ANALIZA ERORILOR DATELOR EXPERIMENTALE

Prof.univ.dr.ing Statescu Florian

- Metoda stintifica pentru cercetarea experimentala
- Elemente de statistica si probabilitati pentru analiza datelor experimentale
- Precizia si acuratetea datelor experimentale
- Erori sistematice si erori aleatoare
- Prelucrarea statistica a datelor experimentale
- Media ponderata. Combinarea datelor cu erori diferite
- Propagarea erorilor
- Analiza preciziei datelor experimentale
- Analiza acuratetii datelor experimentale

Metoda stintifica pentru cercetarea experimentala

- In primul rand se observa un fenomen intersant si/sau neobisnuit;
- Se formuleaza o ipoteza pentru o posibila explicatie a fenomenului observat;
- Experimente sunt (i) proiectate (planificate), and (ii) efectuate pentru a testa ipoteza formulata;
- Rezultatele experimentului sunt analizate pentru a determina daca acestea sustin ipoteza.

Pentru ca un experiment sa fie eficient, acesta trebuie sa permita stabilirea valabilitatii predictiilor teoriilor concurente. Atingerea acestui scop este determinata de capacitatea cercetatorului de a estima “calitatea” rezultatelor experimentale obtinute.

Veridicitatea rezultatelor din
stiintele naturii si inginerie
⇒ existenta rezultatelor experimentale

↓

Experiment ⇒ Proces de masurare

↓

Tema experimentala (de masurare):
ce trebuie facut?

↓

Strategia experimentala (de masurare):
cum trebuie procedat?

↓

Prelucrarea rezultatelor, formularea concluziilor, luarea deciziilor

Factori perturbatori ai procesului de masurare

Cauze:
- principiul sau metoda de masurare;
- mijloacele de masurare;
- mediul ambiant;
- obiectul supus masurarii;
- interactiunea obiect supus masurarii – mijloc de masurare;
- operator.

Conditii de referinta: temperatura, presiune, umiditate, vibratii etc.

Categorii de erori de masurare

- Erori grosolane ⇒ rezultate aberante (eliminate);
- Erori sistematice ⇒ prezinta repetabilitate; partial compensate, partial eliminate;
- Erori aleatoare ⇒ aparitie intamplatoare; pot fi diminuate, dar nu eliminate.

⇓

Prelucrarea rezultatelor masurarilor se face in conditiile prezentei (cel putin a) erorilor aleatoare de masurare.

m ⇒ adevarata valoare a marimii masurate; z_i ⇒ erori aleatoare; x_i ⇒ rezultatele masurarilor

$$x_i = m + z_i$$

Elemente de statistica si probabilitati pentru analiza datelor experimentale

- set de date experimentale (variabile) independente: x_i (i = 1…n)

- media datelor: $\bar{x} = \frac{1}{n}(x_1 + x_2 + ... + x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$

- dispersia (varianţa) datelor: $s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$

- abaterea medie patratica (deviaţia standard): $s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$

Elemente de statistica si probabilitati pentru analiza datelor experimentale

- percentila p: valoarea v_p pentru care $p\%$ din valorile celor n variabile sunt mai mici sau egale cu v_p;

- mediana: valoarea $\widetilde{x}$ pentru care jumatate din valorile variabilelor sunt mai mici sau egale cu $\widetilde{x}$, iar cealalta jumatate a valorilor variabilelor x_i sunt mai mari sau egale cu $\widetilde{x}$.

- modulul unui set de date experimentale este valoarea (valorile) din acel set de date corespunzatoare datei (datelor) cu cea mai mare frecventa de aparitie.

Elemente de statistica si probabilitati pentru analiza datelor experimentale

- densitatea de probabilitate: $P(x)$
  - x - discret: $P(x)$ = frecventa de obtinere a valorii x.

$$\sum P(x_i) = 1$$

  - x - continuu: $P(x)dx$ = probabilitatea de gasire a valorii x in intrevalul dintre x si $x + dx$.

$$\int P(x)dx = 1$$

- distribuţii cumulative (integrale):
  - x - discret: $P(x_1 \leq x \leq x_2) = \sum_{i=1}^{2} P(x_i)$
  - x - continuu: $P(x_1 \leq x \leq x_2) = \int_{x_1}^{x_2} P(x)dx$

Elemente de statistica si probabilitati pentru analiza datelor experimentale

- speranţa matematica a unei variabile (funcţii):
  - x - discret: $E[x] = \sum_i x_i \, P(x_i)$
  - x - continuu: $E[x] = \int x \, P(x) \, dx$
- momente ale distribuţiilor:
  - momentul de ordinul r al variabilei x in raport cu un punct fix x_0 este $E[(x-x_0)^r]$
  - momentul de ordinul 1 in raport cu 0:

    $\mu = E[x] = \int x \, P(x) \, dx =$ media (teoretica) a lui x
  - momentul de ordinul 2 in raport cu μ:

    $\sigma^2 = E[(x-\mu)^2] = \int (x-\mu)^2 \, P(x) \, dx =$ varianta (teoretica) a lui x

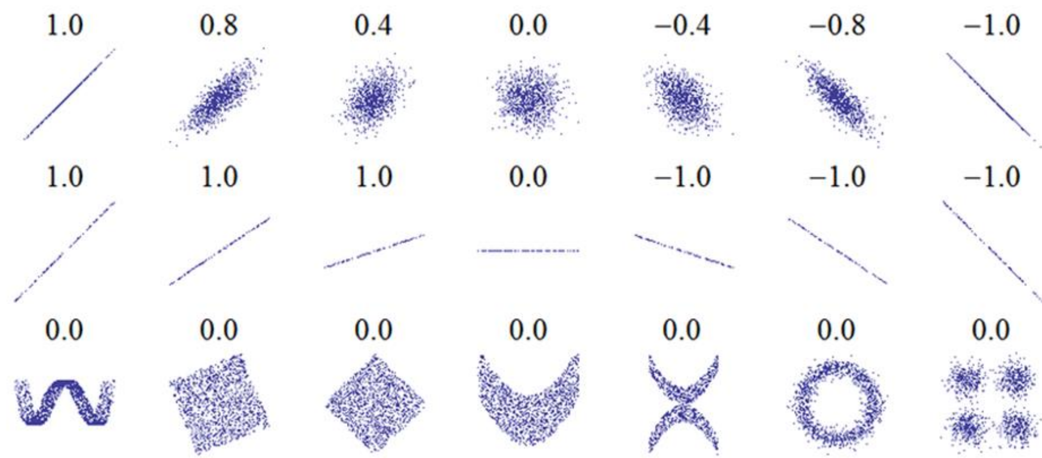
Elemente de statistica si probabilitati pentru analiza datelor experimentale

- covarianţa:

$$\mathrm{cov}(x,y) = E[(x-\mu_x)(y-\mu_y)]$$

- coeficient de corelaţie:

$$\rho = \frac{\mathrm{cov}(x,y)}{\sigma_x \sigma_y}$$



$|\rho| = 1$

x si y = perfect corelate liniar

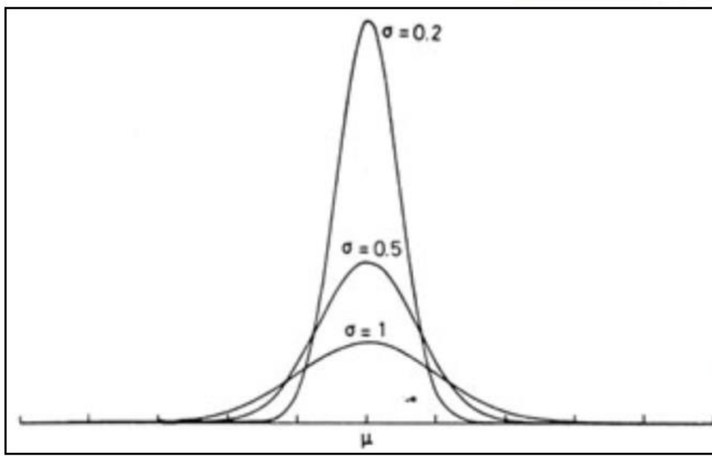
$\rho = 0$

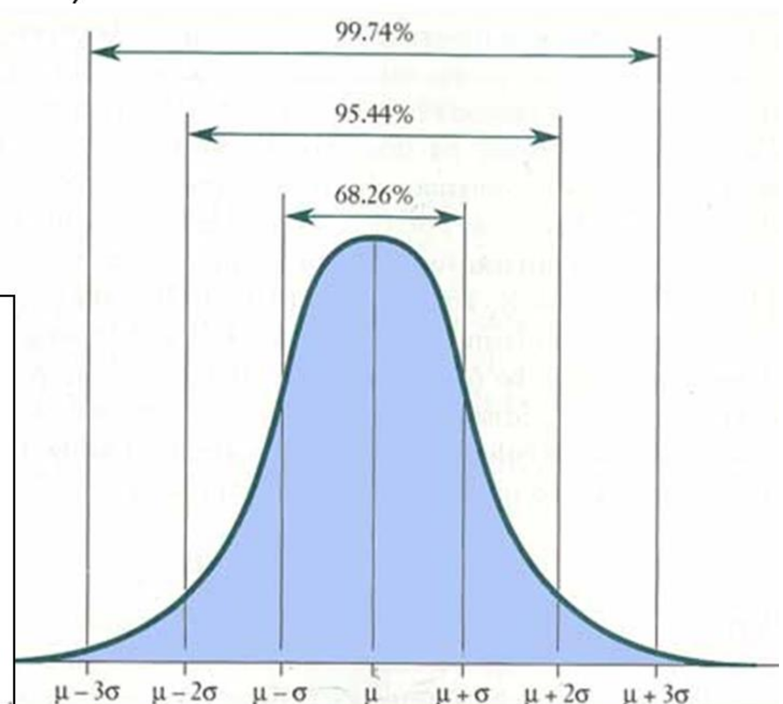
x si y = liniar independente

Elemente de statistica si probabilitati pentru analiza datelor experimentale

- distribuţia normala (Gauss):

