Elemente de teoria erorilor si incertitudinilor
Calcule statistice si modele de aproximare

➔"*Să măsurăm ce se poate măsura şi să facem măsurabil ceea ce nu se poate măsura încă*." **Galileo Galilei**

1. *Introducere în teoria erorilor: erori de măsurare si reprezentare, distribuția erorilor, parametri caracteristici, propagarea erorilor*
2. *Calcule statistice: indicatori statistici, corelații între seturi de măsurători, modele de corelație empirice şi teoretice*

Generalitati despre erori, incertitudini si aproximari

In sens larg cuvantul "*eroare*" inseamna *greseala, incertitudine, nesiguranta*, etc. Prin *greseala* intelegem un fapt realizat de om in activitatea profesionala, sociala, economica, etc. privind un rationament gresit, o metoda aplicata gresit, un instrument utilizat gresit, o atitudine ce contrazice regulile morale, sociale sau legistative, neintelegeri ale unor notiuni, termeni sau concepte din limbajul stiintific, economic, social, etc. Prin *incertitudine* se intelege lipsa de certitudine, indoiala asupra unor rationamente, calcule, sau experimente, iar in domeniul social poate reprezenta starea unei persoane lipsite de siguranta, de hotarare. In doate domeniile exista incertitudini, de exemplu in domeniul stiintific s-au dezvoltat diverse teorii care "controleaza" incertitudinile:

- **logica matematica bivalenta** (cu 2 valori: *true*, *false*; logica propozitiilor, logica predicatelor, logica relatiilor) ofera metode si tehnici certe (logica matematica are aplicatii in electrotehnica-studiul schemelor cu relee, al schemelor electronice-, in cibernetica-teoria automatelor, tehnica programarii-, in neurofiziologie-modelarea sistemelor neuronale-, lingvistica - lingvistica matematica, etc.); sistemele de calcul folosesc limbajul binar pentru procesarea informatiilor; pentru rezolvarea diverselor probleme complexe a fost necesara conceperea unor teorii *de* **logica matematica trivalente si cu mai multe valori** (primele sisteme de logica polivalenta au fost construite de *J. Lukasiewicz* (1920), *E. Post* (1921) si de *Grigore C. Moisil* (1963)); În limbajul de manipulare a datelor SQL (*Structured Query Language*), o stare de adevăr TRUE pentru o expresie (de exemplu într-o clauză WHERE) inițializează o acțiune pe un rând (returnează un rând), în timp ce o stare de adevăr UNKNOWN sau FALSE nu face acest lucru. În acest fel, logica trivalentă este implementată în SQL, şi se comportă ca logică bivalentă pentru utilizatorul SQL; limbajul Prolog (programare in logica), limbaj al Inteligentei artificiale este conceput si elaborat avand la baza logica de ordinul I (cuantificatorii oricare($\forall$) si exista ($\exists$) opereaza doar asupra variabilelor).
- **teoria logicii si multimilor fuzzy** (suport pentru studiul incertitudinii si impreciziei; aplicatii in analiza fenomenelor si proceselor, fiabilitatea sistemelor, uzura produselor, gradul de utilizare a produselor sau masinilor, procesarea imaginilor, etc.). Incompletitudinea unei informații/date se exprimă pe două scări: scara incertitudinii se referă la încrederea care i se acordă informației (dacă sursa de informație, instrumentul de măsură sau expertul sunt siguri, demni de încredere, informația este certă), scara impreciziei se referă la conținutul

informaţional (informaţia este precisă dacă mulţimea valorilor specificate în enunţul corespunzător este o valoare unică). Există fenomene si procese în care gradualitatea şi ambiguitatea joacă un rol important (imprecizie nu este de tip aleator). Problema inseamna faptul de a putea aprecia în ce măsură un obiect dat aparţine unei clase ale cărei margini nu pot fi precizate clar. Clasa de obiecte are grade de apartenenţă continue. O astfel de mulţime este caracterizată de o funcţie de apartenenţă ce atribuie fiecărui obiect un grad de apartenenţă între 0 şi 1.

Sunt cunoscute exemple de oameni de stiinta din matematica, fizica, chimie, etc. ce au facut greseli in cercetarile/teoriile lor (exista cazuri cand s-au facut descoperiri stiintifice in mod intamplator, de ex. *razele X*, *Penicilina*, *Viagra*, etc.):

- exemple relevante pentru matematica sunt prezentate in Alexandru Froda (1894-1973), Eroare şi paradox în matematică, Editura Enciclopedică Română, 1971.
- sute de lucrari stiintifice sunt retrase in fiecare an, din cauza documentarilor superficiale, plagiatului sau analizelor gresite; de exemplu: "*Apendicita se tratează cu antibiotice. The Journal of Gastrointestinal Surgery a publicat în 2009 un studiu al unor cercetători indieni care susţineau că antibioticele sunt o metodă mai sigură decât îndepărtarea chirurgicală a apendicelui. Ei au fost contestaţi de chirurgi italieni, iar studiul a fost retras din publicaţie pe motiv de plagiat*." (**Sursa: LiveScience**);
- inventii atribuite gresit - Conceptul de computer desktop-"oficial": Microsoft (prin Windows), real: Xerox PARC; Razele X- Inventator "oficial": Thomas Edison, real: Wilhelm Rontgen; Becul- Inventator "oficial": Thomas Edison, real: Sir Humphry Davy; Radioul- Inventator "oficial": Guglielmo Marconi, real: Nikola Tesla (**Sursa: http://www.descopera.ro/)**

Analiza datelor experimentale: Tipuri de erori

In Chimie si Fizica (precum si in alte stiinte ingineresti), metodele folosite la *masurarea parametrilor* (marimi fizice sau chimice) sunt în general precise. Totusi, în timpul masuratorilor pot interveni diferiti factori perturbatori care genereaza aparitia *erorilor de masurare*. Pentru determinarea marimilor fizice sau chimice se folosesc instrumente de masura, care au o anumita precizie. Nici o masuratoare nu este absoluta. Masurând de mai multe ori aceeasi marime fizica, în aceleasi conditii, cu aceleasi mijloace, se poate observa ca rezultatele obtinute sunt diferite. *Diferentele* ce apar depind de constructia instrumentelor de masura, de observator, sau de alti factori perturbatori. *Acuratetea* unui experiment arata cât de aproape este rezultatul masuratorii de valoarea adevarata. Prin urmare, acuratetea este o masura a corectitudinii rezultatelor obtinute prin masurare si prin calcul. *Precizia* unui experiment este o masura a exactitatii determinarii rezultatelor.

Procedurile de observare statistica in analiza fenomenelor si proceselor pot fi afectate de *erori*. Prelucrarea statistica a datelor experimentale prin calculele matematice ce urmeaza a fi efectuate cu datele respective, contribuie cu o anumita cantitate de erori. De aceea, specialistii stiu ca atât *erorile de observare statistica* cât si *cele de calcul*, vor afecta rezultatele obtinute din prelucrarea si interpretarea datelor experimentale. De aceea, ne propunem sa examinam în acest capitol atât sursele de erori cât si modul în care acestea influenteaza rezultatele finale.

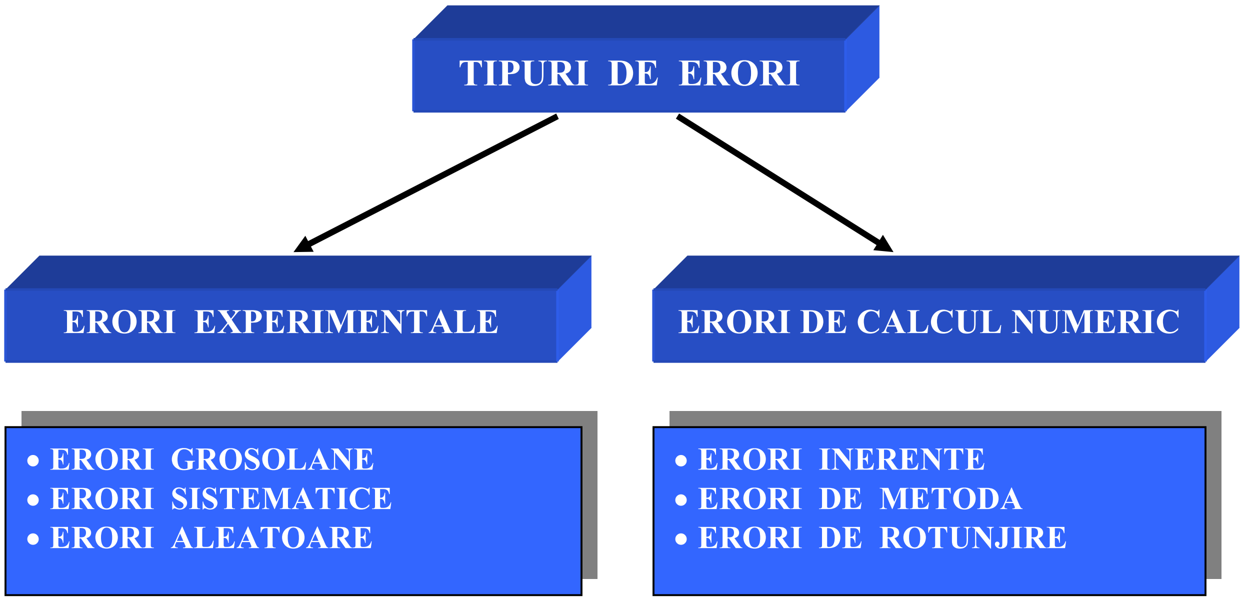


Figura 14. Tipuri de erori

Erorile se clasifica in doua mari categorii:

1. **erori experimentale** – efectuarea masuratorilor pot produce erori care au aceeasi marime, când procesul de masurare se efectueaza în conditii identice, sau erori care au marimi variabile, variatia acestora fiind supusa unei anumite legi de variatie; erorile de masurare se clasifica în:
   - ***erori grosolane (greseli)***: pot proveni din aplicarea unor metode de calcul inexacte, din citiri eronate, din neatentia sau lipsa de instruire a personalului; aceste erori trebuie eliminate si refacute masuratorile;
   - ***erori sistematice***: pot proveni din cauza unor caracteristici constructive ale aparatelor, incorectei etalonari sau uzurii; pot fi erori produse de metoda de masurare sau erori produse de factori externi (erori de influenta), deosebit de greu de evaluat prin calcule, deoarece nu întotdeauna pot fi cunoscute cauzele si legile de variatie în timp a conditiilor de mediu (temperatura, presiunea, umiditatea, câmpuri magnetice, radiatii, etc.) ;
   - ***erori aleatoare (accidentale, întâmplatoare)***: pot proveni ca urmare diversitatii proceselor si fenomenelor precum si a interactiunilor experimentului cu alte procese si fenomene ce se desfasoara simultan; nu este posibila depistarea si înlaturarea lor, efectul global fiind producerea unor erori aleatorii inevitabile ce nu pot fi înlaturate din rezultatele masuratorilor;
2. **erori de calcul numeric** - interpretarea matematica a datelor reprezinta totalitatea operatiilor matematice ce trebuie efectuate pentru obtinerea unui anumit rezultat, în vederea caruia au fost efectuate masurarile respective. În timpul efectuarii acestor calcule, pot interveni anumite erori ce se vor adauga la erorile experimentale, si astfel *valoarea masurata* sa se abata si mai mult fata de *marimea adevarata*; se disting urmatoarele categorii de erori de calcul:

- ***erori inerente***: pot proveni ca urmare a folosirii aproximative a unor valori provenite din masuratori, a utilizarii in calcule a numerelelor irationale (π, e, $\sqrt{2}$) sau ca urmare a calculelor aproximative (*serii numerice*) oferite de calculatoarele numerice; trebuie specificat faptul ca multe valori ale unor functii obisnuite (*sin*, *cos*, *lg*, etc.) sunt obtinute prin calculul aproximativ al valorii unor serii numerice;
- ***erori de metoda***: analiza si interpretarea datelor experimentale depind de experienta specialistilor care efectueaza prelucrarea datelor experimentale; matematica si in special analiza numerica ofera o multitudine de metode si tehnici de rezolvare a problemelor in acest caz; unele din aceste metode sunt mai eficiente sau nu pentru un anumit caz, de aceea, alegerea metodei este foarte importanta pentru rezultatul final care se doreste a fi obtinut cu o anumita *eroare de aproximare*; de remarcat este faptul ca determinarea solutiilor se realizeaza prin procese iterative, *numarul de iteratii* determinand eroarea de aproximare;
- ***erori de rotunjire***: aceste erori sunt inevitabile deoarece depind de posibilitatile limitate de reprezentare a numerelor în memoria calculatoarele numerice; orice calculator, indiferent cat de performant este construit, poate reprezenta numerele cu un numar redus de *cifre semnificative*, depinzând de lungimea *cuvântului de memorie* (numarul de biti: 32 sau 46) utilizat la stocarea unui numar; calculatoarele actuale ofera calcule pentru numerele reale cu maxim *7 cifre semnificative* în *simpla precizie,* si cu maxim *15 cifre semnificative* în *dubla precizie*.

Termeni si concepte despre erori

- *Eroarea reala* este definita ca diferenta dintre valoarea reala (corecta) a unei marimi y si valoarea masurata (aproximativa) y' a marimii, adica $\Delta y = y - y'$. In cazul in care $y' <$ y, marimea respectiva este aproximata prin lipsa, altfel aproximatia este prin exces sau adaos.
- *Eroarea absoluta* - uneori nu se cunoaste semnul erorii $\Delta y = y - y'$, de aceea se foloseste notiunea de *eroare absoluta* care este definita prin relatia $\Delta y = | y - y' |$.
- *Eroarea relativa* se defineste ca raportul dintre eroarea absoluta si valoarea absoluta a marimii exacte, adica

$$\delta_y = \frac{|y - y'|}{|y|} = \frac{\Delta y}{|y|}$$

Eroarea relativa se poate exprima si în procente, adica

$$\delta_{y\%} = \frac{\Delta y}{|y|} \times 100[\%]$$

- *Eroarea absoluta limita* – in cazul in care valoarea marimii y nu este cunoscuta, se introduce notiunea de *eroare absoluta limita* ε_y corespunzatoare valorii aproximative y'; valoarea acestei erori reprezinta cel mai mic numar pozitiv care contine una sau mai multe cifre semnificative, ales în asa fel, încât sa putem fi siguri ca *eroarea absoluta* comisa, în cazul respectiv, nu depaseste acest numar; prin urmare avem urmatoarea relatie

$$\Delta y = | y - y' | \le \varepsilon_y \text{ , adica } \quad y' - \varepsilon_y \le y \le y' + \varepsilon_y ,$$

ceea ce inseamna ca valoarea *y* este aproximata prin lipsa, respectiv adoaos.

- *Incertitudine de masurare* (Δ) reprezinta intervalul în care se estimeaza, cu o anumita probabilitate, ca se afla valoarea adevarata a marimii *y*;
- *Eroarea conventionala* - În realitate valoarea adevarata a unei marimi nu poate fi cunoscuta, de aceea este necesar sa se adopte o valoare de referinta, care are un caracter conventional. Se defineste astfel eroarea conventionala ca diferenta dintre valoarea masurata si valoarea de referinta y_{conv} admisa adica $\Delta y_{conv} = y_{conv} - y'$.

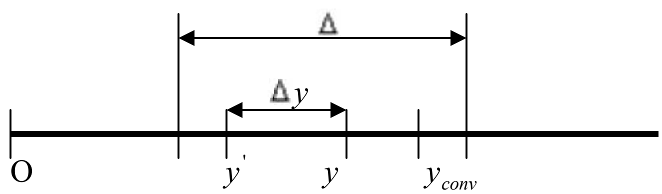


Figura 15. Erori de masurare

Erori de trunchiere si erori de rotunjire

Metodele numerice oferite de analiza matematica impreuna cu implementarea algoritmilor eficienti din domeniul informaticii sunt utilizate cu succes la multe probleme complexe din toate domeniile stiintifice, tehnice, economice, etc. Cu toate acestea, trebuie sa se cunoasca corect gradul de precizie privind obtinerea solutiilor in aceste rezolvari de probleme. Am vazut mai sus ca varietatea si combinarea diverselor erori (*de masurare, de calcul, de aproximare, de rotunjire*, etc.) pot sa conduca la rezultate ce nu raspund exigentelor practice. Acest lucru este si mai complicat cand in diverse situatii (la fizica, chimie, etc.) trebuie sa se realizeze calcule cu valori foarte mari, dar si cu zecimale foarte multe care depasesc performanta calculatoarelor actuale (de exemplu aritmetica modala).

