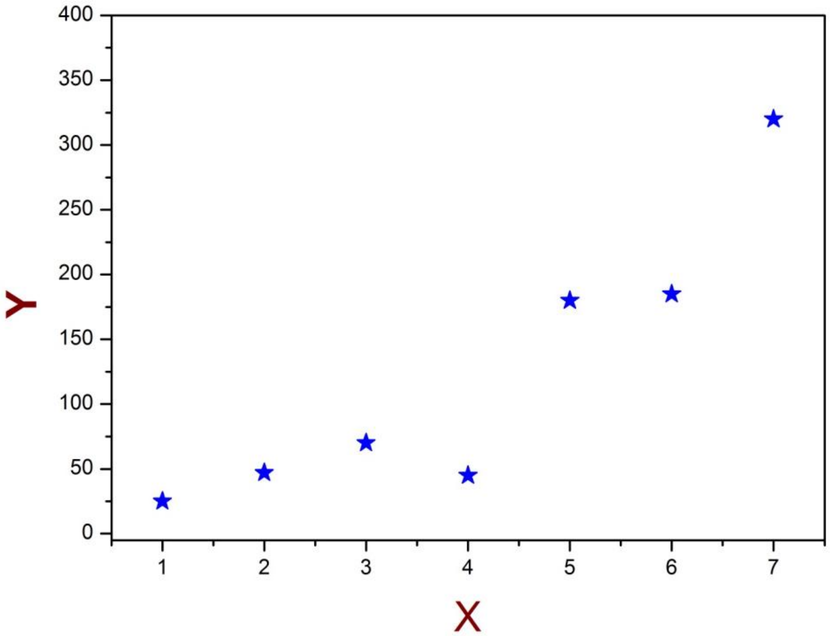
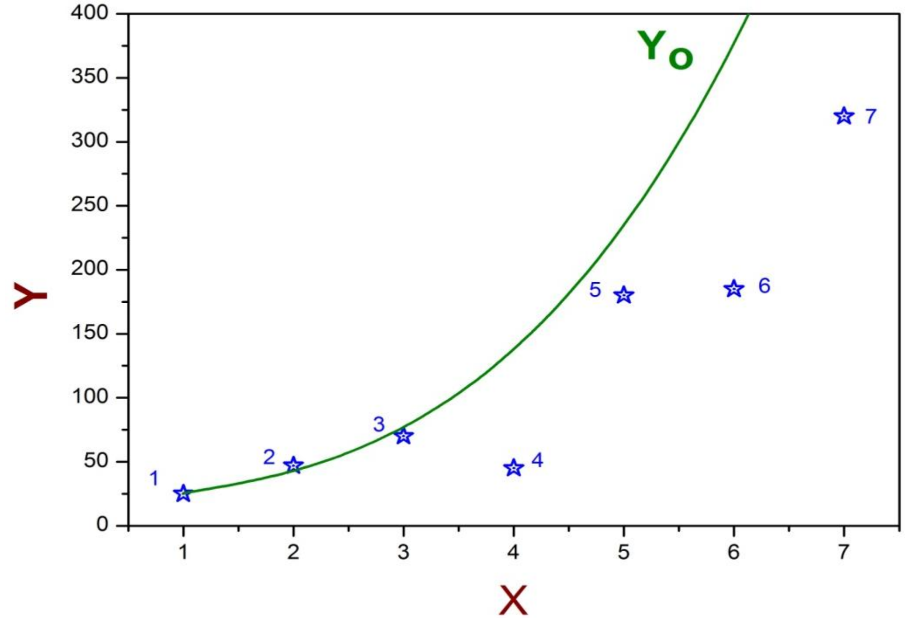
WYKŁAD 8
ANALIZA REGRESJI

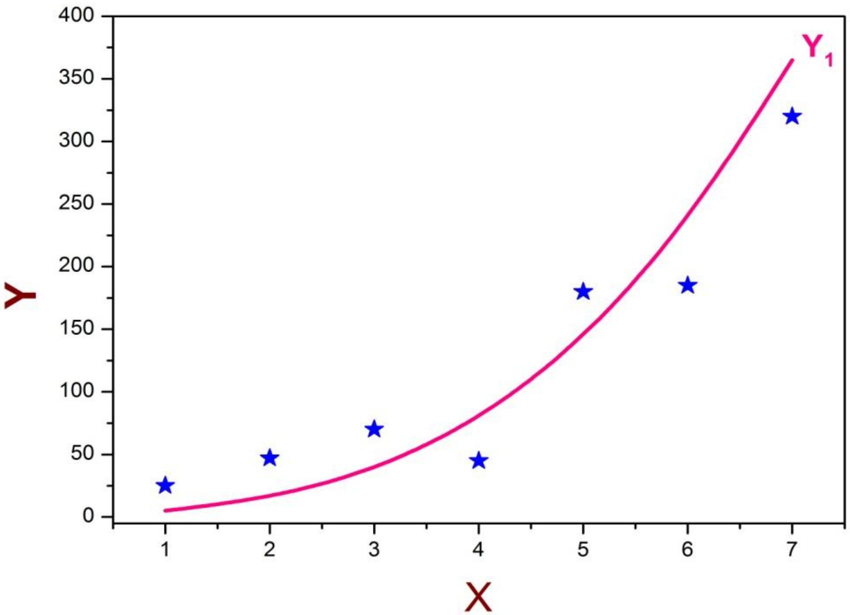
Regresja

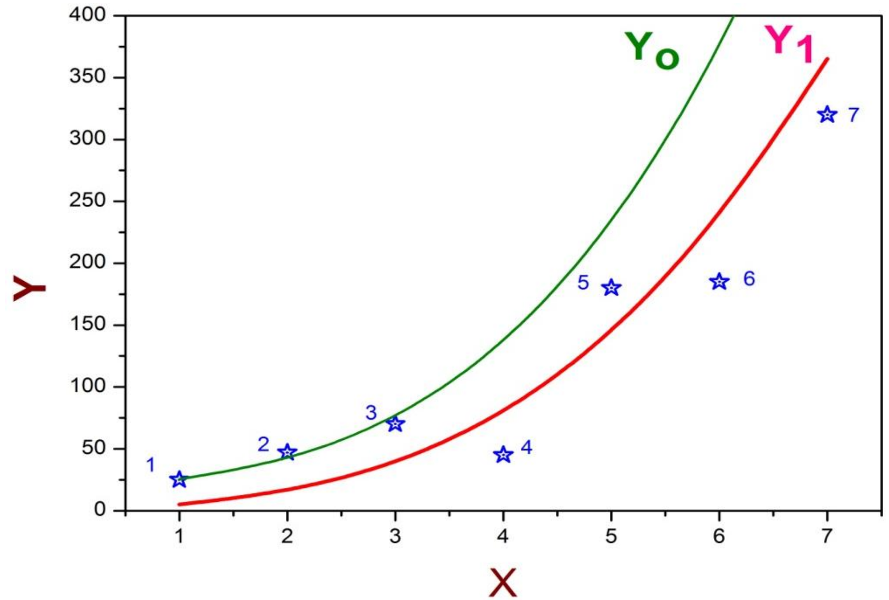
1. Metoda najmniejszych kwadratów-regresja prostoliniowa
2. Regresja krzywoliniowa
3. Estymacja liniowej funkcji regresji
4. Testy istotności współczynnika regresji liniowej
5. Zamiana przypadków nieliniowych na liniowe

