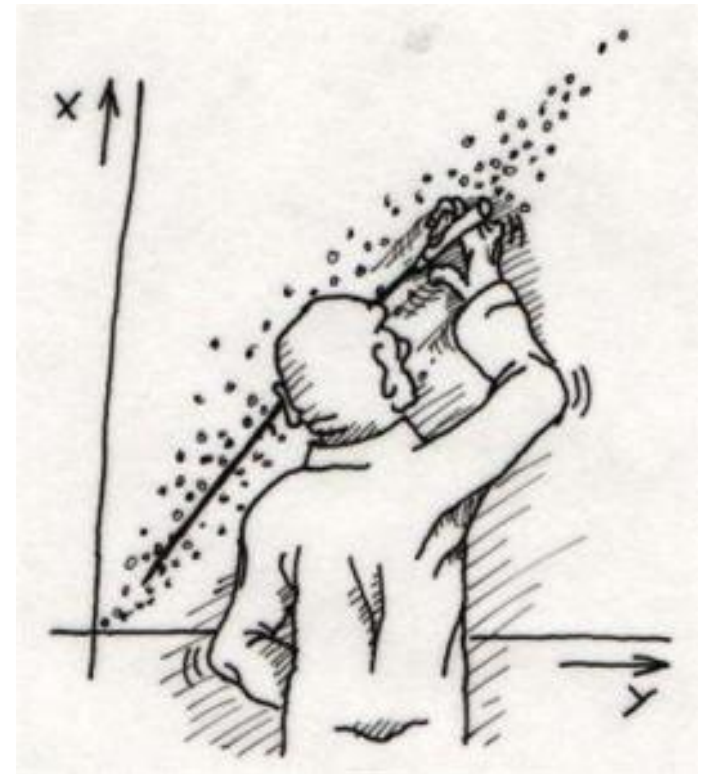




MODELE LINIOWE

KORELACJA I REGRESJA LINIOWA

BIOMATEMATYKA WYKŁAD 2
DR WIOLETA DROBIK-CZWARNO

TYPY ZALEŻNOŚCI

PRZYPOMNIENIE

- Deterministyczna (funkcyjna) – jednej wartości zmiennej odpowiada ściśle określona wartość drugiej zmiennej

$$Y = f(X)$$

- Statystyczna – konkretnej wartości jednej zmiennej może odpowiadać wiele wartości drugiej zmiennej
 - Występuje dodatkowo czynnik zakłócający zwany składnikiem losowym (ε)

$$Y = f(X) + \varepsilon$$



WERYFIKACJA HIPOTEZ STATYSTYCZNYCH

PRZYPOMNIENIE

- **Hipoteza badawcza:** leki A i B różnie wpływają na stan kliniczny pacjentów
 $EXA \neq EXB$ (na tym nam zależy)
- **Hipoteza** statystyczna (H_0): lek A i B tak samo wpływają na stan kliniczny pacjentów
 $EXA = EXB$ (łatwiejsze do zbadania)



TEST STATYSTYCZNY

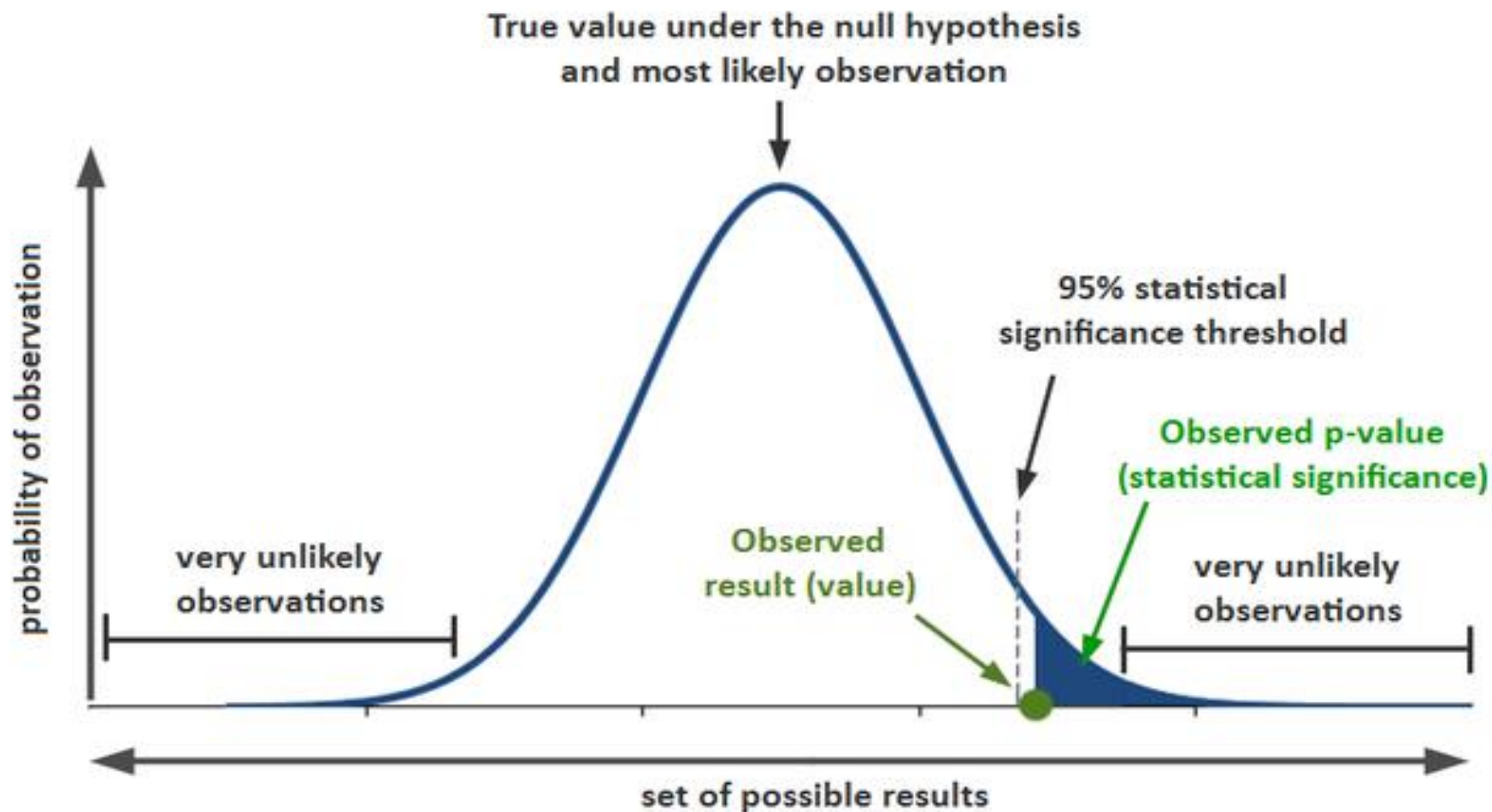
PRZYPOMNIENIE

Zbiór reguł postępowania, w wyniku których podejmujemy decyzję o przyjęciu lub odrzuceniu H_0 .

decyzja \ hipoteza	prawdziwa	fałszywa
przyjmujemy	+	błąd II rodzaju (β) moc testu ($1-\beta$) 0,8-0,9
odrzucaamy	błąd I rodzaju poziom istotności (α) 0,05; 0,01	+

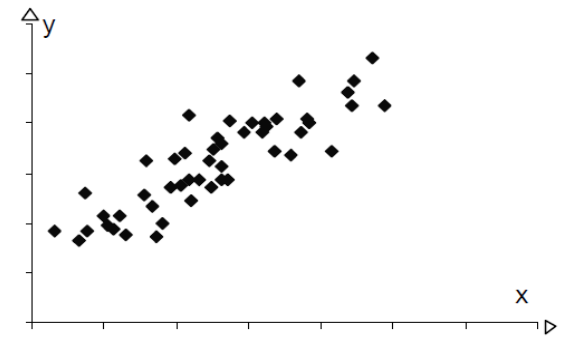
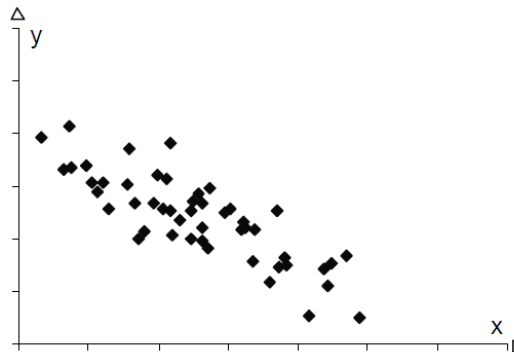
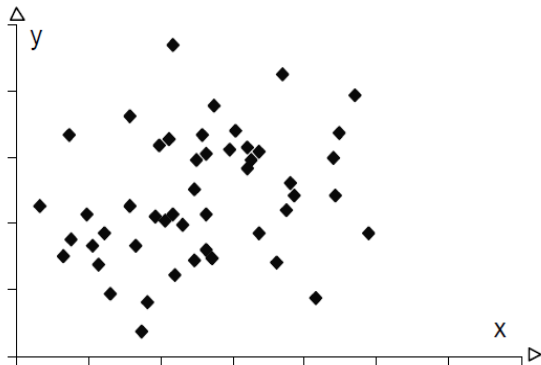


Probability & Statistical Significance Explained



KORELACJA LINIOWA

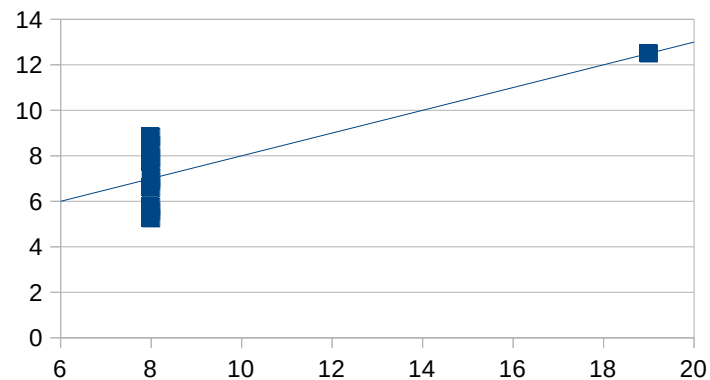
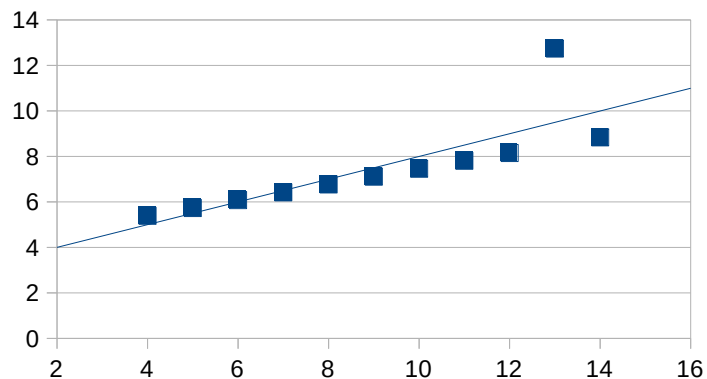
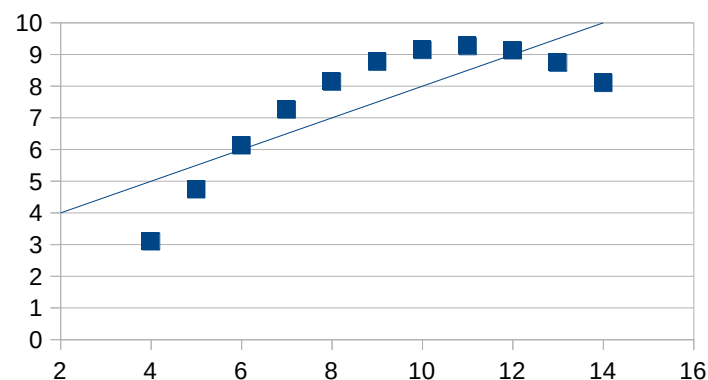
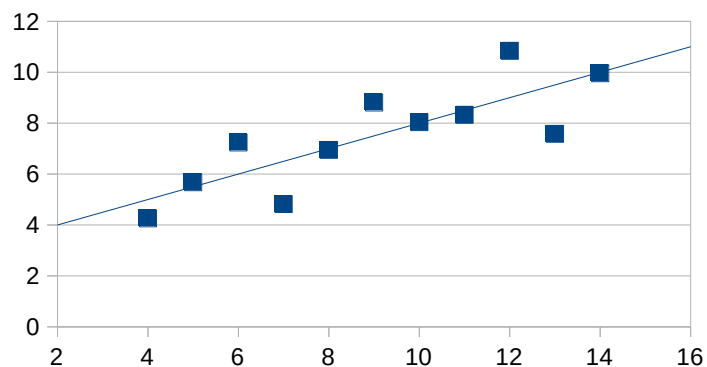
- Korelacja NIE JEST badaniem związku przyczynowo-skutkowego, ale współzmienności cech.
- **Współczynnik korelacji liniowej Pearsona**
 - określa kierunek i siłę zależności
 - przyjmuje wartości od -1 do 1.
- Jakie współczynniki korelacji dla poniższych danych?