Calculele matematice si operatiile implementate in algoritmii de calcul pentru calculatoarele numerice utilizeaza aproximarea cu serii numerice si dezvoltarea functiilor analitice prin descompunere de tip *Taylor* si de tip *Mac-Laurin*. Dezvoltarile in serii numerice se utilizeaza la obtinerea rezultatelor cu mai multe zecimale exacte, si anume se tine seama de precizia dorita 10^{-p} , unde p reprezinta numarul de zecimale exacte. De exemplu, pentru calculul valorii *ln2* cu p=2 zecimale exacte, folosind dezvoltarea in serie alternanta,

$$\ln 2 = \sum_{i=1}^{\infty} (-1)^{i+1} \frac{1}{i}$$

trebuie sa se calculze suma seriei pana la n=99 (*trunchiere de rang 99*). In practica, exista alte reprezentari care sunt mai eficiente decat cazul n=99, si anume trunchierea se realizeaza la un rang mai mic. Ex.: Calculul valorii *sin(2)* cu eroarea 10^{-7} este 0.909297. Folosind programul Excel se obtine valoarea **0.909297427**, cu 9 zecimale exacte si valoarea **0.909297426825682**, cu 15 zecimale exacte.

Programul EXCEL ofera pentru calcule si reprezentarea valorilor reale urmatoarele formate:

- **Number** – decimal places, de exemplu 345.67845634322 cu p=11 zecimale exacte;
- **Scientific** – forma exponentiala $xE \pm nm$, unde nm reprezinta exponentul lui 10, adica $x10^{\pm nm}$, de exemplu 3.45678456343E+02;
- **Fraction** –forma fractionala de diverse tipuri, de exemplu 345 211/311 .

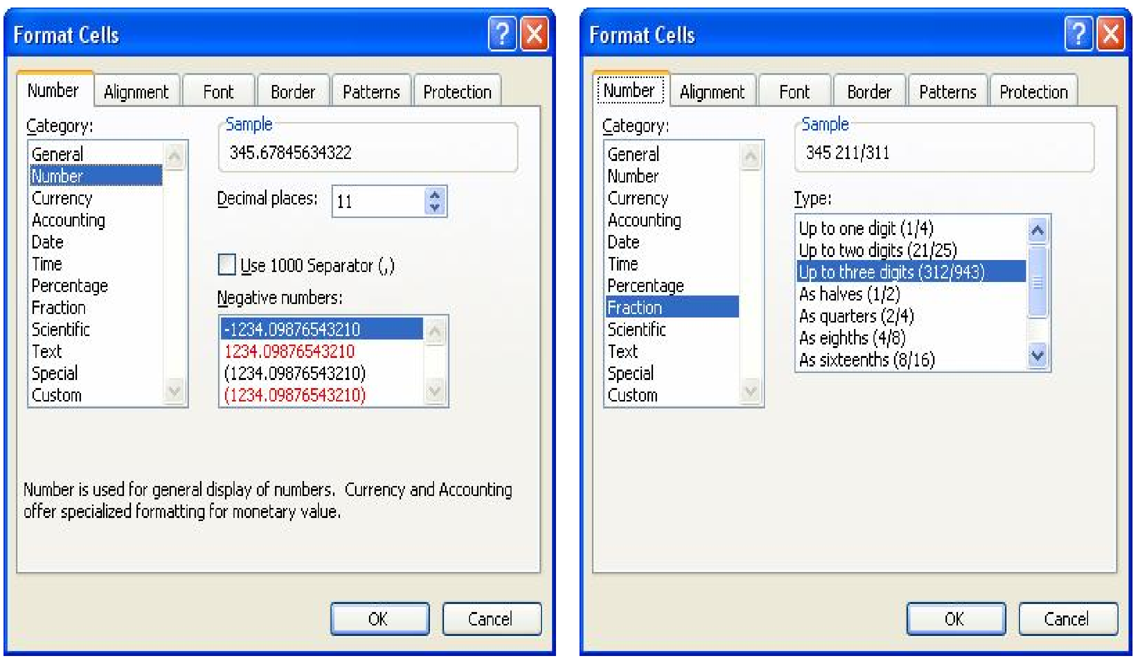


Figura 16. Fereastra "*Format Cells*"

O functie reala $f : I \to R$ derivabila de o infinitate de ori in $x_0 \in I \subset R$ este *analitica* in punctul x_0 daca exista relatia

$$f(x)=f(x_0)+\sum_{i=1}^{\infty}\frac{f^{(i)}(x_0)}{i!}(x-x_0)^i ,$$

pentru $\forall\ x \in (x_0-\alpha\ ,\ x_0+\alpha\ ) \subset I$, unde $\alpha > 0$ este un numar real dat.

Orice functie analitica se descompune in *polinomul Taylor* de ordinul *n* si in *restul seriei Taylor* de ordinul *n*, adica $f(x) = T_n\ (x) + R_n\ (x)$, unde

$$T_n(x)=f(x_0)+\sum_{i=1}^{n}\frac{f^{(i)}(x_0)}{i!}(x-x_0)^i$$, si restul de la rangul *(n+1)*

$$R_n(x) = \sum_{i=n+1}^{\infty} \frac{f^{(i)}(x_0)}{i!}(x - x_0)^i .$$

Restul seriei Taylor de ordinul n se poate reprezenta sub forma *Lagrange*, adica

$$R_n(x) = \frac{f^{n+1}(\theta)}{(n+1)!}(x - x_0)^{n+1}, \text{ unde } \theta \in (x_0, x) \text{ sau } \theta \in (x, x_0).$$

Functiile elementare (*sin*, *cos*, *ln*, etc.) sunt functii reale analitice ce au proprietatea ca restul seriei lui Taylor tinde la 0. Mai jos sunt exemple de dezvoltari de tip Mac-Laurin pentru $x_0 = 0$.

$$\ln 2 = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \ldots + (-1)^{n+1}\frac{1}{n} + \ldots$$

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \ldots + \frac{x^n}{n!} + \ldots$$

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} + \ldots + (-1)^n \frac{x^{2n+1}}{(2n+1)!} + \ldots$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} + \ldots + (-1)^n \frac{x^{2n}}{(2n)!} + \ldots \text{ etc.}$$

Reprezentarea in virgula mobila a numerelor reale

Calculatoarele actuale utilizeaza reprezentarea in virgula mobila a numerelor reale. Daca b este o baza de numeratie (se presupune numar par) si p este o precizie (*numar de cifre semnificative)*, atunci reprezentarea unui numar real in virgula mobila are urmatoarea forma:

$\pm (c_0 + \sum_{k=1}^{p-1} \frac{c_k}{b^k}) b^E$, cu cifrele semnificative $c_k \in \{0,1, ..., b-1\}, \forall k = \{0,1, ..., p-1\}$, E fiind *exponentul marginit* $E_{\min} \leq E \leq E_{\max}$.

Tabelul de mai jos exemplifica cei patru parametri (*baza, precizia, valorile limita ale exponentului*) ce caracterizeaza reprezentarea în virgula mobila în diverse sisteme(IEEE-*Institute of Electrical and Electronics Engineers*):

Sistem reprezentare	Baza b	Precizia p	$E_{\min}$	$E_{\max}$
IEEE single-precission	2	24	-126	127
IEEE double-precission	2	53	-1022	1023
Cray	2	48	-16383	16384
Calculator HP	10	12	-499	499
Mainframe IBM	16	6	-64	63

Tabelul 1. (Ref.: http://www.utgjiu.ro/math/mbuneci/book/mn2007/c04.pdf)

Reprezentarea in virgula mobila in *forma normalizata* este reprezentarea unui numar y sub forma
$y = f\, b^E,\; b^{-1} \le |f| \le 1$, unde *f* reprezinta mantisa, iar *E* exponentul.

Reprezentarea normalizata a numerelor reale are urmatoarele avantaje:

- reprezentarea fiecarui numar este unica;
- nu se pierd cifre pentru reprezentarea primele zerourilor de la dreapta virgulei;
- în sistemul binar (baza b =2) prima cifra poate sa nu mai fie stocata (deoarece este întotdeauna 1).

Un numar real cu mai multe cifre semnificative este rotunjit la numarul de cifre maxim. Acest lucru se realizeaza prin rotunjirea mantisei. Alte rotunjiri se efectueaza în decursul operatiilor. Aproximarea unui numar real cu cele doua forme de reprezentare se numeste *tehnica de rotunjire* ce introduce *eroarea de rotunjire*. Exista mai multe modalitati de rotunjire:

- **trunchiere** (rotunjire prin taiere) – se retin primele p cifre din reprezentarea normalizata;
- **rotunjire la cel mai apropiat in virgula mobila** (rotunjire la par) – forma in virgula mobila este cel mai apropiat numar de numarul aproximat.

Rotunjirea la par determina o acuratete mai mare a reprezentarii. Acuratetea sistemului în virgula mobila este caracterizata de asa-numita *precizie a masinii* ε_{mach}. Daca regula de rotunjire este trunchierea, atunci $\varepsilon_{mach} = b^{1-p}$, iar daca regula de rotunjire este *rotunjirea la par* atunci $\varepsilon_{mach} = \frac{1}{2} b^{1-p}$.

Cazuri speciale: conceperea de metode si algoritmi noi

Exemplul 1: Puterile mari ale lui 2.

Exista cazuri in (in chimie, fizica, etc.) in care trebuie sa se lucreze in calcule cu numere foarte mari. In acest caz, trebuie sa se cunoasca foarte bine limitele oferite de calculatoare privind reprezentarea numerelor si modul de calcul pentru toate operatiile. Pe langa teoriie (aritmetica modala) ce se ocupa de aceste aspecte, exista diverse implementari de algoritmi pentru astfel de situatii. Un alt exemplu este lucrul cu tablouri foarte mari de date (tablouri de tip masive). In acest caz este vorba de matricele rare. *Matricele rare* îşi găsesc aplicabilitatea în modelarea unor procese biologice, neoronale, de natură industrială, economică, tehnică, socială, etc.

a) Utilizarea programului Excel. (Puterile 2^k, k > 30). *Pentru k > 30 să se determine numărul cifrelor şi cifrele puterii 2^k (de exemplu, să se verifice ca 2^{100} are 31 de cifre şi 2^{100} = 1267650600228229401496703205376 , iar 2^{1000} are 302 cifre).*

Evident, problema ar fi simpla (fără sens) dacă s-ar rezolva printr-o singură instrucțiune scrisa intr-un *limbaj de programare*. Acest lucru se poate realiza doar dacă ar exista restricția k < 31. Ținând seama de reprezentarea *tipului* **integer** în memoria internă a calculatorului, astazi *microprocesoarele* şi *limbajele de programare* pot stoca/reprezenta o *valoare întreagă* doar pe **4 bytes** (*32 biți*). Prin urmare **2^{31}-1 = 21474836**47 este cea mai mare *valoare întreagă* pe care o poate stoca. Este necesar să concepem un *algoritm* pentru calculul puterilor 2^k, k>30. Vom lua in consideratie următorul *tabel* (generat printr-un *simplu program*, sau folosind facilitățile unor *programe de calcul*, de exemplu programul Excel inclus în *pachetul* Microsoft Office, vers. 2003-2007 ; vers. 2010 ofera precizie mai mare) :

K	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2^k	2	4	8	16	32	64	128	256	512	1024	2048	4096	8192	16384

Folosind *programul* **Excel** (ce oferă *funcția* **Power** şi *operația de putere* “^ “) se poate constata că 2^{36}= 68719476736 (dacă se utilizează pentru celule *formatul* “**General**”) este *puterea maximă* ce se poate *calcula*, şi 2^{49}= 562949953421312 (dacă se utilizează pentru celule *formatul* “**Number**” cu 0 zecimale) este *puterea maximă* ce se poate *calcula*.

K = 1	2
2	4
3	8
4	16
5	32
6	64
7	128
8	256
9	512
10	**1024**
11	2048
12	4096
13	8192
14	16384
15	32768
16	65536
17	131072
18	262144
19	524288
20	**1048576**
21	2097152
22	4194304
23	8388608
24	16777216
25	33554432
26	67108864
27	134217728

K = 28	268435456
29	536870912
30	**1073741824**
31	2147483648
32	4294967296
33	8589934592
34	17179869184
35	34359738368
36	**68719476736**
37	EROARE 1.37439E+11
38	2.74878E+11
39	5.49756E+11
40	1.09951E+12
...	...
49	Corect 562949953421312
50	1125899906842620
51	2251799813685250
52	4503599627370500
53	9007199254740990
54	18014398509482000
55	36028797018964000
56	72057594037927900
57	144115188075856000
58	288230376151712000

Rezultate eronate !

De la k=50 *rezultatele sunt eronate* (versiunea Excel 2010 ofera precizie mai mare in acest caz), si anume se poate observa ca ultimele cifre din dreapta sunt **eronate**: ptr. k=50, prima cifra din dreapta, ptr. k=51, ultimele 2 cifre, s.a.m.d.

Rezultate corecte calculate cu *Web 2.0 scientific calculator* (http://web2.0calc.com/):

2^{50}= 112589990684262**4** si 2^{51} = 22517998136852**48**.

b**) Utilizarea Web 2.0 scientific calculator**:

Astazi, nu este nevoie sa se apeleze frecvent la algoritmi de calcul care sa utilizeze un limbaj de programare (C++, Java, Visual Basic, etc.), deoarece pana in prezent s-a dezvoltat foarte mult piata sistemelor de programe specializate ce ofera programe eficiente si comode pentru a fi utilizate de elevi, studenti, specialisti. De altfel, dezvoltarea *tehnologiilor Web* si a *sistemului Internet*, a facut posibila aparitia unui numar foarte mare de astfel de programe specializate.Un astfel de program este oferit de site-ul http://web2.0calc.com/ ce ofera un *Web 2.0 Scientific Calculator*.

Rezultate obtinute prin utilizarea acestui program:

2^{100}=1267650600228229401496703205376
2^{300}=2037035976334486086268445688409378161051468393665936250636140449354381299763336706183397376

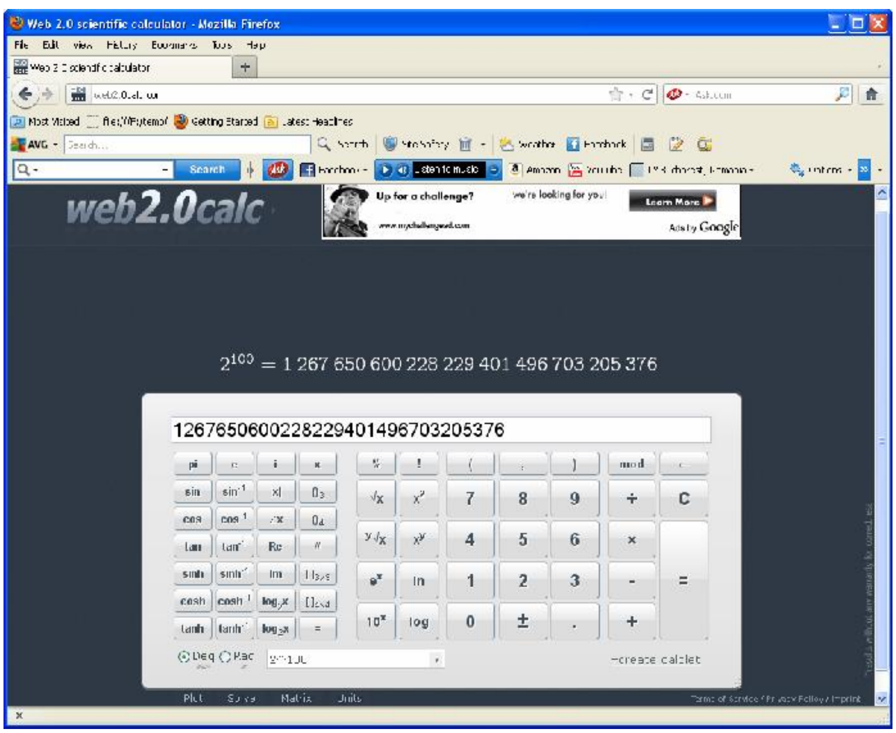


Figura 17. http://web2.0calc.com/
Observatie: programul lucreaza cu 14 zecimale exacte!