$$P(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$





Elemente de statistica si probabilitati pentru analiza datelor experimentale

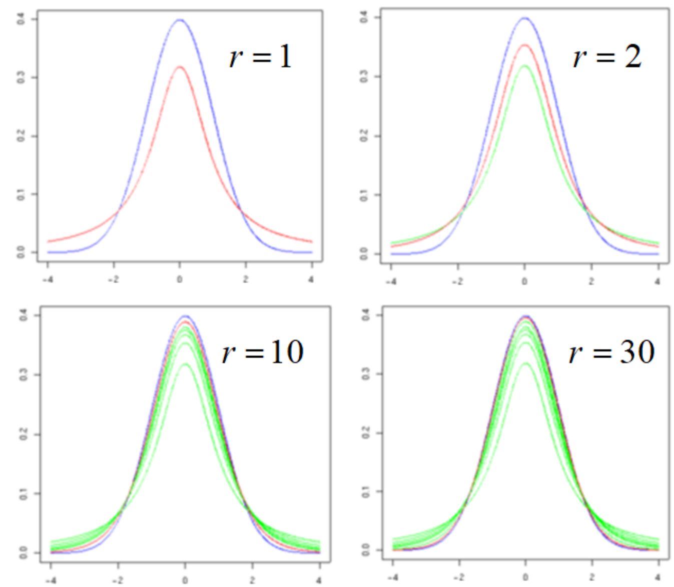
- distribuţia t :
  - utilizarea ei permite estimarea cu precizie mai mare a erorilor cu distributie normala, atunci cand dimensiunea esantionului (numarul de date experimentale dintr-un set) este redus.

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

$$P(t) = \frac{\Gamma\left[\frac{1}{2}(r+1)\right]}{\sqrt{r\pi}\,\Gamma\left(\frac{1}{2}r\right)\left(1+\frac{r^2}{r}\right)^{(r+1)/2}}, \quad r = n-1$$

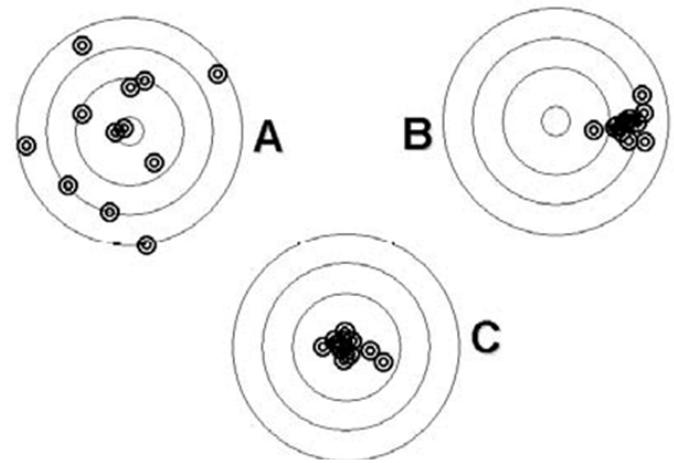
$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$$

$$\Gamma(n) = (n-1)!$$



Precizia si acuratetea datelor experimentale

- Precizia datelor experimentale
  - este o masura a reproductibilitatii masurarii repetate a aceleiasi proprietati.



- Acurateţea datelor experimentale
  - este o masura a apropierii datelor experimentale de valoarea reala a proprietatii masurate.

Erori sistematice

- Deviatii cu o cantitate fixa a rezultatului masuratorii de la valoarea exacta a marimii masurate.
- Erori de valoare constanta la repetarea masuratorii.
- Evidentiaza acuratetea datelor experimentale.
- Pot aparea datorita unor defectiuni ale aparatelor de masura, unor observatii umane gresite, sau unor schimbari ale conditiilor de masurare survenite in timpul masuratorilor.
- Exemple:
  - Instrumentul de masura are o eroare a zero-ului: un voltmetru indica 1 V chiar atunci cand este deconectat;
  - Instrumentul masoara eronat: un cromometru indica 100 s pentru o durata de 99 s.
- Eliminarea (minimizarea) lor se face prin calibrarea instrumentului, un control mai bun asupra conditiilor de masurare, sau imbunatatirea tehnicilor de masurare (calificarea operatorului) si NU prin efectuarea unui numar mare de masuratori.