KORELACJA LINIOWA

KWADRAT ANSCOMBE'A

Współczynnik korelacji na każdym wykresie jest 0.816



KORELACJA LINIOWA

- Jak wyznaczamy współczynnik korelacji?

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

Licznik to kowariancja (cov_{xy})

Mianownik to iloczyn odchyleń standardowych (s_x, s_y)



MODELE LINIOWE

- Jedna z najstarszych i najpopularniejszych metod modelowania
- Zależność między zbiorem **zmiennych objaśniających**, a zmienną ilościową nazywaną **zmienną objaśnianą**
- Zmienne są mierzone na skali ilościowej
- Cele modelowania:
 - Zrozumienie zależności między obserwowanymi zmiennymi: które zmienne objaśniające, i w jaki sposób, są zależne od zmiennej objaśnianej
 - Predykcja: Punktowe lub przedziałowe szacowanie wartości cechy objaśnianej na podstawie zmiennych objaśniających



METODA NAJMNIEJSZYCH KWADRATÓW



Johann Carl Friedrich Gauss (1777-1855)
- Używa metody do prac dotyczących
astronomii od roku 1795

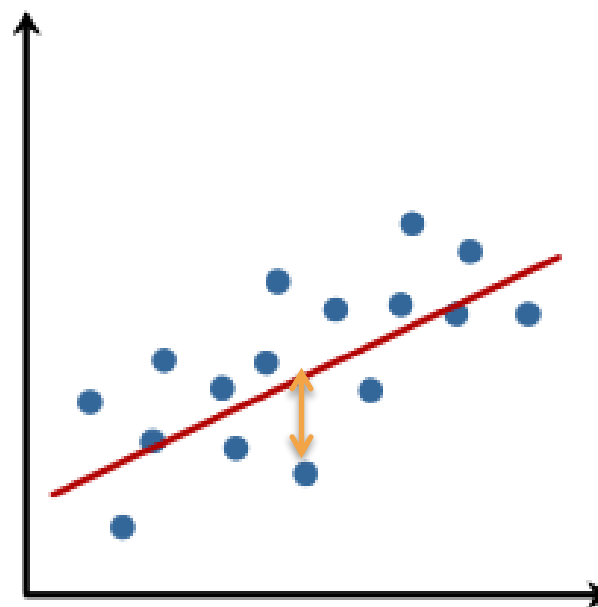
Adrien-Marie Legendre (1752 -1833)
- Pierwsza opublikowana praca w
1805 r

METODA NAJMNIEJSZYCH KWADRATÓW

- Metoda najmniejszych kwadratów – minimalizowana jest suma reszt czyli kwadratów odchyleń wartości rzeczywistych (Y) od wartości szacowanych cechy ($\hat{Y}$)

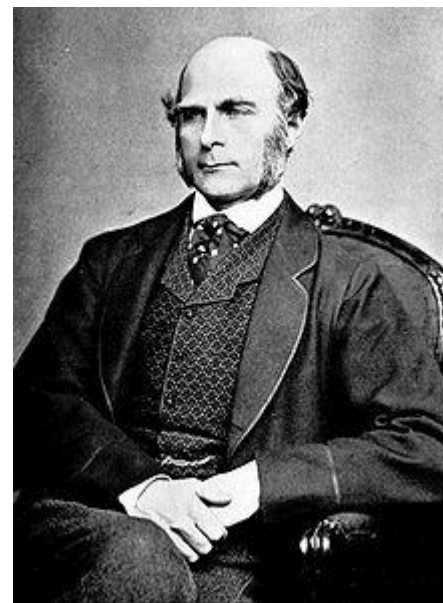
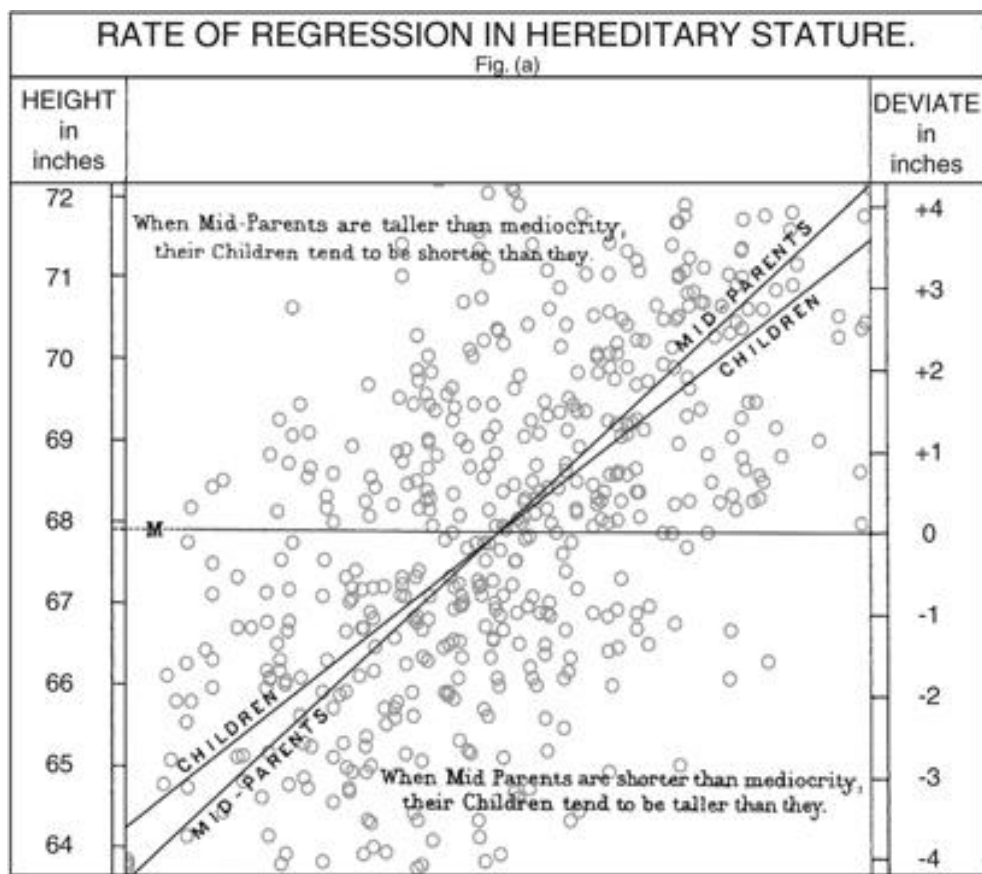
$$\min \left(\sum_i \underbrace{(y_i - \hat{y}_i)^2}_{\text{Reszta (e)}} \right)$$

Reszta (e) – różnica między wartością obserwowaną, a szacowaną



Regresja liniowa

- Pierwsze użycie terminu regresja – badania nad dziedziczeniem wzrostu u ludzi w XIX wieku



Sir Francis Galton
(1822 – 1911)

REGRESJA LINIOWA

- Regresję nazywamy funkcję, w której wartość cechy zależnej (**objaśnianej**) wyrażona jest za pomocą wartości cech niezależnych (**objaśniających**)

- Funkcja liniowa:

$$\hat{y}_i = b_{yx}x_i + a_{yx}$$

- Gdzie:

- b_{yx} – współczynnik regresji (w terminologii matematycznej: współczynnik kierunkowy prostej)
- a_{yx} – wyraz wolny funkcji regresji (punkt przecięcia linii prostej z osią Y)



WSPÓŁCZYNNIK REGRESJI PROSTEJ

- Współczynnik regresji

$$b_{yx} = \frac{cov_{xy}}{S_x^2}$$

Interpretacja: o ile zmieni się wartość zmiennej zależnej (Y) jeżeli wartość zmiennej niezależnej zmieni się o jednostkę

- Stała regresji (wyraz wolny funkcji)

$$a_{yx} = \bar{y} - b_{yx}\bar{x}$$

Punkt przecięcia linii prostej z osią Y
Nie interpretujemy



REGRESJA LINIOWA

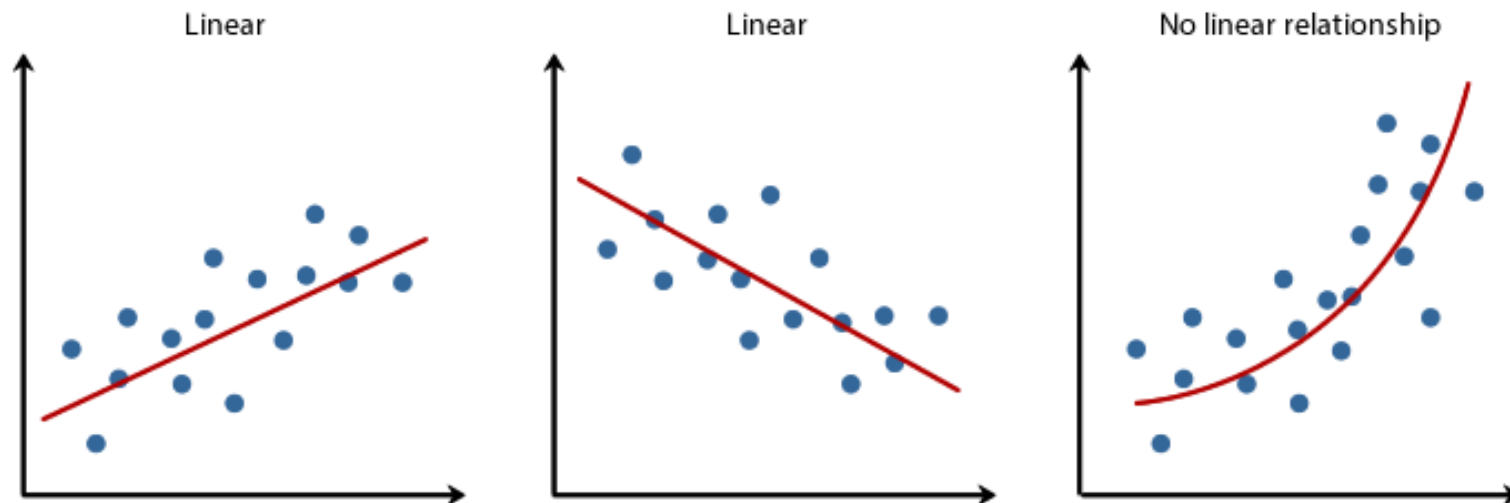
○ Założenia:

1. **Zależność jest liniowa**
2. Brak znaczących obserwacji odstających
3. **Homoscedastyczność** – wariancja reszt, składnika losowego jest taka sama dla wszystkich obserwacji
4. Reszty mają rozkład zbliżony do **rozkładu normalnego**
5. Regresja wielokrotna:
 - ❖ Liczba obserwacji musi być większa, bądź równa liczbie parametrów
 - ❖ Brak współliniowości parametrów
 - ❖ Nie występuje autokorelacja reszt



ZAŁOŻENIA MODELU

Zależność jest liniowa



Copyright 2014. Laerd Statistics.