π = 3.14159265358979, e = 2.71828182845905 (reprezentare cu 14 zecimale exacte)

Se poate utiliza la obtinerea diverselor calcule matematice si ingineresti (cu utilizarea unitatilor de masura: **Units**), rezolvarea de ecuatii (**Solve**), operatii cu matrice (**Matrix**), reprezentarea grafica a functiilor (**Plot**), etc.,

Exemplul 2: Reprezentarea grafica a functiilor

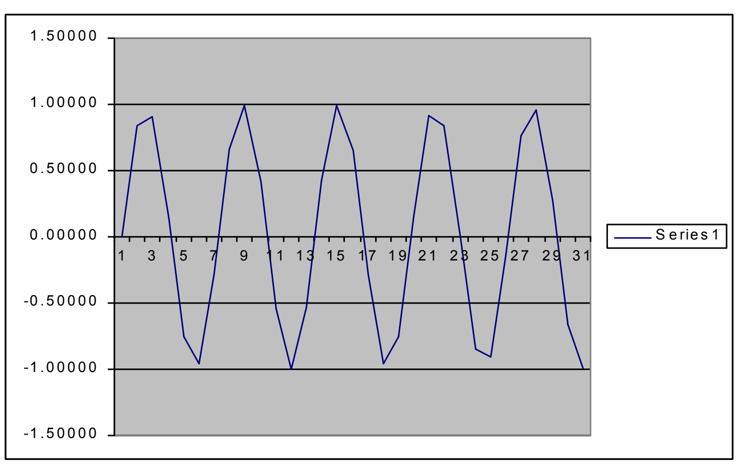
In functie de metoda utilizate, de programul specializat si functie de complexitatea unei functii pot aparea erori frecvente in astfel de situatii. Aceste erori pot aparea in primul rand din cauza neintelegerii notiunilor matematice despre functii sau ca urmare a unei slabe experiente in acest tip de probleme. Vom exemplifica printr-un simplu exemplu.

Sa presupunem ca trebuie sa se reprezinte grafic functia *f(x) = x*sin (x)*, unde *x* apartine intervalului [-50,50]. Evident functia este o compunere de functii, o dreapta si o sinusoida. Metoda matematica invatata de elevi la liceu nu este chiar comoda in acest caz. Nici nu se recomanda se se utilizeze procedura rezultata din metoda matematica. Nici studentul de anul I nu se gandeste mai inainte la metoda matematica. Stie si intuieste ca sunt foarte multe programe care ofera posibilitatea reprezentarii grafice a functiilor. Probleme este aceea a alegerii unui astfel de program tinand seama de licenta de utilizare si functiile acelui produs software. Majoritatea *programelor stiintifice (2D si 3D)* ofera aceasta posibilitate.

a) **cazul programului Excel**

Pentru testarea modului de a utiliza programul Excel in cazul reprezentarii grafice a functiilor, condideram exemplu doar pentru funtia *g(x)=sin(x)* pe intervalul [-50,50]. La activitatile practice de Laborator am avut posibilitatea in ultimii ani sa realizez un sondaj in acest caz. S-a dovedit faptul ca din 20 de studenti, au fost cazuri cand nici un student nu a obtinut rezultatul corect, dar au fost cazuri cand doar unul sau doi au obtinut rezultatul corect. Acest lucru dovedeste ca intelegerea notiunilor, conceptelor si relatiilor intre diversi termeni lasa de dorit la multi studenti din anul I.

Probabil cauzele sunt in invatamantul general si mediu cu multa teorie si cunostinte multiple, fara activitati demonstrative si practice care sa determine obtinerea unor competente utile, importante si oportune. Tot pentru un test sa considaram ca graficul trebuie obtinut pe intervalul [0,30]. Primul lucru care se realizeaza rapid si *fara sa se intuiasca eroarea*, se genereaza valorile naturale 1, 2, 3, ... , 30 pentru argumentul *x*. Evident ca va rezulta graficul unei *linii poligonale* si nu graficul real al functiei *sin(x)*.

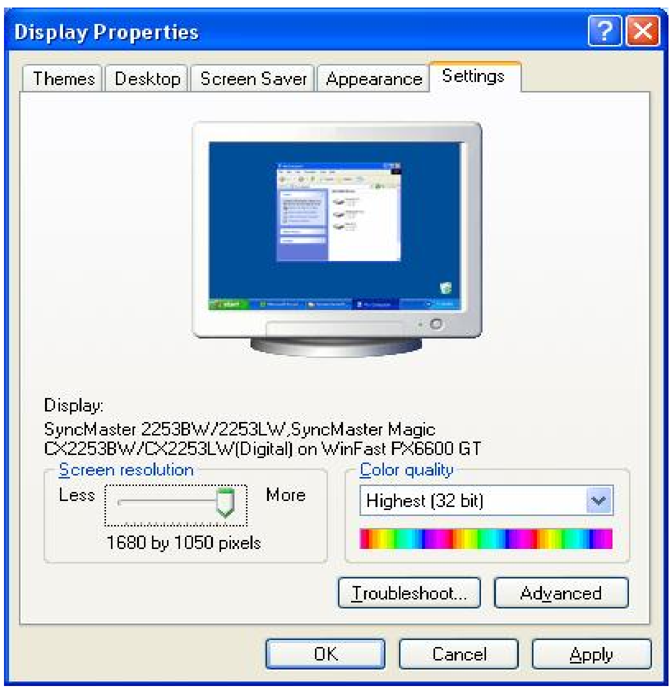


Eroarea provine de la faptul ca trebuie sa se realizeze *discretizarea intervalului* (*tabelarea functie* cu un pas cat mai mic p= 10^{-1} , 10^{-2} , etc. ce are legatura cu functia studiata; trebuie sa "cuprinda" convexitatile si cancavitatile graficului). In cazul functiei *sin(x)* este suficienta discretizarea cu pasul p= 10^{-1}, dar tabelarea va produce 10x50 = 500 puncte pe axa pozitiva si tot atatea pe axa negativa. Acum, daca se tine seama ca mai inainte, trebuie sa se *genereze tabelarea functiei*, se poate trece la realizarea graficului *f(x) = x*sin (x)*, pe intervalul [-50,50]. Va rezulta graficul corect ce este mai fidel si mai realist.

Tabelarea functiei vs. Discretizare-Calculul integral vs. Rezolutia suportului grafic

Sistemul de diviziuni (*proces de discretizare*) din calculul integral este analog *rezoluției* (matricea de pixeli; un „**pixel**" este unitatea grafică indivizibilă a unui display grafic) oferite de un display grafic (CRT sau LCD). Această structură de pixeli reprezintă în informatică, ceea ce reprezintă *calculul integral* în analiza matematică (*Newton, Riemann, Darboux, Leibniz* etc.). Cu cat rezolutia este mai mare cu atat reprezentarea este de buna calitate. Mai jos este rezolutia oferita de un ecran grafic.

Display Properties → Screen Resolution: Less-*800 x 600 pixels*, More-*1680x1050 pixels*.



Odată cu apariția **display-ului grafic** (*Graphic Display*), în anul 1953, s-a trecut la o nouă etapă în dezvoltarea şi răspândirea calculatorului. Utilizarea **bit-ului** prin organizarea eficientă a memoriei calculatorului, nu oferea nici hardware, nici software posibilitatea de modelare spațială a ieşirilor (OUTPUT). Reprezentările grafice folosind caractere (numerice sau alfanumerice) nu era o soluție care să realizeze o reprezentare fidelă a obiectelor reale. Suportul hardware fiind inventat, în perioada 1960-1980 au fost nevoie de cercetări şi experimente, modele, algoritmi si programe care să folosească

aprinderea unui „**pixel**” (unitatea grafică indivizibilă oferită de un display grafic) ce oferea şi culoare, dar mai ales o structură de reprezentare grafică. Atunci s-a născut Grafica pe calculator: trasarea unui segment de dreaptă (algoritmul Bresenham), trasarea cercului şi elipsei, trasarea şi aproximarea curbelor, algoritmi de clipping (decupare) (algoritmul Cohen – Sutherland, algoritmul Suitherland-Hodgman, algoritmul Weiler-Atherton), tehnici de vizualizare 2D şi 3D, modele de iluminare şi reflexie, modele de tip rastru, modele vectoriale, tehnici de textură. Astfel, s-au pus bazele pentru soluții integrate software şi hardware pentru proiectare, analiză şi producție asistată de calculator (CAD/CAM/CAE) - Computer Aided Design.
După anul 1990, s-au obținut rezultate deosebite în domeniul modelării şi simulării obiectelor din lumea reală, atât prin elaborarea de tehnici şi algoritmi specifici, cât prin apariția produselor software care să sprijine acest domeniu. Astfel, **Realitatea Virtuală** (Virtual Reality) este un nou domeniu al Informaticii ce are un impact deosebit în utilizarea calculatorului pe scară largă şi pentru o mare diversitate de teme.

b) **cazul programului Web 2.0 scientific calculator**
Se introduce comanda: **plot(x*sin(x),x=-50..50)** si se obtine imediat graficul corect.

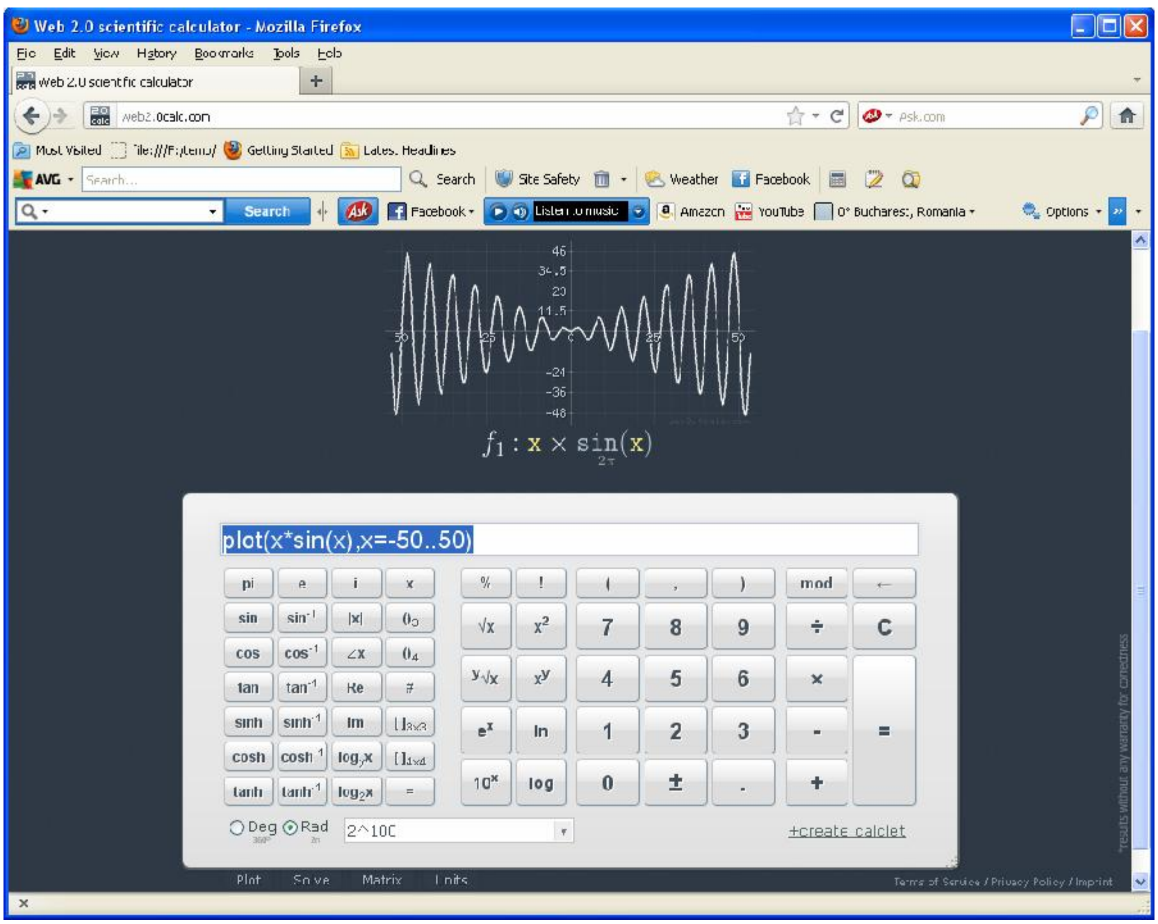


Figura 18. Graficul folosind Web 2.0 scientific calculator

Exemplul 3: Problema lui Gauss. *Un vas conține 2000 litri dintr-un lichid cu o concetrație de 80 % alcool. În fiecare zi se scot din vas 15 litri şi se înlocuiesc cu alți*

12 litri dintr-un lichid a cărui concentrație în alcool este de numai 40 %. După câte zile concentrația lichidului din vas ajunge la 50 % ?

In cele ce urmeaza vom aborda *3 variante de rezolvari* pentru aceasta problema pentru a evidentia atat evolutia *metodelor si tehnicilor de rezolvare* (teorii si metode numerice), cat si obstacole in utilizarea diverselor metode (de exemplu, *problema propagarii erorilor in calcule*) :

1. **Modelarea matematica-metoda matematica** – modelarea matematica va reprezenta o *ecuatie funtionala* ce se poate aborda ca o *ecuatie cu diferente finit de orinul I neomogena*;
2. **Algoritm de calcul-program** intr-un limbaj de programare – conceperea procesului de calcul ce realizeaza un proces iterativ al operatiilor pentru rezolvarea problemei;
3. **Rezolvare cu programul EXCEL** – se vor utiliza faciltatile programului Excel si forma algoritmica oferita de metoda algorimica.

Modelarea matematica si Metoda algoritmica.

Problema este prezentată în [1], enunțul ei , aparent este al unei probleme simple, dar interesantă din punctul de vedere a rezolvării ei, deoarece problema a fost menționată la vremea respectivă chiar de *GAUSS.* În [2] apare *rezolvarea problemei cu calculatorul*.

Rezolvarea problemei nu este evidentă, după cum se va vedea în cele ce urmează. Din punct de vedere *matematic*, rezolvarea necesită noțiuni şi concepte de matematică superioară din *domeniul ecuațiilor funcționale*, şi anume a ecuațiilor cu diferențe finite de ordinul I neomogene. În două articole ştiințifice, problema a fost rezolvată de către W. LOREY (1935) şi A. WALTHER (1936). Din punct de vedere *numeric*, rezolvarea problemei necesită cunoaşterea *metodelor numerice specifice rezolvării ecuațiilor cu diferențe finite*. De altfel, W. LOREY a şi utilizat o *maşină de calcul* pentru rezolvarea numerică a unui ecuații cu diferențe finite, aceasta deoarece a sesizat faptul că soluția se obține după un număr considerabil de iterații.

Din punct de vedere *informatic*, rezolvarea va fi simplă deoarece nu se va utiliza *modelul matematic* (ecuația funcțională) obținut din modelarea analitică a problemei, ci un *proces de calcul* care simulează *operațiile şi stările* unor *locații de memorie* (acesta este de fapt *algoritmul* care codifică rezolvarea problemei), şi care implementat într-un *limbaj de programare* (de exemplu *C* sau *Pascal*) va rezolva problema în cazul general.

Pentru a face comparația dintre soluția algoritmică obținută pentru calculator şi soluția analitică, prezentăm succint rezolvarea dată de A. WALTHER. Vom considera problema în cazul general, de accea vom face următoarele notații :

a - cantitatea de lichid (în litri) conținută inițial în vas;

b - cantitatea de lichid ce se scoate zilnic din vas;

c - cantitatea de lichid ce se adaugă zilnic în vas;

y_0 - cantitatea de alcool pe litru (*concentrația de alcool*) a lichidului din vas la momentul inițial;

y_p - cantitatea de alcool pe litru a lichidului ce se adaugă;

y_f - cantitatea de alcool pe litru a lichidului din vas, la momentul final;

x - numărul de zile (operații de înlocuire a lichidului);

y(x) - cantitatea de alcool pe litru a lichidului din vas după *x* operații de înlocuire a lichidului.