Erori aleatoare

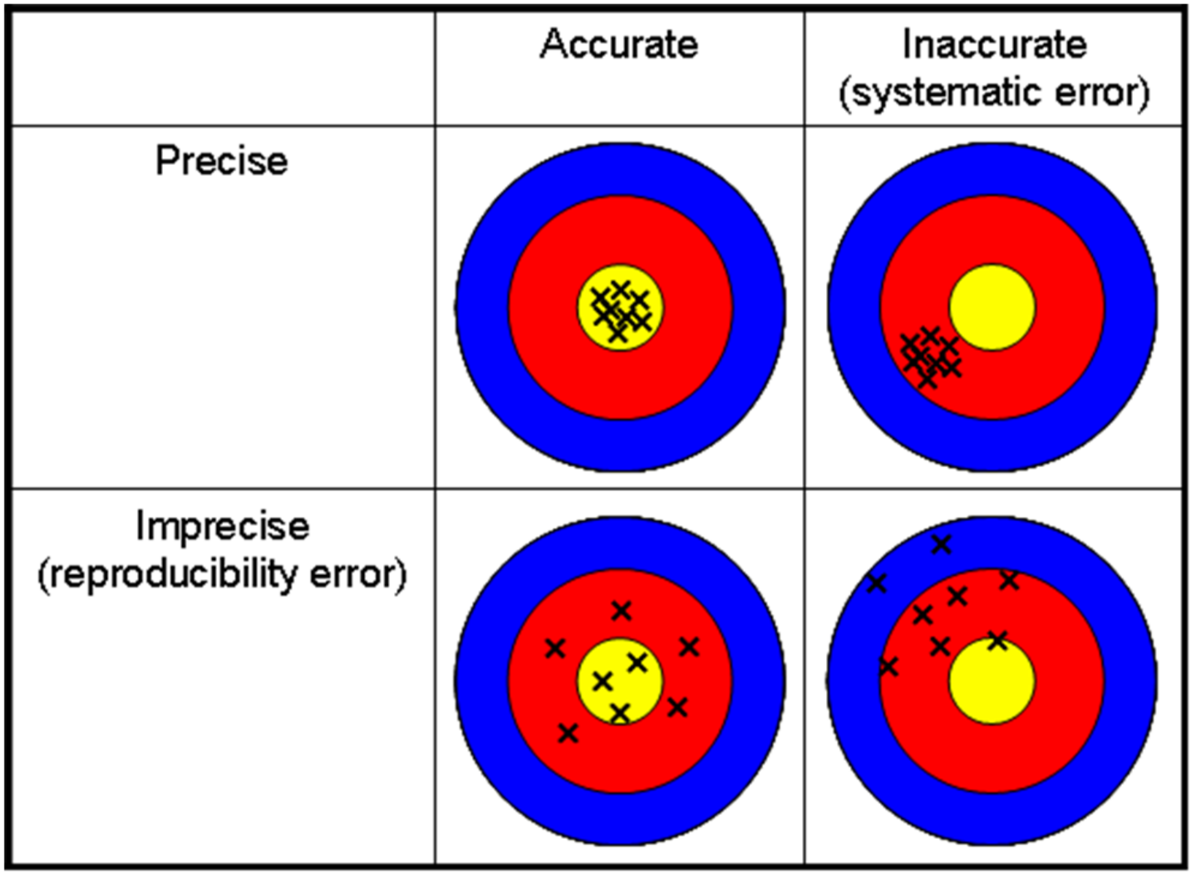
- Fluctuatii statistice (in orice sens) ale datelor masurate. Se mai numesc si *erori statistice*.

- Evidentiaza precizia masuratorilor si implicit reproductibilitatea acestora.

- Cauze:
  - Precizia limitata a instrumentului de masura
  - Inabilitatea experimentatorului de a efectua o masuratoare in acelasi fel pentru a obtine exact acelasi rezultat.

- Eliminarea (minimizarea) lor se face prin efectuarea unui numar mare de masuratori si printr-o analiza statistica a datelor experimentale.

