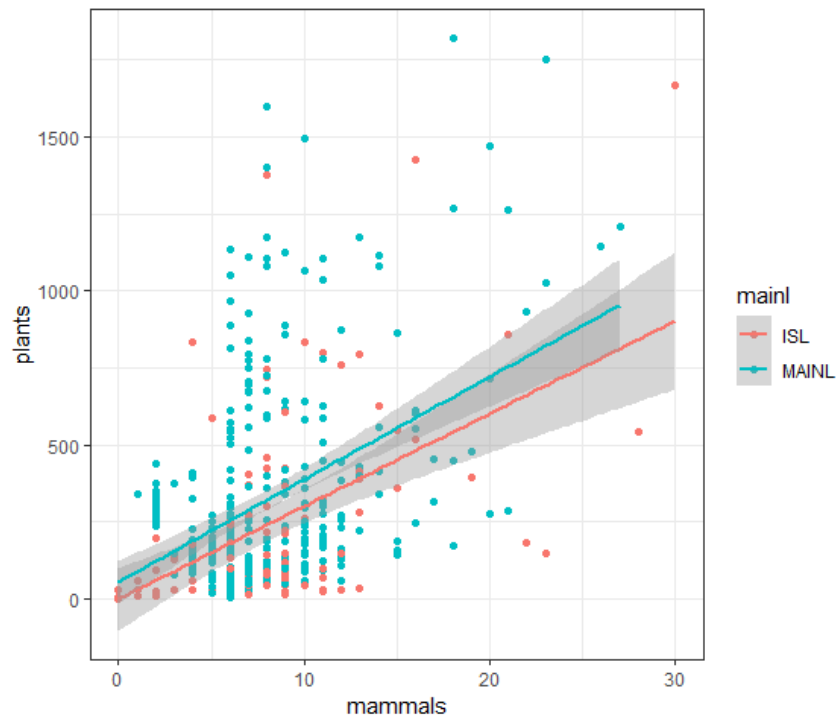
Min	1Q	Median	3Q	Max
-574.32	-203.29	-80.29	130.46	1275.97

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-18.392	39.241	-0.469	0.640
mammals	32.249	3.101	10.400	<2e-16 ***
mainlMAINL	82.437	33.056	2.494	0.013 *

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

Residual standard error: 293 on 446 degrees of freedom
(160 observations deleted due to missingness)
Multiple R-squared: 0.2005, Adjusted R-squared: 0.1969
F-statistic: 55.93 on 2 and 446 DF, p-value: < 2.2e-16



```
ggplot(hotspots, aes(x=mammals,y=plants,col=mainl))  
+geom_point()+theme_bw()+geom_smooth(method='lm')
```

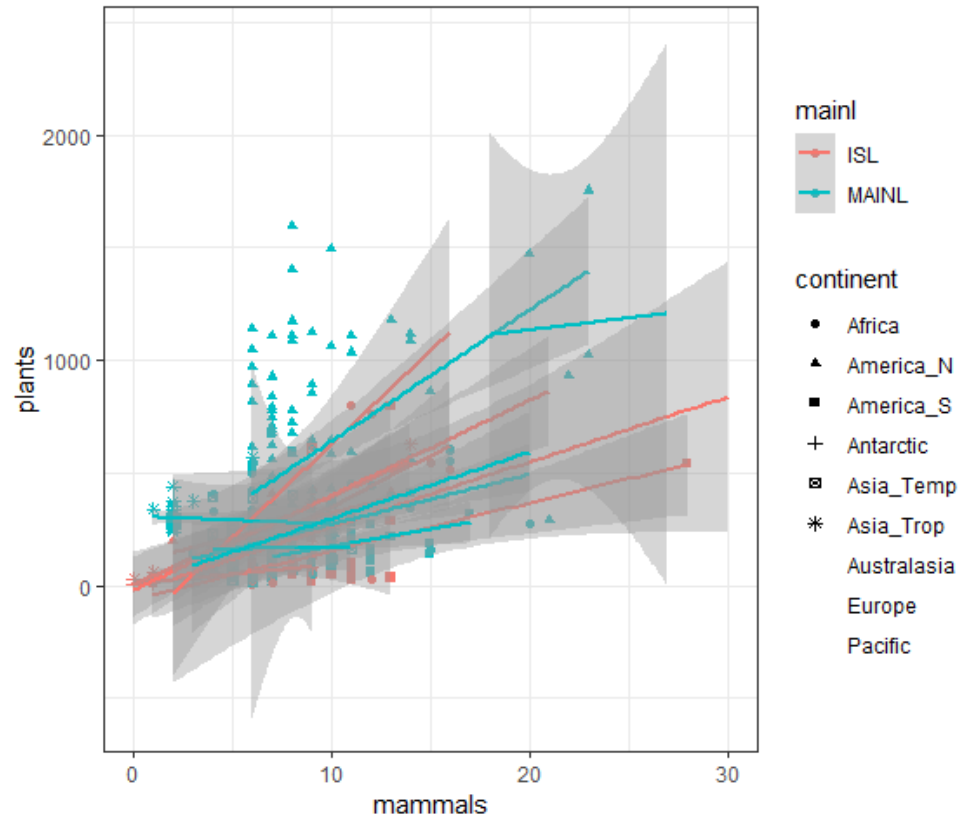
czy dołożyć kontynent?