Źródło: <https://statistics.laerd.com/spss-tutorials/linear-regression-using-spss-statistics.php>

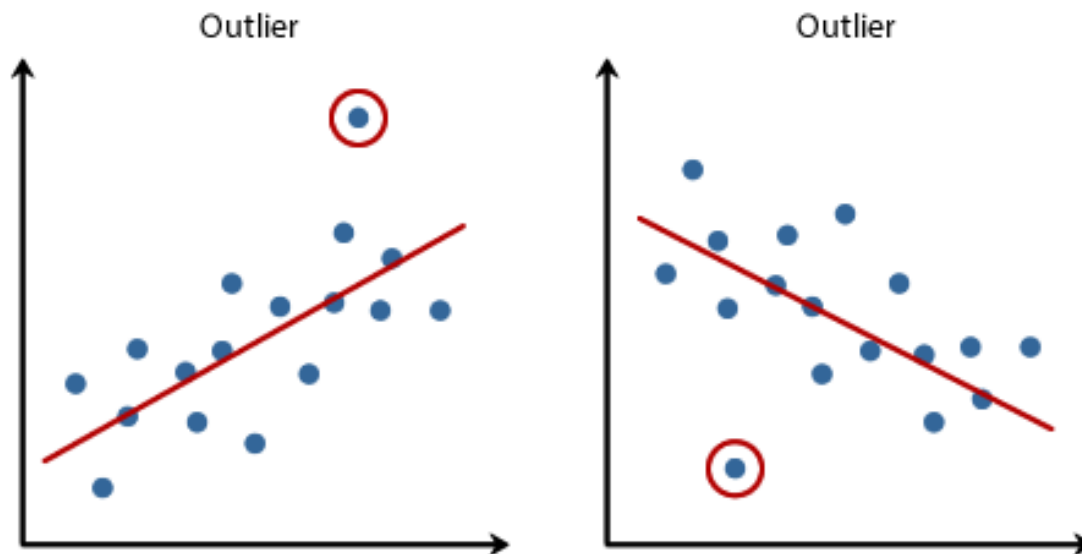
Sposób weryfikacji: Inspekcja wykresów punktowych
Jeżeli założenie nie jest spełnione:

- transformacja zmiennej zależnej
- dopasować model nieliniowy



ZAŁOŻENIA MODELU

Brak znaczących obserwacji odstających



Copyright 2014. Laerd Statistics.

Sposób weryfikacji: Inspekcja wykresów punktowych, statystyka opisowa, odległość Cooka

Źródło: <https://statistics.laerd.com/spss-tutorials/linear-regression-using-spss-statistics.php>



NORMALNOŚĆ RESZT

- Rozkład zakłócenia losowego przekłada się na spodziewany rozkład reszt

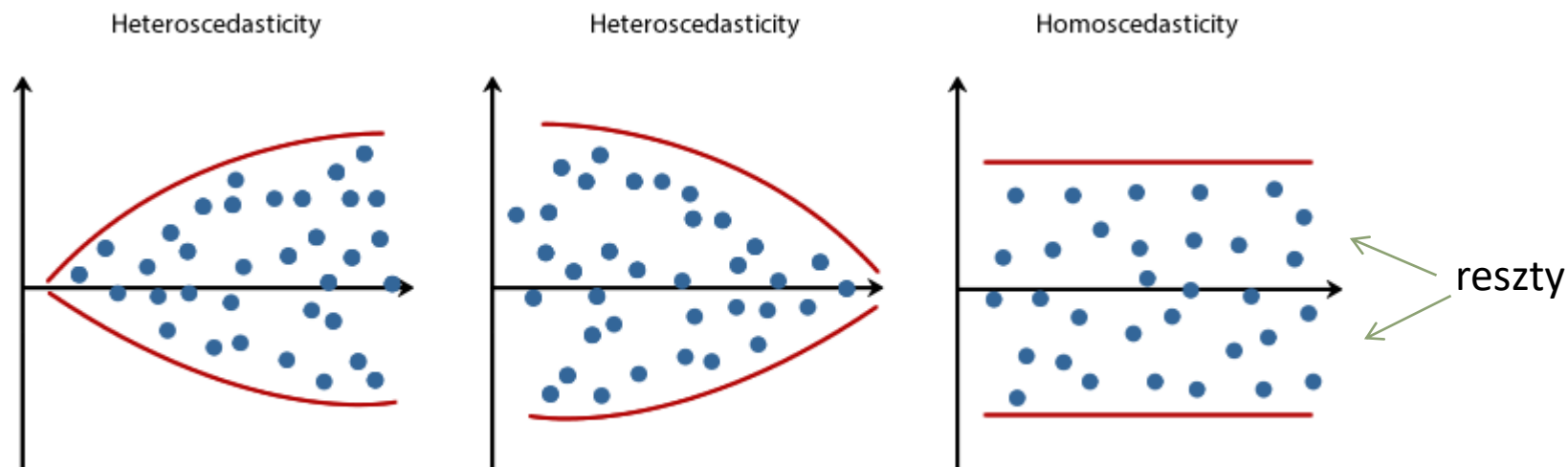
Analiza jakości dopasowania modelu oraz jego diagnostyka opiera się na analizie reszt

- Weryfikujemy:
 - Normalność rozkładu reszt
 - Homoscedastyczność



ZAŁOŻENIA MODELU

Homoscedastyczność



Copyright 2014. Laerd Statistics.

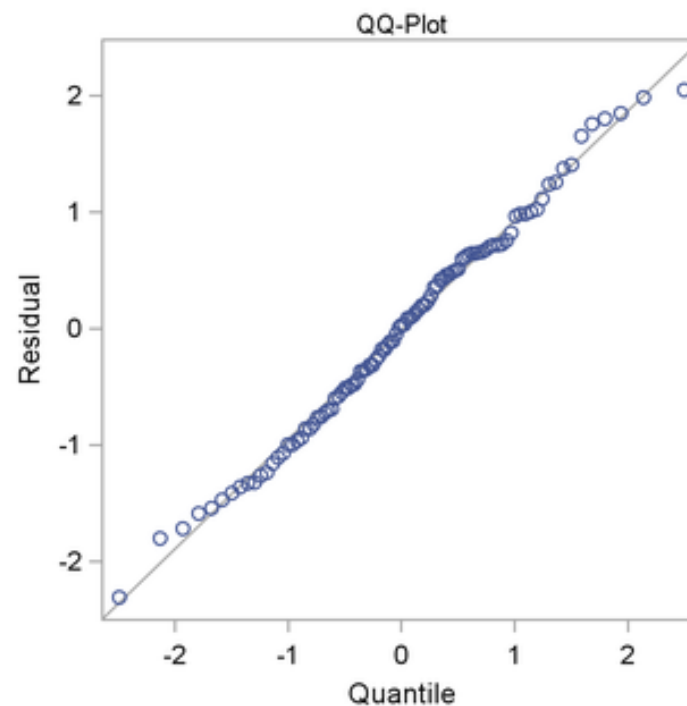
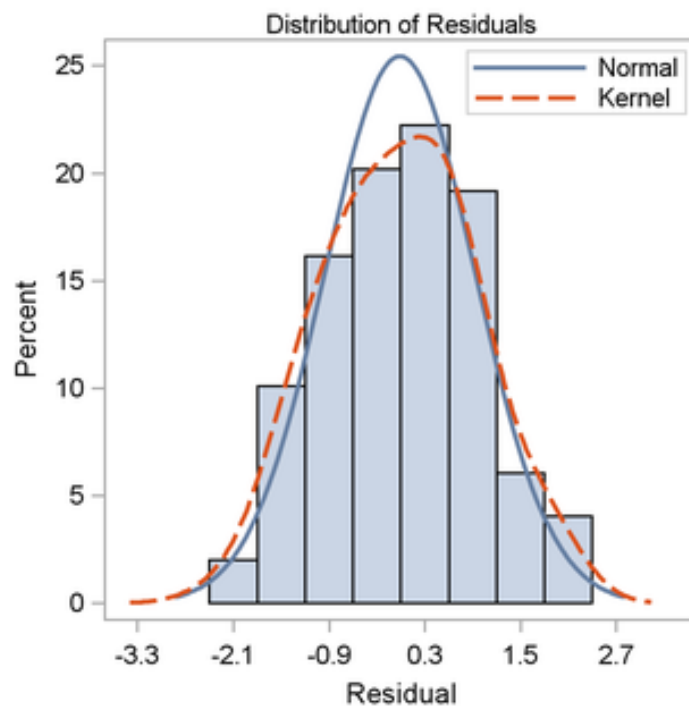
Źródło: <https://statistics.laerd.com/spss-tutorials/linear-regression-using-spss-statistics.php>

homoscedastyczność - stałość wariancji składnika losowego dla poszczególnych wartości zmiennej niezależnej



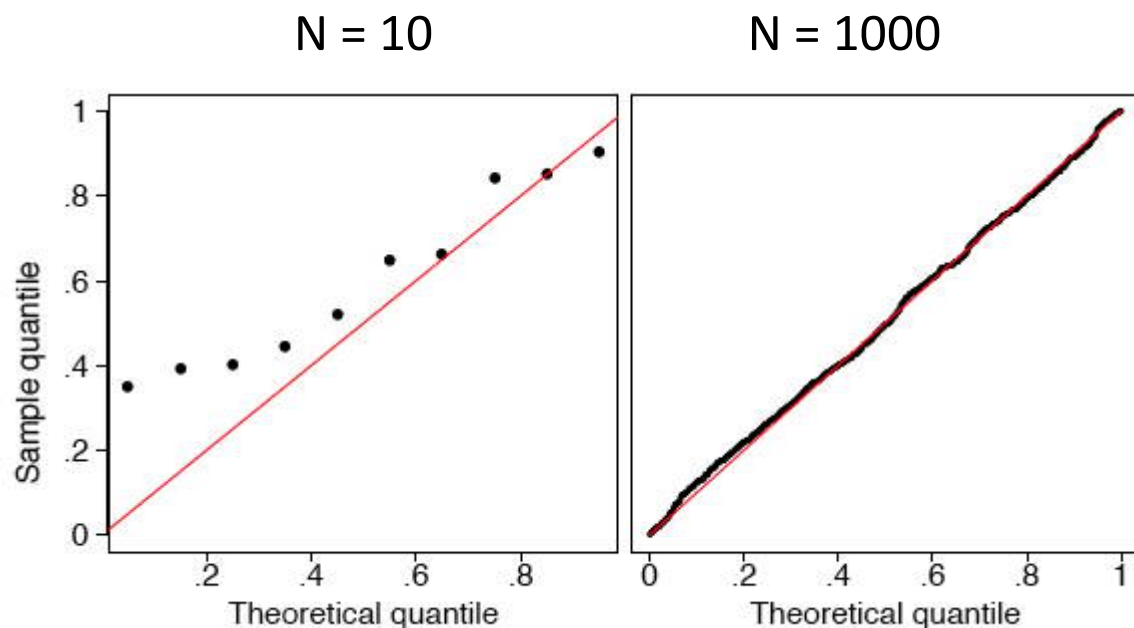
ZAŁOŻENIA MODELU

- Normalność reszt – diagnostyka
 - Histogram oraz wykres kwantyl-kwantyl
 - Testy statystyczne (np. Shapiro-Wilka) tylko jako dodatkowe źródło informacji



NORMALNOŚĆ RESZT

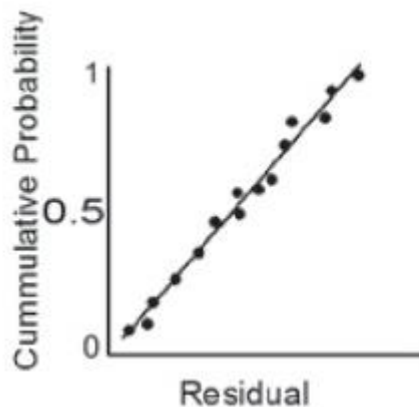
- Wykres kwantyl-kwantyl – Uwaga na małą liczebność próby !