METODA NAJMNIEJSZYCH KWADRATÓW

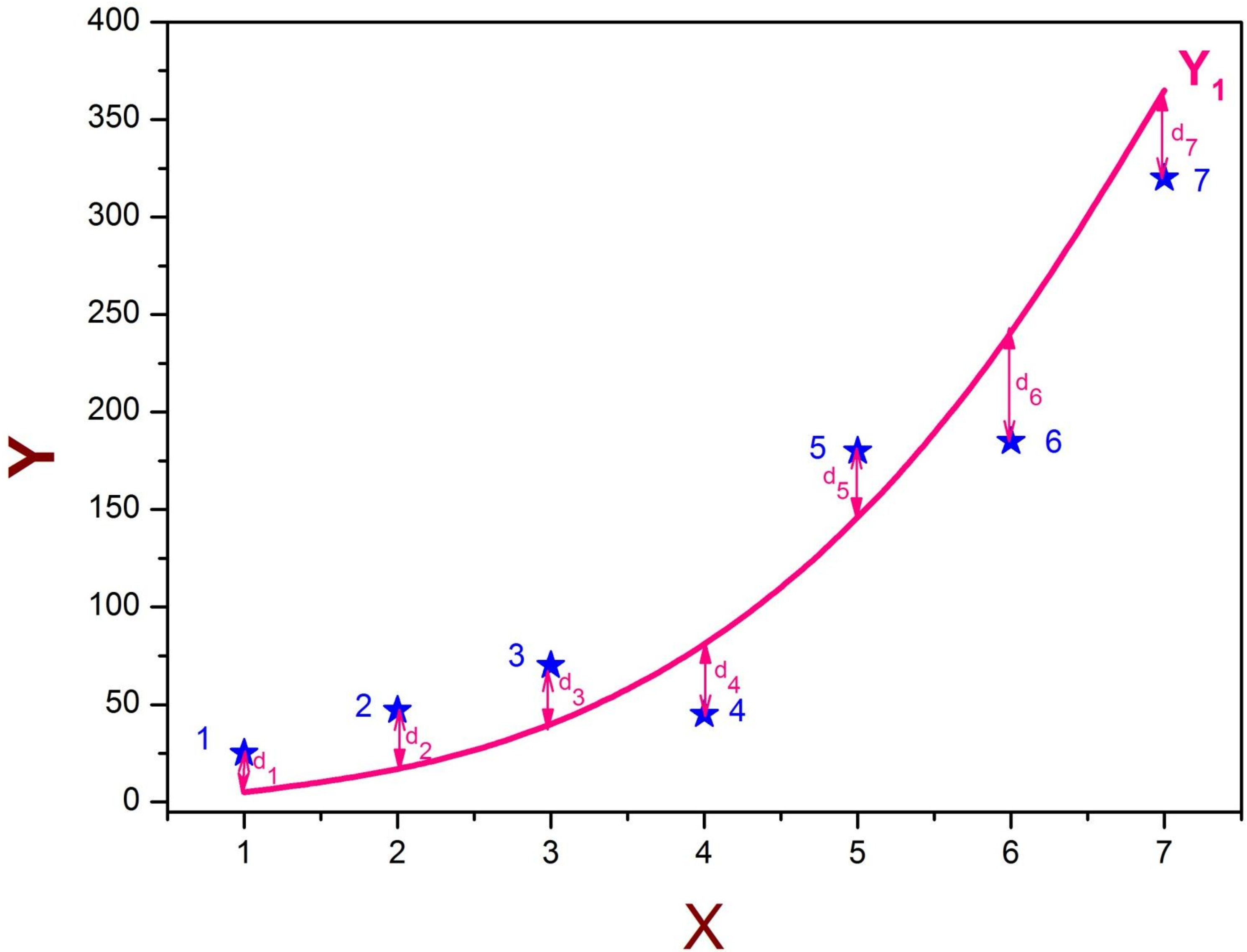




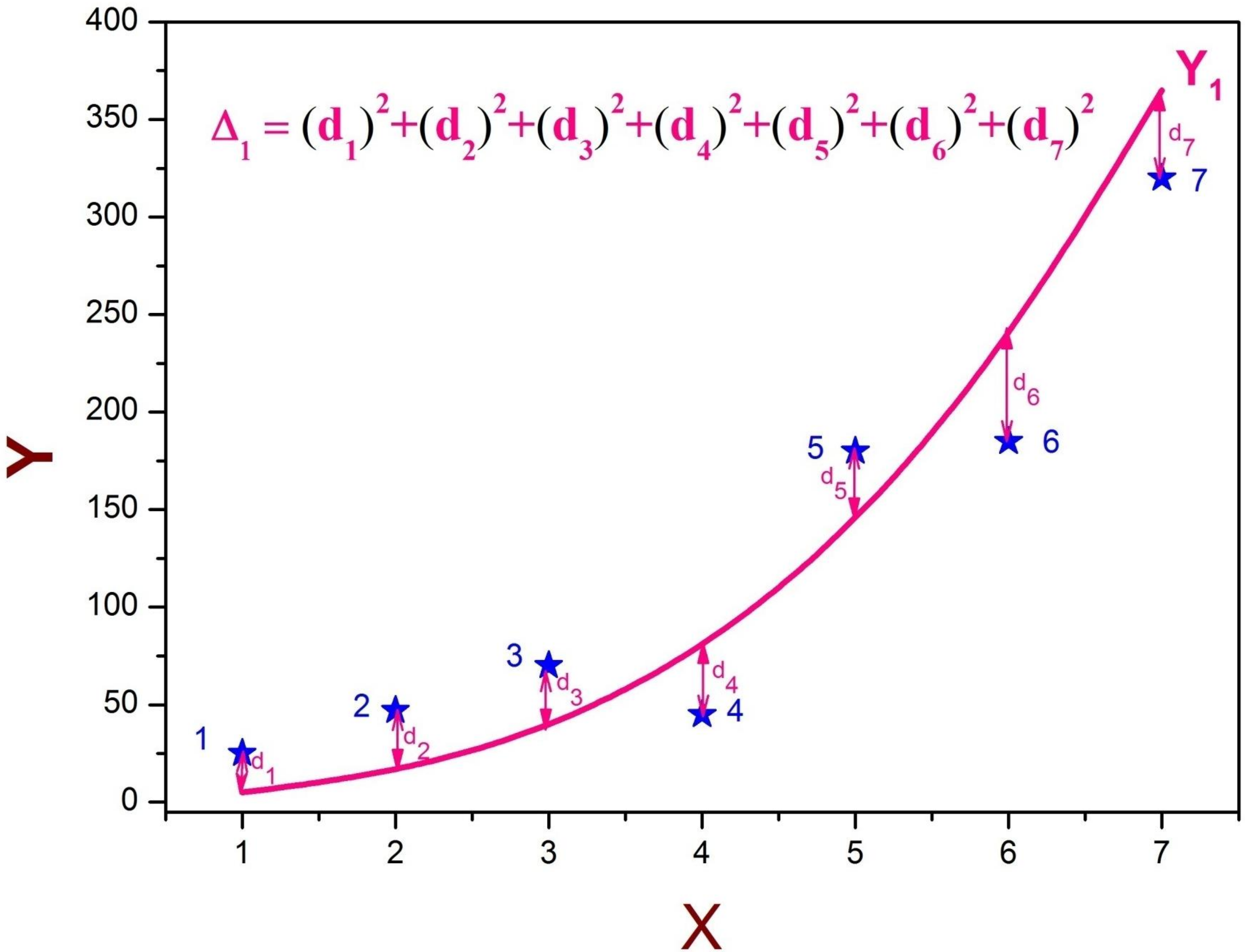




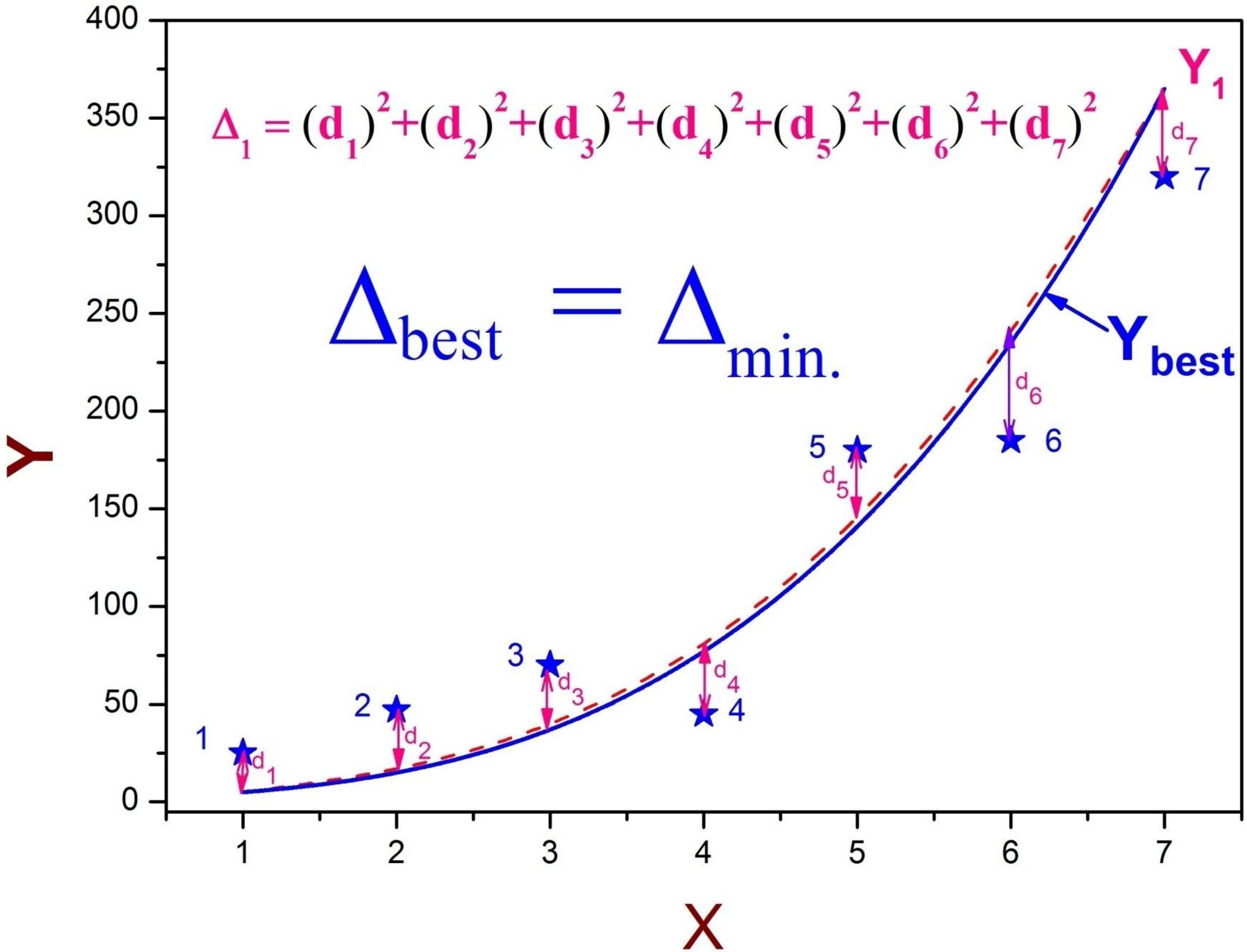
METODA NAJMNIEJSZYCH KWADRATÓW



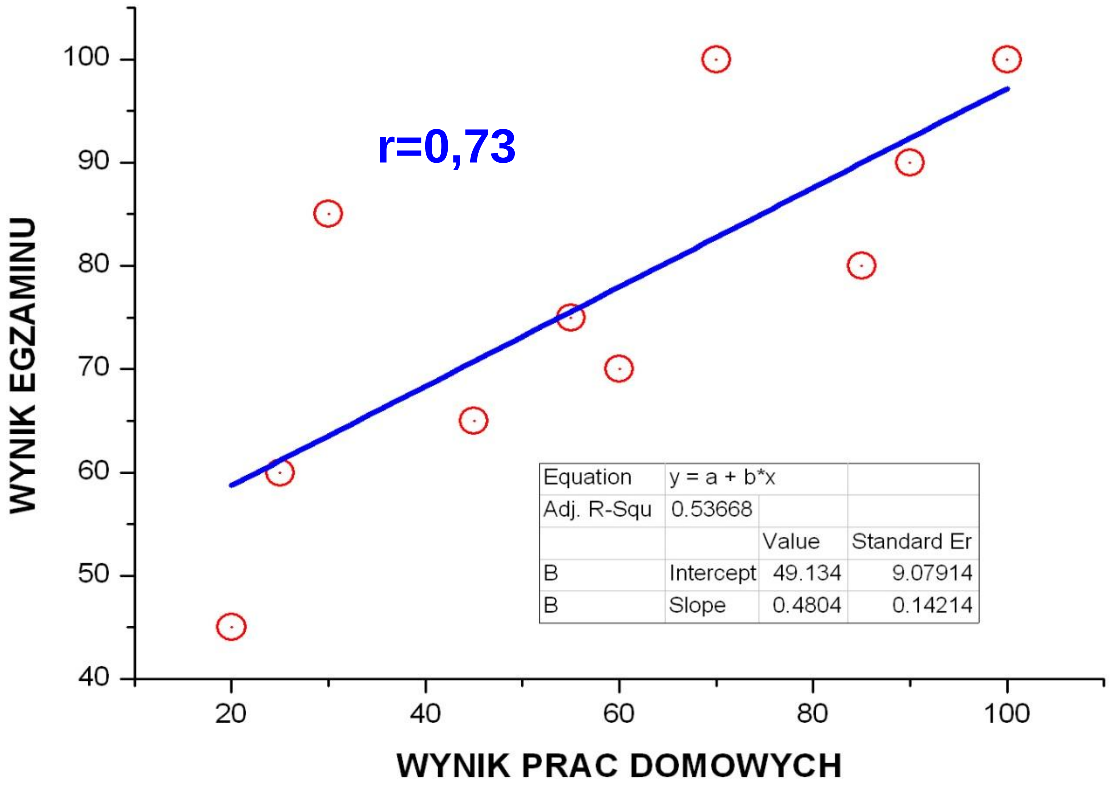
METODA NAJMNIEJSZYCH KWADRATÓW



METODA NAJMNIEJSZYCH KWADRATÓW



WSPÓŁCZYNNIK KORELACJI LINIOWEJ



METODA NAJMNIESZYCH KWADRATÓW

$$y=A+Bx \qquad (1)$$

Zakładamy, że: a) niepewność w określaniu x jest zaniedbywalna ; b) każdy wynik pomiaru y_i podlega rozkładowi normalnemu; c) niepewności wszystkich wartości y mają tę samą wartość: σ_y .

Prawdziwa wartość $y_i = A+Bx_i$ (2)

Wynik pomiaru y_i podlega rozkładowi normalnemu wyśrodkowanemu wokół swojej prawdziwej wartości z odchyleniem σ_y, więc prawdopodobieństwo zmierzonej wartości y_i wynosi:

$$P_{A,B}(y_i) \propto \frac{1}{\sigma_y} \exp\left[-\frac{(y_i-A-Bx_i)^2}{2\sigma_y^2}\right] \qquad (3)$$

Prawdopodobieństwo otrzymania kompletnego zbioru wyników $y_1, \ldots, y_N$ jest iloczynem:

$$P_{A,B}(y_1,\ldots,y_N) = P_{A,B}(y_1)\cdots P_{A,B}(y_N) \propto \frac{1}{\sigma_y^N} \exp\left[-\frac{\chi^2}{2}\right] \qquad (4)$$

gdzie:

$$\chi^2 = \sum_{i=1}^{N} \frac{(y_i-A-Bx_i)^2}{\sigma_y^2} \qquad (5)$$

METODA NAJMNIESZYCH KWADRATÓW