Ecuația funcțională (ecuația cu diferențe finite) pentru determinarea funcției *y(x)*, se obține exprimând cantitatea totală de alcool din vas după *x* zile, în două moduri :

i) $(a - bx + cx)\, y(x)$

ii) $(a - bx + c(x-1))\, y(x-1) + c\, y_p$,

unde cazul ii) se obține adunând cantitatea de alcool din lichidul rămas în vas după *(x-1)* zile, din care s-au scot *b* litri, cu cantitatea de alcool a celor *c* litri care se adaugă.

Prin urmare, se obține următoarea ecuație funcțională:

(1) $(a - bx + cx)\, y(x) - (a - bx + c(x-1))\, y(x-1) = c\, y_p$, *ecuație cu diferențe finite de ordinul I neomogenă*.

Rezolvarea acestei ecuații este prezentă în [1], soluția generală fiind

$$y(x) = y_0 + (y_0 - y_p)\frac{\Gamma((a-b)/(b-c))\Gamma(a/(b-c)-x)}{\Gamma(a/(b-c)\Gamma((a-c)/(b-c)-x)}$$, unde $\Gamma(x)$ este *funcția lui Euler* dată de relația:

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

În cazul particular *a*=2000, *b*=15, *c*=12, y_0=0.8, y_p=0.4, *y(x)* este un polinom de gradul IV :

$$y(x) = 0.4 + 0.4\left(1 - \frac{3x}{1988}\right)\left(1 - \frac{3x}{1991}\right)\left(1 - \frac{3x}{1994}\right)\left(1 - \frac{3x}{1997}\right).$$

de unde, prin aproximare se deduce că $y(194) = 0.50048$, $y(195) = 0.49963$, prin urmare după $x=195$ zile se ajunge la concentrația de 0.5.

Metoda algoritmica- proces de calcul si program

În cazul rezolvării algoritmice, vom abandona metoda obținerii *ecuației funcționale* şi rezolvarea ei *analitică* sau *numerică*, şi vom concepe *algoritmul* ce realizează *procesul de calcul* generat de cerințele problemei.
Pe lângă variabilele x, a, b, c, y_p, y_f cu semnificațiile prezentate mai sus, vom utiliza şi următoarele *variabile*:
z - *cantitatea de alcool din vas la un moment dat* ;
t - *cantitatea de lichid din vas la un moment dat* ;
y_0 - *concentrația de alcool din vas la un moment dat*.

Algoritmul în limbaj pseudo-cod este urmatorul :

```
algorithm Gauss;
int x;
float a,b,c,y0,yp,yf,z,t;

begin // main
  read a,b,c ;   //liquid quantities
  read y0,yp,yf; //concentrations
  // initializations
  x←1; z←(a-b)*y0+c*yp;
  t←a-b+c
  while yf < z/t  do
  begin
    x←x+1;
    y0← z/t; //concentration
    z←(t-b)*y0+c*yp;
    t←t-b+c;
  end
  write x; // solution
end
```

Prin execuția algoritmului/programului de mai sus (in limbaj de programare C, Pascal, etc.), pentru valorile b=15, c=12, y_0 (inițial) = 0.8, y_p= 0.4, y_f = 0.5 se obțin următoarele rezultate :

a = 2000 , yf = 0.5004515, x(days) = 195
a = 5000 , yf = 0.5001438, x(days) = 488
a = 10000 , yf = 0.5000983, x(days) = 976
a = 100000 , yf = 0.5000064, x(days) = 9763

Referinte

[1] GABRIEL SUDAN, *Câteva probleme matematice interesante*, Biblioteca SSM, *Editura Tehnică,* Bucureşti, 1969.

[2] MARIN VLADA, O problemă a lui K.F. Gauss rezolvată cu calculatorul, Gazeta Matematică, nr. 5/1995.

Rezolvare cu programul EXCEL

Pentru a realiza in Excel calculul iterativ din algoritmul de mai sus vom introduce mai inainte, in celulele corespunzatoare valorile datelor cunoscute:

a	b	c	y0	yp	yf
2000.000	15.000	12.000	0.800	0.400	0.500

Calculul iterativ si valorile parametrilor/variabilelor acestui calcul trebuie sa fie implementate intr-un tabel de forma:

x	ycurent	z	t
0	0.800	1600.000	2000.000
1	0.800	1592.800	1997.000
2	0.798	1585.636	1994.000
3	0.795	1578.508	1991.000

Deoarece in algorimul de calcul precedent variabila y_0 este folosita si pentru *concentrația de alcool din vas la un moment initial*, dar si pentru *concentrația de alcool din vas la un moment curect*, von introduce variabila
- *ycurent = concentrația de alcool din vas la un moment curect.*

Din aceste motive, trebuie sa implementam in Excel un calcul iterativ de forma:

```
while yf < z/t  do
  begin
    x←x+1;
    ycurent← z/t; //concentration
    z←(t-b)*ycurent+c*yp;
    t←t-b+c;
  end
```

Trebuie sa se realizeze urmatoarele etape (capul de tabel este *pe randul 6*):

1. *se genereaza* cu **Edit → Fill** valorile pentru variabila (*numar de zile*) x: 0..200 pe coloana A corespunzatoare acesteia, si anume pe randurile 7-207;
2. *se introduc valorile* pentru starea initiala (*x=0*), adica pentru **ycurent**, in B7 valoare 0.800, pentru **z** in C7 formula =A$4*D$4, iar pentru **t**, in celula D7, valoarea 2000;
3. *se introduc formulele pentru prima iteratie* (*x=1*) tinand seama de calcul iterativ de mai sus (a se vedea imaginea capturata din programul Excel), si anume,

- pentru *ycurent*, **B8= =C7/D7**
- pentru *z*, **C8 =(D7-B$4)*B8+C$4*E$4**
- pentru *t*, **D8 =D7-B$4+C$4**

4. *se genereaza formulele (prin Copy sub Excel) pentru iteratiile x= 2..200*, adica se selecteaza domeniul de celule **B8:D8**, se elibereaza butonul de mouse, dupa care se aduce cursorul cruce (mare) al mouse-lui catre coltul *dreapta-jos* al cadrului ce a selectat domeniul de celule, determinad aparitia cursorului de cruce mica; dupa aceea se apasa butonul stanga si se trage pana la randul 207 (*x=200*), realizandu-se astfel calcule corespunzatoare pentru cele 3 coloane din tabel..

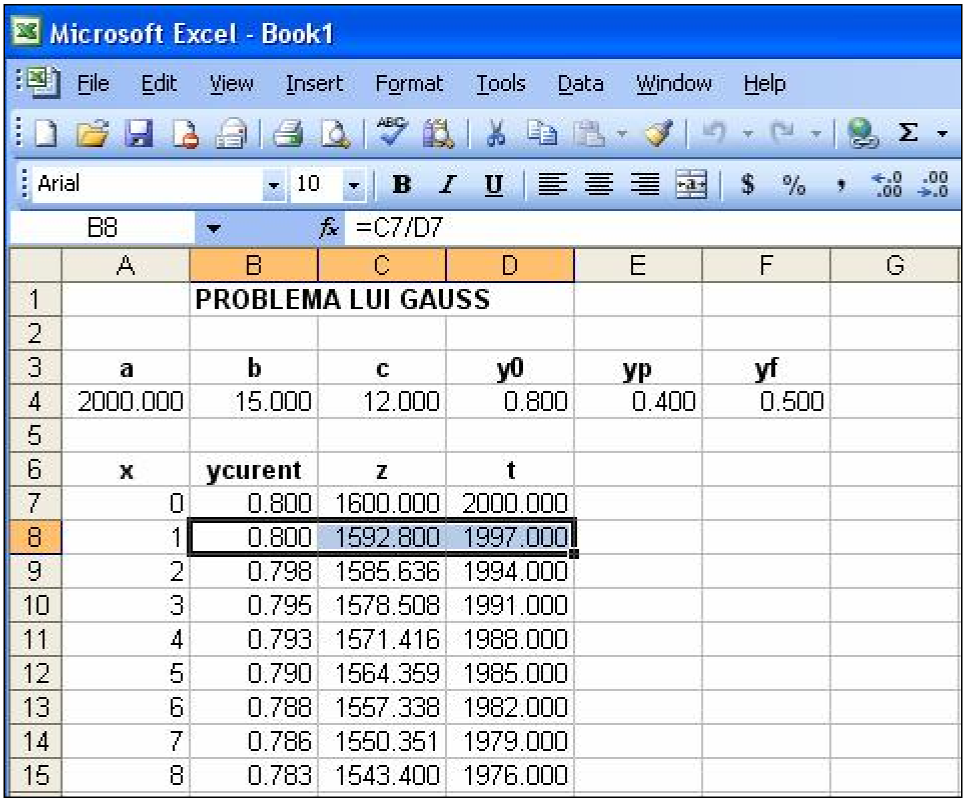


Figura 19. Problema lui Gauss folosind Excel

Valorile generate de calculul iterativ sunt prezentate in continuare. Concluzia este ca solutia in acest caz este ***x= 195*** , adica identica cu solutia determinata prin algoriumul/programul precedent.

x	ycurent	z	t
0	0.800	1600.000	2000.000
1	0.800	1592.800	1997.000
2	0.798	1585.636	1994.000
3	0.795	1578.508	1991.000
4	0.793	1571.416	1988.000
5	0.790	1564.359	1985.000
6	0.788	1557.338	1982.000
7	0.786	1550.351	1979.000
8	0.783	1543.400	1976.000
9	0.781	1536.484	1973.000
10	0.779	1529.603	1970.000
11	0.776	1522.756	1967.000
12	0.774	1515.944	1964.000
13	0.772	1509.166	1961.000
14	0.770	1502.422	1958.000
15	0.767	1495.712	1955.000
16	0.765	1489.036	1952.000
17	0.763	1482.394	1949.000

18	0.761	1475.785	1946.000
19	0.758	1469.209	1943.000
20	0.756	1462.667	1940.000
21	0.754	1456.158	1937.000
22	0.752	1449.681	1934.000
23	0.750	1443.238	1931.000
24	0.747	1436.827	1928.000
25	0.745	1430.448	1925.000
26	0.743	1424.102	1922.000
27	0.741	1417.788	1919.000
28	0.739	1411.505	1916.000
29	0.737	1405.255	1913.000
30	0.735	1399.036	1910.000
31	0.732	1392.849	1907.000
32	0.730	1386.693	1904.000
33	0.728	1380.569	1901.000
34	0.726	1374.475	1898.000
35	0.724	1368.413	1895.000
36	0.722	1362.381	1892.000
37	0.720	1356.380	1889.000
38	0.718	1350.409	1886.000
39	0.716	1344.469	1883.000
40	0.714	1338.559	1880.000
41	0.712	1332.679	1877.000
42	0.710	1326.829	1874.000
43	0.708	1321.008	1871.000
44	0.706	1315.218	1868.000
45	0.704	1309.457	1865.000
46	0.702	1303.725	1862.000
47	0.700	1298.022	1859.000
48	0.698	1292.349	1856.000
49	0.696	1286.704	1853.000
50	0.694	1281.088	1850.000
51	0.692	1275.501	1847.000
52	0.691	1269.942	1844.000
53	0.689	1264.412	1841.000
54	0.687	1258.910	1838.000
55	0.685	1253.436	1835.000
56	0.683	1247.990	1832.000
57	0.681	1242.571	1829.000
58	0.679	1237.181	1826.000
59	0.678	1231.818	1823.000
60	0.676	1226.482	1820.000
61	0.674	1221.174	1817.000
62	0.672	1215.893	1814.000
63	0.670	1210.638	1811.000
64	0.668	1205.411	1808.000
65	0.667	1200.210	1805.000

66	0.665	1195.036	1802.000
67	0.663	1189.889	1799.000
68	0.661	1184.767	1796.000
69	0.660	1179.672	1793.000
70	0.658	1174.603	1790.000
71	0.656	1169.560	1787.000
72	0.654	1164.543	1784.000
73	0.653	1159.552	1781.000
74	0.651	1154.586	1778.000
75	0.649	1149.645	1775.000
76	0.648	1144.730	1772.000
77	0.646	1139.839	1769.000
78	0.644	1134.974	1766.000
79	0.643	1130.134	1763.000
80	0.641	1125.319	1760.000
81	0.639	1120.528	1757.000
82	0.638	1115.762	1754.000
83	0.636	1111.020	1751.000
84	0.635	1106.302	1748.000
85	0.633	1101.609	1745.000
86	0.631	1096.939	1742.000
87	0.630	1092.294	1739.000
88	0.628	1087.672	1736.000
89	0.627	1083.074	1733.000
90	0.625	1078.499	1730.000
91	0.623	1073.948	1727.000
92	0.622	1069.420	1724.000
93	0.620	1064.916	1721.000
94	0.619	1060.434	1718.000
95	0.617	1055.975	1715.000
96	0.616	1051.539	1712.000
97	0.614	1047.126	1709.000
98	0.613	1042.735	1706.000
99	0.611	1038.367	1703.000
100	0.610	1034.021	1700.000
101	0.608	1029.698	1697.000
102	0.607	1025.396	1694.000
103	0.605	1021.116	1691.000
104	0.604	1016.858	1688.000
105	0.602	1012.622	1685.000
106	0.601	1008.408	1682.000
107	0.600	1004.215	1679.000
108	0.598	1000.043	1676.000
109	0.597	995.893	1673.000
110	0.595	991.764	1670.000
111	0.594	987.656	1667.000
112	0.592	983.569	1664.000
113	0.591	979.503	1661.000

114	0.590	975.457	1658.000
115	0.588	971.432	1655.000
116	0.587	967.427	1652.000
117	0.586	963.443	1649.000
118	0.584	959.479	1646.000
119	0.583	955.536	1643.000
120	0.582	951.612	1640.000
121	0.580	947.708	1637.000
122	0.579	943.824	1634.000
123	0.578	939.960	1631.000
124	0.576	936.115	1628.000
125	0.575	932.290	1625.000
126	0.574	928.485	1622.000
127	0.572	924.698	1619.000
128	0.571	920.931	1616.000
129	0.570	917.182	1613.000
130	0.569	913.453	1610.000
131	0.567	909.743	1607.000
132	0.566	906.051	1604.000
133	0.565	902.378	1601.000
134	0.564	898.724	1598.000
135	0.562	895.087	1595.000
136	0.561	891.470	1592.000
137	0.560	887.870	1589.000
138	0.559	884.289	1586.000
139	0.558	880.725	1583.000
140	0.556	877.180	1580.000
141	0.555	873.652	1577.000
142	0.554	870.142	1574.000
143	0.553	866.650	1571.000
144	0.552	863.175	1568.000
145	0.550	859.718	1565.000
146	0.549	856.278	1562.000
147	0.548	852.855	1559.000
148	0.547	849.449	1556.000
149	0.546	846.060	1553.000
150	0.545	842.688	1550.000
151	0.544	839.333	1547.000
152	0.543	835.995	1544.000
153	0.541	832.673	1541.000
154	0.540	829.368	1538.000
155	0.539	826.079	1535.000
156	0.538	822.807	1532.000
157	0.537	819.551	1529.000
158	0.536	816.311	1526.000

159	0.535	813.087	1523.000
160	0.534	809.878	1520.000
161	0.533	806.686	1517.000
162	0.532	803.510	1514.000
163	0.531	800.349	1511.000
164	0.530	797.204	1508.000
165	0.529	794.074	1505.000
166	0.528	790.960	1502.000
167	0.527	787.861	1499.000
168	0.526	784.777	1496.000
169	0.525	781.708	1493.000
170	0.524	778.654	1490.000
171	0.523	775.615	1487.000
172	0.522	772.591	1484.000
173	0.521	769.582	1481.000
174	0.520	766.588	1478.000
175	0.519	763.608	1475.000
176	0.518	760.642	1472.000
177	0.517	757.691	1469.000
178	0.516	754.754	1466.000
179	0.515	751.832	1463.000
180	0.514	748.923	1460.000
181	0.513	746.029	1457.000
182	0.512	743.148	1454.000
183	0.511	740.282	1451.000
184	0.510	737.429	1448.000
185	0.509	734.590	1445.000
186	0.508	731.764	1442.000
187	0.507	728.952	1439.000
188	0.507	726.154	1436.000
189	0.506	723.369	1433.000
190	0.505	720.597	1430.000
191	0.504	717.838	1427.000
192	0.503	715.092	1424.000
193	0.502	712.360	1421.000
194	0.501	709.640	1418.000
195	**0.500**	**706.934**	**1415.000**
196	0.500	704.240	1412.000
197	0.499	701.558	1409.000
198	0.498	698.890	1406.000
199	0.497	696.233	1403.000
200	0.496	693.590	1400.000

Solutia corecta!