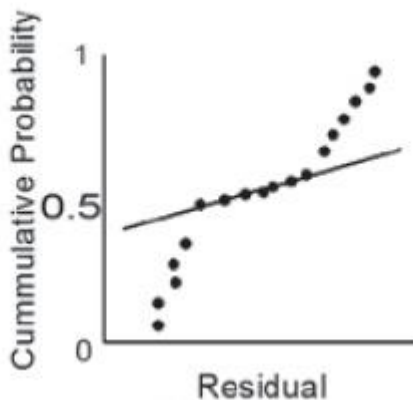
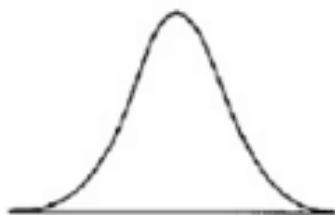
NORMALNOŚĆ RESZT

- Normalność rozkładu reszt

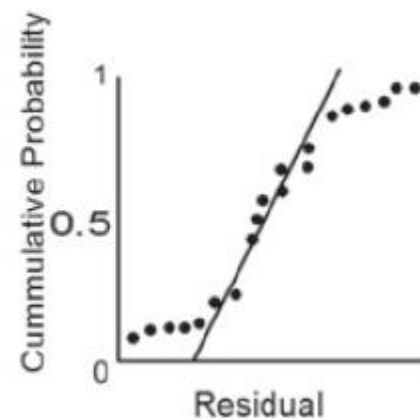
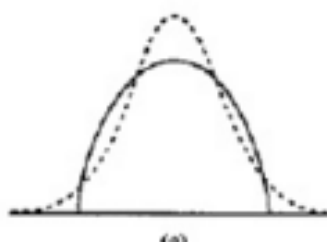
Tzw. **grube ogony** - gęstość prawdopodobieństwa na końcach rozkładów jest większa niż w rozkładzie normalnym



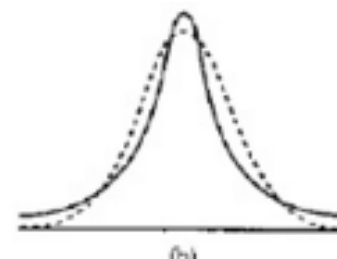
Rozkład normalny



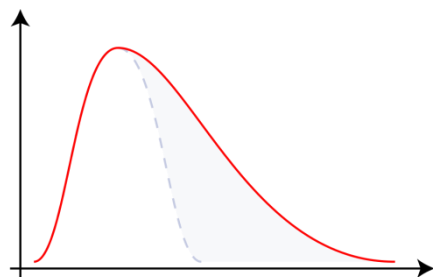
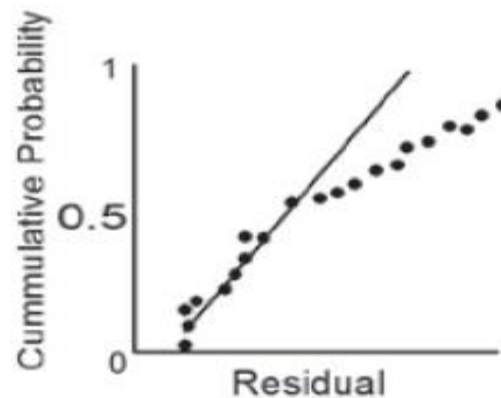
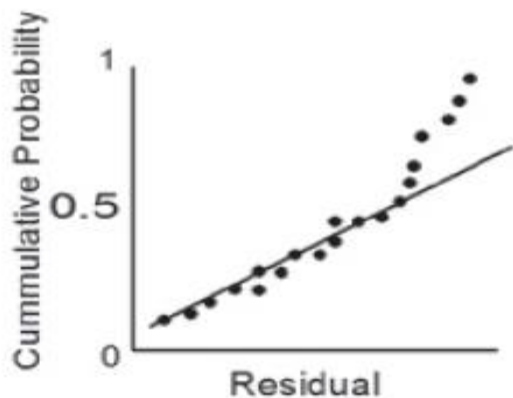
Grube ogony



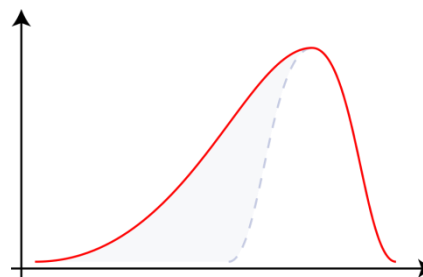
Cienkie ogony



NORMALNOŚĆ RESZT



Asymetria
prawostronna

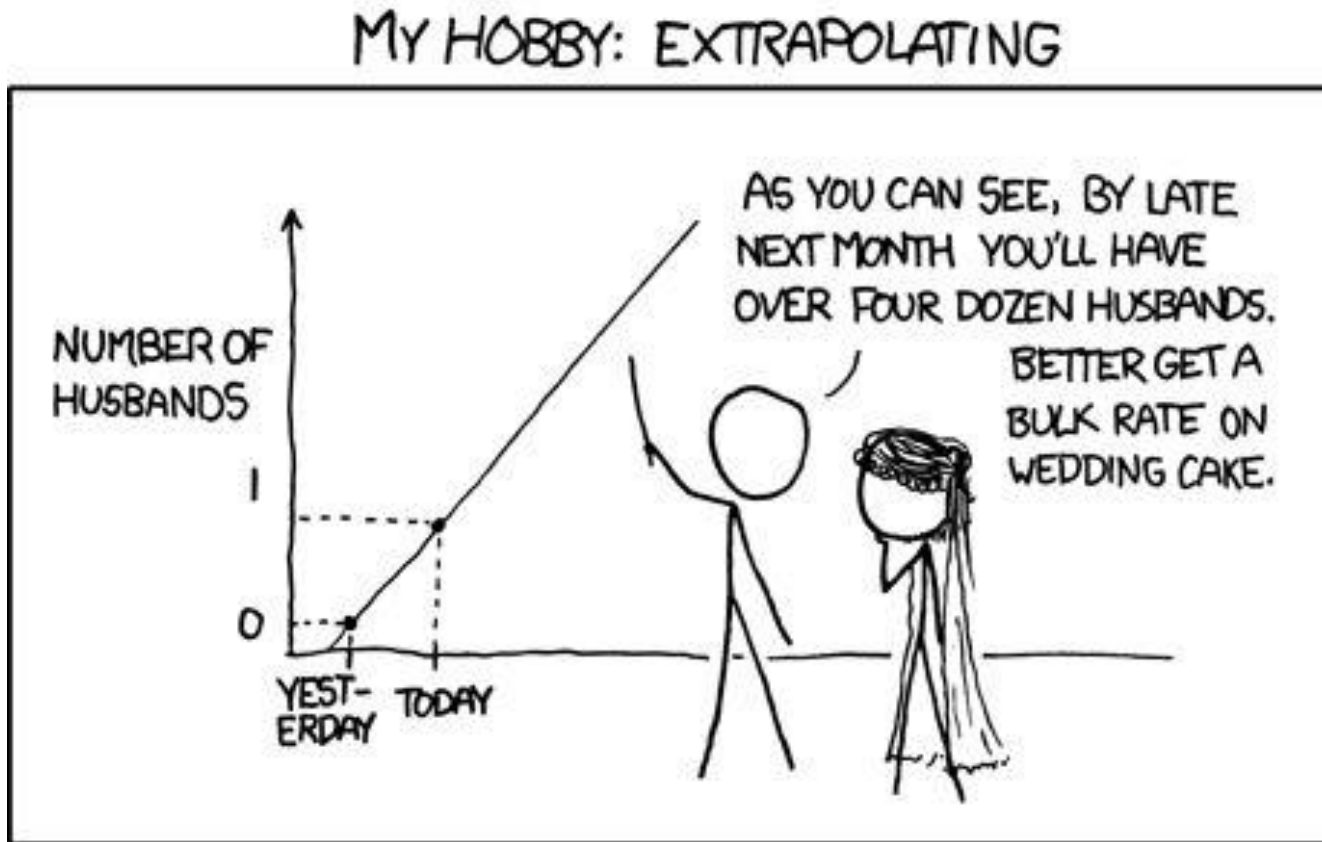


Asymetria
lewostronna



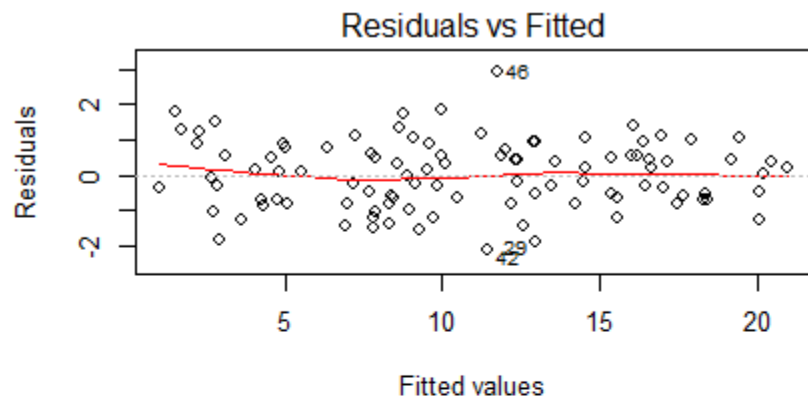
MOŻLIWOŚCI EKSTRAPOLACJI

- Model regresji **nie powinien być ekstrapolowany na dane spoza zakresu** na których został opracowany!

