



**AKADEMIA GÓRNICZO-HUTNICZA
IM. STANISŁAWA STASZICA W KRAKOWIE**

Regresja i korelacja

Statystyka w medycynie

**Dr inż. Janusz Majewski
Katedra Informatyki**

Regresja i korelacja

Dla każdej jednostki (obiektu) mamy wartości dwóch zmiennych losowych: x i y . Chcemy zbadać związek między tymi dwoma zmiennymi dla:

- -uzyskania liczbowych miar pewnych podstawowych cech zależności,
- -dostarczenia możliwości prognozowania (predykcji) wartości jednej ze zmiennych, gdy druga jest znana,
- - stwierdzenia, czy obserwowany kierunek trendu jest istotny.

Mówimy o związku między dwiema zmiennymi, gdy rozkład jednej zmiennej związany jest z wartościami drugiej. Nie znaczy to, że jedna zmienna jest przyczyną drugiej, nie mówimy więc o związku przyczynowo-skutkowym (por. liczba rozwodów versus produkcja papierosów)

Regresja prostoliniowa

Obserwowano zmienne x i y dla dużej liczby obiektów. Interesuje nas, jakiej przeciętnej zmianie ulega y gdy x przyjmuje różne wartości.

Zależność $E(y|x)$ od x nazywamy funkcją regresji. Mówimy o regresji prostoliniowej, jeżeli zmienna y przyjmuje rozkład normalny ze średnią

$$E(y|x) = A + Bx$$

oraz stałą (niezależną od x) wariancją równą σ^2 .

Mamy n par obserwacji (x_i, y_i) . Należy znaleźć liczby a i b będące estymatorami A i B tak, aby zminimalizować sumę kwadratów

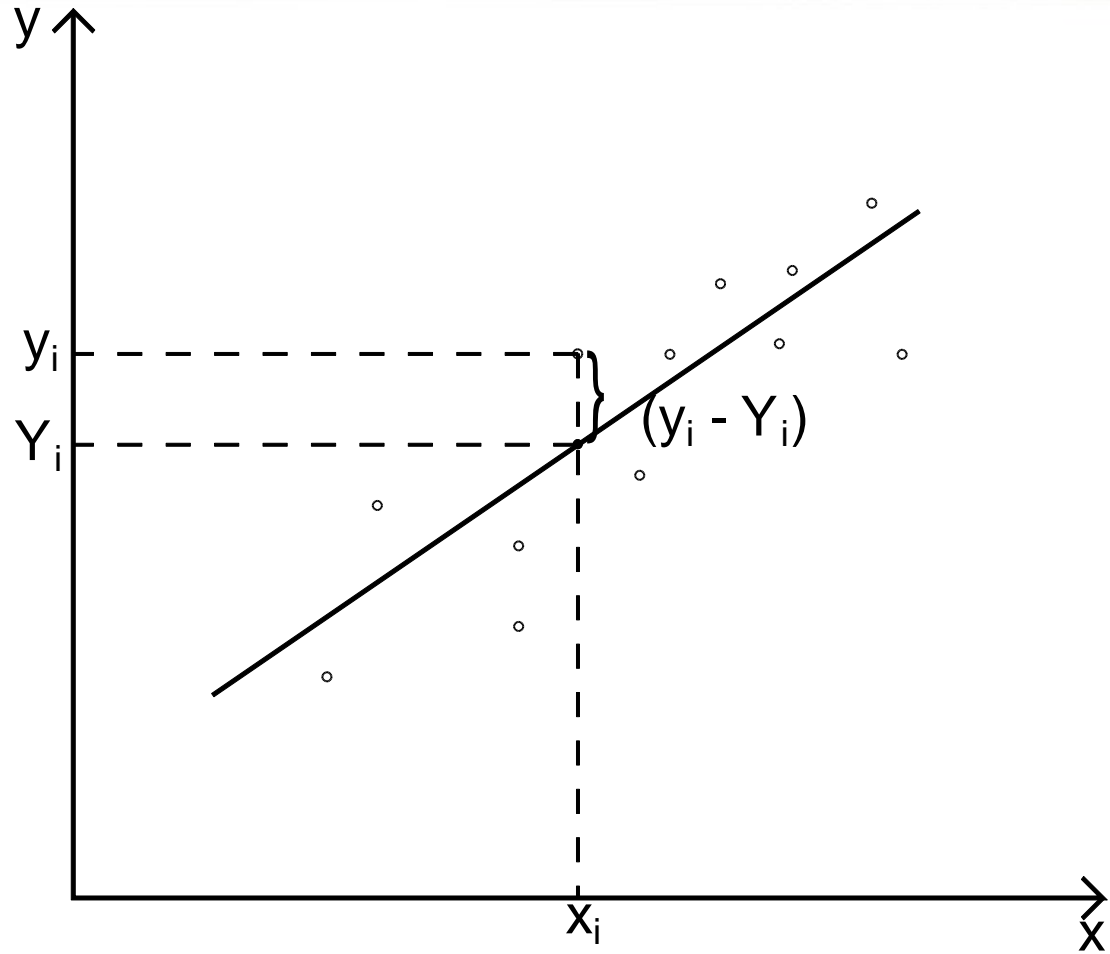
$$\sum_i (y_i - Y_i)^2$$

gdzie Y_i są wartościami wyznaczonymi przez szacowane równanie regresji:

$$Y_i = a + b \cdot x_i$$

Regresja prostoliniowa

y_i - wartość
obserwowana
 Y_i - wartość
teoretyczna

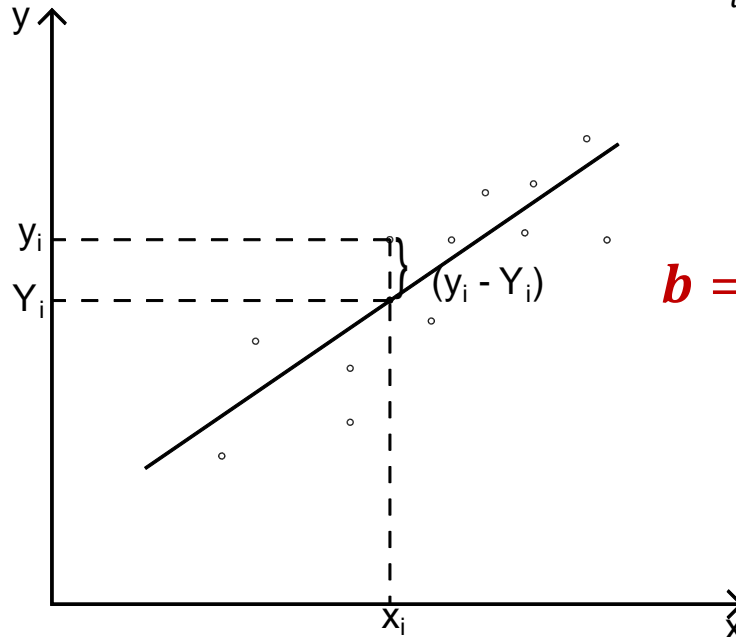


Regresja prostoliniowa

Mamy n par obserwacji (x_i, y_i) . Należy znaleźć liczby a i b będące estymatorami A i B tak, aby zminimalizować sumę kwadratów