Najlepsze przybliżenie nieznanych stałych A, B uzyskamy, gdy prawdopodobieństwo $P_{A,B}(y_1, \ldots, y_N)$ jest największe, co odpowiada:

$$\chi^2 = min:$$

$$\frac{\partial \chi^2}{\partial A} = -\frac{2}{\sigma_y^2}\sum_{i=1}^{N}(y_i - A - Bx_i) = 0 \qquad (6)$$

$$\frac{\partial \chi^2}{\partial B} = -\frac{2}{\sigma_y^2}\sum_{i=1}^{N}x_i(y_i - A - Bx_i) = 0 \qquad (7)$$

stąd otrzymujemy tzw. *równania normalne* o niewiadomych A i B:

$$AN + B\sum x_i = \sum y_i \qquad (8)$$

$$A\sum x_i + B\sum x_i^2 = \sum x_i y_i \qquad (9)$$

rozwiązanie tego układu równań daje:

METODA NAJMNIESZYCH KWADRATÓW

$$A = \frac{(\sum x_i^2)(\sum y_i) - (\sum x_i)(\sum x_i y_i)}{\Delta} \qquad (10)$$

$$B = \frac{N(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{\Delta} \qquad (11)$$

gdzie:

$$\Delta = N(\sum x_i^2) - (\sum x_i)^2 \qquad (12)$$

lub, w równoważnej formie:

$$B = \frac{\sum_{i=1}^{n} x_i y_i - N\overline{x}\,\overline{y}}{\sum_{i=1}^{N} x_i^2 - N\overline{x}^2} \qquad (11')$$

$$A = \overline{y} - B\overline{x} \qquad (10')$$

METODA NAJMNIESZYCH KWADRATÓW

$$P_{A,B}(y_1, \dots, y_N) = P_{A,B}(y_1) \cdots P_{A,B}(y_N) \propto \frac{1}{\sigma_y^N} \exp\left[-\frac{\chi^2}{2}\right] \qquad (4)$$