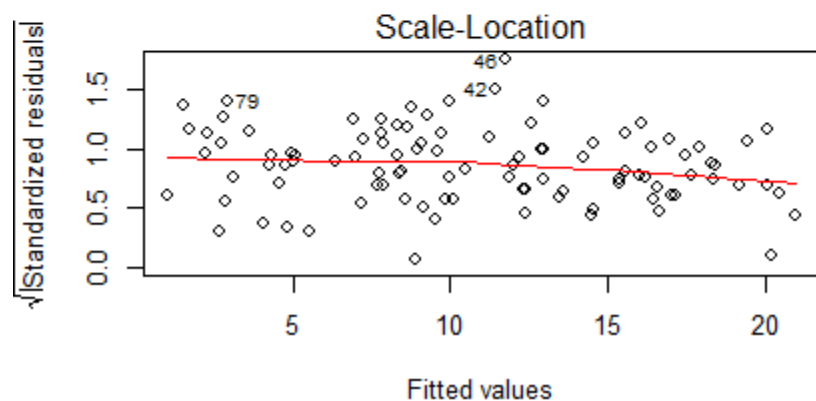
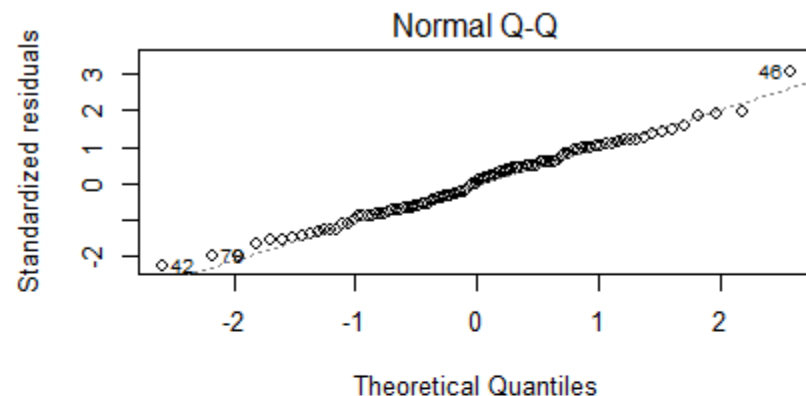


WERYFIKACJA ZAŁOŻEŃ MODELU W R

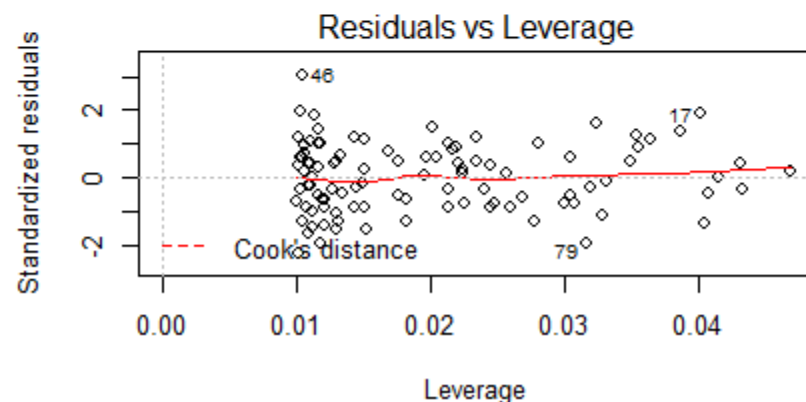
normalność rozkładu reszt,
homoscedastyczność



normalność rozkładu reszt

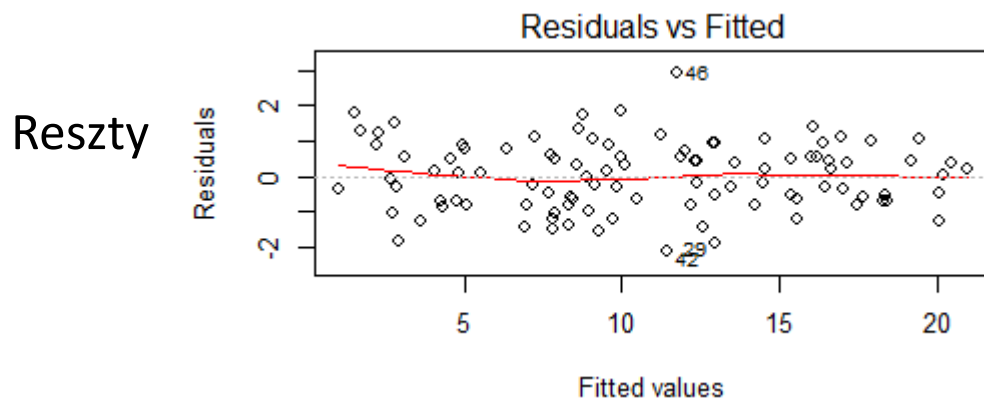


homoscedastyczność



Wpływ pojedynczych obserwacji
na szacowany model

WERYFIKACJA ZAŁOŻEŃ MODELU



Wartości dopasowane przez model

Dla adekwatnego modelu reszty nie powinny zależeć funkcyjnie od zmiennej objaśnianej, powinny mieć średnią warunkową równą zero, niezależnie od wartości y_i

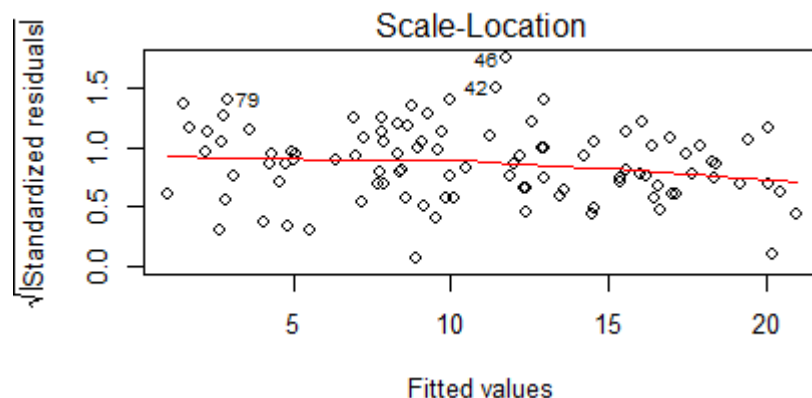


WERYFIKACJA ZAŁOŻEŃ MODELU

- Standaryzowane reszty to reszty dzielone przez oszacowanie odchylenia standardowego reszt

Pierwiastki z modułów
standaryzowanych reszt

Pozioma czerwona linia prosta
wskazuje na brak zależności
wariancji reszt od oszacowanych
wartości



Wartości dopasowane przez model

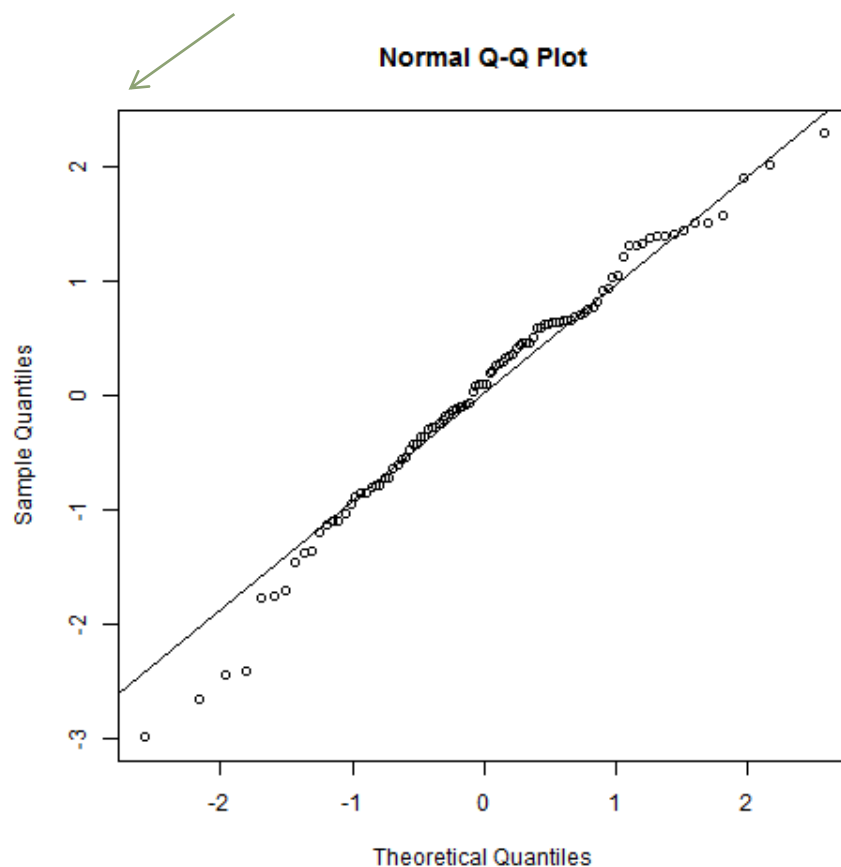
- Wariancja reszt powinna być jednorodna i nie powinna zależeć funkcyjnie od wartości dopasowanych przez model
 - obserwacja jakiegokolwiek trendu wskazuje na odstępstwo od założeń jednorodności wariancji
 - niejednorodną wariancję można zredukować przez odpowiednią transformację zmiennych

WERYFIKACJA ZAŁOŻEŃ MODELU

○ Wykres kwantyl-kwantyl (ang. Q-Q plot)

- Inaczej: wykres kwantylowy dla rozkładu normalnego
- Weryfikacja zgodności z rozkładem teoretycznym
- Dobre dopasowanie rozkładu teoretycznego do zaobserwowanych wartości w przypadku gdy wykreślane wartości będą tworzyły **linię prostą**
- Mała liczebność próby utrudnia lub uniemożliwia prawidłową interpretację

Kwantyle empiryczne dla standaryzowanych reszt



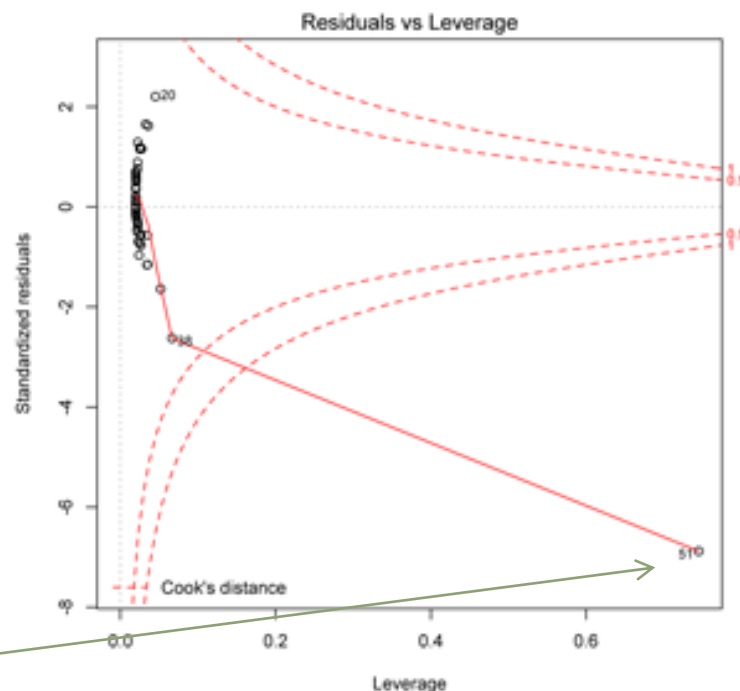
Wartości kwantyli rozkładu normalnego odpowiadające resztom

WERYFIKACJA ZAŁOŻEŃ MODELU

○ Odległość Cooka

- miara stopnia zmiany współczynników regresji, gdyby dany przypadek pominąć w obliczeniach
- miara wpływu poszczególnych obserwacji na prostą regresji
 - wysoka **dźwignia** (ang. leverage) – typowa wartość Y_i i nietypowa wartość X_i
 - wysoka **odmienność** (ang. discrepacy) – nietypowa wartość Y_i dla typowej wartości X_i

