EGYENESILLESZTÉS: A LEGKISEBB NÉGYZETEK MÓDSZERE

Kísérleteket elvégeztük. – Dolgozzuk fel az adatokat!

Cél: mért változók (T, p, I, U…) közötti kapcsolat felderítése.

1. szóródási diagram $\{x_i, y_i\}$ ábra
2. kvantitatív megfogalmazás: minta korrelációs koefficiense.

$$r = \frac{\sum_i (\xi_i - \bar{\xi})(\eta_i - \bar{\eta})}{\left\{\sum_i (\xi_i - \bar{\xi})^2 \sum_i (\eta_i - \bar{\eta})^2\right\}^{1/2}} :$$

ez az eloszlás korrelációs együtthatójának becslését szolgáltatja.

ξ_i = i-ik mérés valószínűségi változója
$r \rightarrow$ valószínűségi változó, várható értékkel, standard deviációval $\rightarrow$ 1 mérési sorozatra r realizációja:

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\left\{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2\right\}^{1/2}}$$

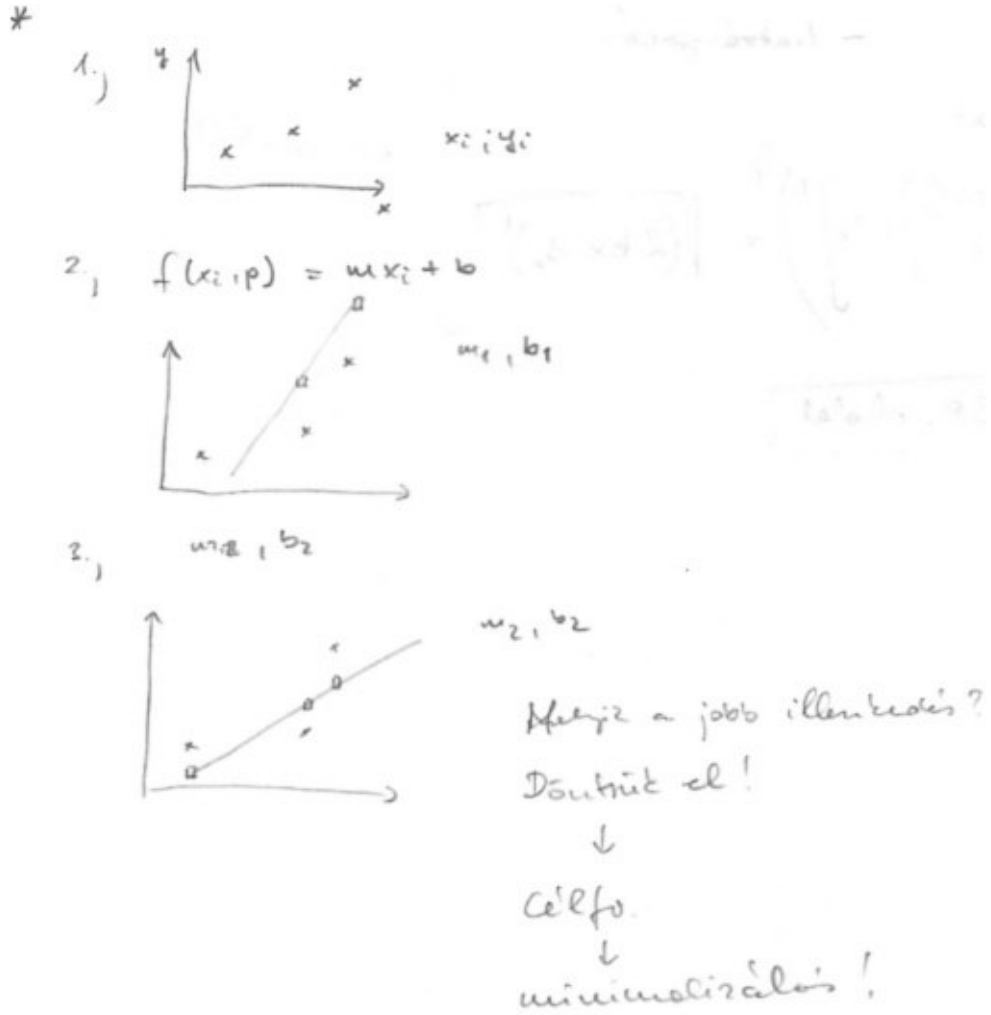
Vigyázat! A laborban ezt számoljuk!

$\lvert r \rvert = 1 \Rightarrow$ egyenes	ez nem feltétlenül oksági kapcsolat
$\lvert r \rvert \sim 1$ közel egyenes	ez nem feltétlenül oksági kapcsolat
$r \sim 0$	nem jelenti, hogy nem függenek egymástól, annyit mond, hogy a függés nem lineáris.

Most pedig nézzük meg a közel lineáris függés esetét! $\rightarrow$ számszerűsítsük ezt a függést!!!

Alapja az ún. paraméterbecslési eljárás.

y_i mérési pontok $f(x_i, p)$ függvény kapcsolat $\rightarrow$ p paraméterek



y_i és $f(x_i, p)$ közötti különbség csökkenésére nagyon jól bevált módszer LKN módszer.