CONCLUZII.

Din analiza celor 3 rezolvari ale problemei lui Gauss se poate exprima concluzia ca **metoda matematica** (rezolvarea unei ecuatii functionale) este *laborioasa si incomoda*, iar **metoda algoritmica** sustinuta de un program scris intr-un limbaj de programare este cea mai comoda si eficienta. De asemenea, rezolvarea folosind facilitatile **programului Excel** este comoda si eficienta, in primul pentru ca se bazeaza pe procesul de calcul iterativ din metoda algoritmica. *Incovenientele* (eliminate in cazul programului scris intr-un limbaj de programare) apar atunci cand in vas cantitatea de lichid este foarte mare (5000, 10000, etc.), caz in care *tabelul de calcul* necesita *dimensiuni mari*. Mai jos vom exemplifica printr-o situatie modul in care **propagarea erorilor** pot denatura obtinerea rezultatului corect in cazul acestei probleme.

Exemplu privind propagarea erorilor.

Pentru cantitatea de lichid de 2000, numarul de iteratii este considerabil (x=195, solutia) si pot determina *procesul de propagare a erorilor*. Formula variabilei/parametrului *z* din algoritmul de calcul, utilizeaza valoarea concentratiei de la pasul precedent

$$\mathbf{z \leftarrow (t-b)*y_{curent} + c*yp\ .}$$

Vom modifica formula astfel ca sa se utilizeze valoare concentratiei la momentul curent, adica formula C8 = (D7-B$4)***B8**+C$4*E$4 va fi modificata astfel:

C8 = (D7-B$4)***B7**+C$4*E$4.

In urma refacerii calculelor obtinem rezultatele de mai jos:

X	ycurent	z	t
0	0.800	1600.000	2000.000
1	0.800	1592.800	1997.000
2	0.798	1590.400	1994.000
3	0.798	1583.243	1991.000
4	0.795	1580.843	1988.000
5	0.795	1573.730	1985.000
6	0.793	1571.330	1982.000
7	0.793	1564.259	1979.000
8	0.790	1561.859	1976.000
9	0.790	1554.831	1973.000
10	0.788	1552.432	1970.000
11	0.788	1545.446	1967.000
12	0.786	1543.047	1964.000

186	0.607	875.596	1442.000
187	0.607	871.634	1439.000
188	0.606	869.466	1436.000
189	0.605	865.531	1433.000
190	0.604	863.367	1430.000
191	0.604	859.459	1427.000
192	0.602	857.300	1424.000
193	0.602	853.418	1421.000
194	0.601	851.263	1418.000
195	**0.600**	847.408	1415.000
196	**0.599**	845.257	1412.000
197	**0.599**	841.428	1409.000
198	**0.597**	839.282	1406.000
199	**0.597**	835.479	1403.000
200	**0.595**	833.337	1400.000

Rezultate eronate !

Solutia, in acest caz are *valoare mai mare* decat valoarea corecta. Influenta propagarii erorilor a determinat obtinerea unor *rezultate eronate*.

Indicatori statistici

Indicatorii statistici sunt definiţi pentru a surprinde (a analiza) variaţii de manifestare a unor valori masurate pentru fenomene si procese si care necesită elaborarea unor metodologii şi tehnici de rafinare, transformare şi aplicare a unor operaţii speciale de calcul pentru obţinerea unor determinări cantitativ-numerice. Indicatorul statistic, în forma sa generală, este expresia numerică a manifestărilor unor fenomene, procese, activităţi sau categorii economice şi sociale, delimitate în timp, spaţiu. Pentru cunoaşterea proceselor si fenomenelor, indicatorii statistici îndeplinesc mai multe funcţii şi anume: de măsurare; de comparare; de analiză sau de sinteză; de estimare; de verificare a ipotezelor şi/sau de testare a semnificaţiei parametrilor utilizaţi.
Indicatorii statistici se pot grupa în:

- **Indicatori primari (mărimi absolute) –** exprimă direct valori initiale (masuratori) pentru obiectivele cercetate; se pot obţine prin înregistrarea directă, centralizarea datelor sau prin însumarea parţială sau totală a datelor individuale; prezintă o capacitate relativ limitată de descriere a fenomenului/procesului analizat, şi nu permite realizarea unor aprecieri calitative;
- **Indicatori derivaţi –** se obţin prin prelucrarea indicatorilor primari şi fac posibilă analiza aspectelor calitative ale fenomenelor şi proceselor analizate (ex: *mărimi relative, mărimi medii, indicatori ai variaţiei, indici, indicatori ai corelaţiei*, etc).

Indicatorii tendinţei centrale

În general, indicatorii tendinţei centrale se determină în general ca indicatori medii sau indicatori de poziţie (ai localizării), în funcţie de natura caracteristicilor urmărite în colectivitatea investigată, de scopul investigaţiei. Sunt multe situaţiile când tendinţa centrală se caracterizează printr-un anumit tip de medie (**aritmetică**, **armonică**, **pătratică**, **geometrică**), dar şi situaţii de utilizare a *indicatorilor sintetici de poziţie* (localizare: *modul, cuantile*).

Diverse tipuri de medii ale valorilor primare:

- **Media aritmetica** - În sens statistic, media aritmetică a valorilor individuale (x_1, x_2, …, x_n) ale variabilei / parametrului $X = (x_1, x_2, \ldots, x_n)$ reprezintă acea valoare $\bar{x}$ care s-ar fi înregistrat dacă toţi factorii de influenţă ar fi acţionat constant (cu aceeaşi intensitate) la nivelul fiecărei valori masurare/înregistrare. Prin urmare, $\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$, sau $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$, si avem $\min_i x_i \le \bar{x} \le \max_i x_i$.
- **Media ponderată** - Într-o colectivitate statistică, suficient de mare (n mare), unde de obicei, multe valori prezintă o anumită *frecvenţă de apariţie*, media aritmetică se calculează ca o medie ponderată:

$\overline{x} = \frac{\sum_{i=1}^{n} f_i x_i}{n}$, unde f_i reprezintă frecvența valorii x_i, şi avem $\sum_{i=1}^{n} f_i = n$.

- **Media armonică** - Media armonică este folosită numai în anumite situații, şi anume atunci când valorile/seturile de date sunt alcătuite din valori exprimate sub formă de rapoarte, cum ar fi prețurile vitezele (în mp/h), prețurile (în u.m./kg), sau productivitatea (produse/oră-om). Media armonică se defineşte ca valoare inversă a mediei aritmetice a inverselor valorilor elementelor individuale înregistrate; relația de calcul a mediei armonice simple a şirului de valori X = (x_1, x_2, …, x_n) este următoarea:

$$m_a = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}};$$

Pentru o serie de distribuții de frecvențe media armonică ponderată se calculează după relația:

$$m_a = \frac{\sum_{i=1}^{n} f_i}{\sum_{i=1}^{n} \frac{1}{x_i} f_i},$$

- **Media geometrică** - Media geometrică este o mărime specializată folosită pentru a calcula media creşterilor procentuale (*media creşterilor procentuale a salariilor sau prețurilor bunurilor*). Media geometrică reprezintă acea valoare a caracteristicii observate care dacă ar înlocui fiecare valoare individuală din serie produsul acestora nu s-ar modifica, adică

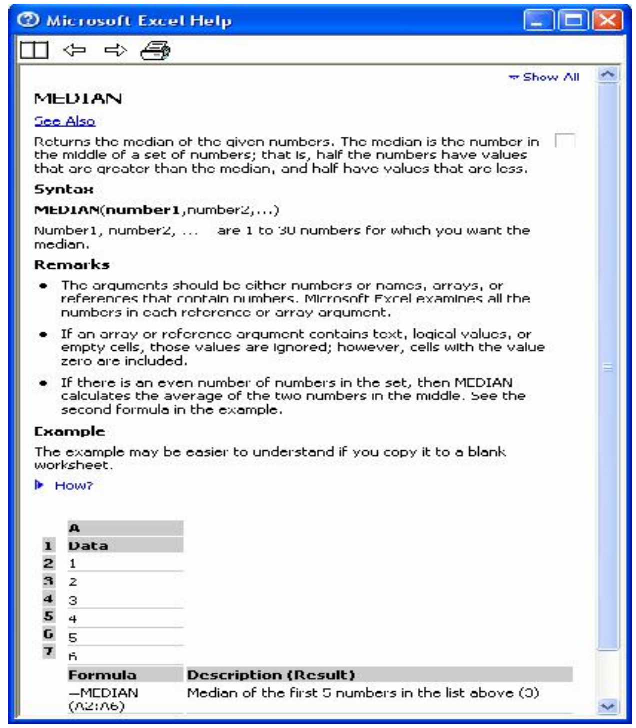
$$m_g = \left[\prod_{i=1}^{n} x_i\right]^{\frac{1}{n}}$$

Indicatori de poziție

Indicatorii de poziție calculează si se identifică în cadrul unui set de valori cu câte o variantă reală, care posedă o anume proprietate, conform căreia respectiva variantă oferă o informație satisfăcătoare despre setul de valori studiat:

- **Mediana** (*Median*)- Me, aceasta reprezintă valoarea centrală a unei serii de date aranjate crescător sau descrescător, si are proprietatea ca imparte seria in 2 grupuri egale, astfel incat jumatate din valori sunt mai mici decat mediana si jumatate sunt mai mari decat mediana. Este cuartila de mijloc, cuartilele fiind valori care impart seria in 4 grupe, sau este percentila de mijloc, percentilele fiind valori care impart seria in 10 grupe egale. Pentru o serie cu *numar impar de valori*, valorile seriei sunt in ordine crescatoare si valoarea care imparte seria in doua parti egale este mediana. Valoarea de mijloc a unei distribuții, este definită drept cel mai mic număr astfel încât jumătate dintre valori să nu fie mai mari decât el. Cu alte cuvinte, jumătate dintre valori sunt mai mici sau egale cu mediana, jumătate sunt mai mari decât mediana. De remarcat că, deşi este utilizat în general ca un indicator de tendință centrală, mediana oferă mai degrabă informații asupra repartizării observațiilor (indicator de împrăştiere). De regulă, mediana este raportată împreună cu quartilele distribuției în aşa-zisa rezumare prin cinci valori. Dacă $x_1, x_2, \ldots, x_n$ sunt valorile observate, mediana este calculată, după ordonarea crescătoare a valorilor, $x_{(1)} <= x_{(2)} <= \ldots <= x_{(n)}$, prin

$$Me = \begin{cases} x_{k+1}, \text{pentru } N = 2k+1 \\ \dfrac{x_k + x_{k+1}}{2}, \text{pentru } N = 2k \end{cases}.$$



Este de notat că mediana realizează minimul sumei abaterilor absolute ale valorilor distribuției de la un punct fixat: $\sum_{i=1}^{n} |x_i - m|$ este minimă pentru m egală cu mediana distribuției (în cazul unui număr par de valori, mediana – aşa cum a fost definită – nu este singura valoare cu această proprietate).

Funcție Excel:

MEDIAN(number1,number2,...)

Number1, number2, ... are 1 to 30 numbers for which you want the median.

Exemplu: **Median** (18,19,20,21,22,23,24,25,26,27,28,29,30,31,32)=25 (nr. impar de valori) si **Median** (18,19,20,21,22,23,24,25,26,27,28,29,30,31) = 24.5

- **Modulul** (*Mode*) – valoarea modala, adica dominanta unei variabile ce reprezintă valoarea care înregistrează cea mai mare frecvență de apariție. Valoarea modală se utilizează ca indicator al tendinței centrale atunci când media nu se poate calcula sau nu are sens să fie calculată.Valoarea mod este cea mai frecventă valoare dintr-o mulțime de valori. Grafic, dintr-o histogramă, o valoare mod este identificată printr-un maxim relativ. O distribuție poate avea astfel mai multe valori mod (distribuții unimodale, bimodale, etc.).

  Funcție Excel: **MODE** (number1,number2,...)

Number1, number2, ... are 1 to 30 arguments for which you want to calculate the mode. You can also use a single array or a reference to an array instead of arguments separated by commas. :

Exemplu: **Mode** (18,19,20,21,22,20,24,20,26,27,20,29,30,31,32)=20,
Mode (18,19,20,18,22,18,24,25,26,27,18,29,30,31) = 18

În Excel, funcțiile corespunzătoare acestor parametri *media arimetica*, *mediana* si *modulul*, sunt: **AVERAGE, MEDIAN, MODE**.

Indicatori ai împrăştierii (variației)

- **Amplitudine** (*Range*) – sau indice de dispersie (Dispersion indexes) - este definită ca $x_{max}–x_{min}$, unde x_{max} şi x_{min} sunt valorile extreme ale unui set de numere observate. Oferă o imagine a raspandirii datelor, dependentă însă de numărul de valori observate. Cu cât se măsoară mai multe elemente, cu atât şansa de a observa valori mai depărtate creşte, deci şansa de a obține o amplitudine mai mare.
- **Abaterea medie** (*Mean Deviation*) – deviatia sau abaterea medie reprezinta media abaterilor valorilor individuale fata de valoarea medie:

$$D_M = \frac{1}{n}\sum_{i=1}^{n}(x_x - \bar{x})$$

- **Abaterea standard** (*Standard Deviation – SD*) este radicalul mediei pătratice a abaterilor datelor față de medie şi se calculează cu formula:

$$s = \sigma_X = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n-1}}$$ (in Excel este functia **STDEV** sau **STDEVP**).
- **Varianța** (*Variance*) sau **dispersia** este pătratul abaterii medii pătratice, $V\left(x\right) = \sigma_x^2$ (in Excel este functia **VAR** sau **VARP**).
- **Intervalul de confidenta** (*Confidence interval*) – interval de incredere (numar de valori in intervalul de incredere) pentru estimarea unui parametru (ex. *media*, *dispersia*, etc) in cazul unei **distributii normale Gauss**:
  a) $x = \bar{x} \pm \sigma$ cu probabilitate de 0.682
  b) $x = \bar{x} \pm 2\sigma$ cu probabilitate de 0.954
  c) $x = \bar{x} \pm 3\sigma$ cu probabilitate de 0.997

  In Excel exista functia **CONFIDENCE**(**alpha**,**standard_dev**,**size**), **Alpha** is the significance level used to compute the confidence level. The confidence level equals 100*(1 - alpha)%, or in other words, an alpha of 0.05 indicates a 95 percent confidence level. **Standard_dev** is the population standard deviation for the data range and is assumed to be known. **Size** is the sample size.

Distribuția şi propagarea erorilor. Estimarea erorilor

Erorile aleatoare (accidentale) produc efecte asupra preciziei datelor si rezultatelor. Acestea nu sunt corelate si afecteaza valorile observate (masuratorile) si se considera ca pentru masuratori de volum foarte mare (*n* tinde catre infinit) aceste erori sunt realizari (sunt distribuite) ale unei *variabile aleatoare normale* (**distributia normala Gauss**) X. Proprietatea importanta a aceste *distributii de probabilitati* este aceea ca valorile observate (masurate) se distribuie aleator la stanga si la dreapta fata de valoarea medie, adica satisface legea densitatii de probabilitate Gauss (numita si *clopotul lui Gauss*), *distributia normala standard N(0,1)*, avand media 0 si dispersia 1:

$$f(x) = \frac{h}{\sqrt{\pi}} \cdot e^{(-h^2 \cdot x^2)}, \quad x \in (-\infty, +\infty), \quad h = \frac{1}{\sqrt{2} \cdot \sigma} \text{ (precizia)},$$

si $\lim\limits_{x\to-\infty} f(x) = \lim\limits_{x\to+\infty} f(x) = 0$. Mai jos este graficul densitatii de probabilitate pe intervalul [-2,2] realizat (pasul discretizarii/diviziunii *p=0.1*) cu programul Excel.