$\sum_i (y_i - Y_i)^2$ gdzie Y_i są wartościami wyznaczonymi przez szacowane równanie regresji:

$$Y_i = a + b \cdot x_i$$



y_i - wartość obserwowana
 Y_i - wartość teoretyczna

$$b = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2} = \frac{\sum_i x_i y_i - n\bar{x}\bar{y}}{\sum_i x_i^2 - n\bar{x}^2}$$

$$a = \bar{y} - b\bar{x}$$

Regresja prostoliniowa

Pojęcie regresji związane jest z kształtem zależności $E(y|x)$ od x , pojęcie korelacji związane jest zaś z siłą tej zależności

Współczynnik korelacji

$$\rho = \frac{cov(x, y)}{\sigma_x \sigma_y}$$

$cov(x, y)$ - kowariancja zmiennych x i y

$$cov(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [x - E(x)] [y - E(Y)] f_{xy}(x, y) dx dy$$

σ_x, σ_y - odchylenie standardowe zmiennych x i y

Regresja prostoliniowa

Współczynnik korelacji

$$\rho = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y}$$

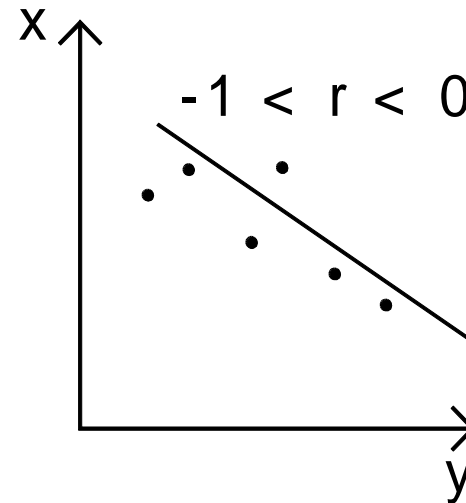
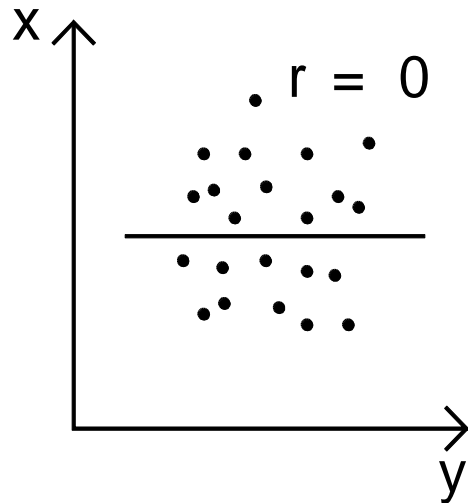
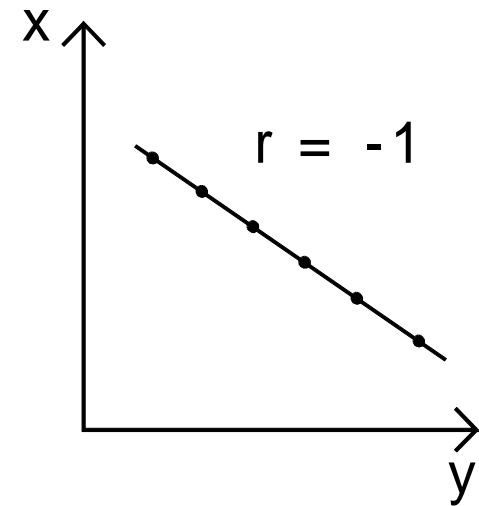
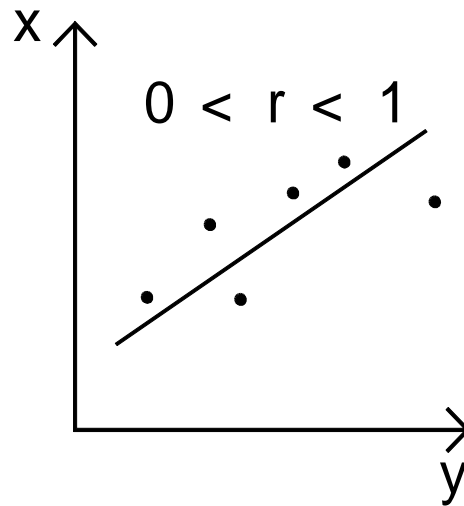
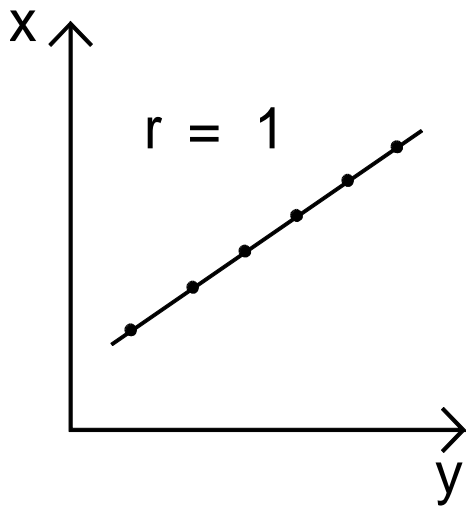
Estymatorem ρ jest współczynnik korelacji Pearsona r

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} = \frac{\sum_i x_i y_i - n\bar{x}\bar{y}}{\sqrt{(\sum_i x_i^2 - n\bar{x}^2)(\sum_i y_i^2 - n\bar{y}^2)}}$$

