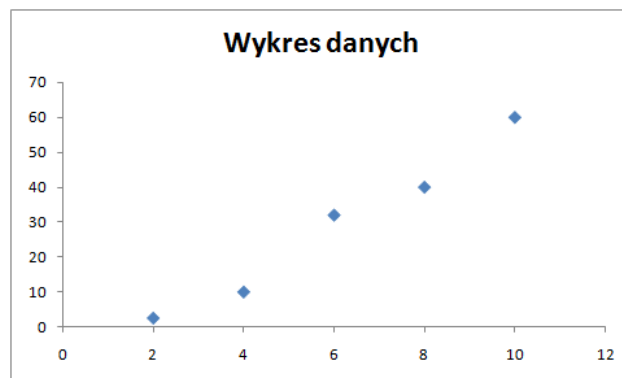


# 1 Metoda najmniejszych kwadratów

Założmy, że mamy pięć punktów doświadczalnych w tabeli

| i | $x_i$ | $y_i$ |
|---|-------|-------|
| 1 | 2     | 2,5   |
| 2 | 4     | 10    |
| 3 | 6     | 32    |
| 4 | 8     | 40    |
| 5 | 10    | 60    |

Wykreślamy te punkty otrzymując rysunek



Punkty są porozrzucane, ale wyraźnie układają się wzdłuż **prostej**. Równanie prostej ma następującą postać

$$y = ax + b,$$

gdzie  $a$  i  $b$  to współczynniki, których chwilowo nie znamy. Poszukiwanie parametrów polega na minimalizacji sumy

$$S(a, b) = \sum_{i=1}^n [y_i - y(x_i)]^2 = \sum_{i=1}^n [y_i - (ax_i + b)]^2,$$

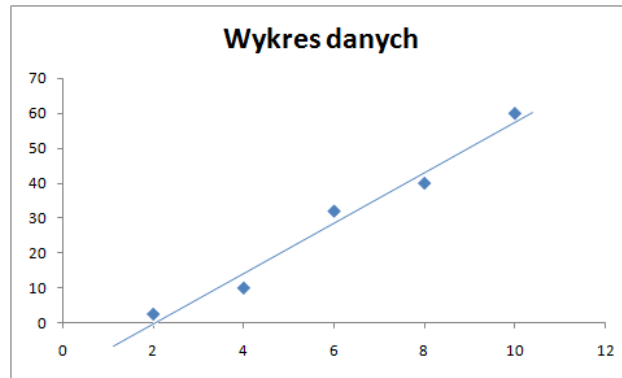
gdzie  $y(x_i)$  to wartości współrzędnej  $y$  obliczonej z równania prostej dla danych  $x_i$ . Różnice między dokładnymi wartościami  $y_i$  oraz wartościami obliczonymi z równania prostej są podniesione do kwadratu, aby uniknąć możliwości, że będą się nawzajem znosiły na skutek różnicy znaków. Z tego też względu przedstawiona metoda postępowania nosi nazwę **metody najmniejszych kwadratów**.

Dla danych z tabeli wielkość  $S$  wynosi

$$S(a, b) = [2, 5 - (2a + b)]^2 + [10 - (4a + b)]^2 + [32 - (6a + b)]^2 + [40 - (8a + b)]^2 + [60 - (10a + b)]^2.$$

Minimalizujemy wielkość  $S(a, b)$ , czyli

$$\begin{aligned} &2[2, 5 - 2a - b](-2) + 2[10 - 4a - b](-4) + 2[32 - 6a - b](-6) \\ &\quad + 2[40 - 8a - b](-8) + 2[60 - 10a - b](-10) = 0 \\ &2[2, 5 - 2a - b](-1) + 2[10 - 4a - b](-1) + 2[32 - 6a - b](-1) \\ &\quad + 2[40 - 8a - b](-1) + 2[60 - 10a - b](-1) = 0. \end{aligned}$$



Po uproszczeniu dostajemy

$$440a + 60b = 2314,$$

$$60a + 10b = 289.$$

Rozwiązaniem są  $a = 7,25$  oraz  $b = -14,6$ . Na rysunku przedstawiono punkty z tabeli oraz linię prostą określoną równaniem  $y = 7,25 \cdot x - 14,6$ .