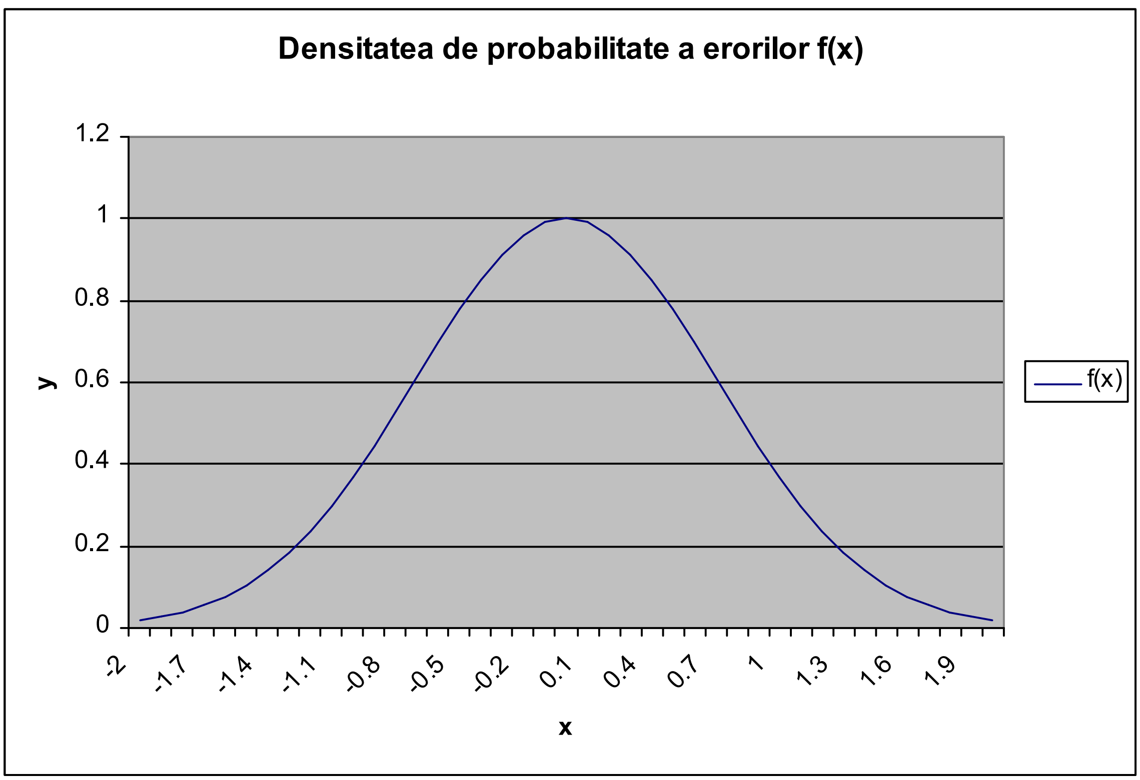


Figura 20. Graficul folosind Excel

Pentru o valoare data $x \in (-\infty,+\infty)$, conform definiției *funcției de repartiție*, probabilitatea ca X < x este data de relatia:

$$F(x) = P\ (X < x) = \int_{-\infty}^{x} f(u)du\ ,$$

adica reprezinta *aria de sub curba normală standard* delimitată de - ∞ şi x .

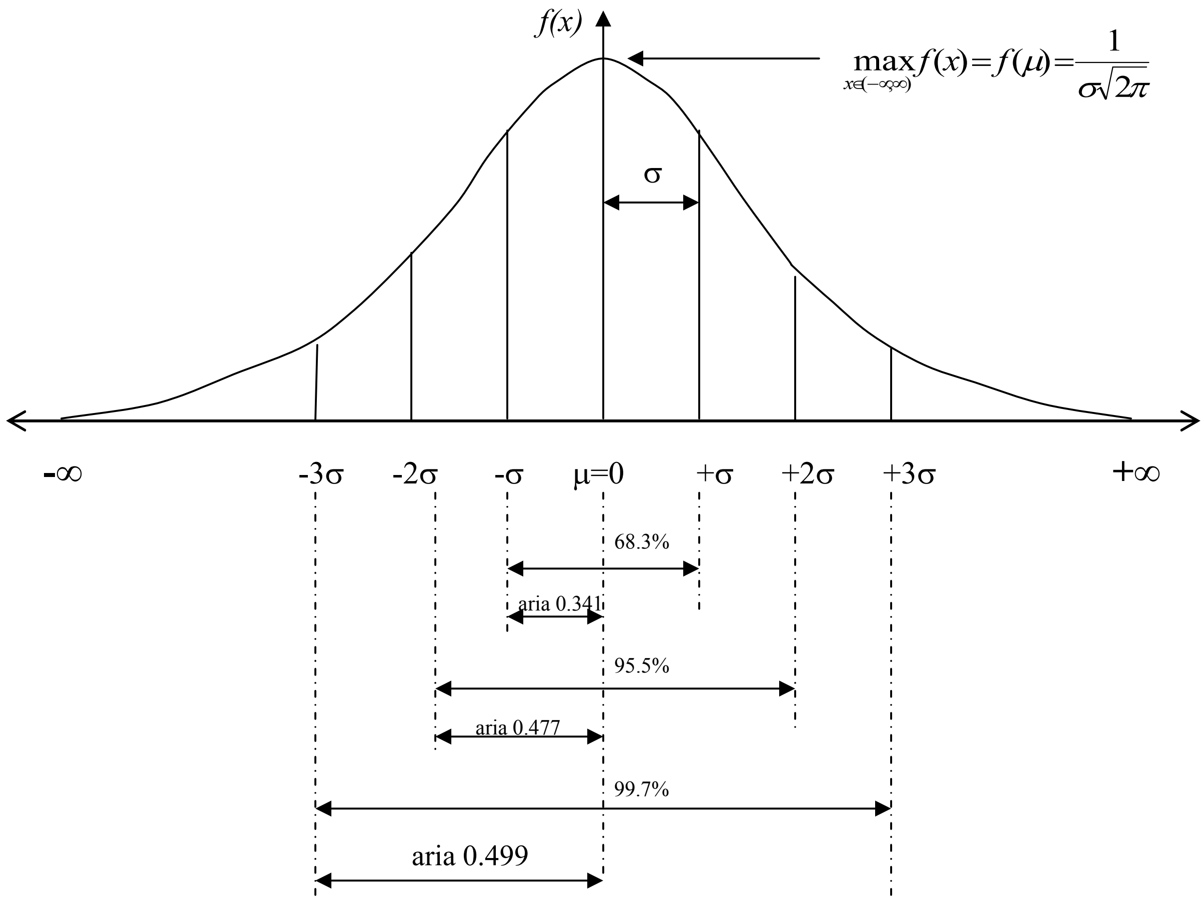


Figura 21. Erorile aleatoare: Distributia probabilitatilor si relatia cu functia de repartitie

Distribuție normală (*Normal Distribution - ND*) – Densitatea de probabilitate Gauss

Prin definiție, o *variabila aleatoare*. X are o repartiție normală cu parametrii μ şi σ dacă densitatea sa de probabilitate este

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad \int_{-\infty}^{+\infty} f(x)dx=1,\ \max_{x\in(-\infty,\infty)} f(x)=f(\mu)=\frac{1}{\sigma\sqrt{2\pi}}$$

Se demonstrează că μ şi σ^2 este media, respectiv dispersia, *variabila aleatoare* X. Conform definiției funcției de repartiție,

$$F(x)=P(X<x)=\int_{-\infty}^{x} f(u)du$$

şi se poate demonstra că pentru orice ***a* ≤ *b***, probabilitatea ca ***a* < (X-m)/s < *b*** este

P(*a* < (X-m)/s < *b*) = aria de sub curba normală standard delimitată de **x = *a*** şi **x = *b***

formulă care permite calcularea probabilităților asociate cu repartiția normală doar cunoscând probabilitățile asociate repartiției normale standard. Notația uzuală este X~N(μ,σ^2). Pentru distribuția normală standard se obține X~N(0,1).

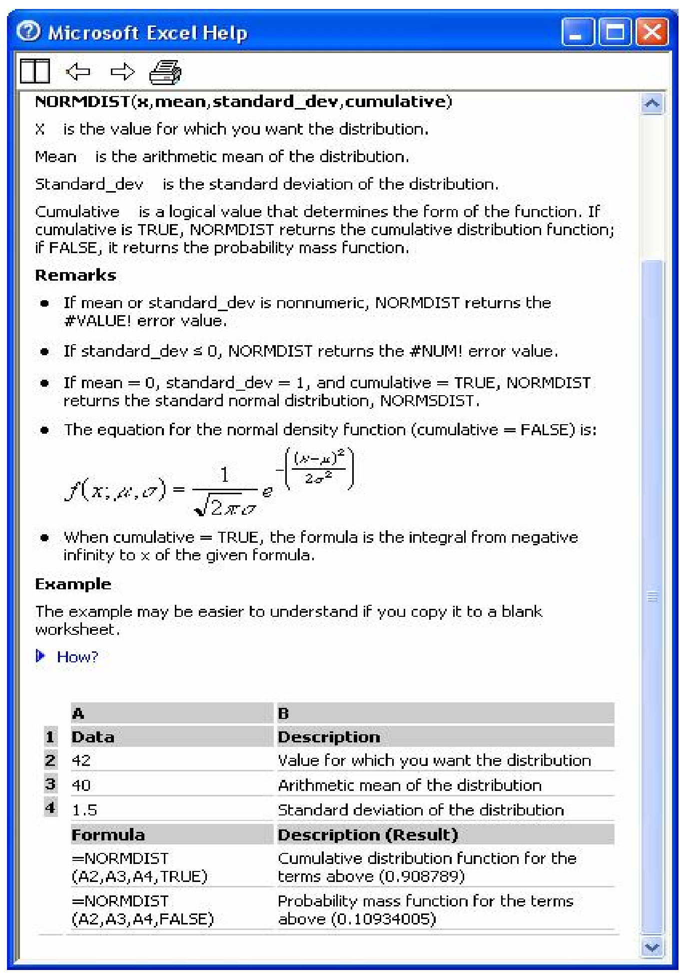
In EXCEL exista functia:
NORMDIST(*x,mean,standard_dev,cumulative*)

- X is the value for which you want the distribution.

- Mean is the arithmetic mean of the distribution. Standard_dev is the standard deviation of the distribution.

- Cumulative is a logical value that determines the form of the function. If cumulative is TRUE, NORMDIST returns the cumulative distribution function; if FALSE, it returns the probability mass function.

- The equation for the normal density function (cumulative = FALSE) is:
- When cumulative = TRUE, the formula is the integral from negative infinity to x of the given formula.


Microsoft Excel Help

NORMDIST(x,mean,standard_dev,cumulative)

X is the value for which you want the distribution.

Mean is the arithmetic mean of the distribution.

Standard_dev is the standard deviation of the distribution.

Cumulative is a logical value that determines the form of the function. If cumulative is TRUE, NORMDIST returns the cumulative distribution function; if FALSE, it returns the probability mass function.

Remarks

- If mean or standard_dev is nonnumeric, NORMDIST returns the #VALUE! error value.
- If standard_dev ≤ 0, NORMDIST returns the #NUM! error value.
- If mean = 0, standard_dev = 1, and cumulative = TRUE, NORMDIST returns the standard normal distribution, NORMSDIST.
- The equation for the normal density function (cumulative = FALSE) is:

$$f(x;\mu,\sigma)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\left(\frac{(x-\mu)^2}{2\sigma^2}\right)}$$

- When cumulative = TRUE, the formula is the integral from negative infinity to x of the given formula.

Example

The example may be easier to understand if you copy it to a blank worksheet.

How?

	A	B
1	**Data**	**Description**
2	42	Value for which you want the distribution
3	40	Arithmetic mean of the distribution
4	1.5	Standard deviation of the distribution
	Formula	**Description (Result)**
	=NORMDIST (A2,A3,A4,TRUE)	Cumulative distribution function for the terms above (0.908789)
	=NORMDIST (A2,A3,A4,FALSE)	Probability mass function for the terms above (0.10934005)

Este remarcat faptul ca pentru o curba a distributiei erorilor cu o medie data μ si cu diverse dispersii σ_1 ,σ_2 şi σ_3 crescatoare. atunci cele trei curbe au "baza" crescatoare asa cum se vede in figura urmatoare:

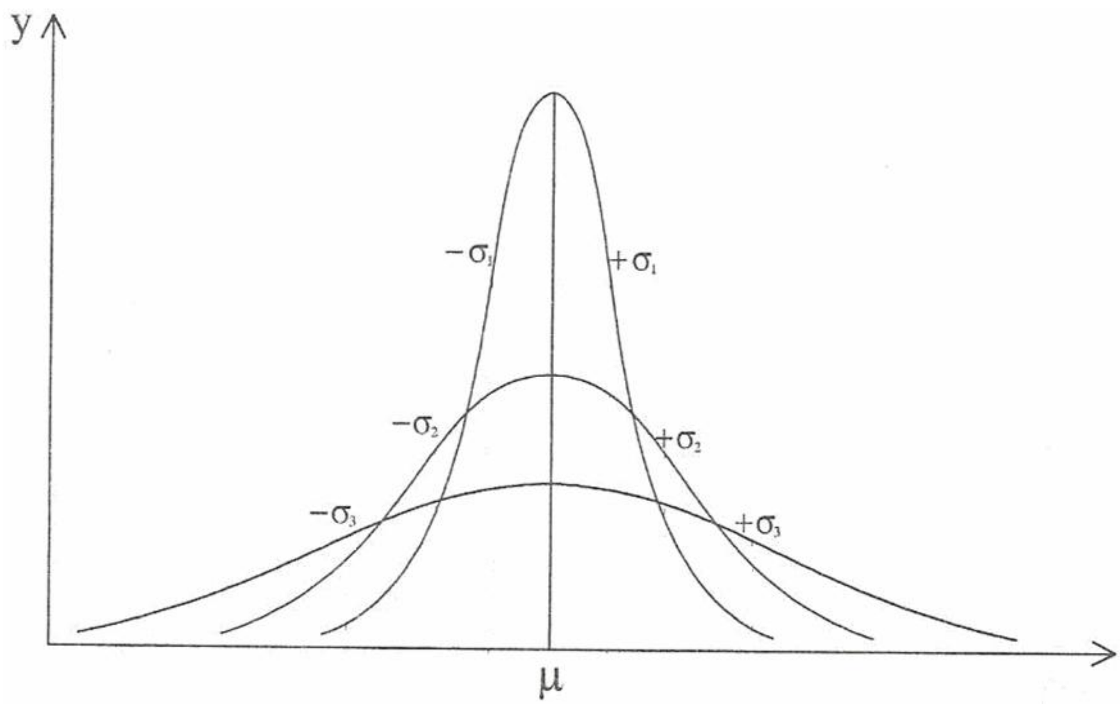


Figura 22. Curbele distributiei pentru diverse dispersii crescatoare σ_1 ,σ_2, σ_3

Modelul teoretic al distributiei erorilor (*curba lui Gauss: distributia normala standard*) se refera la un numar infinit de masuratori pentru valorile masurate (observate). In practica, numarul observatiilor este finit, si uneori acest numar este mic asa cum este cazul domeniilor chimie, fizica, etc. Sa presupunem ca se fac masuratori pentru marimea **Y**. Daca se repeta masurarea marimii Y in conditii identice se constata ca valorile masurate difera intre ele, si atat pentru un numar foarte mare de masuratori (teoretic infinit), cat si pentru un numa mic de masuratori (finit) se obtin doua siruri (seturi) distincte de valori masurate. Daca pentru ambele seturi de valori masurate se reprezinta grafic frecventele de aparitie (distributia probabilitatilor) a valorii masurate in functie de valorile masurate, se obtin doua curbe diferite (a se vedea figura de mai jos). Vom nota:

$\mathbf{Y_r}$ = valoarea adevarata (reala, corecta) a marimii **Y**;

m = media valorilor masurate pentru un numar infinit de masuratori

$\overline{Y}$ = media valorilor masurate pentru un numar mic (finit) de masuratori

Eroarea sitematica (obiectiva) este data de diferenta dintre media valorilor masurate pentru un numar infinit de masuratori si valoarea adevarata a marimii Y , adica m - Y_r . *Eroarea aleatoare* (accidentala)) este data de diferenta dintre media valorilor masurate pentru un numar finit de masuratori si media valorilor masurate pentru un numar infinit de masuratori, adica $\overline{Y}$ - m.