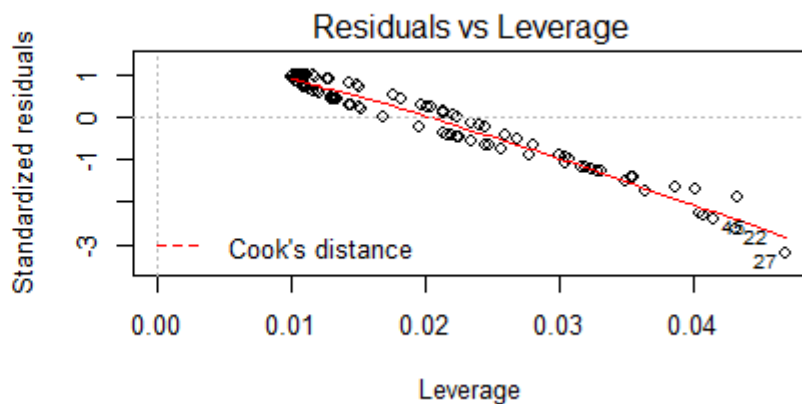
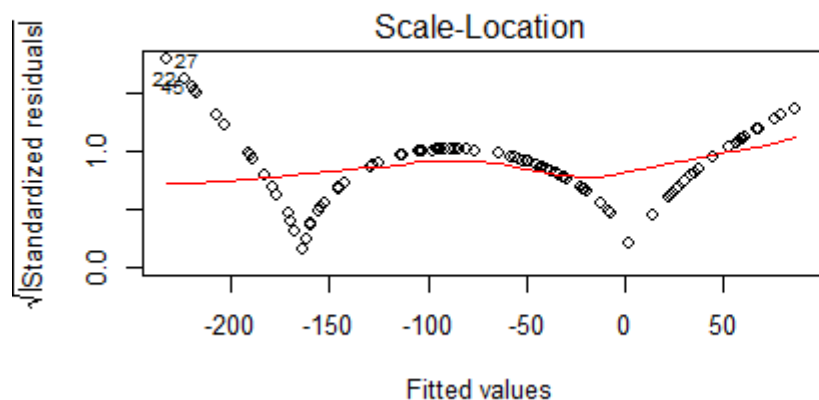
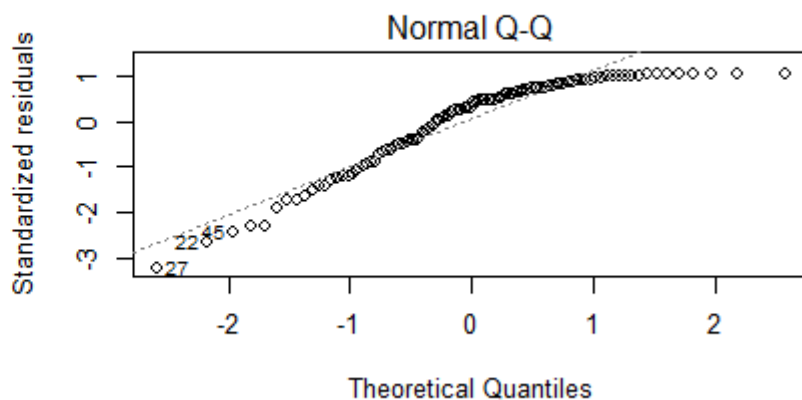
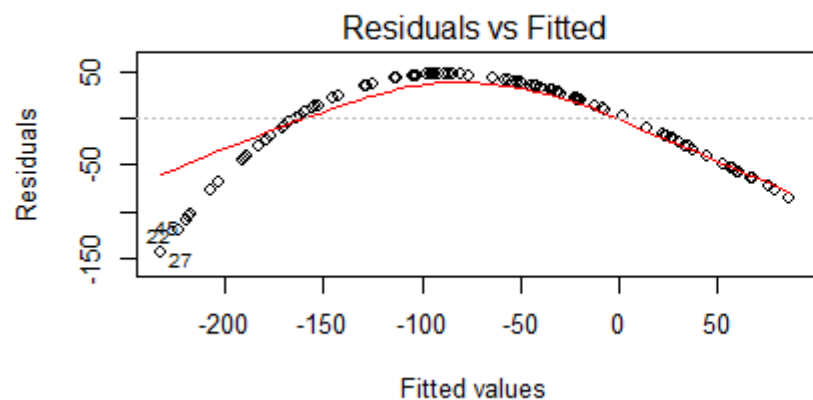
Wartość wpływowa =
wysoka odmienność + wysoka dźwignia



Duża odległość Cooka (> 1) –
duży wpływ na obciążenie
równania regresji

Dźwignia – miara wpływu obserwacji na
ocenę współczynników modelu

Weryfikacja założeń modelu w R



OCENA DOPASOWANIA MODELU

$$\sum_{i=1}^N (y_i - \bar{y})^2 = \sum_{i=1}^N (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

$$SST = SSR + SSE$$

- SST – Suma kwadratów odchyłeń całkowitych
 - SSE – suma kwadratów odchyłeń nie wyjaśnionych regresją (**reszt**)
 - SSR – suma kwadratów odchyłeń wyjaśnionych regresją
-
- Współczynnik determinacji

$$R^2 = \frac{SSR}{SST} = \frac{SSR}{SSR + SSE}$$



WSPÓŁCZYNNIK DETERMINACJI

Procent wariancji wyjaśnionej przez model

- Jaki **procent zmienności** zmiennej zależnej (Y - objaśnianej) jest wyjaśniany za pomocą zmiennej niezależnej (X - objaśniającej)
 - Na ile nasz model wyjaśnia zgromadzone dane pomiarowe?
- Dla prostej regresji liniowej otrzymujemy podnosząc do kwadratu współczynnik korelacji liniowej Pearsona (r)
- Przyjmuje wartości **od 0 do 1**
 - Im bliżej jedynki tym lepsze dopasowanie modelu do danych
- Silna zależność od liczby parametrów modelu – aby tego uniknąć liczymy zmodyfikowany R^2



ISTOTNOŚĆ MODELU I WSPÓŁCZYNNIKÓW

○ Istotność modelu

- Test F-Snedecora
- H_0 : Funkcja regresji jest nieistotna
- H_1 : Funkcja regresji jest istotna

○ Istotność współczynników regresji

- $H_0: \beta_{yx} = 0$ – istotność współczynnika kierunkowego
- $H_0: \alpha_{yx} = 0$ – istotność wyrazu wolnego
- Hipotezy weryfikujemy przy pomocy testu t-studenta

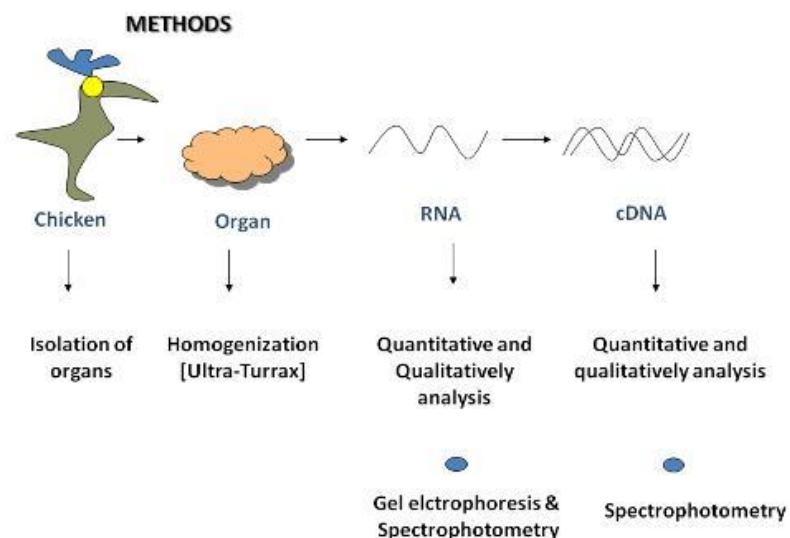


REGRESJA LINIOWA - PRZYKŁAD

<i>Gen_1</i>	<i>Gen_2</i>
-1.06	-1.08
-0.81	-1.02
-0.48	-0.39
-0.42	-0.48
-0.30	-0.58
-0.35	-0.24
-0.31	-0.05
-0.18	-0.33
-0.20	0.51
-0.11	-0.53
-0.09	-0.47
0.16	0.10
0.45	0.39
0.53	0.11
0.67	0.52
0.80	0.34
0.87	1.08
0.92	1.21

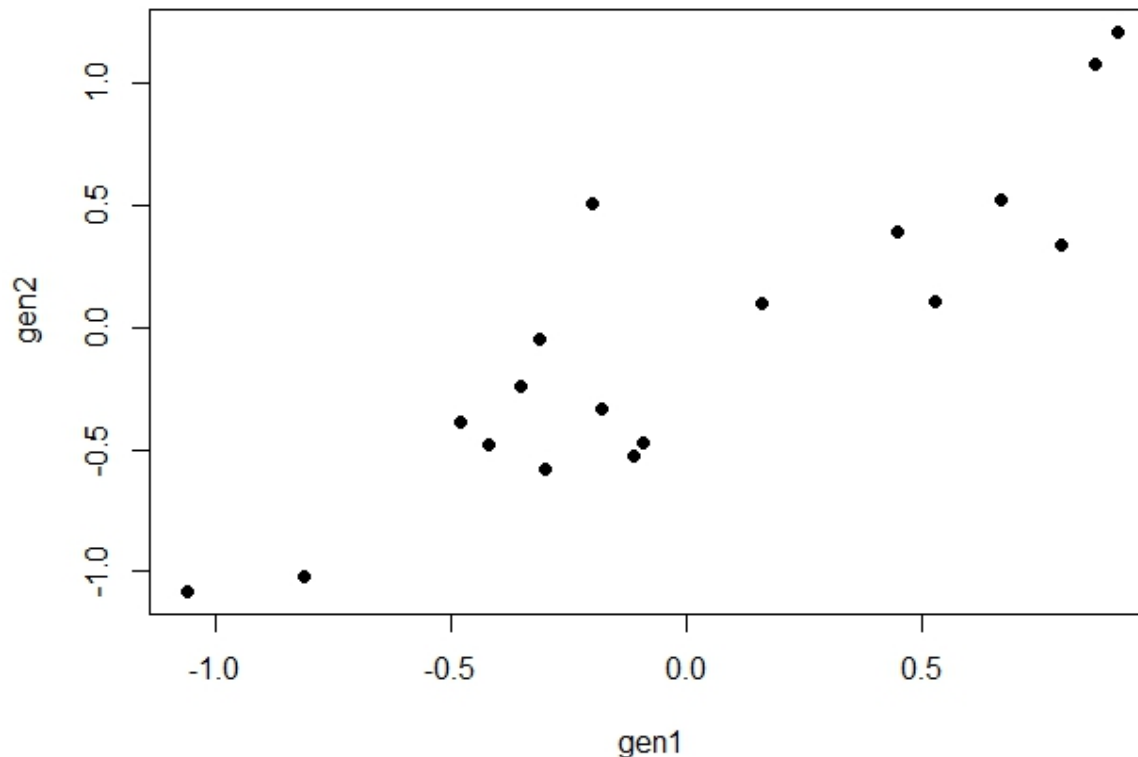
Zmiana w poziomie ekspresji dwóch genów

Dopasuj model, który opisuje zmianę w poziomie ekspresji genu 2 w zależności od poziomu ekspresji genu 1



REGRESJA LINIOWA PRZYKŁAD W R

- Czy istnieje zależność pomiędzy poziomem ekspresji analizowanych genów? Czy jest ona prostoliniowa?