Erori ale datelor experimentale



Prelucrarea statistica a datelor experimentale

- valoare masurata = valoare exacta ± eroare

- in general: distributia valorilor masurate ≈ distributie normala

- $x = \bar{x} \pm eroare$

  - $eroare = s$ ⟷ 68% interval de incredere
  - $eroare = 2s$ ⟷ 95% interval de incredere
  - $eroare = 3s$ ⟷ 99% interval de incredere

Prelucrarea statistica a datelor experimentale

- $\mu = \bar{x} \pm eroare\ \ a\ mediei$
  - deviaţia standard a mediei: $s_{\bar{x}} = \dfrac{s}{\sqrt{n}}$
  - $eroare\ a\ mediei = s_{\bar{x}}$ ⟷ 68% interval de incredere
  - $eroare\ a\ mediei = 2s_{\bar{x}}$ ⟷ 95% interval de incredere
  - $eroare\ a\ mediei = 3s_{\bar{x}}$ ⟷ 99% interval de incredere

Prelucrarea statistica a datelor experimentale

Set de date cu un numar redus de variabile :

$$\mu = \bar{x} \pm \lambda$$

$$\lambda = \frac{t\,s}{\sqrt{n}}$$

n	t (90%)	t (95%)	t (99%)
2	6.31	12.7	63.7
3	2.92	4.30	9.92
4	2.35	3.18	5.48
5	2.13	2.78	4.60
6	2.01	2.57	4.03
7	1.94	2.45	3.71
9	1.86	2.31	3.36
11	1.81	2.23	3.17
16	1.75	2.13	2.95
21	1.72	2.09	2.85
31	1.70	2.04	2.75
∞	1.64	1.96	2.58

Media ponderata. Combinarea datelor cu erori diferite

$$\bar{x}_p = \frac{\sum x_i / \sigma_i^2}{\sum 1/\sigma_i^2}$$

$$s(\bar{x}_p) = \sqrt{\frac{1}{\sum 1/\sigma_i^2}}$$

2,198 ± 0,001
2,197 ± 0,005
2,1948 ± 0,0010
2,203 ± 0,004
2,198 ± 0,002
2,202 ± 0,003
2,1966 ± 0,0020

$$\bar{x}_p = 2{,}19696$$

$$s(\bar{x}_p) = 0{,}00061$$

$$\mu = 2{,}1970 \pm 0{,}0006$$

Propagarea erorilor

- Date: x, σ_x, y, σ_y, $u = f(x, y)$
- Scop: calcularea σ_u ca functie de σ_x si de σ_y.

$$\sigma_u^2 = E[(u - \overline{u})^2]$$

- Se admite: $\overline{u} \cong f(\overline{x}, \overline{y})$

$$(u - \overline{u}) \cong (x - \overline{x}) \left.\frac{\partial f}{\partial x}\right|_{\overline{x}} + (y - \overline{y}) \left.\frac{\partial f}{\partial y}\right|_{\overline{y}}$$

$$E\left[(u - \overline{u})^2\right] \cong E\left[(x - \overline{x})^2 \left(\frac{\partial f}{\partial x}\right)^2 + (y - \overline{y})^2 \left(\frac{\partial f}{\partial y}\right)^2 + 2(x - \overline{x})(y - \overline{y}) \frac{\partial f}{\partial x} \frac{\partial f}{\partial y}\right]$$

Propagarea erorilor

$$\sigma_u^2 \cong \left(\frac{\partial f}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial f}{\partial y}\right)^2 \sigma_y^2 + 2\operatorname{cov}(x,y)\frac{\partial f}{\partial x}\frac{\partial f}{\partial y}$$

Cazuri particulare:

$$u = x + y \qquad \sigma_u^2 = \sigma_x^2 + \sigma_y^2 + 2\operatorname{cov}(x,y)$$

$$u = x - y \qquad \sigma_u^2 = \sigma_x^2 + \sigma_y^2 - 2\operatorname{cov}(x,y)$$

$\frac{\sigma_u}{u} =$ mare !!!

Propagarea erorilor

$$\sigma_u^2 \cong \left(\frac{\partial f}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial f}{\partial y}\right)^2 \sigma_y^2 + 2\,\mathrm{cov}(x,y)\frac{\partial f}{\partial x}\frac{\partial f}{\partial y}$$

Cazuri particulare:

$$u = xy$$

$$\sigma_u^2 \cong y^2\sigma_x^2 + x^2\sigma_y^2 + 2\,\mathrm{cov}(x,y)xy \qquad \frac{\sigma_u^2}{u^2} \cong \frac{\sigma_x^2}{x^2} + \frac{\sigma_y^2}{y^2} + 2\frac{\mathrm{cov}(x,y)}{xy}$$

$$u = x/y$$

$$\sigma_u^2 \cong y^{-2}\sigma_x^2 + x^2y^{-4}\sigma_y^2 - 2\,\mathrm{cov}(x,y)xy^{-3} \qquad \frac{\sigma_u^2}{u^2} \cong \frac{\sigma_x^2}{x^2} + \frac{\sigma_y^2}{y^2} - 2\frac{\mathrm{cov}(x,y)}{xy}$$

Analiza preciziei datelor experimentale

- **Masuri ale preciziei:**

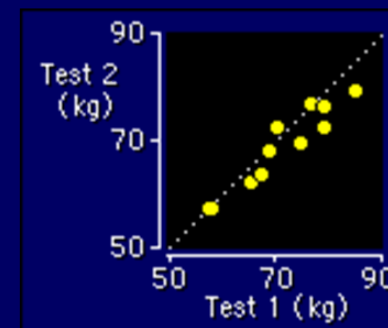
- modificarea mediei: diferenta dintre mediile rezultatelor a doua masuratori (teste) succesive ale aceleiasi proprietati.