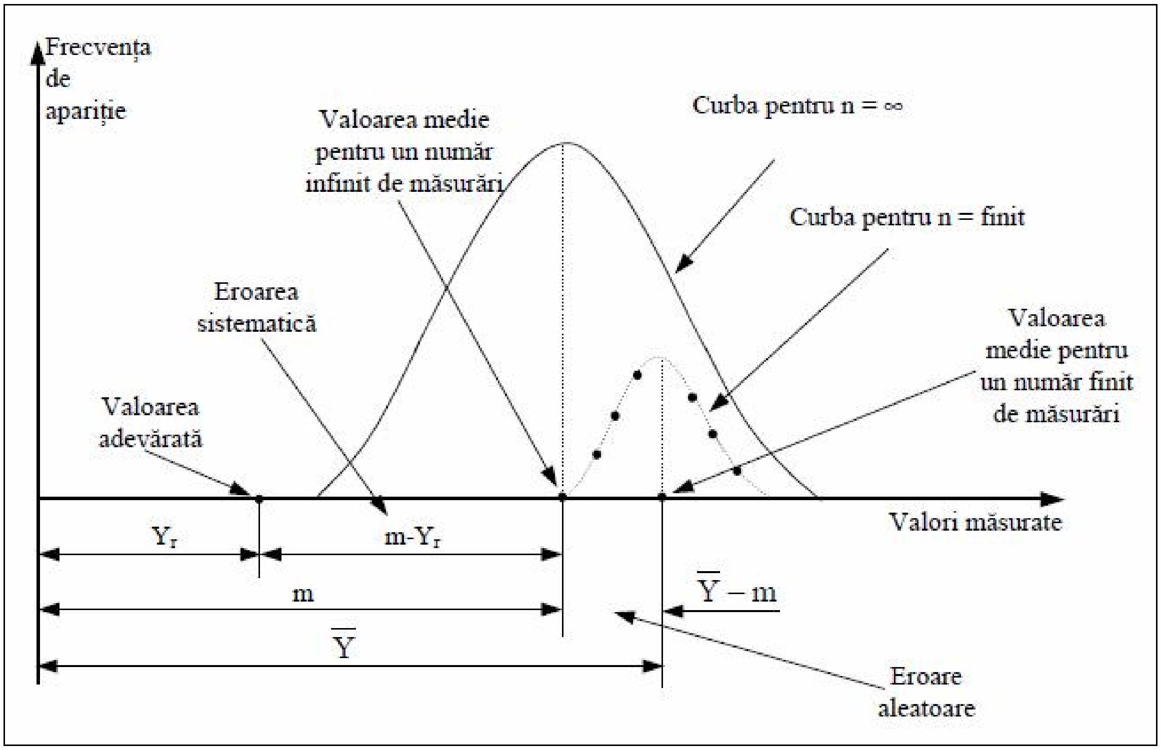


Figura 23. Erori de masurare sistematice si aleatoare
(**Sursa:** M. Miron, L. Miron, *Masurari electrice si electronice*, Brasov, 2003, http://www.afahc.ro/invatamant/electro/mee.pdf)

Propagarea erorilor

Atunci când un *rezultat experimental* depinde de unul sau mai multe masuratori nesigure, este necesar să se analizeze propagarea erorilor (incertitudinile: *propagation of error* or *propagation of uncertainty*) acestor măsurători în rezultat final al cercetarii (experimentului).
In sens statistic, daca X este o *variabila aleatoare* data ce are o distributie cunoscuta a erorilor si asupra ei actioneaza un sistem de prelucrare (experiment system), se doreste sa sa cunoasca propagarea erorilor (distributia erorilor) pentru variabila aleatoare rezultat Y:

(input) **X** ⟶ **SISTEM** (experimet system) ⟶ **Y** (Output)

Trebuie sa se determine *distributia functiei de iesire* pentru variabila Y, adica Y = f(X), unde f este cunoscuta si distributia erorilor pentru varaiabila aleatoare X este cunoscuta.

Presupunem ca variabila X (input) este normal distribuita N(μ_x , σ_x) cu media μ_x si abaterea standar σ_x si se doreste sa se determine cum se propaga intervalul cu probabilitatea 68% [μ_x - σ_x , μ_x + σ_x] prin sistemul de prelucrarea in rezultatul final, adica in variabila iar Y (output). Daca f este o functie complexa, din figura urmatoare se poate observa ca aceste interval depinde de aceasta functie sa determine o anumita distributie a erorilor pentru rezultatul final Y. In cazul normal distribuit pentry Y, avem notatia N (μ_y , σ_y).

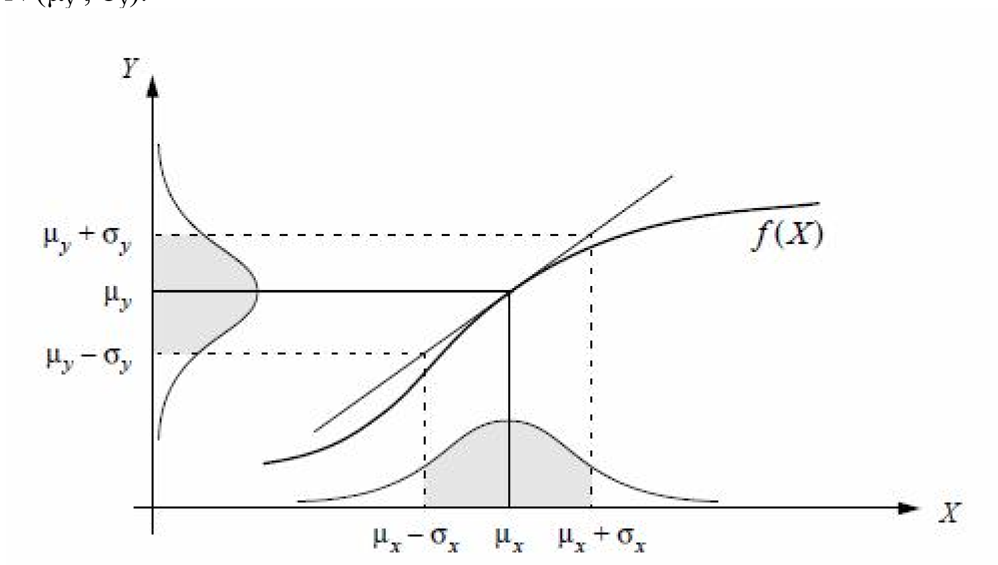


Figura 24. Propagarea erorilor pentru cazul neliniar al rezultatului

Pentru cazul general cand avem n varaibila aleatoaea la intrare (input) X_1 , X_2, ... X_n , avem urmatoarea schema generala:

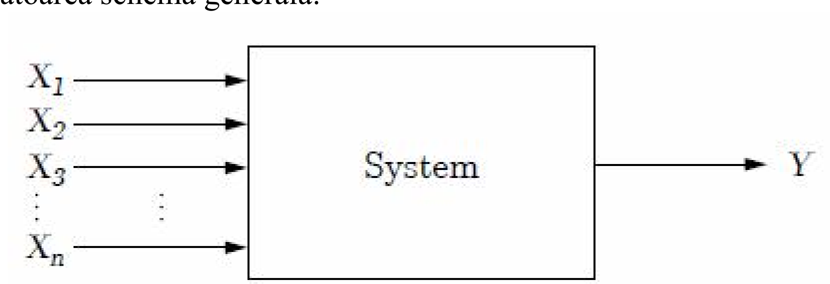


Figura 25. Schema generala pentru n intrari

In acest caz avem $Y = f(X_1 , X_2, \ldots X_n)$, unde $X_1 , X_2, \ldots X_n$ sunt variabile aleatore de intrare (input) avand distributia normala N(μ_i , σ_i), unde $i \in \{1,2,\ldots,n\}$.
In acest caz, reprezentarea lui Y sub forma dezvoltatii in serie Tayloy de ordinul I (se utilizeaza doar deriva de ordinul I)) in punctul (μ_1 , μ_2, ... , μ_n) este

$$Y \approx f(\mu_1, \mu_2, \ldots, \mu_n) + \sum_{i=1}^{n} \left[\frac{\partial f}{\partial X_i}(\mu_1, \mu_2, \ldots, \mu_n)\right][X_i - \mu_i]$$

Daca pentru **medie** utilizam notatia din statistica (probabilitati), **E (.)**, atunci avem urmatoarele calcule:

$$Y \approx a_0 + \sum a_i(X_i - \mu_i)$$, cu notatiile

$$a_0 = f(\mu_1, \mu_2, \ldots, \mu_n),$$

$$a_i = \frac{\partial f}{\partial X_i}(\mu_1, \mu_2, \ldots, \mu_n)$$

Vom presupune ca functia *f* este liniara si astfel *Y* este o variabila aleatore distribuita normal N(μ_y , σ_y) cu media μ_y si abaterea standar σ_y . sa calculam μ_y si σ_y^2 :

$$\begin{aligned}\mu_Y &= E[Y] = E[a_0 + \sum_i a_i(X_i - \mu_i)] \\ &= E[a_0] + \sum_i E[a_i X_i] - E[a_i\mu_i] \\ &= a_0 + \sum_i a_i E[X_i] - a_i E[\mu_i] \\ &= a_0 + \sum_i a_i\mu_i - a_i\mu_i\end{aligned}$$

adica

$$\begin{aligned}&= \sum_i a_i^2 E[(X_i - \mu_i)^2] + \sum_{i \neq j}\sum a_i a_j E[(X_i - \mu_i)(X_i - \mu_j)] \\ &= \sum_i a_i^2\sigma_i^2 + \sum_{i \neq j}\sum a_i a_j \sigma_{ij} \\ \sigma_Y^2 &= \sum_i \left(\frac{\partial f}{\partial X_i}\right)^2 \sigma_i^2 + \sum_{i \neq j}\sum \left(\frac{\partial f}{\partial X_i}\right)\left(\frac{\partial f}{\partial X_j}\right)\sigma_{ij}\end{aligned}$$

si daca vom considera ca variabilele aleatoare X_1 , X_2, ... X_n sunt independente, atunci covarianta $\boldsymbol{\sigma_{ij}}$ este zero si avem

$$\sigma_Y^2 \approx \sum_i \left(\frac{\partial f}{\partial X_i}\right)^2 \sigma_i^2 .$$

Pentru exemplificare vom da cateva exemple de operatii asupra intrarilor. Calculul erorii rezultatului final va fi analilat in cele ce urmeaza.

Input: **a, b, c** obtinute din masuratori directe cu erorile $\mathbf{s_a}$, $\mathbf{s_b}$, $\mathbf{s_c}$

Output: rezultatul final **x**, cu eroarea $\mathbf{s_x}$

Nr. crt.	**Rezultatul final**	**Propagarea erorilor**
1	$x = a + b - c$	$s_x = \sqrt{s_a^2 + s_b^2 + s_c^2 + ...}$
2	$x = a * b/c$	$s_x = x \cdot \sqrt{\left(\frac{s_a}{a}\right)^2 + \left(\frac{s_b}{b}\right)^2 + \left(\frac{s_c}{c}\right)^2 + ...}$
3	$x = a^{bc}$	$s_x = x \cdot \sqrt{\left(\frac{s_a}{a}\right)^2 + \left(\frac{s_b}{b}\right)^2 + \left(\frac{s_c}{c}\right)^2 + ...}$

Tabelul 2. Propagarea erorilor

De exemplu, se poate calcula eroarea la *etalonul de curent* pe baza legii lui Ohm, sau in general la masurarea indirecta a curentului, prin masurarea *caderii de tensiune* pe o rezistenta etalon. In Chimie si Fizica sunt diverse formule de calcul pentru care trebuie sa se calculeze eroarea.

Analiza datelor experimentale. Modele matematice si statistice

In cercetare si in analiza datelor experimentale din diverse domenii stiintifice trebuie sa se realizeze proceduri de calcul si modele care sa conduca la concluzii privind interpretarea masuratorilor, calculelor si rezultatelor *modelelor teoretice* sau *empirice (aproximative)*.

Presupunem ca trebuie sa se studieze variabila Y (dependenta) in functie de variabila X (independenta), adica dependenta Y = f(X), de exemple daca X reprezita parametrul "*temperatura*", iar Y parametrul "*presiune*". In acest caz variabila Y se exprima ca o functie de o singura variabila. Considerăm că s-au determinat n perechi de valori (x_i,y_i), i=1,…,n corespunzătoare celor două variabile pentru care se doreste să se studieze *asocierea* şi *relația* dintre ele. O primă apreciere asupra *distribuției comune* o vom avea dacă realizăm *diagrama de împrăştiere a valorilor*, de fapt reprezentarea într-un sistem de axe XOY pentru punctele având coordonatele (x , y). Analiza vizuală a organizării şi formei *norului de puncte* obținut poate oferi indicii importante asupra relației dintre variabile. Datele vor susține ipoteza asocierii între variabile dacă forma norului de puncte se apropie de o curbă data cu expresie analitica cunoscuta. Astfel, se pot aprecia asocieri *liniare*, *curbilinii*, etc. Dacă în norul de puncte nu se poate distinge o tendință, se va spune că variabilele nu sunt corelate. Diversitatea priceselor si fenomenelor studiate determina obtinerea unei mari diversitati de tendinte: *liniare* si *neliniare* (curbilinii).

În figuririle următoare sunt ilustrate câteva tendințe ale acestor asocieri.

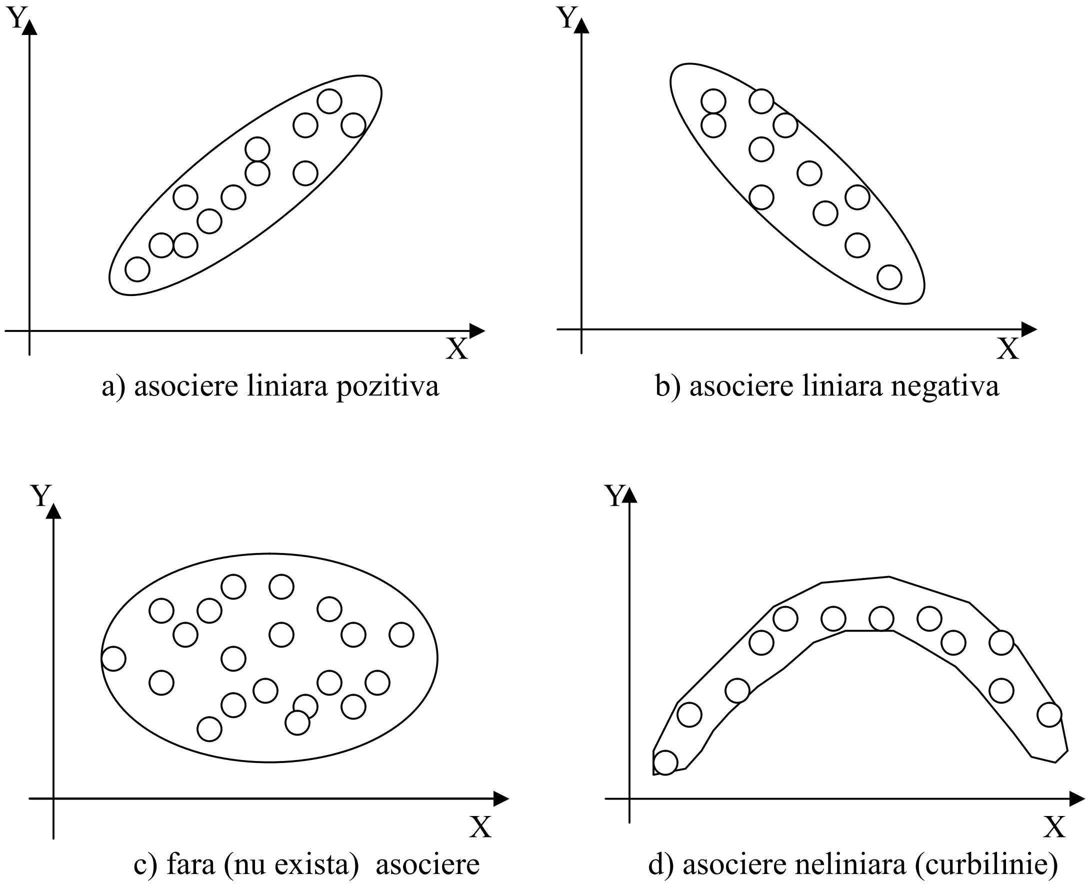


Figura 26. Diferite tipuri de asociere pentru variabilel X si Y

Pentru a sintetiza (estima) modul în care schimbările variabilei Y sunt asociate cu schimbările variabilei X, se utilizeaza metoda matematică "*metoda celor mai mici pătrate - MCMMP*" (conceputa de Legendre, 1806). Aplicată în cazurile a) si b), asocierea dintre X şi Y este reprezentată printr-o dreaptă trasată printre punctele diagramei de împrăştiere. Dreapta estimată (*dreapta de regresie*) este "*cea mai bună*" în sensul că exprimă cel mai central drum printre puncte: linia pentru care *suma pătratelor distanțelor* (pe verticală) dintre puncte şi dreaptă este *minimă*.