Poza tym $b = r \frac{S_y}{S_x}$

S_x, S_y -estymatory odchyłeń standardowych

Regresja prostoliniowa



Testowanie parametrów regresji i współczynnika korelacji

$$S^2(b) = \frac{S_o^2}{\sum_i (x_i - \bar{x})^2}$$

$$S^2(a) = S_o^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_i (x_i - \bar{x})^2} \right]$$

$$S_o^2 = \frac{\sum_i (y_i - Y_i)^2}{n - 2} = \frac{\sum_i (y_i - \bar{y})^2 (1 - r)^2}{n - 2}$$

$$\begin{array}{l} H_0: A = 0 \\ H_1: A \neq 0 \\ a \\ t = \frac{a}{\sqrt{S^2(a)}} \end{array}$$

Odrzucamy H_0 , gdy
 $|t| \geq \alpha t_{(n-2)}$

$$\begin{array}{l} H_0: B = 0 \\ H_1: B \neq 0 \\ b \\ t = \frac{b}{\sqrt{S^2(b)}} \end{array}$$

Odrzucamy H_0 , gdy
 $|t| \geq \alpha t_{(n-2)}$

$$\begin{array}{l} H_0: \rho = 0 \\ H_1: \rho \neq 0 \\ r \\ t = \frac{r}{\sqrt{S^2(r)}} \end{array}$$

Odrzucamy H_0 , gdy
 $|t| \geq \alpha t_{(n-2)}$

$$S^2(r) = \frac{1 - r^2}{n - 2}$$

Nie wolno wykorzystać tak obliczonego $S^2(r)$ dla estymacji przedziałowej ρ !

Zadanie predykcji

Dana jest pewna wartość x_o . Chodzi o znalezienie przewidywanej wartości Y zmiennej y odpowiadającej danemu x_o . Najlepszym oszacowaniem wartości przewidywanej jest wartość Y_o wynikająca z prostej regresji, czyli

$$Y_o = a + b \cdot x_o$$

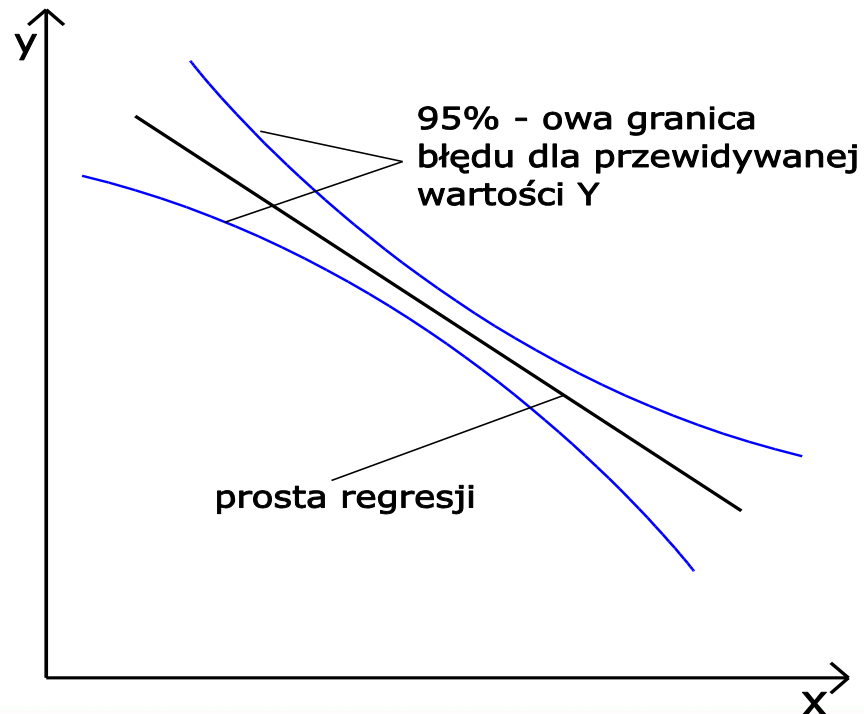
Granice błędu dla Y_o związane z losowością samego y oraz niedokładnością określenia parametrów prostej regresji wyrażają się wzorem:

$$Y_o \pm \alpha t_{(n-2)} \cdot \sqrt{S_o^2} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_o - \bar{x})^2}{\sum_i (x_i - \bar{x})^2}}$$

Zadanie predykcji

Granice błędu dla Y_o związane z losowością samego y oraz niedokładnością określenia parametrów prostej regresji wyrażają się wzorem:

$$Y_o \pm \alpha t_{(n-2)} \cdot \sqrt{S_o^2} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_o - \bar{x})^2}{\sum_i (x_i - \bar{x})^2}}$$



Analiza regresji - przykład

Przykład: Przez dwa tygodnie karmiono szczury dietą ubogą w witaminę D dla wywołania krzywicy. Następnie przez dalsze dwa tygodnie podawano szczurom preparat zawierający witaminę D. Po upływie tego czasu określono stopień wyleczenia przez radiografię prawego kolana każdego zwierzęcia doświadczalnego. Porównano analizowane zdjęcie radiologiczne ze standardowym zestawem fotografii opatrzonej numerami od 0 do 12 (stopniowanie w kierunku wzrastającego wyleczenia). Każdą dawkę preparatu podawano grupie kilku szczurów i późniejsze zdjęcia analizowało kilku radiologów. Zbadać regresję między **logarytmem** dawki (x_i) preparatu a średnim efektem dla każdej dawki (y_i)

dawka	2.5	5	10	20	40	80	160	320
x_i	0,398	0,699	1,000	1,301	1,602	1,903	2,204	2,505
y_i	0,250	1,0833	1,6667	2,8333	3,5833	4,3333	4,9167	5,5555