Różniczkując równanie (4) względem σ_y, a następnie przyrównując uzyskaną pochodną do zera, ponadto uwzględniając fakt, że liczba stopni swobody wynosi w tym przypadku N-2 (N- niezależnych pomiarów minus dwa wyznaczone parametry: A i B) , uzyskujemy następujące wyrażenie na niepewność pomiarów y:

$$\sigma_y^2 = \frac{1}{N-2} \sum_{i=1}^{N} (y_i - A - Bx_i)^2 \qquad (13)$$

METODA NAJMNIESZYCH KWADRATÓW

DEFINICJA

W obliczeniach statystycznych **LICZBA STOPNI SWOBODY** wynosi: liczba **niezależnych pomiarów (N) minus liczba parametrów obliczonych z tych pomiarów**

z prawa przenoszenia błędów otrzymujemy:

$$\sigma_A^2 = \sigma_y^2 \sum \frac{x_i^2}{\Delta} \qquad (14)$$

$$\sigma_B^2 = N \frac{\sigma_y^2}{\Delta} \qquad (15)$$

gdzie Δ dane jest równaniem (12) $\Delta = N(\sum x_i^2) - (\sum x_i)$ (12)

METODA NAJMNIEJSZYCH KWADRATÓW DLA INNYCH ZALEŻNOŚCI FUNKCYJNYCH

Analogicznie można postępować w dopasowaniu innych krzywych przedstawioną metodą najmniejszych kwadratów np. w przypadku wielomianu:

$$y = A + Bx + Cx^2 + \cdots + Zx^n \qquad (16)$$

uzyskujemy następującą sumę kwadratów:

$$\chi^2 = \sum_{i=1}^{N} \frac{(y_i - A - Bx_i - Cx_i^2 - \cdots - Zx_i^n)^2}{\sigma_y^2} \qquad (17)$$

Następnie obliczamy pochodne cząstkowe względem A,B,C,…,Z przyrównujemy je do zera, otrzymując układ równań normalnych na parametry A, B,C,…Z

Podobnie postępujemy w przypadku funkcji wielu zmiennych np. z=A+Bx+Cy (regresja wielokrotna)

METODA NAJMNIEJSZYCH KWADRATÓW (ogólna zasada) (1)

Dana jest funkcja:

$$y = f(x; a_o, a_1, \dots, a_m) \qquad (1)$$

zmiennej niezależne x oraz m+1 parametrów $a_o, a_1, \dots, a_m$. Parametry te są stałe, które należy wyznaczyć. W tym celu przeprowadza się n pomiarów x i y otrzymując n par (x_i, y_i) i=1,2,...,n. Wstawiając wartości (x_i , y_i) do (1) otrzymujemy:

$$y_i = f(x_i; a_o, a_1, \dots, a_m) \qquad i = 1, 2, \dots, n \qquad (2)$$

Jeśliby wartości x i y można było wyznaczyć bardzo dokładnie, to do znalezienia m+1 parametrów wystarczyłoby przeprowadzić m+1 pomiarów. W rzeczywistości wartości x i y są obarczone błędami, co nie pozwala na dokładne wyznaczenie prawdziwych wartości tych parametrów. Z reguły przeprowadza się większą liczbę pomiarów czyli n> m+1. W tym przypadku układ równań (2) jest sprzeczny.

METODA NAJMNIEJSZYCH KWADRATÓW (ogólna zasada) (2)

Zadaniem naszym jest dobór takich parametrów a_o, a_1,....,a_m aby równania (2) były spełnione w możliwie najlepszym sposobie (czyli chcemy określić najbardziej prawdopodobne wartości tych parametrów). Ponieważ. równania (2) nie są ściśle spełnione, więc:

$$y_i - f(x_i; a_o, a_1,\ldots,a_m) = \varepsilon_i \qquad i = 1, 2,\ldots, n \qquad (3)$$

gdzie ε_i są odchyleniami mierzonych wartości od obliczonych ze wzoru (1).

Zasada najmniejszych kwadratów:
Najbardziej prawdopodobne parametry są takie, dla których suma kwadratów odchyleń ε_i jest najmniejsza:

$$\sum_{i=1}^{n}\left[y_i - f(x_i; a_o, a_1, \ldots, a_n)\right]^2 = \Delta = \min \qquad (4)$$

METODA NAJMNIEJSZYCH KWADRATÓW

(ogólna zasada) (3)

Przypadek liniowy:

$$f(x: a_o, a_1, \ldots, a_{m)} = \phi_o(x)*a_o + \phi_1(x)*a_1 + \ldots + \phi_m(x)*a_m \qquad (5)$$

Korzystając z warunku koniecznego na minimum Δ ($\frac{\partial \Delta}{\partial a_j} = 0 \quad dla\ j = 0, 1, \ldots, m$)
po wstawieniu (5) do (4) mamy:

$$2\sum_{i=1}^{n}[y_i - \varphi_o(x_i)a_o - \varphi_1(x_i)a_1 - \cdots - \varphi_m(x_i)a_m](-\varphi_o(x_i) = 0$$

$$2\sum_{i=1}^{n}[y_i - \varphi_o(x_i)a_o - \varphi_1(x_i)a_1 - \cdots - \varphi_m(x_i)a_m](-\varphi_1(x_i) = 0 \qquad (6)$$

. . . .

$$2\sum_{i=1}^{n}[y_i - \varphi_o(x_i)a_o - \varphi_1(x_i)a_1 - \cdots - \varphi_m(x_i)a_m](-\varphi_m(x_i) = 0$$

METODA NAJMNIEJSZYCH KWADRATÓW
(ogólna zasada) (4)

Wprowadzając oznaczenia:

$$\sum_{i=1}^{n} y_i \varphi_r(x_i) = [y, \varphi_r] \qquad \sum_{i=1}^{n} \varphi_r^2(x_i) = [\varphi_r, \varphi_r] \qquad \sum_{i=1}^{n} \varphi_r(x_i)\varphi_s(x_i) = [\varphi_r, \varphi_s]$$

Układ równań (6) możemy zapisać (tzw. **postać normalna)**:

$$
\begin{aligned}
&[\phi_o, \phi_o]\, a_o + [\phi_o, \phi_1]\, a_1 + \ldots + [\phi_o, \phi_m]\, a_m = [y, \phi_o] \\
&[\phi_1, \phi_o]\, a_o + [\phi_1, \phi_1]\, a_1 + \ldots + [\phi_1, \phi_m]\, a_m = [y, \phi_1] \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
&[\phi_m, \phi_o]\, a_o + [\phi_m, \phi_1]\, a_1 + \ldots + [\phi_m, \phi_m]\, a_m = [y, \phi_m]
\end{aligned}
\qquad (7)
$$

Jest to układ m+1- równań liniowych z m+1 niewiadomymi: $a_o, a_1, \ldots, a_m$. Ponieważ: $[\phi_r, \phi_s] = [\phi_s, \phi_r]$, więc macierz współczynników układu równań (7) jest macierzą symetryczną.

W szczególnym przypadku gdy m=1 oraz $\phi_o(x) = 1$; $\phi_1(x) = x$ mamy przypadek regresji prostoliniowej, w której $a_o = A$ (odcięta) oraz $a_1 = B$ (nachylenie)

ZMIANA NIEKTÓRYCH PRZYPADKÓW NIELINIOWYCH FUNKCJI REGRESJI NA LINIOWE

Przypadek ogólny:

$$y = f(x, a_1, a_2, ..., a_n)$$

zmienne: x, y ; parametry: a_1, a_2, ..., a_n

linearyzacja:

$$y \approx f(x, a_1^o, a_2^o, ..., a_n^o) + (\frac{\partial f}{\partial a_1})_o * \Delta a_1 + (\frac{\partial f}{\partial a_2})_o * \Delta a_2 + ... + (\frac{\partial f}{\partial a_n})_o * \Delta a_n$$

Gdzie: a_i^o przybliżona wartość parametru a_i

„Dokładna" wartość $a_i = a_i^o + \Delta a_i$

WYZNACZANIE A i B

- **EXCEL| statystyczne| NACHYLENIE ,**
- **znane_y *(wstawić wartości y_i)***
- ***znane_x (wstawić wartości x_i)***

→ ***B***

EXCEL| statystyczne| ODCIĘTA ,

- **znane_y *(wstawić wartości y_i)***
  znane_x (wstawić wartości x_i)