Czy robi nam różnice?

O co nam chodzi?

różne nachylenia i różne położenia

slope i intercept zależny od continent



```
ggplot(hotspots, aes(x=mammals,y=plants,col=mainl,shape=continent))  
+geom_point()+theme_bw()+geom_smooth(method='lm')
```


Dodajemy losowy intercept

normalnie $\text{plants} = a * \text{mammals} + b$
teraz
 $\text{plants} = a * \text{mammals} + b + u_2$
 u_2 - losowy intercept

```
mod2<-lmer(plants~mammals+(1|continent),hotspots)
```

```
summary(mod2)
```

Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']

Formula: $\text{plants} \sim \text{mammals} + (1 \mid \text{continent})$

Data: hotspots

REML criterion at convergence: 6242.7

Scaled residuals:

Min	1Q	Median	3Q	Max
-2.8175	-0.5247	-0.0337	0.3143	4.3142

Random effects:

Groups	Name	Variance	Std.Dev.
continent	(Intercept)	21998	148.3
Residual		62794	250.6

Number of obs: 449, groups: continent, 9

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(> t)
(Intercept)	30.684	57.399	12.012	0.535	0.603
mammals	36.700	3.055	441.698	12.011	<2e-16 ***

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

Correlation of Fixed Effects:

(Intr)
mammals -0.421

O co chodzi z zapisem z kreską?

mammals|continent oznacza że mammals ma losowy slope i intercept dla każdego continent dla mammals

normalnie $\text{plants} = a * \text{mammals} + b$

teraz

$\text{plants} = (a + u_1) * \text{mammals} + b + u_2$

u_1 - losowy slope, u_2 - losowy intercept

```
mod3<-lmer(plants~mammals+(mammals|continent),hotspots)
```

```
summary(mod3)
```

Linear mixed model fit by REML. t-tests use Satterthwaite's method

['lmerModLmerTest']

Formula: plants ~ mammals + (mammals | continent)

Data: hotspots

REML criterion at convergence: 6220.9

Scaled residuals:

Min	1Q	Median	3Q	Max
-3.9728	-0.5086	-0.0979	0.3215	4.4747

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
continent	(Intercept)	7088.0	84.19	
	mammals	357.5	18.91	-0.41
Residual		59102.2	243.11	

Number of obs: 449, groups: continent, 9

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(> t)
(Intercept)	42.905	40.798	8.532	1.052	0.3218
mammals	33.215	7.429	5.960	4.471	0.0043 **

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

Correlation of Fixed Effects:

(Intr)
mammals -0.549

Jak wyciągnąć wartości

```
coef(mod2)
```

```
$continent
```

```
(Intercept) mammals
```

```
Africa      -52.99065 36.69969
America_N    223.33035 36.69969
America_S   -179.55621 36.69969
Antarctic    -58.33926 36.69969
Asia_Temp    -34.16470 36.69969
Asia_Trop    161.22696 36.69969
Australasia  159.63777 36.69969
Europe       -80.81611 36.69969
Pacific      137.82662 36.69969
```

```
attr("class")
```

```
[1] "coef.mer"
```

```
coef(mod3)
```

```
$continent
```

```
(Intercept) mammals
```

```
Africa      25.878318 26.99799
America_N    52.495373 57.20669
America_S     4.901014 16.86633
Antarctic    19.770299 22.02693
Asia_Temp    58.166601 19.92525
Asia_Trop    186.721396 22.31839
Australasia  17.782051 48.76824
Europe       24.924300 27.16040
Pacific      -4.492198 57.66827
```

```
attr("class")
```

```
[1] "coef.mer"
```

Co otrzymujemy?

Model bez efektów losowych - część uogólnioną

Efekty losowe - czynniki warunkujące przebieg krzywych dla kontynentów

Informacje o zmienności tych efektów w ramach modelu:

SD i variance

Który lepszy?

```
> AIC(mod, mod2, mod3)
      df    AIC
mod    4 6380.127
mod2   4 6250.657
mod3   6 6232.945
```

Ile procent wyjaśnia?

```
> library(MuMIn)  
> r.squaredGLMM(mod2)
```

```
      R2m      R2c
```

```
[1,] 0.2410538 0.4379521
```

```
> r.squaredGLMM(mod3)
```

```
      R2m      R2c
```

```
[1,] 0.2039087 0.4537038
```

R2m - marginal R2; R2c - conditional R2

R2m - % zmienności wyjaśnionej przez fixed effects

R2c - % zmienności wyjaśnionej przez fixed + random effects

random effects - $R2c - R2m$

LMM w układzie hierarchicznym

4 obiekty, po 50 poletek w każdym

`lmer(biomasa~land.use.history+invasion+(1|variant:plot), dane)`

plot zagnieżdżony w variant - czyli efekty losowe na poziomie plot i variant
zakładamy że dla każdego mamy random intercept

SD tych modyfikacji interceptów wynosi 0,0837

R2c i R2m - z `r.squaredGLMM()`



Table 2

Differences in dry biomass and biomass allocation of *V. myrtillus* shoots (ramets) between types of pine forest assessed using linear mixed effect models.

