Lineáris regresszió

Az **$y = a\,x + b$** alakú egyenes paramétereire
a legkisebb négyzetek módszerének alkalmazásával a következő összefüggéseket kapjuk:

A meredekség

$$a = \frac{\sum (x_i y_i) - \frac{\sum x_i \sum y_i}{N}}{\sum \left(x_i^2\right) - \frac{\left(\sum x_i\right)^2}{N}} \; ,$$

a tengelymetszet

$$b = \frac{\sum y_i}{N} - a\frac{\sum x_i}{N} \; ;$$

avagy az átlagok

$$\overline{x} = \frac{\sum x_i}{N} \; , \quad \overline{y} = \frac{\sum y_i}{N} \; , \quad \overline{xy} = \frac{\sum (x_i y_i)}{N} \; , \qquad \overline{x^2} = \frac{\sum \left(x_i^2\right)}{N}$$

bevezetésével:

$$\boxed{a = \frac{\overline{xy} - \overline{x}\cdot\overline{y}}{\overline{x^2} - \overline{x}^2} \qquad \text{és} \qquad b = \overline{y} - a\,\overline{x}}$$

A fenti a, b paraméterek **varianciáját** a következőképp becsülhetjük:

$$\mathrm{Var}[a] = \frac{\sigma_y^{\,2}}{\sum (x_i - \overline{x})^2} \; , \qquad \mathrm{Var}[b] = \frac{\sigma_y^{\,2}}{N}\left[1 + \frac{N\,\overline{x}^2}{\sum (x_i - \overline{x})^2}\right] \; ;$$

avagy felhasználva, hogy

$$\sum (x_i - \overline{x})^2 = \sum \left(x_i^2\right) - \sum (2x_i\overline{x}) + \sum \left(\overline{x}^2\right) = \sum \left(x_i^2\right) - 2\overline{x}\sum (x_i) + N\overline{x}^2 =$$

$$= \sum \left(x_i^2\right) - 2N\overline{x}^2 + N\overline{x}^2 = \sum \left(x_i^2\right) - N\overline{x}^2 = N\overline{x^2} - N\overline{x}^2 \; ,$$

$$\boxed{\mathrm{Var}[a] = \frac{\sigma_y^{\,2}}{N}\cdot\frac{1}{\overline{x^2} - \overline{x}^2} \qquad \text{és} \qquad \mathrm{Var}[b] = \frac{\sigma_y^{\,2}}{N}\cdot\frac{\overline{x^2}}{\overline{x^2} - \overline{x}^2}} \; .$$

$\sigma_y^{\,2}$ az úgynevezett reziduális szórásnégyzettel, $s_r^{\,2}$ -tel közelíthető:

$$\boxed{s_r^{\,2} = \frac{\sum \left(y_i - (ax_i + b)\right)^2}{N - 2}} \; .$$

A **szórás** a variancia négyzetgyöke.

A fenti képletek origón átmenő egyenes esetén nem érvényesek!