Funkcja  $S(a, b)$  dla danych  $(x_i, y_i)$ , gdzie  $i = 1, 2, \dots, n$  ma postać

$$S(a, b) = \sum_{i=1}^n (y_i - b - ax_i)^2,$$

gdzie  $n$  oznacza liczbę punktów.

Układ **równań normalnych**

$$2 \sum_{i=1}^n (y_i - b - ax_i)(-1) = 0$$

$$2 \sum_{i=1}^n (y_i - b - ax_i)(-x_i) = 0.$$

Po uporządkowaniu otrzymamy układ równań

$$bn + a \sum_{i=1}^n x_i = \sum_{i=1}^n y_i,$$

$$b \sum_{i=1}^n x_i + a \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i.$$

Ogólne wzory na współczynniki definiujące prostą

$$a = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2},$$

$$b = \frac{\sum_{i=1}^n y_i \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i \sum_{i=1}^n x_i y_i}{n \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2}.$$

Średnia

$$s = \sqrt{\frac{1}{n-2} \sum_{i=1}^n \varepsilon_i^2},$$

gdzie  $\varepsilon_i = y_i - b - ax_i$ .

Dla danych z tabeli mamy

$$b = -14,6, \quad a = 7,25, \quad s = 4,151305.$$

Wzory na błędy standardowe wyznaczonych parametrów  $a$  oraz  $b$

$$s_b = s \sqrt{\frac{\sum_{i=1}^n x_i^2}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}}$$

oraz

$$s_a = s \sqrt{\frac{n}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}},$$

gdzie  $s$  jest średnim odchyleniem standardowym od prostej.

Dla danych z tabeli mamy

$$s_b = 4,353925, \quad s_a = 0,656379.$$

Wartość **współczynnika korelacji** jest miarą korelacji między zmiennymi. Jeśli jest bliski  $\pm 1$  oznacza to, że wielkości są dobrze skorelowane i istnieje między nimi zależność funkcyjna.

Dla regresji liniowej jest dany wzorem

$$r = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} \sqrt{n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i\right)^2}}$$

Po podstawieniu danych z tabeli dostajemy

$$r = 0,976.$$

## 2 Zależność wykładnicza

Rozpad jąder promieniotwórczych można zapisać równaniem

$$N(t) = N_0 e^{-t\tau},$$

gdzie  $\tau$  jest średnim czasem życia związanym z okresem połowicznego rozpadu relacją

$$T_{1/2} = 0,693\tau.$$

Po zlogarytmowaniu otrzymujemy wzór na funkcję liniową we współrzędnych logarytmicznych

$$\ln N(t) = -\frac{1}{\tau}t + \ln N_0,$$

gdzie zmienną jest  $t$ ,  $-\frac{1}{\tau}$  odpowiada współczynnikowi  $a$ , a  $\ln N_0$  - współczynnikowi  $b$ .

### 3 Regresja liniowa w Excelu

Excel wyposażony jest w narzędzia i funkcje statystyczne służące do obliczeń związanych z regresją liniową i wykładniczą. Najwygodniejszą jest formuła tablicowa REGLINP oraz formuły działające na jednej komórce

- REGBLSTD - liczy  $s$ ,
- NACHYLENIE - liczy wsółczynnik  $a$ ,
- ODCIĘTA - liczy współczynnik  $b$ ,
- WSP.KORELACJI - liczy  $r$ ,
- PEARSON - liczy  $r$ ,
- R.KWADRAT - liczy  $r^2$ .

**UWAGAW** wyniku działania formuły tablicowej REGLINP otrzymujemy kwadrat współczynnika korelacji liniowej, a nie sam współczynnik.

Poniżej, tabelki i wykres otrzymane przy pomocy funkcji statystycznych.