Dry shoot mass (n = 400)	Random effects	Variance	SD	Mixed model parameters	–
–	Plots nested in variant	0.0070	0.0837	R _m ²	0.1159
–	Residuals	0.1557	0.3946	R _c ²	0.1540
–	Fixed effects	Estimate	SE	t	Pr(> t)
–	(Intercept)	0.745	0.041	18.110	< 0.001
–	Land use history – recent forest	–0.243	0.048	–5.110	< 0.001
–	Invasion – <i>Q. rubra</i>	–0.162	0.048	–3.401	0.002

więcej o modelach mieszanych

<https://cran.r-project.org/web/packages/lme4/vignettes/lmer.pdf>

<https://www.r-bloggers.com/linear-mixed-models-in-r/>

<https://www.r-bloggers.com/getting-started-with-mixed-effect-models-in-r/>

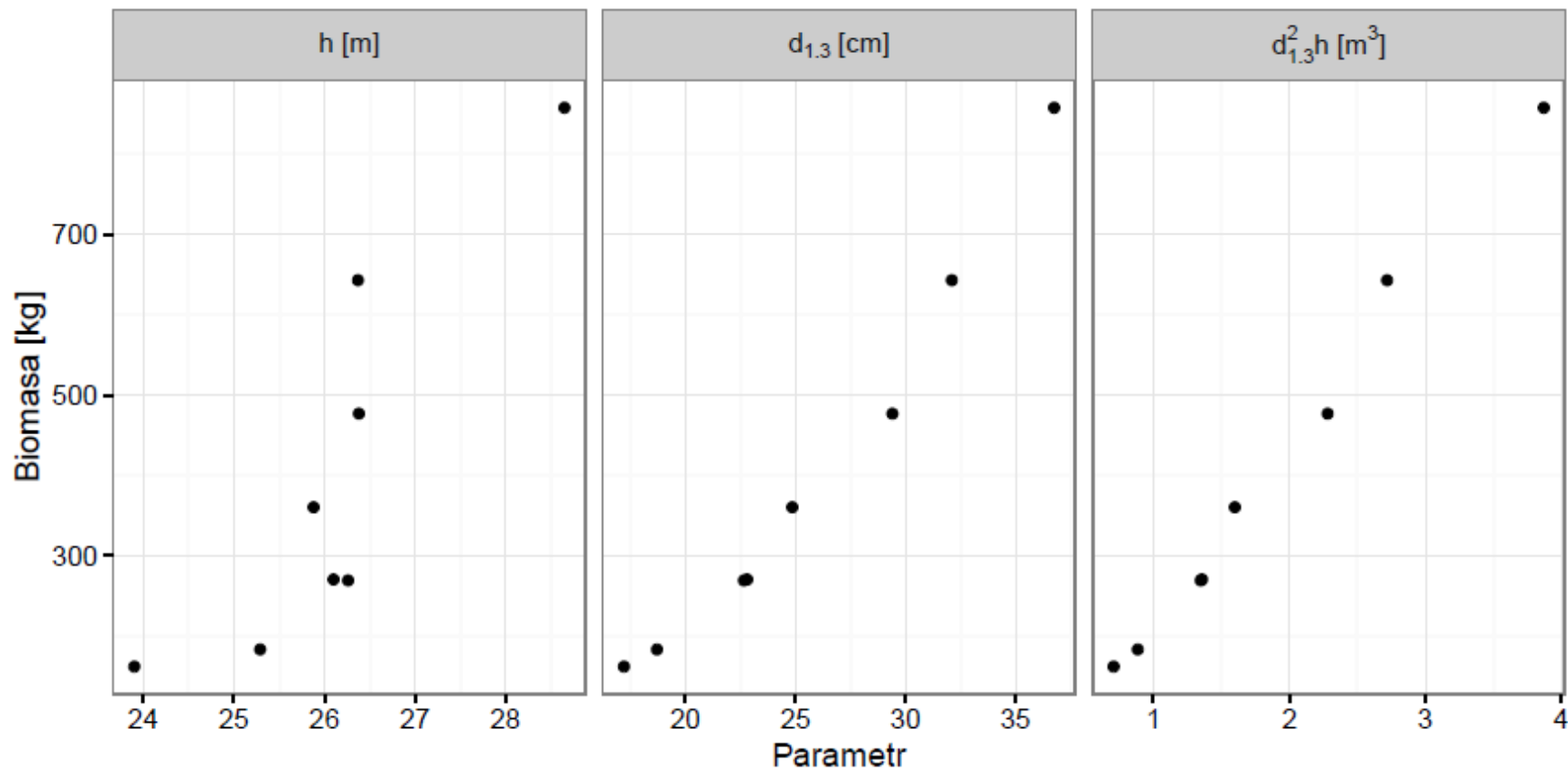
<http://www.biecek.pl/WZUR/PrzemekBiecek2009.pdf>