Analiza regresji - przykład

dawka	2.5	5	10	20	40	80	160	320
x_i	0,398	0,699	1,000	1,301	1,602	1,903	2,204	2,505
y_i	0,250	1,0833	1,6667	2,8333	3,5833	4,3333	4,9167	5,5555

$$\bar{x} = 1,4515$$

$$b = 2,5115$$

$$\bar{y} = 3,00$$

$$a = -0,6454$$

$$n = 8$$

$$r = 0.9943$$

Przykładowo przeprowadzimy test dla r :

$$H_0: \rho = 0$$

$$H_0: \rho \neq 0$$

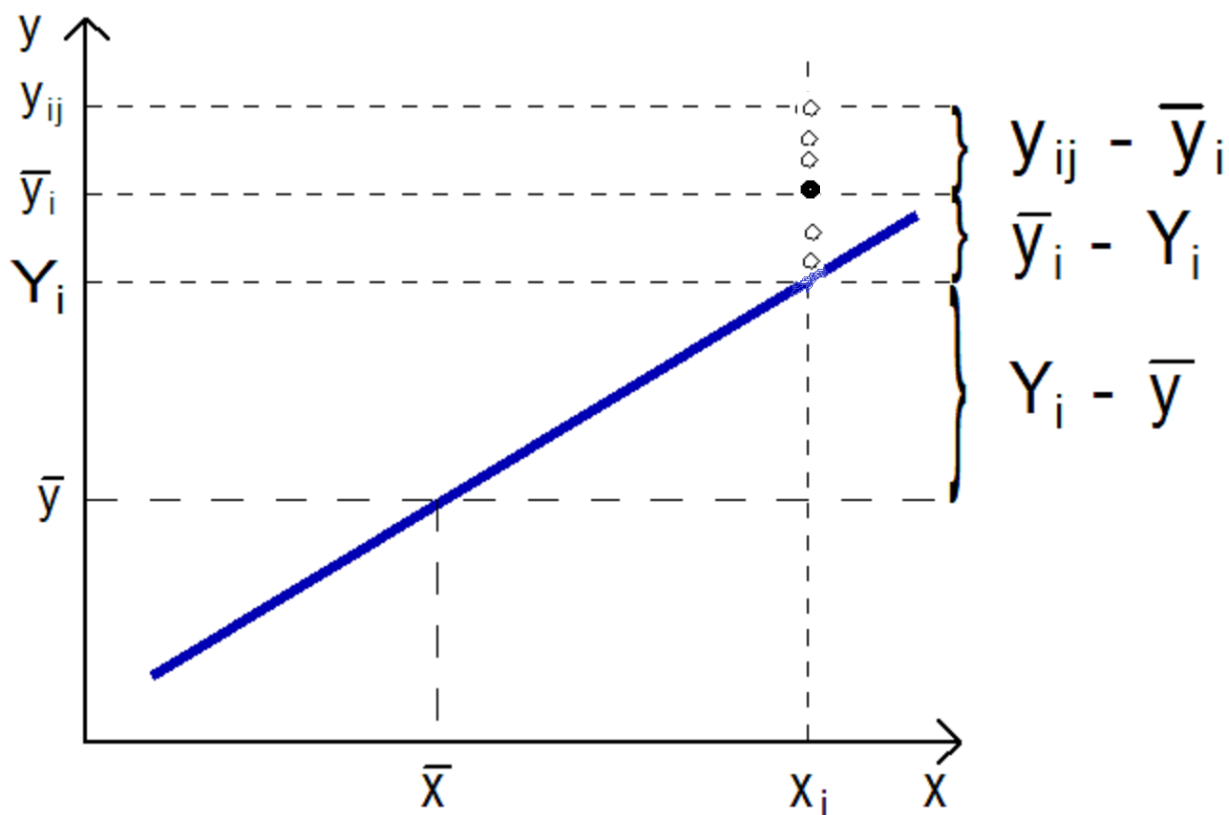
$$t = r \sqrt{\frac{n-2}{1-r^2}} = 22,8296$$

$$t > t_{kryt}$$

$$0,05 t_{(6)} = 2,447$$

H_0 – odrzucamy
przyjmując $\rho \neq 0$

Test na liniowość



$$\sum_{ij} (y_{ij} - \bar{y})^2 = \sum_{ij} (y_{ij} - y_i)^2 + \sum_{ij} (y_i - Y_i)^2 + \sum_{ij} (Y_i - \bar{y})^2$$

Test na liniowość

x_i	x_1	x_2	...	x_i	...	x_k	
y_{ij}	y_{11}	y_{21}		y_{i1}		y_{k1}	
	y_{12}	y_{22}		y_{i2}		y_{k2}	
	
	y_{1n_1}	y_{2n_2}		y_{in_i}		y_{kn_k}	
	T_1	T_2	...	$T_i = \sum_{j=1}^{n_i} y_{ij}$...	T_k	$T = \sum_{i=1}^k T_k$
	s_1	s_2	...	$s_i = \sum_{j=1}^{n_i} y_{ij}^2$...	s_k	$S = \sum_{i=1}^k s_i$
	\bar{y}_1	\bar{y}_2	...	$\bar{y}_i = \frac{T_i}{n_i}$...	\bar{y}_k	$\bar{y} = \frac{T}{N}$
	n_1	n_2	...	n_i	...	n_k	$N = \sum_{i=1}^k n_i$

Test na liniowość

$$\begin{array}{cccc}
 \sum (y_{ij} - \bar{y})^2 & = & \sum (y_{ij} - \bar{y}_i)^2 & + & \sum (\bar{y}_i - Y_i)^2 & + & \sum (Y_i - \bar{y})^2 \\
 \text{SK} & = & \text{SKR} & + & \text{SKL} & + & \text{SKB} \\
 \text{całkowita suma} & & \text{suma kwadratów} & & \text{suma kwadratów} & & \text{suma kwadratów} \\
 \text{kwadratów} & & \text{odchyleń wartości} & & \text{odchyleń średnich} & & \text{odchyleń wartości} \\
 \text{odchyleń wartości} & & \text{obserwacji od} & & \text{serii od wartości} & & \text{teoretycznych od} \\
 \text{obserwacji od} & & \text{średniej serii} & & \text{teoretycznych} & & \text{średniej (świadczy} \\
 \text{średniej} & & \text{(zmiennosc resztowa)} & & \text{wynikających z} & & \text{o nachyleniu} \\
 & & & & \text{prostej regresji} & & \text{prostej regresji-} \\
 & & & & \text{(świadczy o} & & \text{por. współczynnik} \\
 & & & & \text{dopasowaniu} & & \text{B)} \\
 & & & & \text{obserwacji do} & & \\
 & & & & \text{prostej regresji)} & &
 \end{array}$$

