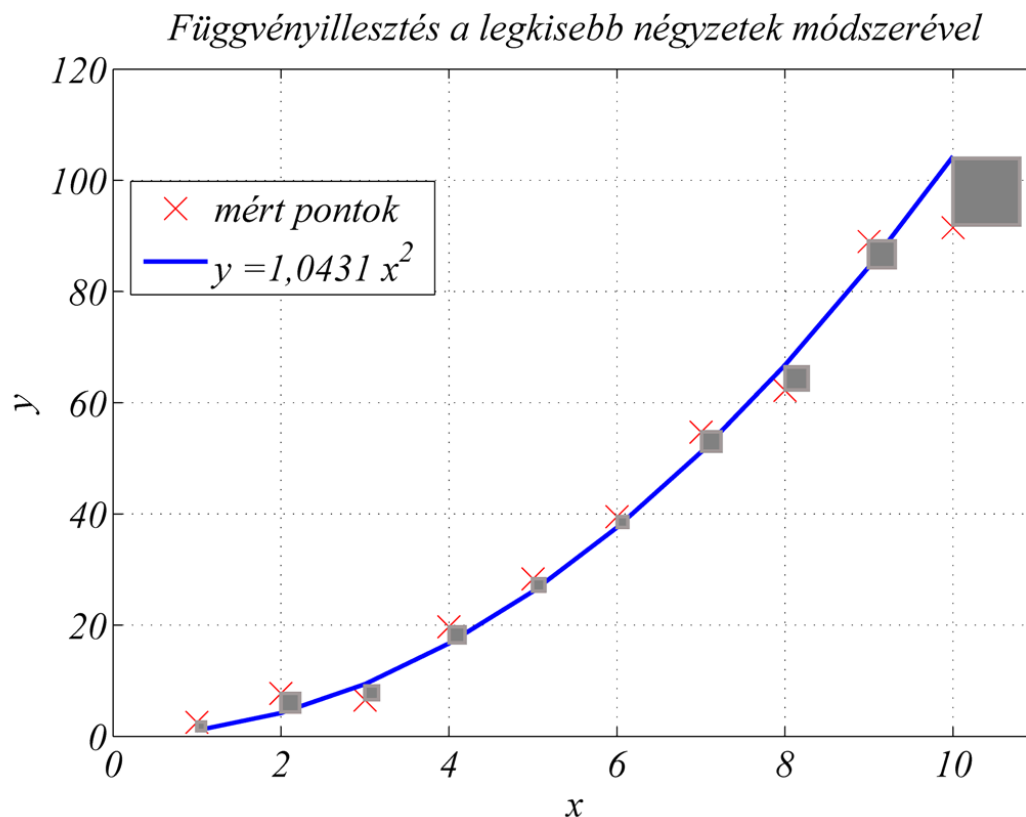


A legkisebb négyzetek módszere

Dr. Berta Miklós
Fizika és Kémia Tanszék
Széchenyi István Egyetem

A módszer alapelve



$$\begin{aligned}\Phi(f_{opt.}) &= \\ &= \sum_{i=1}^N (y_i - f_{opt.}(x_i))^2 = \\ &= \min.\end{aligned}$$

Keressük azt az optimális $y = f_{opt.}(x)$ függvényt, amelyre az eltérésnégyzetek összege a lehető legkisebb!

Parametrizálás

- A keresett optimális f_{opt} függvény a független változón kívül általában még néhány paramétert is tartalmaz.

$$y = f(x, p_1, \dots, p_M)$$

- Az optimális függvény meghatározása ilyenkor az optimális paraméterek meghatározását jelenti.
- Az optimális paramétereket a „normál – egyenletek” alapján határozzuk meg.

$$\frac{\partial \Phi}{\partial p_i} = 0, \forall i = 1, \dots, M$$

Lineáris regresszió, vagy egyenes-illesztés – Arányosság esete

$$y = m.x$$

- Itt egyedül az m paraméter fordul elő.
- Az eltérésnégyzetek összege:

$$\Phi(m) = \sum_{i=1}^N (y_i - m.x_i)^2 = \sum_{i=1}^N y_i^2 - 2m \sum_{i=1}^N x_i y_i + m^2 \sum_{i=1}^N x_i^2$$

- Ebből a normálegyenlet:

$$\frac{d\Phi}{dm} = 2m \sum_{i=1}^N x_i^2 - 2 \sum_{i=1}^N x_i y_i = 0$$

- Tehát az optimális paraméter:

$$m^* = \frac{\sum_{i=1}^N x_i y_i}{\sum_{i=1}^N x_i^2}$$

- Az optimális paraméter bizonytalansággal terhelt, mért értékek alapján kerül kiszámításra, így *maga is valószínűségi változó*, tehát *várható értéke* mellett, *szórása* is lesz.
- Bebizonyítható, hogy a *várható értéke maga a normál egyenletből származtatott optimális érték*,
- míg szórása a következő módon határozható meg:

$$S_0 = \sum_{i=1}^N (y_i - m^* x_i)^2$$

$$s = \sqrt{\frac{S_0}{N-1}}$$

$$S_{m^*} = \frac{s}{\sqrt{\sum_{i=1}^N x_i^2}}$$

Lineáris regresszió, vagy egyenes-illesztés – Általános egyenes esete

$$y = m.x + b$$

- Ez esetben két paraméter fordul elő, m - meredekség és b – offset
- Az eltérésnégyzetek összege:

$$\Phi(m, b) = \sum (y_i - m.x_i - b)^2$$

- Ebből a normálegyenletek:

$$\frac{\partial \Phi}{\partial m} = 0 \Rightarrow m \sum x_i^2 + b \sum x_i = \sum x_i y_i$$

$$\frac{\partial \Phi}{\partial b} = 0 \Rightarrow m \sum x_i + b.N = \sum y_i$$

- Tehát az optimális paraméterek:

$$m^* = \frac{N \sum x_i y_i - \sum x_i \sum y_i}{N \sum x_i^2 - (\sum x_i)^2}$$
$$b^* = \frac{\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i}{N \sum x_i^2 - (\sum x_i)^2}$$

- Az optimális paraméterek szórásai pedig:

$$S_0 = \sum (y_i - m^* \cdot x_i - b^*)^2, \quad s = \sqrt{\frac{S_0}{N-2}}$$
$$s_{m^*} = s \sqrt{\frac{N}{N \sum x_i^2 - (\sum x_i)^2}}, \quad s_{b^*} = s \sqrt{\frac{\sum x_i^2}{N \sum x_i^2 - (\sum x_i)^2}}$$

Korrelációs együttható

- Az r korrelációs együttható annak a mértéke, hogy a mért $[x_i, y_i]$ adatpárok sora mennyire jól illeszkedik egy egyenesre:

$$r = \frac{N \sum x_i y_i - \sum x_i \sum y_i}{N \sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$
$$\bar{x} = \frac{\sum x_i}{N}, \quad \bar{y} = \frac{\sum y_i}{N}$$

- Ha $|r|=1$, akkor az adatpárok tökéletesen illeszkednek egy egyenesre. Általában, ha $|r|>0,8$, akkor az illeszkedést „jó” tekintjük.