<https://libra.ibuk.pl/book/39524> - podręcznik P. Biecka

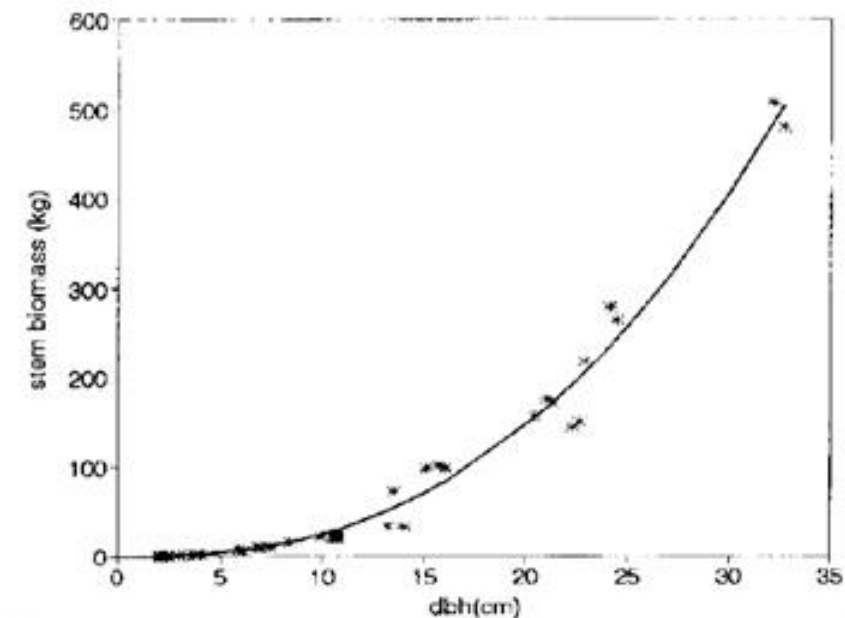
Modele nieliniowe

- Zależności krzywoliniowe
- Znane wcześniej z literatury, np. modele logistyczne, modele potęgowe
- Relacje allometryczne – różne wymiary zmiennych
- Sens biologiczny – np. niezerowy wynik

Modele biomasowe



Równania allometryczne



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

39

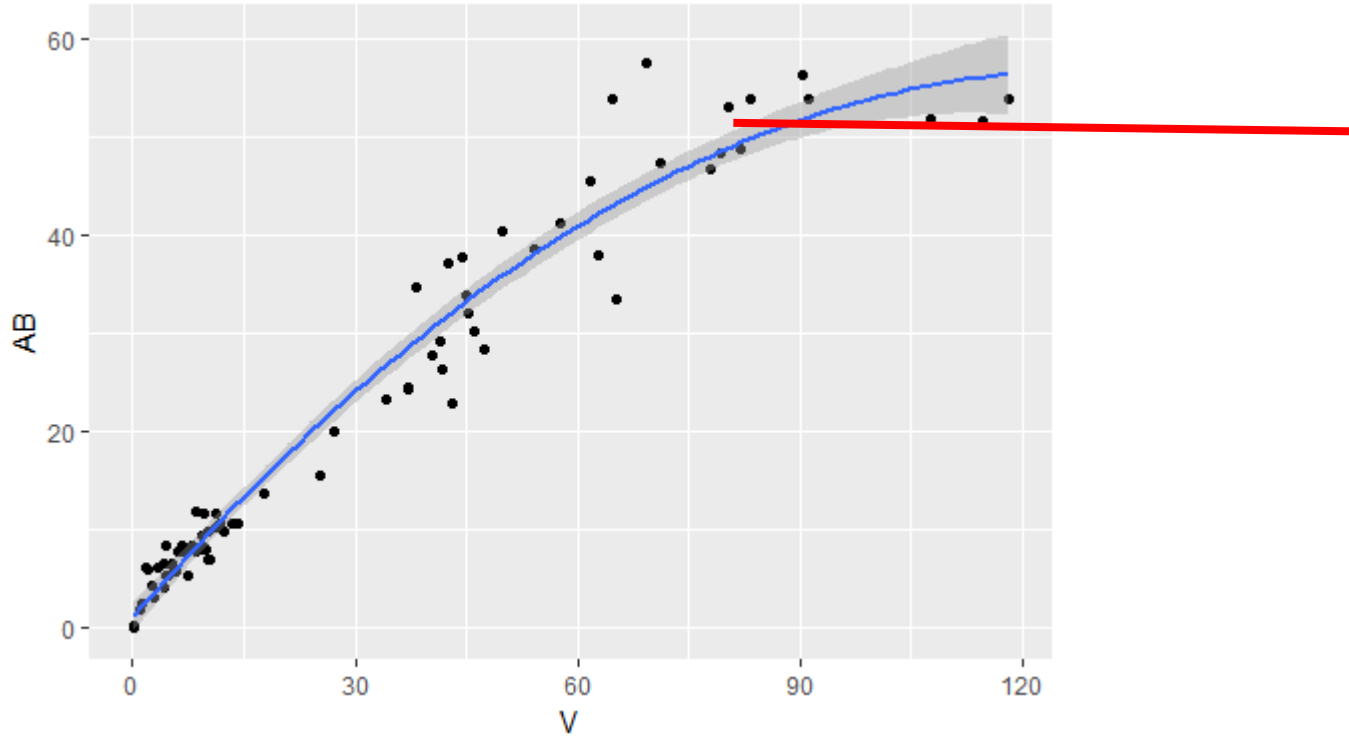
Original article

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

HH Bartelink



Parabola $\text{poly}(x,2)$ $f(x)=ax^2+bx+c$



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

```
> 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
```