- eroarea standard: deviatia standard a valorilor pentru fiecare variabila, de la o masuratoare (*test*) la alta (*retest*)

- corelatia test-retest: coeficientul de corelatie intre o prima masuratoare (test) si o a doua masuratoare (retest):

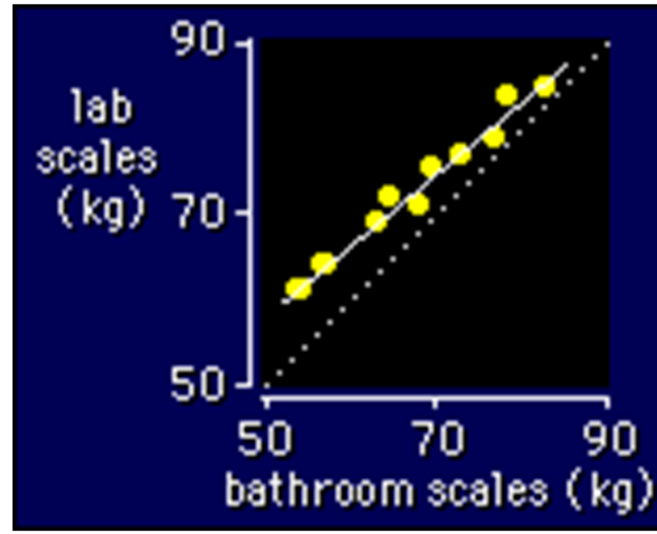
$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y},$$

Test 1	Test 2
57.5	57.4
65.6	63.2
67.0	66.5
68.5	69.9
70.8	72.8
72.2	70.1
74.9	75.6
76.0	75.2
76.1	72.8
83.1	79



Analiza acuratetii datelor experimentale

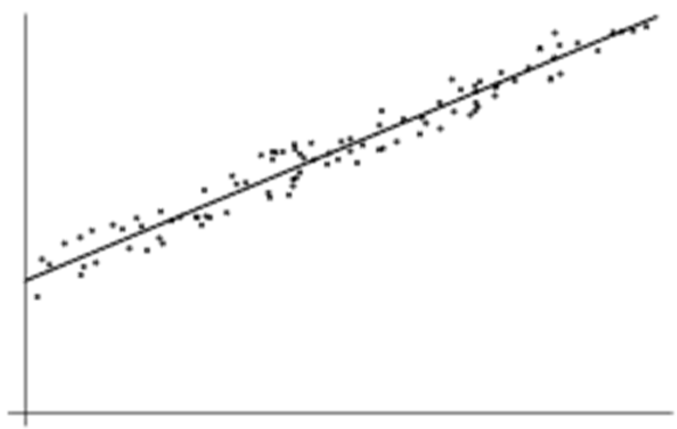
- Masuri ale acuratetii:
- ecuatia de estimare: ecuatia care aproximeaza cel mai bine valorile experimentale fata de valorile reale.
- eroarea standard: deviatia standard a valorilor masurate fata de valorile date de ecuatia de estimare.
- coeficientul de corelatie: coeficientul de corelatie intre valorile masurate si valorile reale:

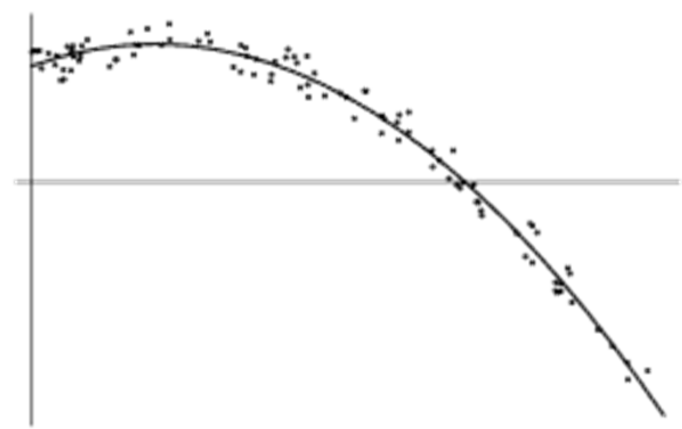


Analiza acuratetii datelor experimentale

- **Metoda celor mai mici patrate pentru determinarea ecuatiei de estimare**

- Scopul metodei: gasirea unei curbe care aproximeaza un set de puncte prin minimizarea sumei patratelor reziduurilor punctelor fata de curba