Test na liniowość

$$SKR = S - \sum_{i=1}^k \frac{T_i^2}{n_i}$$

$$SK = S - \frac{T^2}{N}$$

$$SKB = \frac{\sum_{i=1}^k x_i T_i - \frac{T_i \sum_{i=1}^k n_i x_i}{N}}{\sum_{i=1}^k n_i x_i^2 - \frac{(\sum_{i=1}^k n_i x_i)^2}{N}}$$

$$SKL = SK - SKR - SKB$$

Test na liniowość

Źródło zmienności	Suma kwadratów	Liczba st. swobody	Średni kwadrat	Stosunek wariancji
Odchylenie wartości teoret. od średniej	SKB	1	$S_1^2 = \frac{SKB}{1}$	$F_1 = \frac{S_1^2}{S_o^2}$
Odchylenie średnich serii od prostej regr.	SKL	$k-2$	$S_2^2 = \frac{SKL}{k-2}$	$F_2 = \frac{S_2^2}{S_o^2}$
Reszta wewnątrz serii	SKR	$N-k$	$S_o^2 = \frac{SKR}{N-k}$	
Ogółem	SK	$N-1$		

Testowane hipotezy:

1) $H_o: B = 0$
 $H_1: B \neq 0$
(stosunek F_1)

2) $H_o: \text{Funkcja regresji jest liniowa}$
 $H_1: \text{Funkcja regresji nie jest liniowa}$
(stosunek F_2)

Test na liniowość

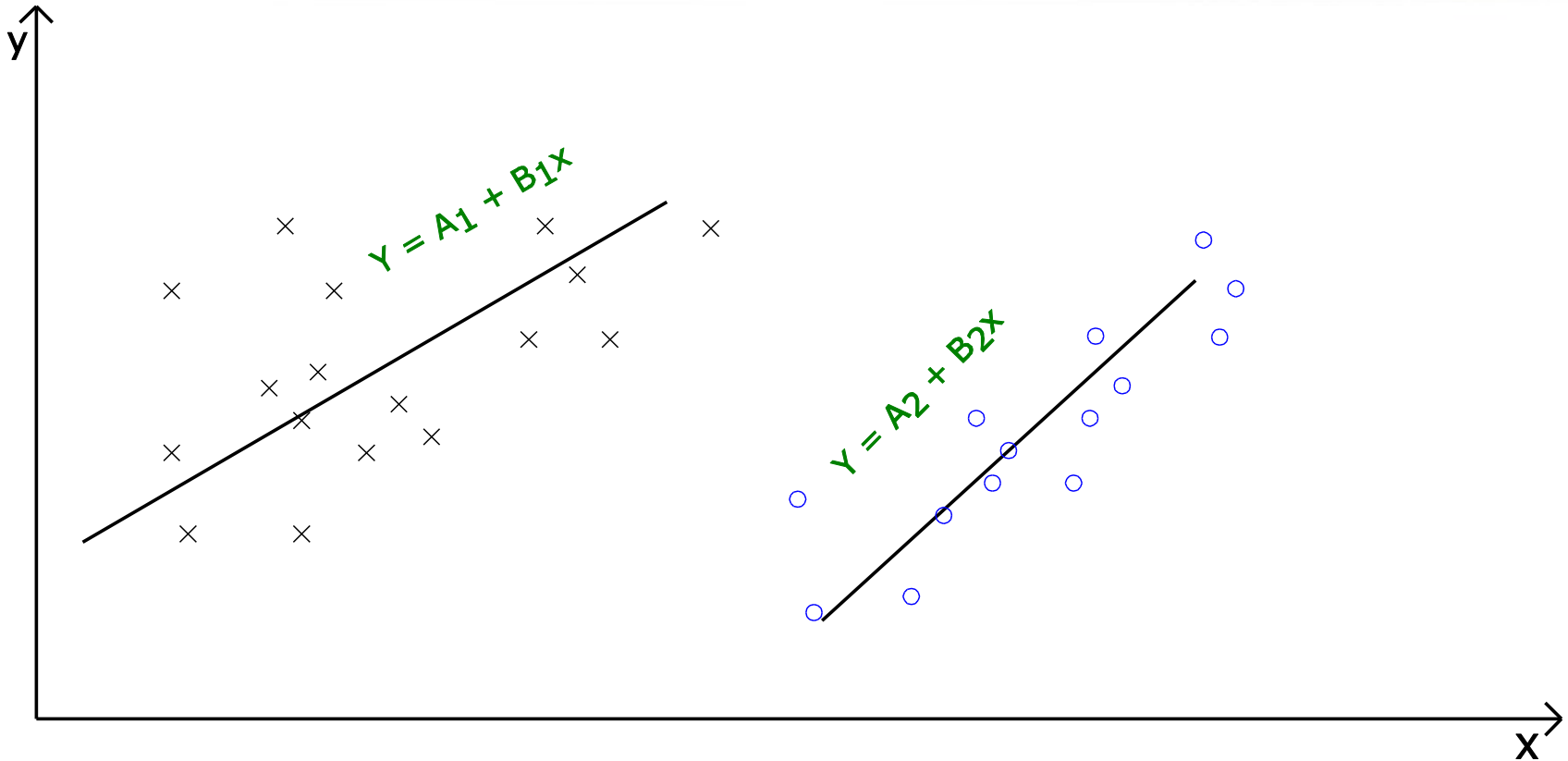
Przykład: W poprzednim przykładzie dopasowaliśmy linię regresji dla zależności wzajemnej logarytmu dawki leku (x_i) i średniego efektu terapeutycznego (y_i). Tutaj zamiast efektu średniego uwzględniamy, że każdą dawkę leku (x_i) podawano sześciu szczurom. Wobec tego dysponujemy sześcioma ocenami efektu leczenia dla każdego x_i

dawka x_i	2,5	5	10	20	40
	0,398	0,699	1,000	1,301	1,602
y_{ij}	0	1,0	1,5	3,0	6,5
	0	1,5	1,0	3,0	3,5
	0	1,5	2,0	5,5	4,5
	0	1,0	3,5	2,5	3,5
	0	1,0	2,0	1,0	3,5
	0,5	0,5	0	2,0	3,0

Test na liniowość

Źródło zmienności	Suma kwadratów	Liczba st. swobody	Średni kwadrat	Stosunek wariancji	Istotność
Odchyl. wart. teoretycznych od średniej	57,0375	1	57,0375	52,248	P<0,005
Odchyl. śr. efektu dawki od prostej	0,8458	3	0,2819	0,258	nieistotne
Reszta wewnątrz dawki	27,2917	25	1,0971		
Ogółem	85,1750	29			

Regresja w dwóch grupach



- (a) Czy można uważać, że obie proste regresji są równoległe?
- (b) Jeśli tak, to czy można uważać, że proste te pokrywają się?