for all biomass components and volume. For each equation we calculated ten regression models:

$$W = a \times D^b$$

$$W = a + b \times D^2$$

$$W = a + b \times \log(D)$$

$$W = a + (b/D)$$

$$W = a \times (D^2 H)^b$$

$$W = a \times D^b \times H^c$$

$$W = a + b \times \log(D^2 H)$$

$$W = a + b \times D^2 + c \times H$$

$$W = a + b \times (D^2 H)$$

$$W = a + b \times D^2 + c \times H^2$$



Contents lists available at ScienceDirect

Forest Ecology and Management

journal homepage: www.elsevier.com/locate/foreco



How do tree stand parameters affect young Scots pine biomass? – Allometric equations and biomass conversion and expansion factors



Andrzej M. Jagodziński^{a,b,*}, Marcin K. Dyderski^{a,b}, Kamil Gęsikiewicz^a, Paweł Horodecki^a, Agnieszka Cysewska^b, Sylwia Wierczyńska^b, Karol Maciejczyk^b

^a Institute of Dendrology, Polish Academy of Sciences, Parkowa 5, 62-035 Kórnik, Poland

^b Poznań University of Life Sciences, Faculty of Forestry, Department of Game Management and Forest Protection, Wojska Polskiego 71c, 60-625 Poznań, Poland

(/)

(8)

(9)

(10)

implementacja

liniowy - `lm()` nieliniowy - `nls()`

`model.liniowy<-lm(masa~D,data=dane)`

`model.liniowy<-lm(masa~log(D), data=dane)`

`modelnieliniowy<-nls(masa~a*D^b, data=dane, start=list(a=1,b=-1))`

formuła, dane, (start)

model nieliniowy

```
nls(AB~a*V^b,data=sosny,start=list(a=1,b=1))
```

Nonlinear regression model

model: $AB \sim a * V^b$

data: sosny

a b

2.0119 0.7233

residual sum-of-squares: 1297

Number of iterations to convergence: 6

Achieved **convergence** tolerance: 1.784e-06

czasem model nie osiąga konwergencji -
można kazać R zmienić maksymalną
liczbę iteracji w nls:

```
nls.control(maxiter=500)
```

#domyślne 50 może nie dać rady
#uwaga - wydłuża czas obliczeń!

inna opcja to robust regression

```
library(robustbase)  
nlrob(AB~a*V^b,data=sosny,start=list(a=  
1,b=1))
```

pomaga przy heteroskedastyczności

```
nls(AB~a*V^b,data=sosny,start=list(a=1,b=exp(99999999)))
```

Error in numericDeriv(form[[3L]], names(ind), env) :

Brakuje wartości lub wyprodukowano wartości nieskończone podczas wyliczania modelu

Parametry startowe

literatura

brute force - dopasowanie metodą prób i błędów

dobra rada - zacząć od 1, -1, potem małe i duże cyfry

*przy funkcjach potęgowych - wykorzystać postać linearyzowaną

$$y=a*x^b$$

$$\log(y)=\log(a)+b*\log(x) \quad \text{czyli} \quad \text{lm}(\log(y) \sim \log(x), \text{data=dane})$$

summary(model.nieliniowy)

```
liczania modelu
> summary(nls(AB~a*v^b,data=sosny,start=list(a=1,b=-2)))

Formula: AB ~ a * v^b

Parameters:
  Estimate Std. Error t value Pr(>|t|)
a  2.01191    0.24523   8.204 4.88e-12 ***
b  0.72327    0.02917  24.793  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.159 on 75 degrees of freedom

Number of iterations to convergence: 11
Achieved convergence tolerance: 1.965e-06

>
```

policzę AIC i porównam z modelem zerowym

```
> 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
```