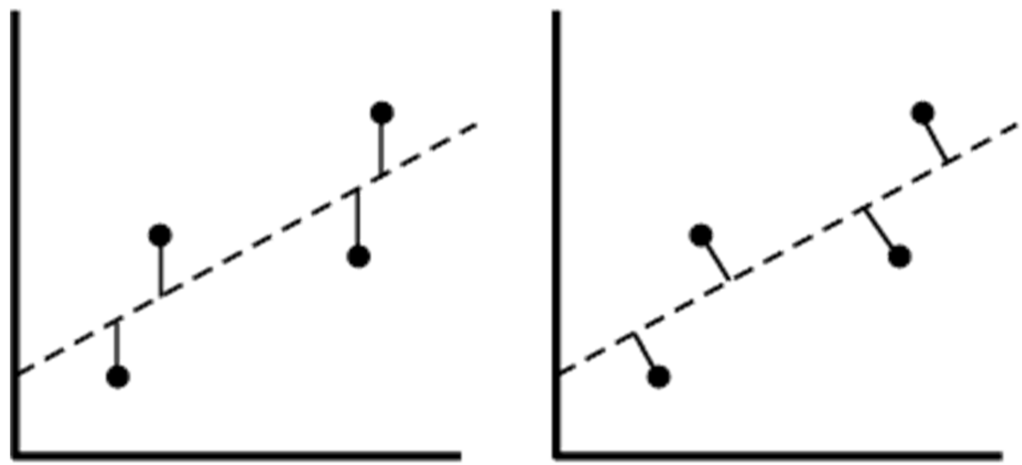




Analiza acuratetii datelor experimentale

- **Metoda celor mai mici patrate pentru determinarea ecuatiei de estimare**

- Scopul metodei: gasirea unei curbe care aproximeaza un set de puncte prin minimizarea sumei patratelor reziduurilor punctelor fata de curba



reziduuri verticale reziduuri perpendiculare

Analiza acuratetii datelor experimentale

- Metoda celor mai mici patrate pentru determinarea ecuatiei de estimare

- Se cauta $f(x)$ astfel incat expresia E sa fie minima:

$$E = \sum_{i=1}^{n} (y_i - f(x_i))^2$$

Metoda celor mai mici patrate pentru determinarea ecuatiei de estimare

$$f(x) = \sum_{i=1}^{p} c_i \varphi_i(x)$$

$p < n$

$\varphi_i(x)$ se aleg, se numesc functii de baza

$\Rightarrow$ $E(c_1, c_2, \ldots, c_p)$ minima $\Rightarrow$

$$\frac{\partial E}{\partial c_1} = 0,\quad \frac{\partial E}{\partial c_2} = 0,\quad \vdots \quad \frac{\partial E}{\partial c_p} = 0.$$

$\Rightarrow$ c_i

Metoda celor mai mici patrate pentru determinarea ecuatiei de estimare

- Regresia liniara:

$$f(x) = a + bx \qquad p = 2, \quad c_1 = a, \quad c_2 = b, \quad \varphi_1(x) = 1, \quad \varphi_2(x) = x$$

$$E = \sum_{i=1}^{n} \left[y_i - (a + bx_i) \right]^2$$ minima

$$\Rightarrow \quad \frac{\partial E}{\partial a} = 0, \quad \frac{\partial E}{\partial b} = 0 \quad \Rightarrow \quad -2\sum_{i=1}^{n} \left[y_i - (a + bx_i) \right] = 0, \quad -2\sum_{i=1}^{n} \left[y_i - (a + bx_i) \right] x_i = 0 \quad \Rightarrow$$

Metoda celor mai mici patrate pentru determinarea ecuatiei de estimare

$$-2\sum_{i=1}^{n}\left[y_i-(a+bx_i)\right]=0 \quad\Rightarrow\quad \sum_{i=1}^{n} y_i-\sum_{i=1}^{n} a-\sum_{i=1}^{n} bx_i=0 \quad\Rightarrow$$

$$-2\sum_{i=1}^{n}\left[y_i-(a+bx_i)\right]x_i=0 \qquad \sum_{i=1}^{n} y_i x_i-\sum_{i=1}^{n} ax_i-\sum_{i=1}^{n} bx_i^{2}=0$$

$$na+\left(\sum_{i=1}^{n} x_i\right)b=\sum_{i=1}^{n} y_i$$

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

$$\bar{x}=\frac{\sum_{i=1}^{n} x_i}{n},\quad \bar{y}=\frac{\sum_{i=1}^{n} y_i}{n} \quad\Rightarrow\quad \boxed{b=\frac{\sum_{i=1}^{n} x_i y_i-n\bar{x}\,\bar{y}}{\sum_{i=1}^{n} x_i^{2}-n\,\bar{x}^{2}}}$$