- → **A**

ESTYMACJA LINIOWEJ FUNKCJI REGRESJI

y=A+Bx

Parametry A i B wyznacza się z próby, czyli są one **estymatorami** następujących parametrów zależności:

y=α+βx

w populacji generalnej. Stwierdzono, że są one estymatorami nieobciążonymi i zgodnymi. Obszar ufności dla prostej regresji **y=α+βx,** ograniczony tzw. k*rzywymi ufności ,* wyznacza się ze wzoru:

$$p\{\breve{y}_i - t_\gamma s_{\breve{y}_i} < \tilde{y}_i < \breve{y}_i + t_\gamma s_{\breve{y}_i}\} = 1 - \gamma \qquad (21)$$

Gdzie $\breve{y}_i$ oznacza wartość funkcji $\breve{y} = A + Bx$; wyznaczonej ze wzorów (10), (11) $\tilde{y}_i$ oznacza wartość szacowanej funkcji regresji y= α +βx, t_γ jest wartością zmiennej o rozkładzie t-Studenta, wyznaczonej dla określonego 1-γ poziomu ufności dla k=N-2 stopni swobody oraz:

$$s_{\breve{y}_i} = s_r \sqrt{\frac{1}{N} + \frac{(x_i - \bar{x})^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}} \qquad (22)$$

Gdzie s_r jest odchyleniem przeciętnym do prostej regresji, obliczanym ze wzoru:

$$s_r = \sqrt{\frac{1}{N-2}\sum_{i=1}^{N}(y_i - \breve{y}_i)^2} \qquad (23)$$

ESTYMACJA LINIOWEJ FUNKCJI REGRESJI

y=A+Bx

Obszar ufności dla prostej regresji **y=α+βx,** ograniczony tzw. k*rzywymi ufności* wyznacza się ze wzoru:

(21)

$$p\{\breve{y}_i - t_\gamma s_{\breve{y}_i} < \tilde{y}_i < \breve{y}_i + t_\gamma s_{\breve{y}_i}\} = 1 - \gamma$$

$$s_{\breve{y}_i} = s_r \sqrt{\frac{1}{N} + \frac{(x_i - \bar{x})^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}} \qquad s_r = \sqrt{\frac{1}{N-2}\sum_{i=1}^{N}(y_i - \breve{y}_i)^2}$$

ESTYMACJA LINIOWEJ FUNKCJI REGRESJI

P.G. $y = \alpha + \beta * x$; **Próba** (N-elementowa): $y = A + B * x$

B- NACHYLENIE; A- ODCIĘTA

1. KRZYWE UFNOŚCI

$$p\{\breve{y}_i - t_\gamma s_{\breve{y}_i} < \hat{\tilde{y}}_i < \breve{y}_i + t_\gamma s_{\breve{y}_i}\} = 1 - \gamma$$

$$s_{\breve{y}_i} = s_r \sqrt{\frac{1}{N} + \frac{(x_i - \bar{x})^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}} \qquad s_r = \sqrt{\frac{1}{N-2}\sum_{i=1}^{N}(y_i - \breve{y}_i)^2}$$

t_γ -ROZKŁAD.T.ODW; k=n-2

$$\hat{y}_1 = A + B * x_i$$

2. ESTYMACJI WSPÓŁCZYNNIKA (NACHYLENIA) REGRESJI LINIOWEJ

$$p\left\{B - t_\gamma \frac{s_r}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}} < \beta < B + t_\gamma \frac{s_r}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}}\right\} = 1 - \gamma$$

t_γ - ROZKŁAD.T.ODW (prawdopodobieństwo= γ ; stopnie swobody: k=n-2)

TESTY ISTOTNOŚCI DLA WSPÓŁCZYNNIKA REGRESJI LINIOWEJ

1. H_o : $B = \beta_o$

$$t = \frac{B - \beta_o}{s_r}\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

$$s_r = \sqrt{\frac{1}{N-2}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$$

$$\hat{y}_i = A + B * x_i$$

k= N-2

2. H_o : $\beta_1 = \beta_2$ (test równoległości)

$$t = \frac{\beta_1 - \beta_2}{s_{\beta_1 - \beta_2}}$$

$$s_{\beta_1-\beta_2} = \sqrt{\frac{\sum_{i=1}^{N1}(y_{i1} - \widehat{y_{i1}})^2 + \sum_{i=1}^{N2}(y_{i2} - \widehat{y_{i2}})^2}{N1 + N2 - 4}\left(\frac{1}{\sum_{i=1}^{N1}(x_{i1} - \overline{x_1})^2} + \frac{1}{\sum_{i=1}^{N2}(x_{i2} - \overline{x_2})^2}\right)}$$

k= N1+N2 -4

ESTYMACJA LINIOWEJ FUNKCJI REGRESJI

Przykład: Celem uzyskania spieku materiału o kontrolowanej porowatości, do wyjściowego proszku przed spiekaniem dodawano grafitu. Po wypaleniu uzyskano następujące wyniki (wartości x_i , y_i w Tabeli) :

Przyjmując poziom ufności 1-γ=0,95 oszacować metodą przedziałową zarówno całą funkcję regresji , jak i sam współczynnik regresji porowatości względem zawartości użytego grafitu.

x_i [% wag. grafitu]	y_i [porowatość, %]	$\hat{y}_i$	$(x_i-\bar{x})^2$	$\sqrt{\frac{1}{N}+\frac{(x_i-\bar{x})^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2}}$	$(y_i-\hat{y}_i)^2$				
1	8	9,5	9	0,681	2,25				
2	13	11,7	4	0,535	1,69				
3	14	13,8	1	0,423	0,04				
4	17	16,0	0	0,378	1,00				
5	18	18,1	1	0,423	0,01				
6	20	20,2	4	0,535	0,04				
7	22	22,4	9	0,681	0,16				
$\bar{x}=4$			28		5,19				

Rozwiązanie: $\hat{y}=A+Bx$

EXCEL| statystyczne| NACHYLENIE ,
znane_y *(wstawić wartości y_i)*
znane_x (wstawić wartości x)

- ***B*=2,14**

EXCEL| statystyczne| ODCIĘTA ,
znane_y *(wstawić wartości y_i)*
znane_x (wstawić wartości x)

- ***A*=7,4** $\hat{y_1}=A+Bx_1=7,4+2,14*1=9,5$

$\bar{x}=4$

ESTYMACJA LINIOWEJ FUNKCJI REGRESJI

Przyjmując poziom ufności 1-γ=0,95 oszacować metodą przedziałową zarówno całą funkcję regresji , jak i sam współczynnik regresji porowatości względem zawartości użytego grafitu.

x_i [% wag. grafitu]	y_i [porowatość, %]	$\widehat{y}_i$	$(x_i-\bar{x})^2$	$\sqrt{\frac{1}{N}+\frac{(x_i-\bar{x})^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2}}$	$(y_i-\widehat{y}_i)^2$	$s_{\widehat{y}_i}$	$t_\gamma s_{\widehat{y}_i}$	$\widehat{y}_i - t_\gamma s_{\widehat{y}_i}$	$\widehat{y}_i + t_\gamma s_{\widehat{y}_i}$
1	8	9,5	9	0,681	2,25	0,695	1,8	7,7	11,3
2	13	11,7	4	0,535	1,69	0,546	1,4	10,3	13,1
3	14	13,8	1	0,423	0,04	0,431	1,1	12,7	14,9
4	17	16,0	0	0,378	1,00	0,386	1,0	15,0	17,0
5	18	18,1	1	0,423	0,01	0,431	1,1	17,0	19,2
6	20	20,2	4	0,535	0,04	0,546	1,4	18,8	21,6
7	22	22,4	9	0,681	0,16	0,695	1,8	20,6	24,2
$\bar{x}=4$			28		5,19				

$$s_r = \sqrt{\frac{1}{N-2}\sum_{i=1}^{N}(y_i-\breve{y}_i)^2} \qquad s_r = \sqrt{\frac{5,19}{5}} = 1,02 \qquad s_{\breve{y}_i} = s_r\sqrt{\frac{1}{N}+\frac{(x_i-\bar{x})^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2}}$$

EXCEL| statystyczne|ROZKŁAD.T.ODW 1-γ=0,95 => γ=0,05 , k=N-2=5,

- **t_γ = 2,571**

ESTYMACJA LINIOWEJ FUNKCJI REGRESJI
y=A+Bx

Obszar ufności dla prostej regresji **y=α+βx,** ograniczony tzw. k*rzywymi ufności* wyznacza się ze wzoru:

(21)