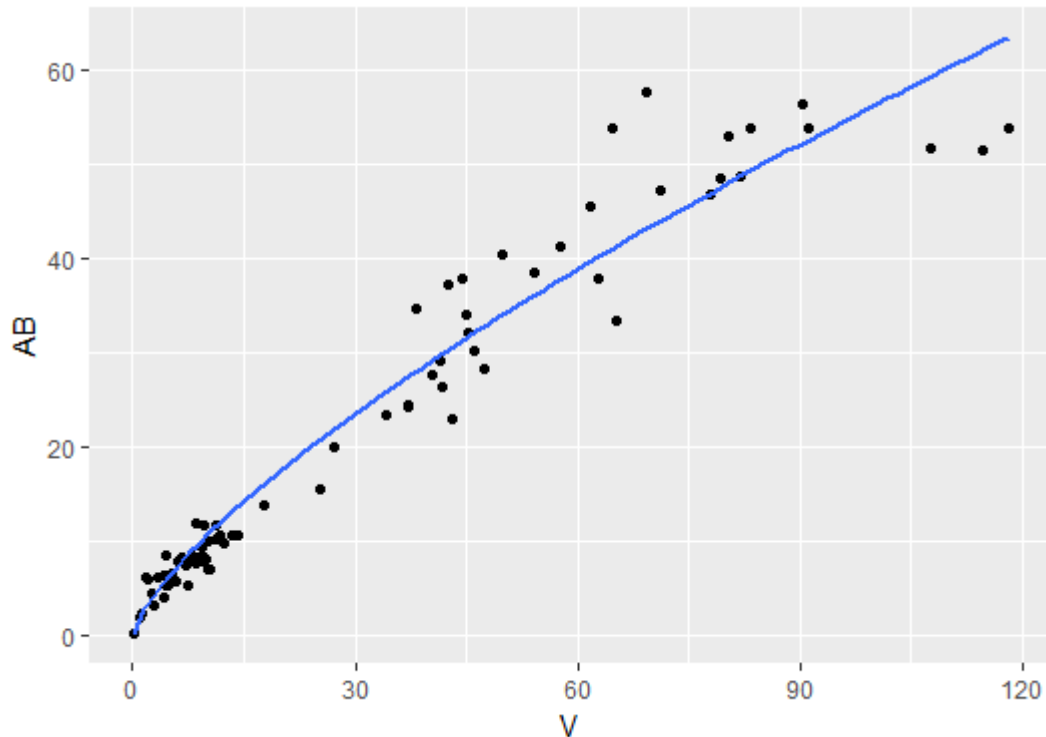
```
[1] 422.4431
```

```
> AIC(nls(AB~a*V^b,data=sosny,start=list(a=1,b=-2)))
```

```
[1] 441.9817
```

diagnostyka modelu:

```
ggplot(sosny, aes(y=AB,x=V))+geom_point()  
+geom_smooth(method='nls',formula=y~a*x^b,  
method.args=list(start=list(a=1,b=1)),se=F)
```



se=F to coś bez czego nie ruszy
zawsze zwróci error!
parametry startowe - jak
wcześniej;)

dlaczego tak?
za dużo punktów małych!
więcej nie znaczy lepiej;)

Czy zawsze warto?

- Słabe wsparcie modeli mieszanych
- Problemy z konwergencją
- Dla zmiennych o dziwnych rozkładach lub znanych nieliniowych przebiegach:
 - transformacje zmiennych, np. $\text{lm}(\log(\text{zmienna}) \sim \log(\text{zmienna}^2), \text{dane})$
 - zastosowanie modeli o innych rozkładach (GLM) - jutro
 - zastosowanie modeli addytywnych (GAM) – jutro

Podsumowanie

obrazki, obrazki, obrazki!

-pomogą dobrać narzędzie i typ rozkładu

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

effect size - o ile rośnie nam Y na jednostkę X?

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

nie $p < 0.05$, tylko dokładna wartość – podawać całe summary!

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

Schemat pracy z modelami

- Czy może być model liniowy? założenia
- Model globalny – wszystkie zmienne które mają **sens biologiczny**
- Sprawdzamy VIFy – czy nie są współliniowe *car::vif()*
- Selekcja modeli w oparciu o AIC/AICc *stats::step()* lub *MuMIn::dredge()*
- Model finalny – *summary()* do tabeli + ocena effect size (wykresy)
- Czy nas to satysfakcjonuje?



ORIGINAL PAPER

Impacts of invasive trees on alpha and beta diver of temperate forest understories