Metoda celor mai mici patrate pentru determinarea ecuatiei de estimare

$$a = \frac{\sum_{i=1}^{n} y_i - b\sum_{i=1}^{n} x_i}{n} = \bar{y} - b\,\bar{x}$$

$$\boxed{a = \frac{\bar{y}\sum_{i=1}^{n} x_i^2 - \bar{x}\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2}}$$

Scrierea se simplifica daca se folosesc notatiile sumelor patratelor : *sum of square* ss

$$ss_{xx} = \sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 = \sum_{i=1}^{n}\left(x_i^2 - 2x_i\,\bar{x} + \bar{x}^2\right) = \sum_{i=1}^{n} x_i^2 - n\,\bar{x}$$

$$ss_{yy} = \sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 = \sum_{i=1}^{n} y_i^2 - n\,\bar{y}$$

$$ss_{xy} = \sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right) = \sum_{i=1}^{n} x_i y_i - n\,\bar{x}\,\bar{y}$$

$$\sigma_x^2 = \frac{ss_{xx}}{n}$$

$$\sigma_y^2 = \frac{ss_{yy}}{n}$$

$$\operatorname{cov}(x,y) = \frac{ss_{xy}}{n}$$

$$\boxed{\begin{aligned} b &= \frac{ss_{xy}}{ss_{xx}} = \frac{\operatorname{cov}(x,y)}{\sigma_x^2} \\ a &= \bar{y} - b\,\bar{x} \end{aligned}}$$

Elemente de statistica si probabilitati pentru analiza datelor experimentale

- covarianţa:

$$\mathrm{cov}(x,y) = E[(x-\mu_x)(y-\mu_y)]$$

- coeficient de corelaţie:

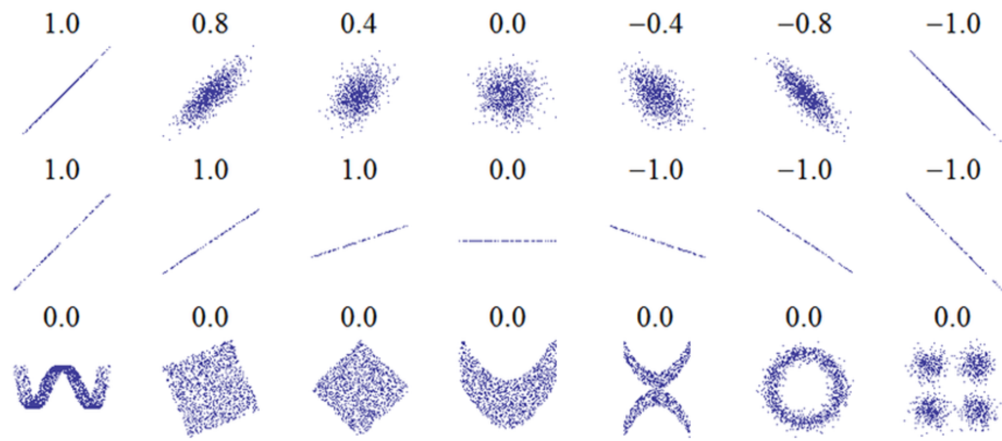
$$\rho = \frac{\mathrm{cov}(x,y)}{\sigma_x \sigma_y}$$

$$|\rho| = 1$$

x si y = perfect corelate liniar

$$\rho = 0$$

x si y = liniar independente



Metoda celor mai mici patrate pentru determinarea ecuatiei de estimare

- Calitatea aproximarii este data de coeficientul de corelare *r*

$$\rho = \frac{\mathrm{cov}(x,y)}{\sigma_x \sigma_y}$$

$$\mathrm{cov}(x,y) = \frac{ss_{xy}}{n}$$

$$y = a + bx \Rightarrow b = \frac{ss_{xy}}{ss_{xx}}$$

$$x = a' + b'y \Rightarrow b' = \frac{ss_{xy}}{ss_{yy}}$$

$$\boxed{r^2 = bb' = \frac{ss_{xy}^2}{ss_{xx} ss_{yy}}}$$

$$\boxed{r = \sqrt{\frac{ss_{xy}^2}{ss_{xx} ss_{yy}}} = \frac{ss_{xy}}{\sqrt{ss_{xx} ss_{yy}}}}$$

Metoda celor mai mici patrate pentru determinarea ecuatiei de estimare

- Erorile standard pentru coeficientii *a* si b

$$SE(a) = s\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{ss_{xx}}}$$

$$SE(b) = \frac{s}{\sqrt{ss_{xx}}}$$