REGRESJA LINIOWA PRZYKŁAD W R

- Współczynnik korelacji Pearsona

```
> cor.test(gen1,gen2)
```

```
Pearson's product-moment correlation
```

```
data:  gen1 and gen2
```

```
t = 7.5105, df = 16, p-value = 1.246e-06
```

```
alternative hypothesis: true correlation is not equal to 0
```

```
95 percent confidence interval:
```

```
0.7071560 0.9556856
```

```
sample estimates:
```

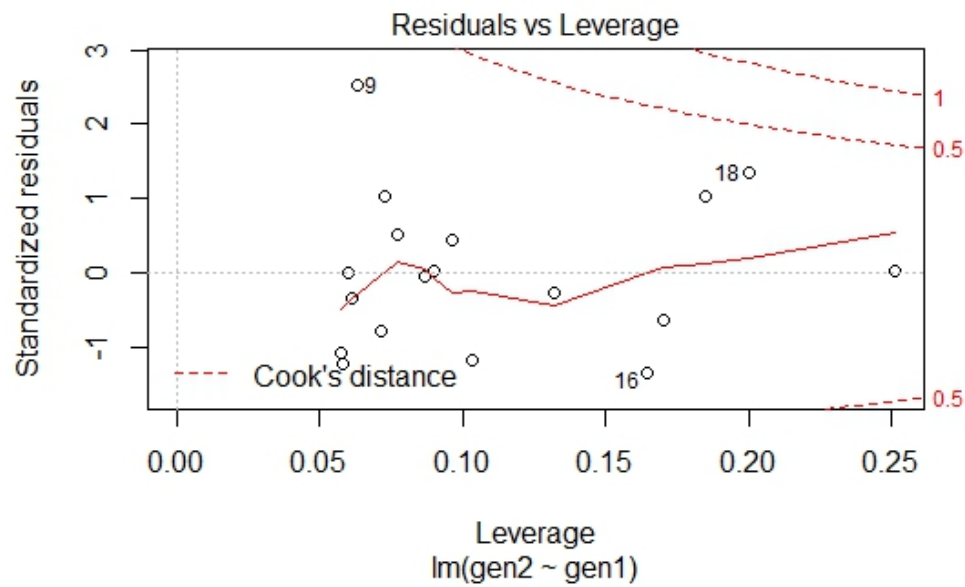
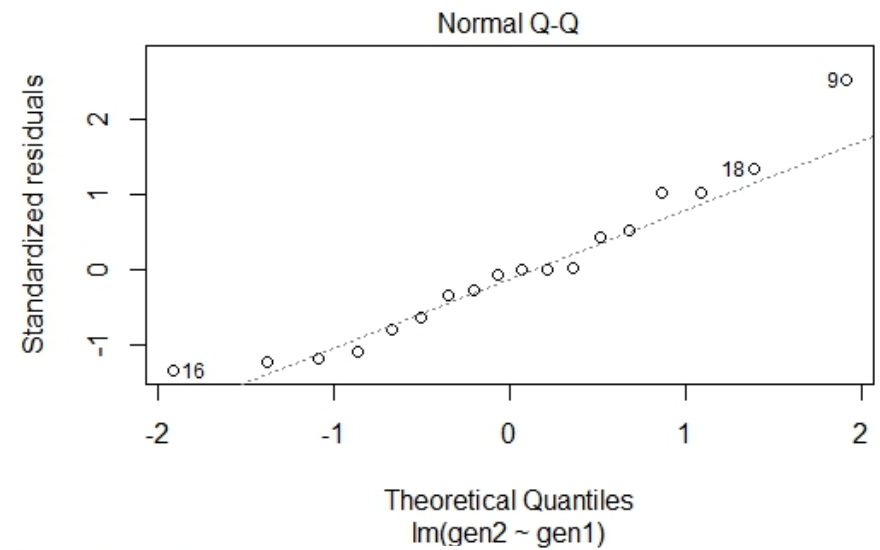
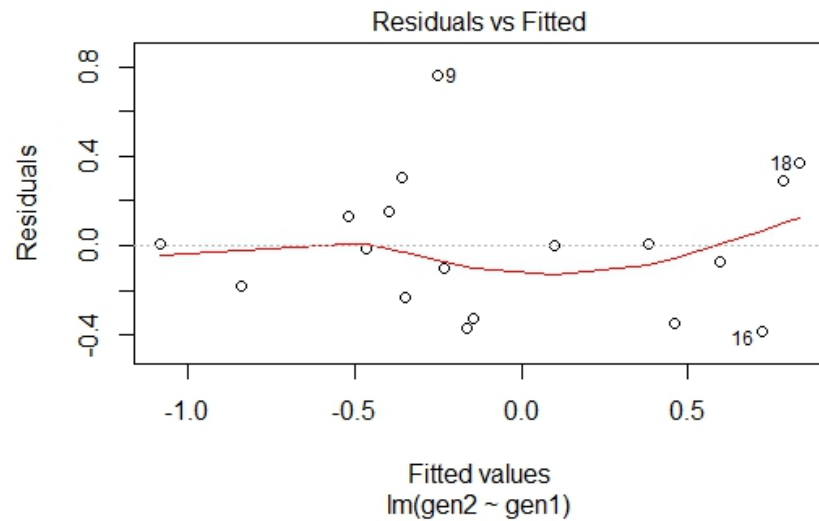
```
cor
```

```
0.8826268
```

- Istnieje istotna zależność ($r=0,882$; $p<0,001$) pomiędzy poziomem ekspresji pierwszego genu, a genu drugiego
- Jak możemy tę zależność scharakteryzować?



DIAGNOSTYKA



MODEL LINIOWY W R

```
> geneLM<-lm(gen2~gen1)
> geneLM
```

```
Call:
lm(formula = gen2 ~ gen1)
```

```
Coefficients:
(Intercept)      gen1
   -0.05541      0.97070
```

```
> summary(geneLM)
```

```
Call:
lm(formula = gen2 ~ gen1)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-0.3811 -0.2196 -0.0084  0.1492  0.7595
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.05541    0.07330  -0.756   0.461
gen1         0.97070    0.12925   7.511 1.25e-06 ***
```

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

```
Residual standard error: 0.311 on 16 degrees of freedom
Multiple R-squared:  0.779, Adjusted R-squared:  0.7652
F-statistic: 56.41 on 1 and 16 DF, p-value: 1.246e-06
```

○ Funkcja **lm**:

model <- lm(y~x) – wykonuje dopasowanie modelu liniowego, wyznacza oceny współczynników β oraz wylicza wartości reszt

summary(model) – funkcja summary dla modelu regresji liniowej. Zwraca dodatkowe informacje i testuje istotność funkcji regresji i współczynników



MODEL LINIOWY W R

formuła modelu

```
> summary(geneLM)
```

Call:

```
lm(formula = gen2 ~ gen1)
```

statystyki opisowe
dla reszt

Residuals:

Min	1Q	Median	3Q	Max
-0.3811	-0.2196	-0.0084	0.1492	0.7595

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.05541	0.07330	-0.756	0.461
gen1	0.97070	0.12925	7.511	1.25e-06 ***

Ocena wartości
współczynników
regresji i ich
istotność

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

Współczynnik
determinacji

Residual standard error: 0.311 on 16 degrees of freedom

Multiple R-squared: 0.779, Adjusted R-squared: 0.7652

F-statistic: 56.41 on 1 and 16 DF, p-value: 1.246e-06

Błąd
standardowy
dla reszt

Istotność funkcji regresji

Poprawiony
współczynnik
determinacji

WSPÓŁCZYNNIK DETERMINACJI

OBLICZENIA

X	Y	$\hat{y}_i$	$(y_i - \hat{y}_i)^2$	$\hat{y}_i - \bar{y}$
-1,06	-1,08	-1,084	0,000	1,176
-0,81	-1,02	-0,842	0,032	0,708
-0,48	-0,39	-0,521	0,017	0,272
-0,42	-0,48	-0,463	0,000	0,214
-0,3	-0,58	-0,347	0,054	0,120
-0,35	-0,24	-0,395	0,024	0,156
-0,31	-0,05	-0,356	0,094	0,127
-0,18	-0,33	-0,230	0,010	0,053
-0,2	0,51	-0,250	0,577	0,062
-0,11	-0,53	-0,162	0,135	0,026
-0,09	-0,47	-0,143	0,107	0,020
0,16	0,1	0,100	0,000	0,010
0,45	0,39	0,381	0,000	0,145
0,53	0,11	0,459	0,122	0,211
0,67	0,52	0,595	0,006	0,354
0,8	0,34	0,721	0,145	0,520
0,87	1,08	0,789	0,085	0,623
0,92	1,21	0,838	0,139	0,702
			SSE	SSR
			1.547	5.500

$$\sum_{i=1}^N (y_i - \bar{y})^2 = \sum_{i=1}^N (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

$$R^2 = \frac{SSR}{SST} = \frac{SSR}{SSR + SSE}$$

$$\hat{y}_i = -0,05541 + 0,97070x_i$$



MODELE LINIOWE W R

- Po podstawieniu oszacowanych współczynników:

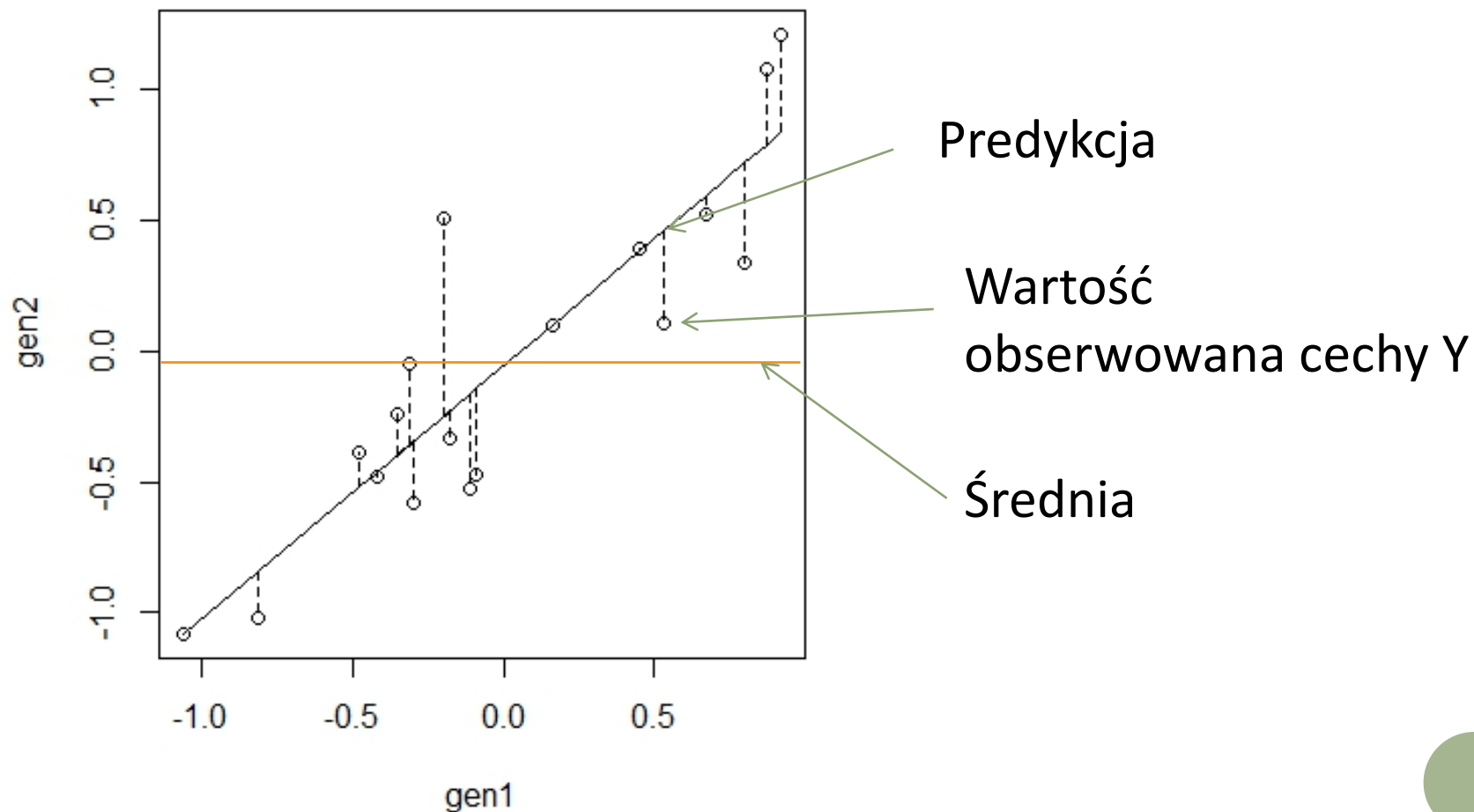
$$\hat{y}_i = -0,05541 + 0,97070x_i$$

- Jeżeli funkcja regresji jest istotna, możemy szacować wartości zmiennej zależnej (y) na podstawie wartości zmiennej niezależnej (x)
- Oszacuj poziom ekspresji genu 2 na podstawie poziomu ekspresji dla genu 1