Regresja w dwóch grupach

n_1 - liczebność par obserwacji (x_{i1}, y_{i1}) w grupie 1

n_2 - liczebność par obserwacji (x_{i2}, y_{i2}) w grupie 2

$\bar{x}_1, \bar{y}_1, \bar{x}_2, \bar{y}_2$ - odpowiednie średnie w grupach 1 i 2

Estymowane proste regresji:

$$Y_{i1} = a_1 + b_1 \cdot x_{i1}$$

$$Y_{i2} = a_2 + b_2 \cdot x_{i2}$$

$$s_o^2 = \frac{\sum_{i=1}^{n_1} (y_{i1} - Y_{i1})^2 + \sum_{i=1}^{n_2} (y_{i2} - Y_{i2})^2}{n_1 + n_2 - 4}$$

$$s^2(b_1 - b_2) = s_o^2 \left[\frac{1}{\sum_{i=1}^{n_1} (x_{i1} - \bar{x}_1)^2} + \frac{1}{\sum_{i=1}^{n_2} (x_{i2} - \bar{x}_2)^2} \right]$$

Regresja w dwóch grupach

$$Y_{i1} = a_1 + b_1 \cdot x_{i1}$$

$$Y_{i2} = a_2 + b_2 \cdot x_{i2}$$

$$s_o^2 = \frac{\sum_{i=1}^{n_1} (y_{i1} - Y_{i1})^2 + \sum_{i=1}^{n_2} (y_{i2} - Y_{i2})^2}{n_1 + n_2 - 4}$$

$$s^2(b_1 - b_2) = s_o^2 \left[\frac{1}{\sum_{i=1}^{n_1} (x_{i1} - \bar{x}_1)^2} + \frac{1}{\sum_{i=1}^{n_2} (x_{i2} - \bar{x}_2)^2} \right]$$

Badamy, czy można uważać obie proste regresji za równoległe

$$(a) \quad \begin{array}{l} H_o: B_1 = B_2 \\ H_1: B_1 \neq B_2 \end{array} \quad t = \frac{b_1 - b_2}{\sqrt{s^2(b_1 - b_2)}}$$

H_o odrzucamy, gdy $|t| \geq \alpha t_{(n_1+n_2-4)}$

Regresja w dwóch grupach

Jeżeli nie znajdziemy podstaw do odrzucenia H_0 , to obliczamy estymowaną wartość wspólnego współczynnika nachylenia równoległych prostych regresji:

$$b = \frac{\sum_{i=1}^{n_1} (x_{i1} - \bar{x}_1)(y_{i1} - \bar{y}_1) + \sum_{i=1}^{n_2} (x_{i2} - \bar{x}_2)(y_{i2} - \bar{y}_2)}{\sum_{i=1}^{n_1} (x_{i1} - \bar{x}_1)^2 + \sum_{i=1}^{n_2} (x_{i2} - \bar{x}_2)^2}$$

Granice przedziału ufności dla b można oszacować jako

$$b \pm \alpha t_{(n_1+n_2-4)} \sqrt{s^2(b)}$$

gdzie:

$$s^2(b) = \frac{s_o^2}{\sum_{i=1}^{n_1} (x_{i1} - \bar{x}_1)^2 + \sum_{i=1}^{n_2} (x_{i2} - \bar{x}_2)^2}$$

Regresja w dwóch grupach

Przykład: Przeprowadzono porównanie dwu preparatów przeciwnowiciowych:
 pierwszego z poprzedniego przykładu oraz drugiego:

dawka	2,5	5	10	20	40
x_{i1}	0,398	0,699	1,000	1,301	1,602
y_{i1}	0	1,0	1,5	3,0	6,5
	0	1,5	1,0	3,0	3,5
	0	1,5	2,0	5,5	4,5
	0	1,0	3,5	2,5	3,5
	0	1,0	2,0	1,0	3,5
	0,5	0,5	0	2,0	3,0

dawka	2,5	5	10	20
x_{i2}	0,398	0,699	1,000	1,301
y_{i2}	3,25	2,50	3,75	5,0
	3,0	2,75	5,25	7,0
	1,75	2,25	6,0	3,0
	2,0	2,25	6,5	3,0
	0,5	3,75	2,25	6,0
	1,5		3,5	

Regresja w dwóch grupach

$$s_o^2 = 1,2384$$

$$s^2(b_1 - b_2) = 0,7257$$

$$b_1 = 3,2392 \quad b_2 = 3,3663$$

$$t = \frac{3,2302 - 3,3663}{\sqrt{0,7257}} = -0,1492$$

$$0,05t_{(48)} = 2,013$$

Nie ma podstaw do odrzucenia H_o w tym przypadku, więc proste regresji dla obu preparatów są równoległe, wspólny współczynnik nachylenia obu równoległych prostych regresji będzie równy:

$$b = 3,2791$$

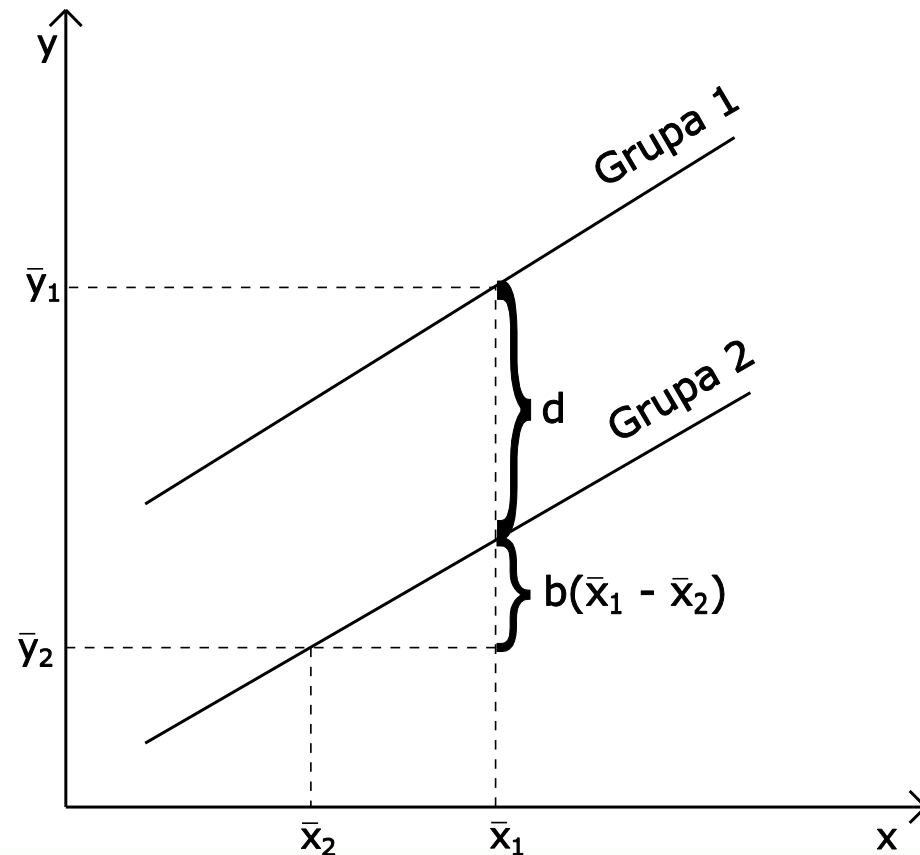
$$s^2(b) = 0,1563$$

więc

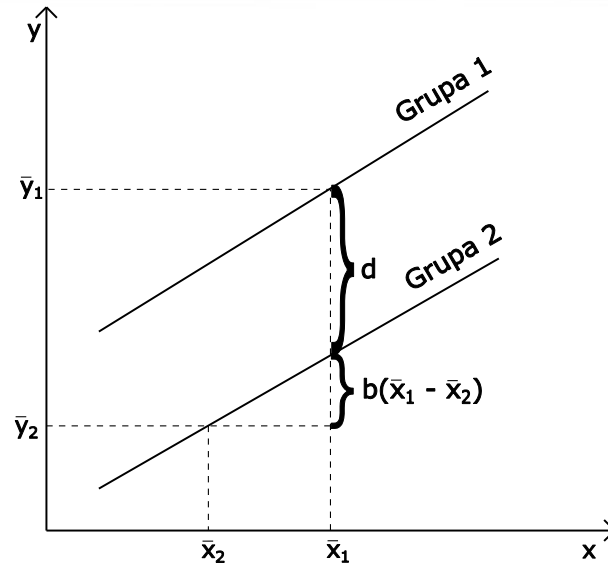
$$B = 3,2791 \pm 0,7958$$

Regresja w dwóch grupach

(b) Gdy nie znaleźliśmy podstaw do odrzucenia hipotezy o równoległości dwu prostych regresji (na czym w istocie nam zależało), pytamy czy te proste równoległe pokrywają się, czy też nie, a jeżeli nie, to jaka jest ich „pionowa odległość” wzajemna.



Regresja w dwóch grupach



Obliczamy więc:

$$d = \bar{y}_1 - \bar{y}_2 - b(\bar{x}_1 - \bar{x}_2)$$

$$s^2(d) = s_c^2 \left[\frac{1}{n_1} + \frac{1}{n_2} + \frac{(\bar{x}_1 - \bar{x}_2)^2}{\sum_{i=1}^{n_1} (x_{i1} - \bar{x}_1)^2 + \sum_{i=1}^{n_2} (x_{i2} - \bar{x}_2)^2} \right]$$

$$s_c^2 = \frac{\sum_{i=1}^{n_1} (y_{i1} - \bar{y}_1)^2 + \sum_{i=1}^{n_2} (y_{i2} - \bar{y}_2)^2 - \frac{[\sum_{i=1}^{n_1} (x_{i1} - \bar{x}_1) \cdot (y_{i1} - \bar{y}_1) + \sum_{i=1}^{n_2} (x_{i2} - \bar{x}_2) \cdot (y_{i2} - \bar{y}_2)]}{\sum_{i=1}^{n_1} (x_{i1} - \bar{x}_1)^2 + \sum_{i=1}^{n_2} (x_{i2} - \bar{x}_2)^2}}{n_1 + n_2 - 3}$$

Regresja w dwóch grupach

Badamy, czy równoległe proste regresji nie pokrywają się

$$H_0: d = 0$$

$$H_1: d \neq 0$$

$$t = \frac{d}{\sqrt{s^2(d)}}$$

Odrzucamy H_0 , gdy $|t| \geq t_{\alpha}(n_1+n_2-3)$

Przedział ufności dla D

$$d \pm t_{\alpha}(n_1+n_2-3)\sqrt{s^2(d)}$$

Regresja w dwóch grupach

Przykład c.d. Obliczamy i szacujemy pionową odległość pomiędzy dwoma równoległymi prostymi regresji badanymi poprzednio dla preparatów przeciwwkrzywicznych 1 i 2.

$$d = -2,0315$$

$$s_c^2 = 1,2137$$

$$s^2(d) = 0,0998$$

$$t = \frac{-2,0315}{\sqrt{0,0998}} = -6,4323$$

$$0,05t_{(49)} = 2,012$$

Ponieważ $|t| \geq 0,05t_{(49)}$ hipotezę o pokrywaniu się prostych regresji odrzucamy. Szacując odległość pionową d mamy:

$$D = -2,0315 \pm 0,6355$$

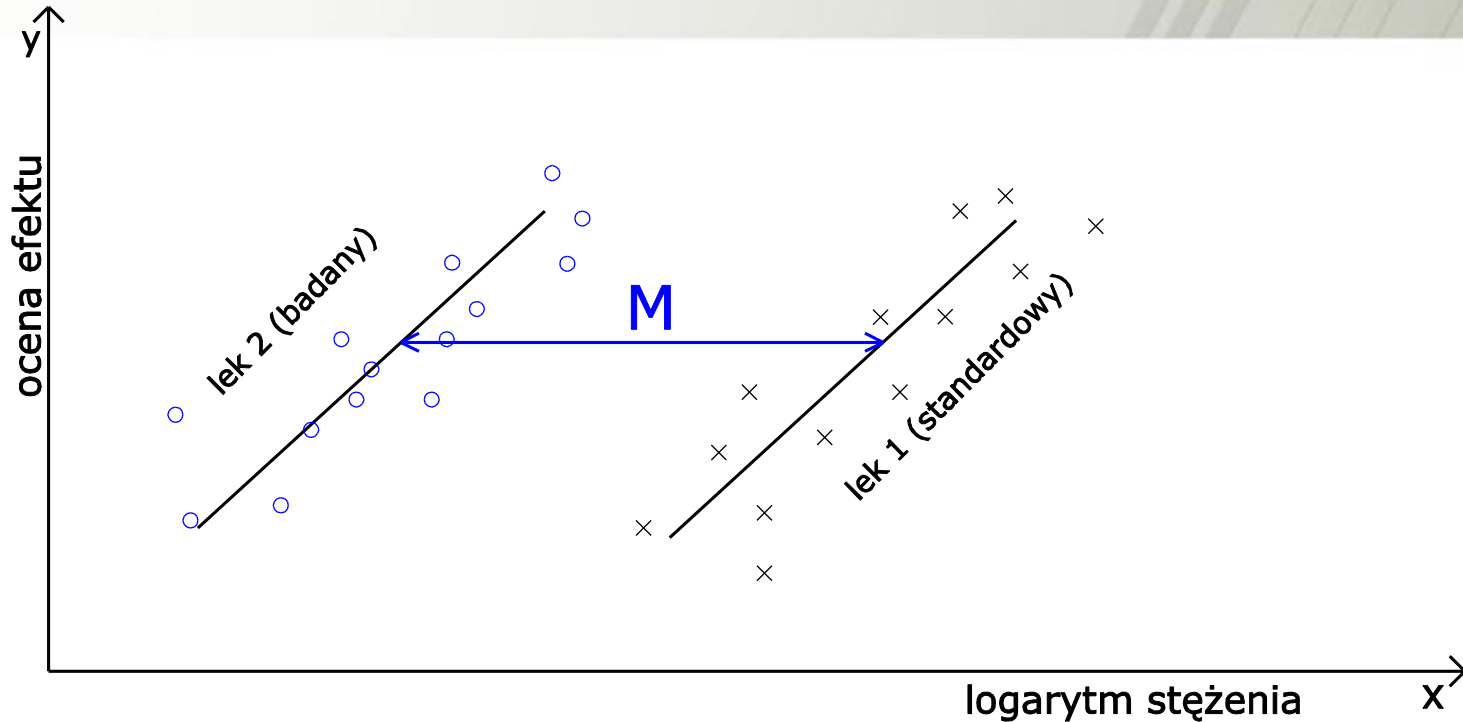
Regresja w dwóch grupach

(c) Jeżeli stwierdziliśmy, że zależność między efektami a logarytmem ich stężeń:

- są prostoliniowe,
- linie regresji obu leków można uznać za równoległe,
- proste regresji nie pokrywają się,

to szczególnie interesujące będzie wyznaczenie i oszacowanie „poziomej odległości” pomiędzy prostymi regresji dla obu leków, gdyż ma ona oczywistą interpretację logarytmu „stosunku mocy” porównywanych preparatów oraz eliminuje ewentualne nieliniowości, jeśli ich charakter jest jednolity dla obu linii regresji

Regresja w dwóch grupach



$$M = \log R$$

R - stosunek mocy leku 2 w porównaniu do leku 1 (iloraz stężenia leku badanego do stężenia leku standardowego dającego ten sam efekt leczenia – w sytuacji, jak na rysunku, jest to ułamek < 1)

Regresja w dwóch grupach

$$M = \bar{x}_1 - \bar{x}_2 - \frac{\bar{y}_1 - \bar{y}_2}{b}$$

$$s^2(M) = \frac{s_c^2}{b^2} \left[\frac{1}{n_1} + \frac{1}{n_2} + \frac{(M - \bar{x}_1 + \bar{x}_2)^2}{\sum_{i=1}^{n_1} (x_{i1} - \bar{x}_1)^2 + \sum_{i=1}^{n_2} (x_{i1} - \bar{x}_2)^2} \right]$$

$$g = \frac{\alpha t_{(n_1+n_2-3)}^2 \cdot s_c^2}{b^2 \left[\sum_{i=1}^{n_1} (x_{i1} - \bar{x}_1)^2 + \sum_{i=1}^{n_2} (x_{i1} - \bar{x}_2)^2 \right]}$$

Regresja w dwóch grupach

Gdy $g < 0,1$, to można oszacować M z zależności:

$$M \pm \alpha t_{(n_1+n_2-3)} \sqrt{s^2(M)}$$

Gdy $0,1 \leq g < 1$ wartość M szacujemy z zależności

$$M \pm \frac{\alpha t_{(n_1+n_2-3)} \cdot \sqrt{s_c^2}}{b} \cdot \frac{\sqrt{(1-g) \left(\frac{1}{n_1} + \frac{1}{n_2} \right) + \frac{(M - \bar{x}_1 + \bar{x}_2)^2}{\sum_{i=1}^{n_1} (x_{i1} - \bar{x}_1)^2 + \sum_{i=1}^{n_2} (x_{i2} - \bar{x}_2)^2}}}{1-g}$$

Regresja w dwóch grupach

Przykład c.d. Badamy odległość poziomą równoległych prostych regresji dla dwóch rozważanych poprzednio preparatów przeciwnowiciwicznych.

$$s_c^2 = 1,2137$$

$$b = 3,2791$$

$$g = 0,0577$$

$$\alpha = 0,05$$

$$0,05t_{(49)} = 2,012$$

$$M = -0,6195$$

$g < 0,1$ więc obliczamy

$$s^2(M) = 0,01185$$

$$t = 5,692$$

$$M = -0,6195 \pm 0,2190$$

$|t| > 0,05t_{(49)} \rightarrow$ odrzucamy

$$H_0: M = 0$$

Regresja w dwóch grupach

$$M = -0,6195 \pm 0,2190$$

Czyli granice przedziału ufności dla M wynoszą:

$$-0,8385 \text{ i } -0,4005$$

Ponieważ $M = -0,6195 = \log 0,240$ więc $R = 0,240$. Przedział ufności dla R jest określony jako:

$$-0,4005 = \log 0,398$$

$$-0,6195 = \log 0,240$$

$$-0,8385 = \log 0,145$$

$$0,145 < R < 0,398$$