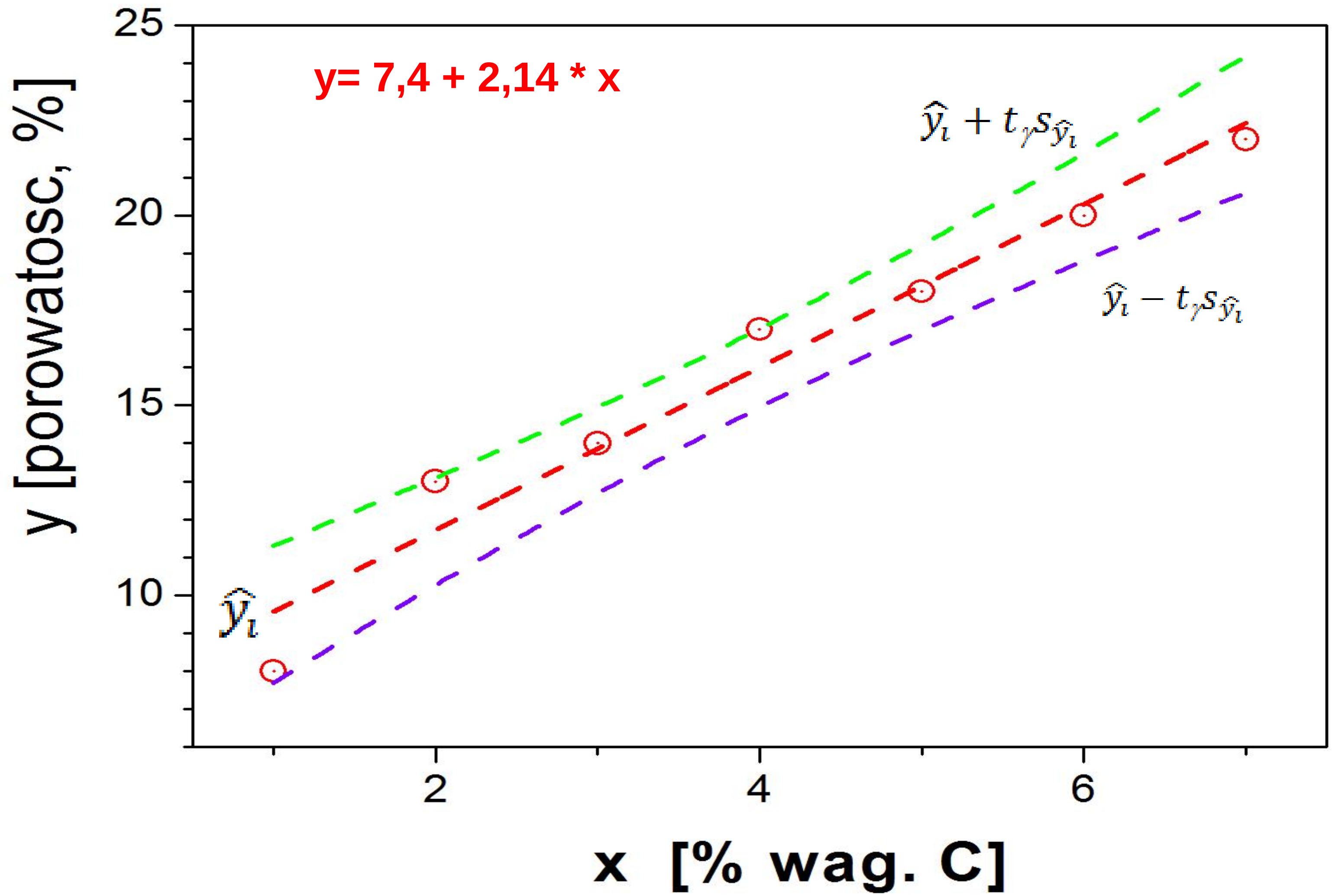
$$p\{\breve{y}_i - t_\gamma s_{\breve{y}_i} < \tilde{y}_i < \breve{y}_i + t_\gamma s_{\breve{y}_i}\} = 1 - \gamma$$

$$s_{\breve{y}_i} = s_r\sqrt{\frac{1}{N} + \frac{(x_i - \bar{x})^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}} \qquad s_r = \sqrt{\frac{1}{N-2}\sum_{i=1}^{N}(y_i - \breve{y}_i)^2}$$

$$p\left\{B - t_\gamma\frac{s_r}{\sqrt{\sum_{i=1}^{N}(x_i - \overline{x})^2}} < \beta < B + t_\gamma\frac{s_r}{\sqrt{\sum_{i=1}^{N}(x_i - \overline{x})^2}}\right\} = 1 - \gamma$$

t_γ - ROZKŁAD.T.ODW **(prawdopodobieństwo= γ ; stopnie swobody: k=n-2)**

ESTYMACJA LINIOWEJ FUNKCJI REGRESJI



ESTYMACJA LINIOWEJ FUNKCJI REGRESJI

PRZEDZIAŁ UFNOŚCI DLA WSPÓŁCZYNNIKA β (NACHYLENIA)

y=α+βx

Dwuwymiarowy rozkład cech X i Y w populacji jest normalny, lub zbliżony do normalnego. Z populacji tej wylosowano n-elementową próbę, która dała wyniki (x_i, y_i) i=1, 2,…,N. Przedział ufności dla współczynnika regresji **β** funkcji regresji **y=α+βx** w populacji wynosi:

$$p\left\{B - t_\gamma \frac{s_r}{\sqrt{\sum_{i=1}^{N}(x_i - \overline{x})^2}} < \beta < B + t_\gamma \frac{s_r}{\sqrt{\sum_{i=1}^{N}(x_i - \overline{x})^2}}\right\} = 1 - \gamma \tag{24}$$

Gdzie B jest współ. regresji, otrzymanym metodą najmniejszych kwadratów (równ. 11 lub 11'), s_r dane jest równ. (23), a t_γ jest wartością zmiennej o rozkładzie t-Studenta, wyznaczonej dla określonego 1-γ poziomu ufności dla k=N-2 stopni swobody.

ESTYMACJA LINIOWEJ FUNKCJI REGRESJI

PRZEDZIAŁ UFNOŚCI DLA WSPÓŁCZYNNIKA β (NACHYLENIA)

$$p\left\{B - t_\gamma \frac{s_r}{\sqrt{\sum_{i=1}^{N}(x_i - \overline{x})^2}} < \beta < B + t_\gamma \frac{s_r}{\sqrt{\sum_{i=1}^{N}(x_i - \overline{x})^2}}\right\} = 1 - \gamma$$

t_γ = 2,571 B=2,14 $s_r = \sqrt{\frac{5,19}{5}} = 1,02$ $\sqrt{\sum_{i=1}^{N}(x_i - \overline{x})^2} = \sqrt{28} = 5,29$

2,14- 2,571*1,02/5,29< β < 2,14+ 2,571*1,02/5,29

1,64< β < 2,64

TESTY ISTOTNOŚCI DLA WSPÓŁCZYNNIKA B (NACHYLENIA) REGRESJI LINIOWEJ

1. H_o : $B = \beta_o$

$$t = \frac{B - \beta_o}{s_r}\sqrt{\sum_{i=1}^{N}(x_i - \overline{x})^2}$$

$$s_r = \sqrt{\frac{1}{N-2}\sum_{i=1}^{N}(y_i - \breve{y}_i)^2}$$

$$\hat{y}_1 = A + B * x_i$$

k= N-2

2. H_o : $\beta_1 = \beta_2$ (test równoległości)

$$t = \frac{\beta_1 - \beta_2}{s_{\beta_1 - \beta_2}}$$

$$s_{\beta_1-\beta_2} = \sqrt{\frac{\sum_{i=1}^{N1}(y_{i1} - \widehat{y_{i1}})^2 + \sum_{i=1}^{N2}(y_{i2} - \widehat{y_{i2}})^2}{N1 + N2 - 4}\left(\frac{1}{\sum_{i=1}^{N1}(x_{i1} - \overline{x_1})^2} + \frac{1}{\sum_{i=1}^{N2}(x_{i2} - \overline{x_2})^2}\right)}$$

k= N1+N2 -4

TEST ISTOTNOŚCI DLA WSPÓŁCZYNNIKA REGRESJI LINIOWEJ

PROBLEM: Dwuwymiarowy rozkład cech X i Y w populacji jest normalny, lub zbliżony do normalnego. Z populacji tej wylosowano n-elementową próbę, która dała wyniki (x_i, y_i) i=1, 2,…,N. Na podstawie wyników próby sprawdzić hipotezę, że współczynnik **β** liniowej funkcji regresji: **y=α+βx** w populacji ma określoną wartość, tj: $H_o : \beta = \beta_o$ wobec $H_1 : \beta \neq \beta_o$ (w szczególności gdy $\beta_o = 0$, hipoteza głosi, że brak jest zależności między badanymi cechami X i Y).