Próbka	Gen 1	Predykcja	Gen 2
1	-0,35		-0,24
2	0,67		0,52
3	0,92		1,21



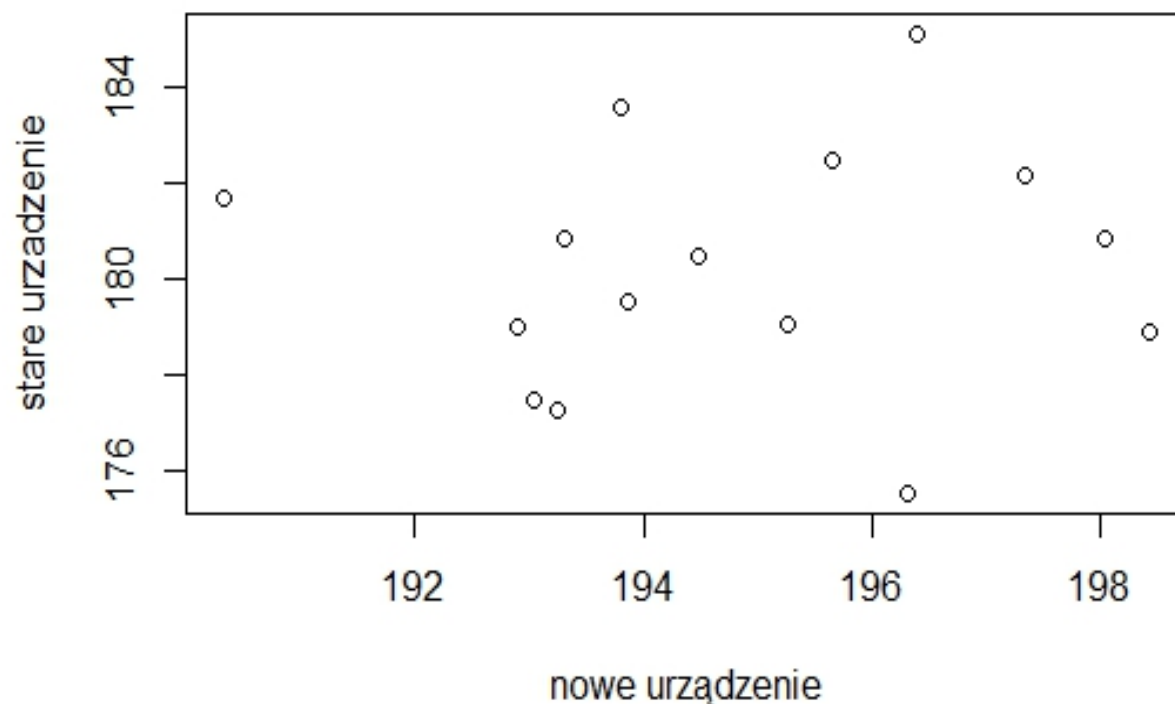
REGRESJA LINIOWA



GLUKOZA

Za pomocą dwóch różnych urządzeń zmierzono poziom glukozy u 15 pacjentów,

Czy uzyskane wyniki są skorelowane?



```
> cor.test(glucose$New,glucose$Old)
```

Pearson's product-moment correlation

data: glucose\$New and glucose\$Old

t = 0.25008, df = 13, p-value = 0.8064

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

-0.4593508 0.5615523

sample estimates:

cor

0.06919402

GLUKOZA

```
> summary(glukoza)
```

```
Call:
```

```
lm(formula = New ~ Old)
```

```
Residuals:
```

	Min	1Q	Median	3Q	Max
	-4.5822	-1.4761	-0.3589	1.5217	3.6892

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	183.91042	43.64673	4.214	0.00101	**
Old	0.06055	0.24210	0.250	0.80643	

```
---
```

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

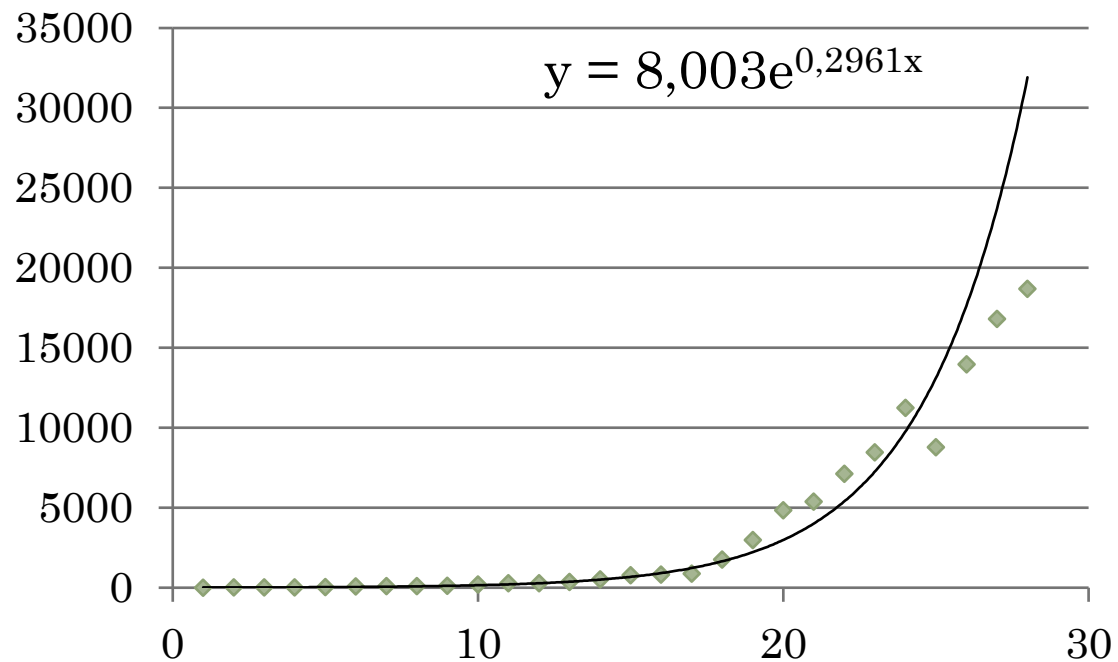
```
Residual standard error: 2.298 on 13 degrees of freedom
```

```
Multiple R-squared:  0.004788, Adjusted R-squared:  -0.07177
```

```
F-statistic: 0.06254 on 1 and 13 DF,  p-value: 0.8064
```

WERYFIKACJA MODELU WYKŁADNICZEGO

- Liczba potwierdzonych przypadków COVID-19 w USA wg ECDC zaczynając od 1.03.2020 do 28.03.2020



MODEL WYKŁADNICZY W R

```
> model <- nls(cases ~ N0*exp(r*Lp),
+             data = dane,
+             start = list(N0 = 1, r = 0.2),
+             trace = TRUE)
1187005190 : 1.0 0.2
1109944407 : 39.7201303 0.1047517
1051265594 : 19.4586007 0.1500131
870358675 : 15.2804362 0.1894517
638506736 : 22.5426373 0.1967932
362693045 : 36.3082920 0.1977626
106658536 : 57.1121609 0.1977482
21228709 : 77.9117969 0.1977524
21228708 : 77.9130348 0.1977507
> summary(model)
```

Formula: cases ~ N0 * exp(r * Lp)

Parameters:

	Estimate	Std. Error	t value	Pr(> t)	
N0	77.9130	21.1203	3.689	0.00105	**
r	0.1978	0.0104	19.011	< 2e-16	***

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

Residual standard error: 903.6 on 26 degrees of freedom

Number of iterations to convergence: 8

Achieved convergence tolerance: 4.748e-06

Dobór
współczynników
modelu w R



KORELACJA LINIOWA

```
> cor.test(Lp, log10(cases))
```

Pearson's product-moment correlation

data: Lp and log10(cases)

t = 33.244, df = 26, p-value < 2.2e-16

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

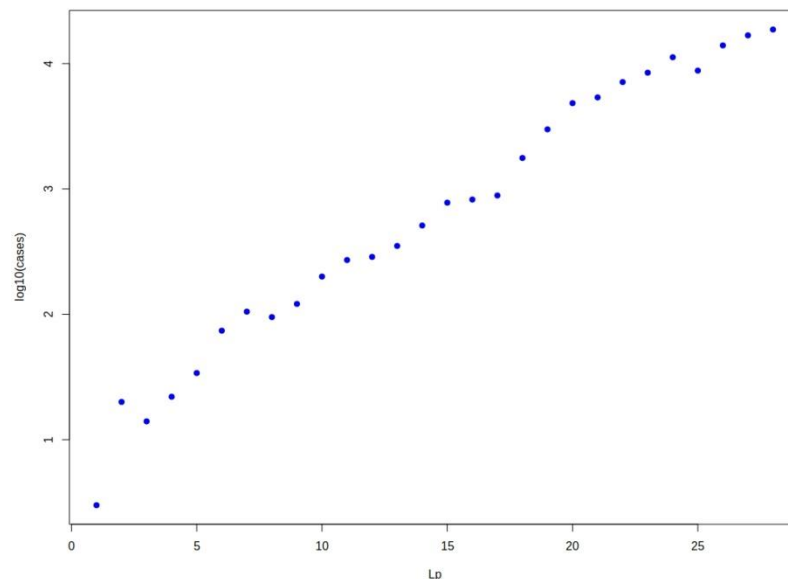
0.9748553 0.9947055

sample estimates:

cor

0.9884404

Dane na
współrzędnych
półlogarytmicznych



REGRESJA LINIOWA

```
> model<-lm(log10(cases)~Lp)
> summary(model)
```

```
Call:
lm(formula = log10(cases) ~ Lp)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.55473	-0.08187	0.02556	0.11641	0.21775

Coefficients:

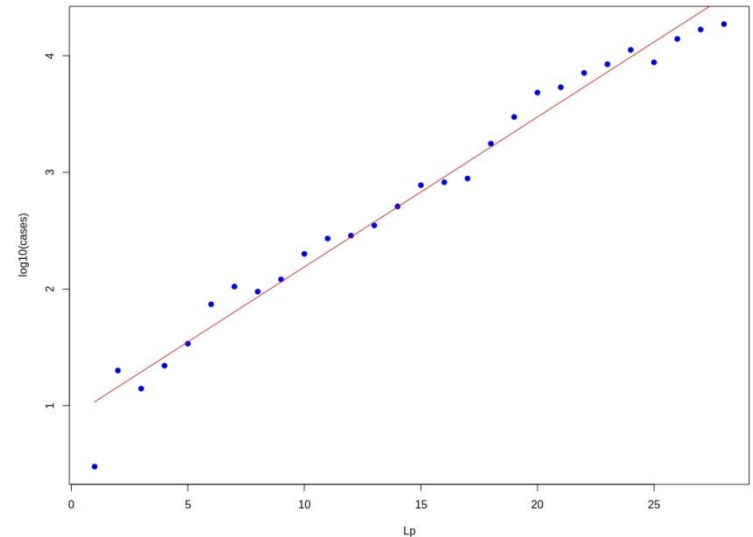
	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	0.903251	0.064208	14.07	1.15e-13	***
Lp	0.128598	0.003868	33.24	< 2e-16	***

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

Residual standard error: 0.1653 on 26 degrees of freedom

Multiple R-squared: 0.977, Adjusted R-squared: 0.9761

F-statistic: 1105 on 1 and 26 DF, p-value: < 2.2e-16



OGRANICZENIA MODELU REGRESJI LINIOWEJ

- Małe możliwości ekstrapolacji
- Zmienna objaśniana musi być na skali ilościowej
- Prosta regresja liniowa – analiza tylko dwóch zmiennych
- Tylko zależność prostoliniowa – bardzo częsty błąd!

- Co zamiast prostej regresji liniowej?
 - Regresja wielokrotna – więcej zmiennych objaśniających
 - Uogólnione modele linowe – zmienna objaśniająca Y nie jest cechą o rozkładzie normalnym (np. regresja logistyczna, regresja Poissona)



ŹRÓDŁA

- Biecek P. 2013. Analiza danych z programem R. Wydawnictwo naukowe PWN. Warszawa
- Olech W., Wieczorek M. 2010. Zastosowanie metod statystyki w doświadczałnictwie zootechnicznym. Wydawnictwo SGGW.