Marcin K. Dyderski · Andrzej M. Jagodziński

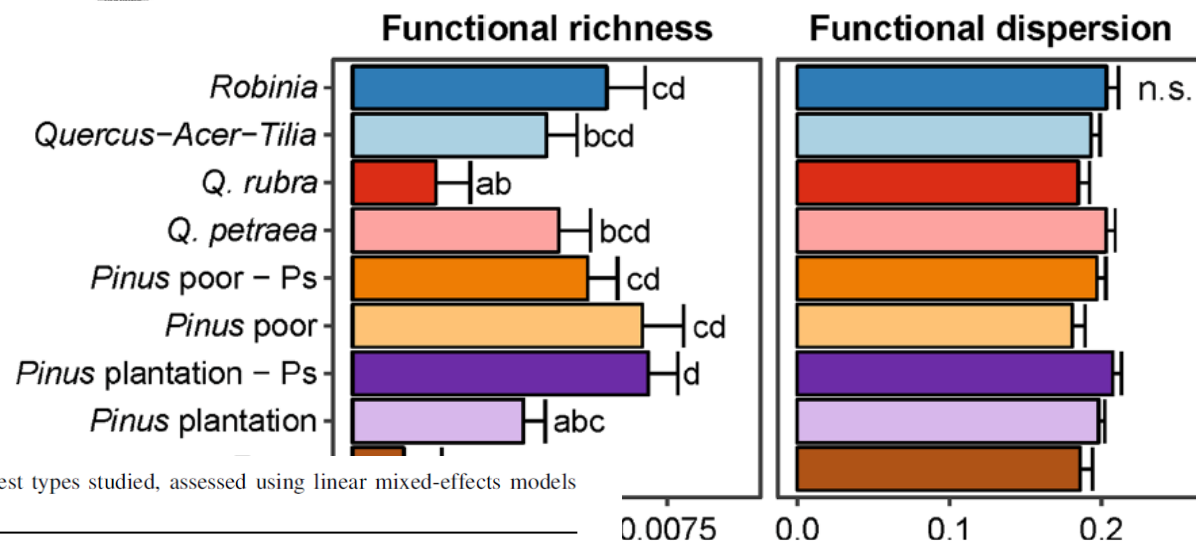


Table 4 Analysis of variance for alpha diversity indices among forest types studied, assessed using linear mixed-effects models (with study plot block as a random intercept)

Response	Variable	df	Sum of Squares	Mean Square	F	Pr(> F)	Block random effects SD
Species richness	Forest type	8	160.3678	20.0460	20.046	< 0.0001	0.265
Shannon diversity index	Forest type	8	8.2356	1.0294	5.720	< 0.0001	0.146
Faith's phylogenetic diversity	Forest type	8	17,473,654.8014	2,184,206.8502	10.821	< 0.0001	286.098
Mean pairwise distance	Forest type	8	184,069.1143	23,008.6393	3.327	0.0014	62.986
Functional dispersion	Forest type	8	0.0086	0.0011	1.978	0.0525	0.008
Functional richness	Forest type	8	0.0003	0.0000	6.703	< 0.0001	0.002

RESEARCH

Open Access



Forest land use discontinuity and northern red oak *Quercus rubra* introduction change biomass allocation and life strategy of lingonberry *Vaccinium vitis-idaea*

Beata Wozniwoda^{1*} , Marcin K. Dyderski² and Andrzej M. Jagodziński²

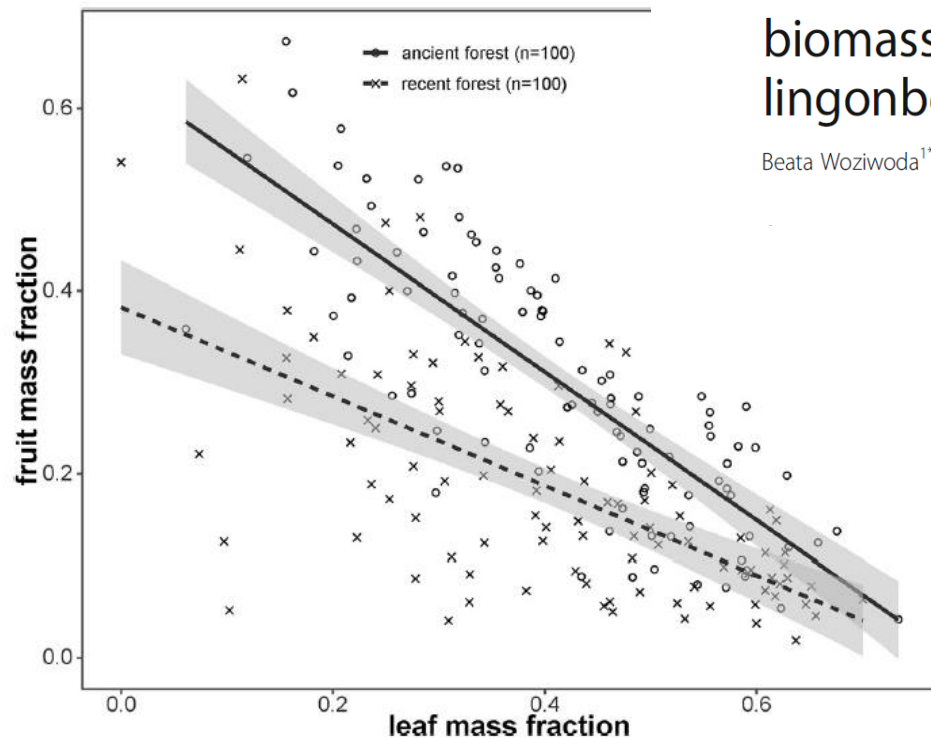


Table 5 Linear model of fruit mass fraction, as a function of leaf mass fraction and forest type (adjusted $R^2 = 0.590$)

Variable	Estimate	SE	t	Pr(> t)
(Intercept)	0.5581	0.0202	27.6620	< 0.0001
Leaf mass fraction	−0.6235	0.0433	−14.3820	< 0.0001
Land use: recent forest	−0.1203	0.0131	−9.2140	< 0.0001