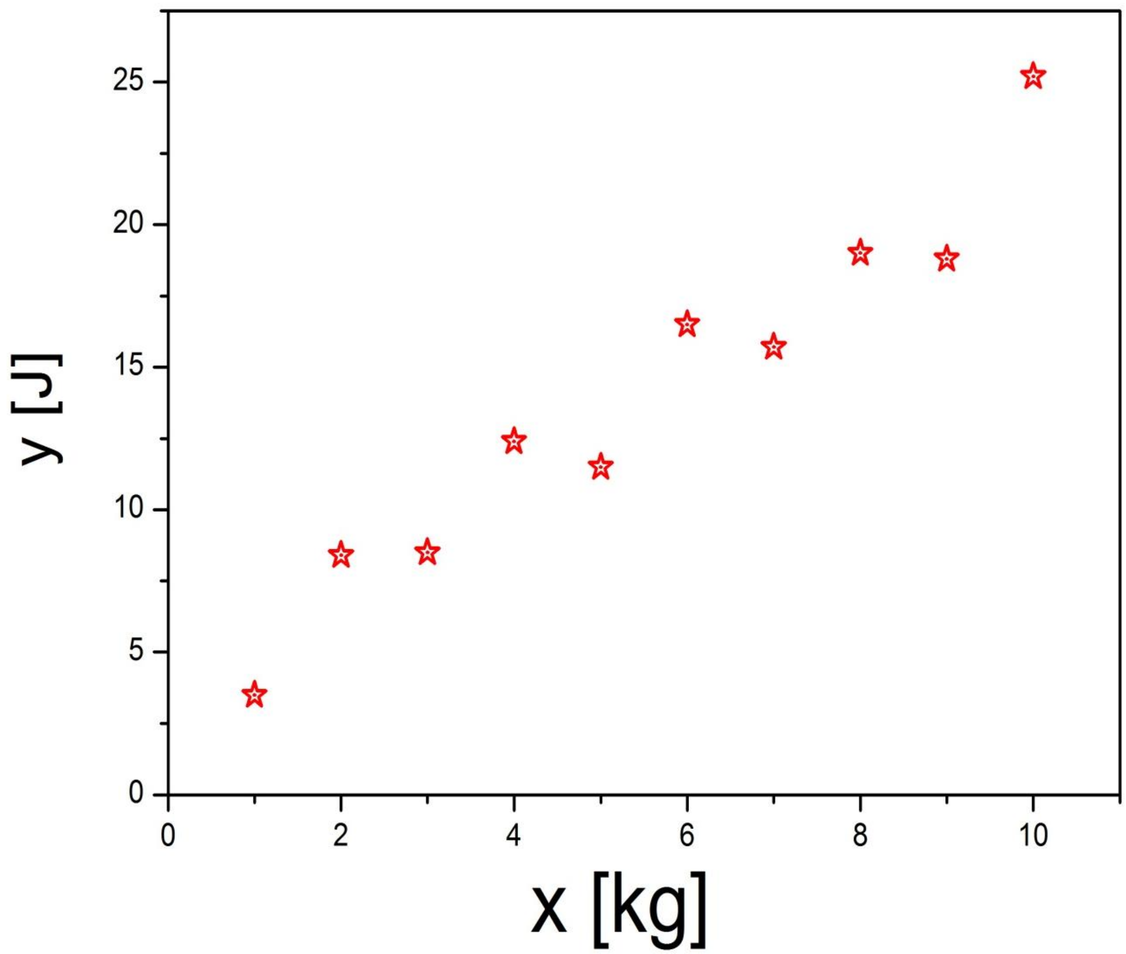
Rozwiązanie: Metodą najmn. kwadratów szacujemy parametry próby (A i B), obliczamy wg (23) s_r oraz statystykę:

$$t = \frac{B - \beta_o}{s_r} \sqrt{\sum_{i=1}^{N} (x_i - \bar{x})^2} \qquad s_r = \sqrt{\frac{1}{N-2} \sum_{i=1}^{N} (y_i - \breve{y}_i)^2}$$

Statystyka ta ma rozkład t-Studenta z k=n-2 stopniami swobody, odczytujemy dla danego γ , t_γ , gdy $|t| \geq t_\gamma$ to H_o odrzucić

ĆWICZENIE

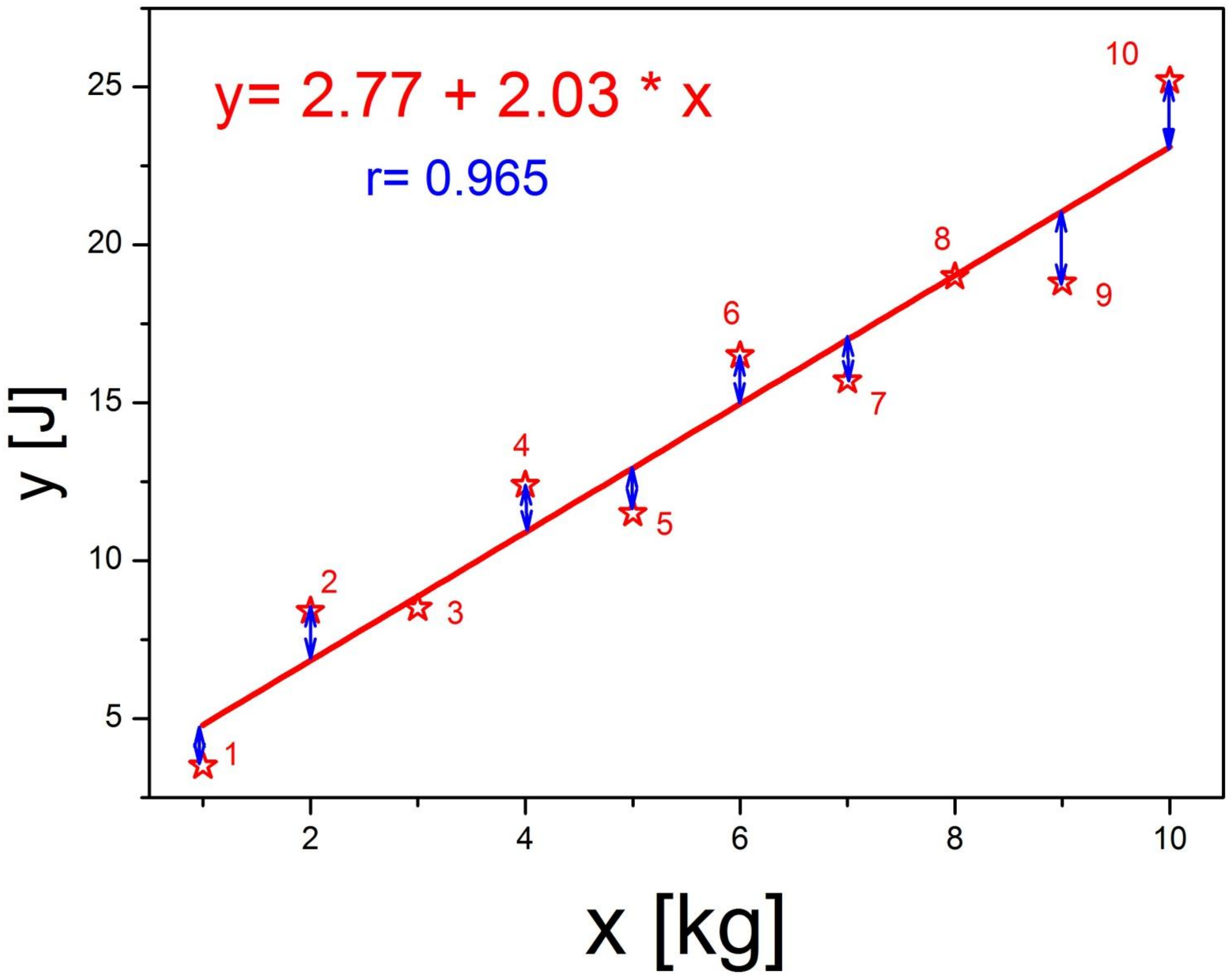
x_i [kg]	y_i [J]
1	3.5
2	8.4
3	8.5
4	12.4
5	11.5
6	16.5
7	15.7
8	19.0
9	18.8
10	25.2
$x_{sr}=5.5$	$y_{sr}=13.95$