Ebben egy Q_p-célfüggvényt minimalizálunk, ahol

$$Q_p = \sum_i (y_i - f(x_i, p))^2 w_i \qquad w_i \text{ súlyok}$$

Lineáris vagy nem lineáris regresszió lehet a feladat (p-ben lineáris-e a függvény vagy nem)

Lineáris regresszió

Egyenes általános egyenlete: $y = mx + b$

Célfüggvény $\quad Q = \sum_{i=1}^{n} [y_i - mx_i - b]^2 \Rightarrow$ minimum

$$\frac{\partial Q}{\partial m} = 0 \qquad \frac{\partial Q}{\partial b} = 0$$

(többváltozós függvények szélsőérték helyére vonatkozó feltételek szerint)

Eredmény:

$$m = \frac{\overline{(xy)} - \overline{x}\,\overline{y}}{\overline{(x^2)} - (\overline{x})^2}$$

$$b = \overline{y} - m\overline{x}$$

Illeszkedés jellemzése (illeszkedés jóságát jellemzi):

s_r reziduális szórással:

$$s_r = \left\{ \frac{\sum_{i}^{n} [y_i - (mx_i + b)]^2}{n-2} \right\}^{1/2}$$

A paraméterek megbízhatóságának jellemzése:

$$s_m^2 = s_r^2 \frac{n}{n\sum x_i^2 - \left(\sum x_i\right)^2}$$

$$s_b^2 = s_r^2 \frac{\sum x_i^2}{n\sum x_i^2 - \left(\sum x_i\right)^2}$$

m és b konfidencia intervallumait az

$$m \pm t_{1-\alpha} s_m \quad \text{és} \quad b \pm t_{1-\alpha} s_b$$

tartományok definiálják az n-2 szabadsági fokhoz tartozó Student-féle t-eloszlás kritikus értékeinek felhasználásával.

Kiegészítés: egy adott x_j értékhez tartozó y_j érték hibájának becslése y_j konfidencia intervalluma

$$y_j + t_{1-\alpha} \cdot s_y \quad \text{ahol} \quad s_y = s_r \cdot \left(\frac{1}{n} + \frac{(x'_j - \bar{x})^2}{\sum_{j=1}^{u} (x_j - \bar{x})^2} \right)^{1/2}$$

($\bar{x}$ közelében a legszűkebb ez az intervallum…)

Sok esetben nemlineáris egyenletek is linearizálhatók!

Ilyen pl. Arrhenius egyenlet:

Mit is ír le? $k = A \cdot \exp\left(-\frac{\Delta E^{\ddagger}}{RT} \right)$

Hogyan linearizáljuk? $\ln k = \ln A - \frac{\Delta E^{\ddagger}}{RT}$

Ábrázolunk: mit minek a függvényében?

Mi micsoda lesz a kiértékelés során? (meredekség, tengelymetszet)

Mérési eredmények ábrázolásának szabályai: Szalma J.

Amikor a két változó kapcsolatát derékszögű koordináta- rendszerben ábrázoljuk, azt a következő szempontok figyelembe- vételével célszerű megtenni.

a. Az x és y tengelyt közelítőleg egyforma hosszúra választjuk (a grafikon méreteit tekintve), így nem kapunk torz ábrát Az x tengelyen balról jobbra, az y tengelyen alulról felfelé nőnek az értékek.

b. A változóknak nemcsak a számértékeit, hanem a nevét vagy jelét és a mértékegységeit (dimenzió) is fel kell tüntetni a tengelyeken. A grafikonon szövegek is elhelyezhetők, cím (fejléc), az illesztett függvény egyenlete, az illesztésre vonatkozó statisztikai mutatók stb.

c. A tengelyeket úgy osztjuk be (olyan léptéket választunk), hogy az x tengely teljes hossza a független változó (x) értelmezési tartományából, az y tengely pedig a függő változó (y) értékkészletéből azt az intervallumot tartalmazza, amelybe a mérési eredményeink esnek. Kivétel az az eset amikor az x = O értékre extrapolálunk, ilyenkor az x tengely beosztása az x = O értéknél kezdődik, amennyiben a grafikonról akarjuk leolvasni az extrapolált értéket. Ha a mérési pontokra illesztett függvényből számítjuk az x = O-hoz tartozó függő változó értéket, akkor nincs szükség ilyen tengelybeosztásra.

d. A tengelyeken egész értékeket tüntetünk fel, nem pedig az egyes pontok koordinátáit.

e. Ha milliméter papíron készítjük a grafikont, akkor a tengelyeken az ábrázolandó egységeknek 1 2, 5, vagy 10 mm-t célszerű megfeleltetni. így könnye átszámítani az egyes pontok koordinátáit mm-be, és a grafikonról is könnyebb visszaolvasni valamely pont koordinátáit Nem célszerű pl. 2.5; 3; vagy 4 mm-t rendelni egy egységhez.

f. Amikor a mérési eredmények grafikus kiértékelése mm-papíron történik (az ábra nemcsak szemléltetési célokat szolgál) a tengelyek beosztásánál lényeges szempont, hogy a mérési adatokat ugyanannyi számjeggyel tudjuk ábrázolni, mint ahány értékes (helyes) jeggyel azok rendelkezne. Ez előfeltétele annak, hogy az eredményt a mérési adatok által biztosított pontossággal megkapjuk. Ebben az esetben előfordulhat, hogy A3-as méretben kell dolgozni.
Amikor számítógéppel illesztünk valamilyen függvényt mérési pontjainkra, akkor az ábra csak szemléltetési célokat szolgál. Viszont jelentkezik egy más szempont, az illesztett függvény paraméterei számunkra részeredmények, vagyis két jeggyel többre kell őket megadni, mint a kiindulási adataink értékes jegyeinek száma.

g. Gyakorlatilag teljesen mindegy milyen szimbólummal (x, Δ, o, stb.) jelöljük mérési pontjaink helyét a koordináta-rendszerben. Az azonban nagyon fontos, hogy az egyenes vagy görbe behúzása után is egyértelműen láthatók legyenek a mérési pontok.