Table 4 Models of shrub layer alpha diversity, estimated using Gaussian distribution GLMMs (for species richness – Poisson distribution and z statistic instead of t). Significant variables were bolded. AICc – Akaike Information Criterion, AICc₀ – AIC of null model (intercept-only), AICc_{full} – AIC of full model (including all hypoth-

esized variables), R²_m – marginal coefficient of determination, R²_c – conditional coefficient of determination; *for Shannon’s index model block random effect SD=0.43, AICc=198.4, AICc_{full}=275.0, AICc₀=199.5, R²_m=0.005, R²_c=0.577

Response	Variable	Estimate	SE	t value	Pr(> t)
Species richness	(Intercept)	0.2992	0.2806	1.0660	0.2863
Block SD=0.464	type= <i>Pinus</i> poor	1.1945	0.3506	3.4070	0.0007
R ² _m =0.185	type= <i>Pinus</i> poor—Ps	1.0788	0.2885	3.7390	0.0002
R ² _c =0.611	type= <i>Pinus</i> rich	1.1043	0.2562	4.3110	<0.0001
AICc= 735.0	type= <i>Pinus</i> rich—Ps	1.3298	0.2789	4.7690	<0.0001
AICc _{full} =749.7	type= <i>Q. petraea</i>	0.8720	0.3016	2.8910	0.0038
AICc ₀ =757.4	type= <i>Q. rubra</i>	0.5930	0.3084	1.9230	0.0545
	type= <i>Quercus-Acer-Tilia</i>	1.2184	0.2769	4.3990	<0.0001
	type= <i>Robinia</i>	1.0889	0.2878	3.7830	0.0002
	plot area	2.2364	1.1208	1.9950	0.0460
Shannon’s index*	(Intercept)	0.6744	0.1174	5.7430	<0.0001
	plot area				
Faith’s phylogenetic diversity	(Intercept)				
Block SD= 161.4	type= <i>Pinus</i> poor				
R ² _m =0.200	type= <i>Pinus</i> poor—Ps				
R ² _c =0.458	type= <i>Pinus</i> rich				
AICc= 2134.2	type= <i>Pinus</i> rich—Ps				
AICc _{full} = 2135.1	type= <i>Q. petraea</i>				
AICc ₀ =2317.8	type= <i>Q. rubra</i>	153.7750	113.7170	1.3520	0.1786
	type= <i>Quercus-Acer-Tilia</i>	471.3920	121.1010	3.8930	0.0002
	type= <i>Robinia</i>	400.9820	133.4240	3.0050	0.0031



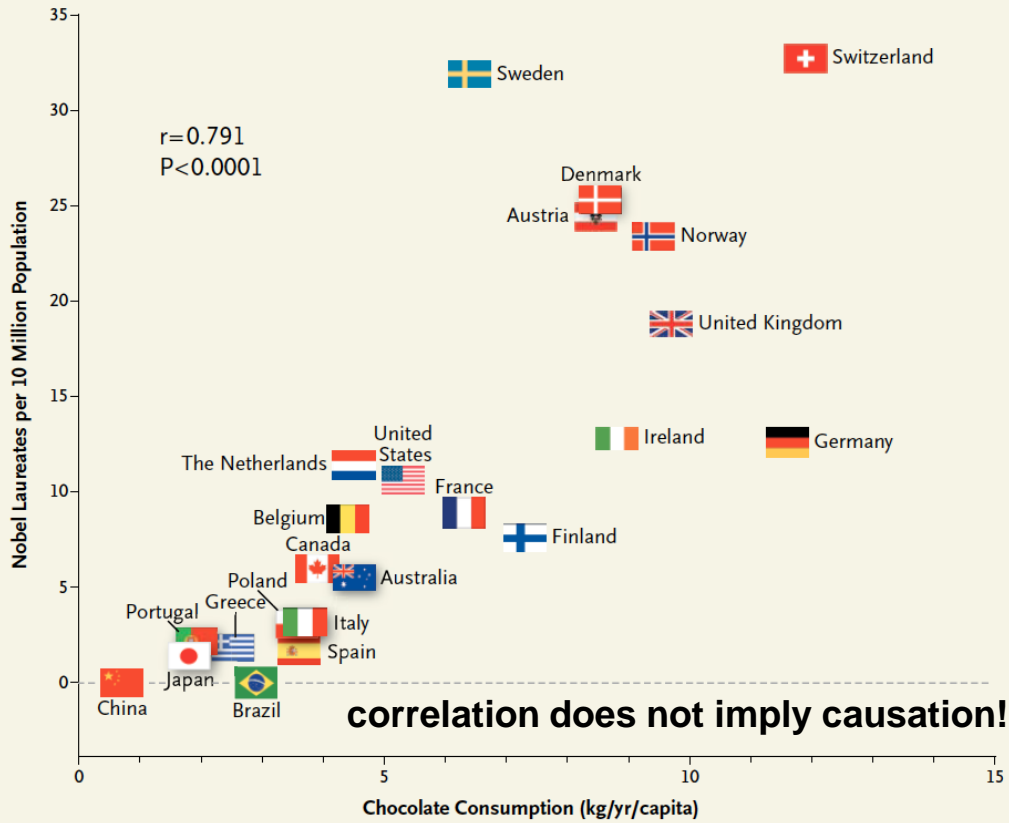


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

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