EXCEL => A (ODCIETA) i B (NACHYLENIE) dla x_i, y_i A=2.77 ; B= 2.03

ĆWICZENIE (2)

x_i [kg]	y_i [J]
1	3.5
2	8.4
3	8.5
4	12.4
5	11.5
6	16.5
7	15.7
8	19.0
9	18.8
10	25.2
$x_{sr}=5.5$	$y_{sr}=13.95$



A= 2.77

B= 2.03

ĆWICZENIE (3)

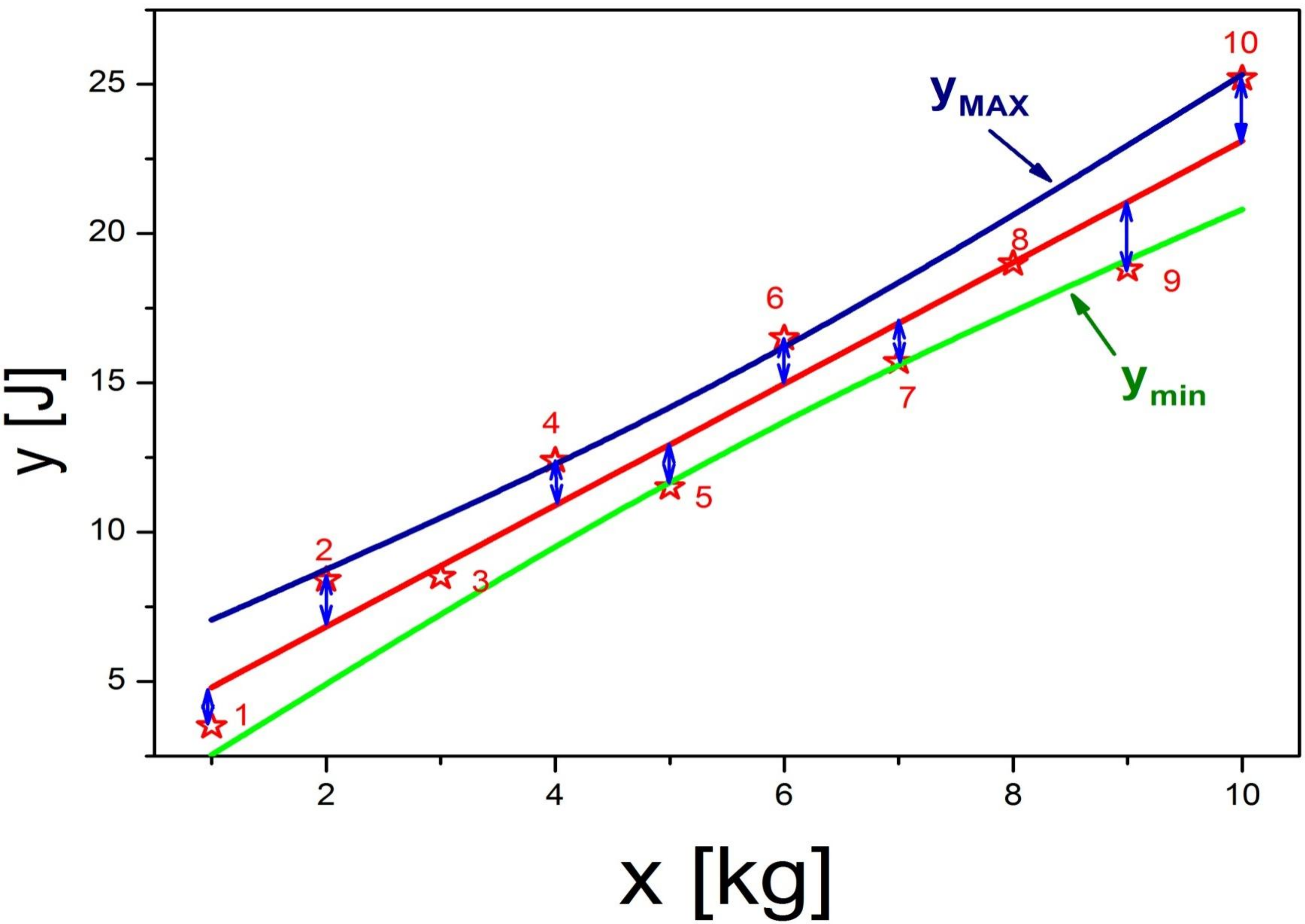
x_i [kg]	y_i [J]	$\hat{y}_i$	$(x_i - x_{sr})^2$	$(y_i - \hat{y}_i)^2$	$\frac{(x_i - x_{sr})^2}{\sum_{i=1}^{n}(x_i - x_{sr})^2} = S_i$	$\sqrt{\frac{1}{n} + S_i}$	$s_{\hat{y}_i}$	$s_{\hat{y}_i} * t_\gamma$	y_{min}	y_{MAX}
1	3.5	4.8	20.25	1.690	0.2455	0.588	0.979	2.257	2.543	7.057
2	8.4	6.83	12.25	2.465	0.1485	0.498	0.830	1.914	4.916	8.744
3	8.5	8.86	6.25	0.130	0.0758	0.419	0.698	1.610	7.250	10.470
4	12.4	10.89	2.25	2.280	0.0273	0.357	0.594	1.370	9.520	12.260
5	11.5	12.92	0.25	2.016	0.0030	0.321	0.534	1.232	11.688	14.152
6	16.5	14.95	0.25	2.403	0.0030	0.321	0.534	1.232	13.718	16.182
7	15.7	16.98	2.25	1.638	0.0273	0.357	0.594	1.370	15.610	18.350
8	19.0	19.01	6.25	0.000	0.0758	0.419	0.698	1.610	17.400	20.620
9	18.8	21.04	12.25	5.018	0.1485	0.498	0.830	1.914	19.126	22.954
10	25.2	23.07	20.25	4.537	0.2455	0.588	0.979	2.257	20.813	25.327
x_{sr}= 5.5	y_{sr}= 13.95		SUMA= 82.5	SUMA= 22.177						

$$\hat{y}_i = A + B * x_i \qquad s_r = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2}}$$

$$\hat{y}_i = 2.77 + 2.03 * x_i \qquad s_r = \sqrt{\frac{22.177}{8}} = 1.665$$

$$s_{\hat{y}_i} = s_r\sqrt{\frac{1}{n} + s_i} \qquad y_{min} = \hat{y}_i - s_{\hat{y}_i} * t_{\gamma} \qquad y_{MAX} = \hat{y}_i + s_{\hat{y}_i} * t_\gamma$$

ĆWICZENIE (4)



ZMIANA NIEKTÓRYCH PRZYPADKÓW NIELINIOWYCH FUNKCJI REGRESJI NA LINIOWE

(układy prostujące)

a) <u>Adsorpcja promieniowania</u>: $I = I_o e^{-\mu x}$

zmienne I, x ; parametry I_o, μ

Układ prostujący: Y= log I = f(X):

$$Y = \log I = \log I_o - \mu \log e * x = A - B * X$$

<u>**b) Przewodnictwo elektryczne półprzewodników**</u> $\sigma = \sigma_o * \exp[-\frac{E_{act}}{RT}]$

zmienne: σ, T ; parametry: σ_o, E_{act} ; stałe: R

Układ prostujący: Y= logσ= f(T^{-1})= f(X):

$$Y = \log \sigma = \log \sigma_o - \frac{E_{act} \log e}{R} * T^{-1} = A - B * X$$

<u>**c) Przewodnictwo elektryczne elektrolitów**</u> $\sigma = \frac{\sigma_o}{T} * \exp[-\frac{E_{act}}{RT}]$

zmienne: σ, T ; parametry: σ_o, E_{act} ; stałe: R

Układ prostujący: Y= log(σT) = f(T^{-1})= f(X):

$$Y = \log(\sigma T) = \log \sigma_o - \frac{E_{act} \log e}{R} * T^{-1} = A - B * X$$

ZMIANA NIEKTÓRYCH PRZYPADKÓW NIELINIOWYCH FUNKCJI REGRESJI NA LINIOWE

(układy prostujące)

d) Izoterma BET

$$V = \frac{v_m c p}{(p_o - p)[1 + \frac{p}{p_o}(c-1)]}$$

zmienne: V, p ; parametry: p_o, c, v_m

Układ prostujący:

$$Y = \frac{p}{(p_o - p)V} = f(\frac{p}{p_o}) = f(X)$$

$$Y = \frac{p}{(p_o - p)V} = \frac{1}{c v_m} + \frac{c-1}{c v_m} * (\frac{p}{p_o}) = A + B * X$$

$$A + B = \frac{1}{v_m}$$