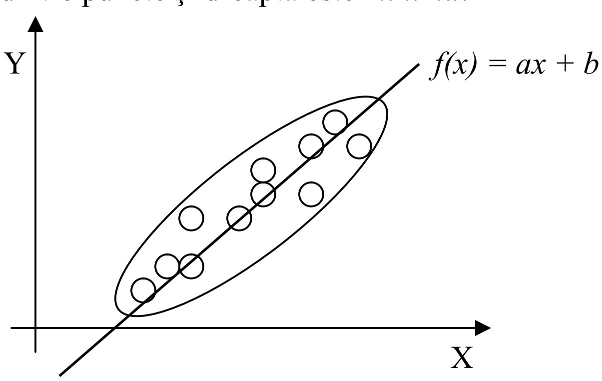


Figura 27. Dreapta de regresie in cazul a)

Distanțele $y_i - f(x_i)$, $i=1,\ldots,n$ sunt considerate ca erori (*reziduuri*) intre valorile masurate si valorile estimate. Dreapta de regresie $f(x) = ax + b$ realizează *valoarea minimă a pătratelor erorilor* (parametri dreptei a si b urmeaza a fi determinati prin MCMMP),

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

în sensul că orice altă dreaptă produce o sumă de pătrate mai mare. Este de amintit că o proprietate a *mediei aritmetice* este aceea că suma pătratelor diferențelor de la medie are o valoare minimă. Astfel se poate spune că după cum *media* reprezintă *punctul de echilibru* pentru o *distribuție univariată* de scoruri, la fel dreapta de regresie reprezintă *punctul de echilibru* într-o *distribuție bivariată*. Utilitatea dreptei de regresiei este aceea că serveşte ca bază pentru *predicția valorilor* lui *Y* asociate valorilor lui *X*.

In cazul *asocierii neliniare* (curbilinie), curba care estimeaza asocierea dintre varabilele Y si X va fi exprimata prin intermediul unor parametri ce urmeaza a fi determinati prin MCMMP. In practica, in functie de natura datelor experimentale si procesul analizat trebuie sa se determine '*evolutia*" procesului pe baza datelor experimentale. Aceasta este reprezentata si estimata de *modele matematice* date de *functii liniare* sau *neliniare* (curbe).

Modelele matematice (liniare sau neliniare) ce estimeaza evolutia proceselor sau fenomenelor sunt exprimate de:

- *Modele teoretice* - acestea se bazeaza pe diverse legi si principii ale domeniului teoretic; sunt *modele rationale* ce se determina prin functii si legi obtinute prin rationamente teoretice ce exprima *functii si ecuatii* ale unor teorii studiate in domeniul respectiv: chimie, fizica, biologie, etc.
- *Modele empirice* (de aproximare) - acestea au la baza un suport teoretic pentru a utiliza *observatii (masuratori) empirice* ale unor *parametri* ce definesc procesele si fenomenele in vederea realizarii de *calcule si aproximari (fitare)* ale datelor.

Modele teoretice

Exemple.

a) Legea densitatii de probabilitate Gauss privind distributia erorilor de masurare (numita si *clopotul lui Gauss*), *distributia normala standard N(0,1)*, avand media 0 si dispersia 1:

$$f(x) = \frac{h}{\sqrt{\pi}} \cdot e^{(-h^2 \cdot x^2)}, \; x \in (-\infty,+\infty), \; h = \frac{1}{\sqrt{2}} \text{ (precizia)},$$

si $\lim_{x \to -\infty} f(x) = \lim_{x \to +\infty} f(x) = 0$.

b) Exemplu din chemical kinetics (*teoria starii de trazitie* – "**'transition state theory'**") - ecuatia **Eyring–Polanyi** (1935) ce descrie dependența de temperatură a ratei de reacție intr-o reactie bimoleculara. Principiile teoriei starii de tranzitie: există un echilibru termodinamic între starea de tranzitie şi starea de reactanți în partea de sus a barierei de energie; rata de reactie chimica este proporțională cu concentrația de particule în stare de tranziție de înaltă energie. **Modelul dat de ecuația Eyring** este folosit în studiul gazelor prin reacții condensate şi mixte (**Sursa**: Peter Keusch, University of Regensburg, http://www.demochem.de/eyr-e.htm):

$$k = \frac{k_B \cdot T}{h} \cdot e^{-\frac{\Delta H^{\ddagger}}{RT}} \cdot e^{\frac{\Delta S^{\ddagger}}{R}}$$

, unde

variabila dependenta k este functie de temperatura T si de parametri $\Delta S^{\ddagger}$ (entropia de activare), $\Delta H^{\ddagger}$ (entalpia de activare) si

k_B = Boltzmann's constant [$1.381 \cdot 10^{-23}$ $J \cdot K^{-1}$]
T = absolute temperature in degrees Kelvin [K]
h = Pank constant [$6.626 \cdot 10^{-34}$ $J \cdot s$]
R = Universal Gas Constant = 8.3144621 [$J \cdot mol^{-1} \cdot K^{-1}$]
$\Delta S^{\ddagger}$ = activation entropy [$J \cdot mol^{-1} \cdot K^{-1}$]
$\Delta H^{\ddagger}$ = activation enthalpy [$kJ \cdot mol^{-1}$]

Observatii:

$$k^{\ddagger} = \frac{k_B \cdot T}{h}$$

is given by statistical thermodynamics,
$k^{\ddagger}$ is known as a **universal rate constant for a transition state**.

$$\Delta G^{\ddagger} = \Delta H^{\ddagger} - T \cdot \Delta S^{\ddagger}$$

, **$\Delta G^{\ddagger}$** = free activation enthalpy [$kJ \cdot mol^{-1}$] (Gibbs energy),
$\Delta G^{\ddagger}$ is also described by the Legendre transformation of the Gibb's free energy function.
$\Delta G^{\ddagger}$ poate fi considerată a fi *forța motrice a unei reacții chimice*, ce determină spontaneitatea de reacție: reacția este spontană (< 0), sistem in echilibru (= 0), reacția nu este spontana (> 0).
Prin logaritmare, **ecuația Eyring** se transforma intr-un model liniar:

$$\ln k = \ln \frac{k_B}{h} \cdot T - \frac{\Delta H^{\ddagger}}{R} \cdot \frac{1}{T} + \frac{\Delta S^{\ddagger}}{R}$$

$$\ln \frac{k}{T} = -\frac{\Delta H^{\ddagger}}{R} \cdot \frac{1}{T} + \ln \frac{k_B}{h} + \frac{\Delta S^{\ddagger}}{R}$$

Modele empirice (de aproximare)
Exemple.
a) **Ecuația Arrhenius** – ecuația se poat aplica numai la *cinetica reacțiilor de gaz* si se bazează pe *observația empirică* a faptului că o reacție se desfăşoară cu o creştere a ratei de reacție la o temperatură mai ridicată:

$k = A \cdot e^{-\frac{E_a}{RT}}$, unde A factor si E_a este energia de activare.

$$E_a = \Delta H^{\ddagger} + RT \qquad \ln k = -\frac{E_a}{RT} + \ln A$$

(forma liniara)

b) **Legea lui Beer** (Spectrofotometrie): $A = \varepsilon L C$, unde A este absortia, ε este cantitate este de absorbție molară, L este lungimea de undă a luminii folosite la măsurare, iar C este C este concentrația analitului (Sursa: David N. Blauch, Beer's Law: http://www.chm.davidson.edu/vce/spectrophotometry/beerslaw.html,si http://teaching.shu.ac.uk/hwb/chemistry/tutorials/molspec/beers1.htm).

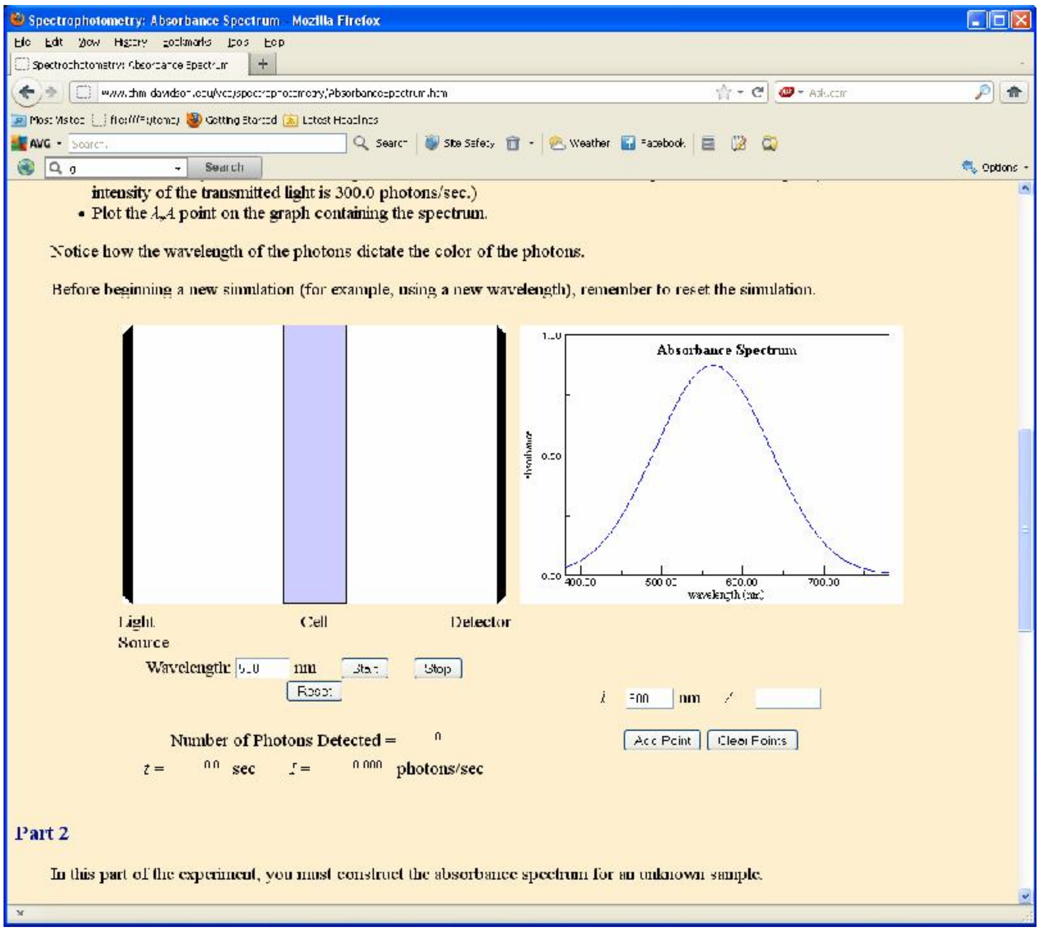


Figura 28. Virtual Chemistry Experiments by David N. Blauch - http://www.chm.davidson.edu/vce/

Coeficientul de corelație (*Correlation coefficient*)

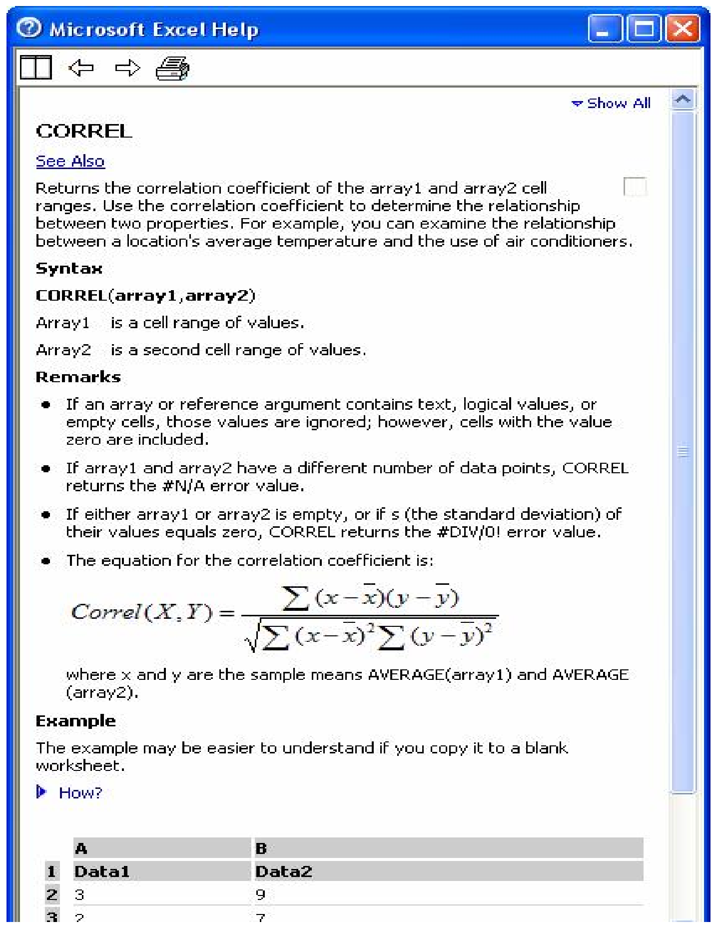
Coeficientul de corelație (Pearson) este o măsură a asocierii liniare dintre două variabile, cu alte cuvinte a gradului în care reprezentarea bivariată sub forma unei diagrame de împrăştiere se apropie de o dreaptă. Notând cu X şi Y cele două variabile şi cu x_i, y_i, $i=1,\ldots,n$, valorile variabilelor, formula de calcul este

$$r_{XY} = \frac{\sum (x_i - \overline{X})(y_i - \overline{Y})}{\sqrt{\left(\sum (x_i - \overline{X})^2\right)\left(\sum (y_i - \overline{Y})^2\right)}}.$$

Coeficientul de corelaţie ia valori în [–1,+1] cu semnificaţia de *asociere pozitivă/negativă* după semnul coeficientului şi de *lipsă de asociere* pentru $r_{XY} = 0$.

Exercitiu. Pentru un set de date ce reprezinta valorile a doua variabile aleatoare X şi Y vom calcula in trei moduri coeficientul de corelatie r_{XY} : a) folosind functia **CORREL (X,Y)** din Excel, b) folosind Excel pentru calculele directe ale formulei de mai sus, si c) folosind covarianta **COVAR (X,Y)** din Excel.

X	Y
12.6	0.42365
12.7	1.692047
12.8	2.963326
12.9	4.22442
13	5.462171
13.1	6.663465
13.2	7.81537
13.3	8.905278
13.4	9.921037
13.5	10.85109
13.6	11.6846
13.7	12.41158
13.8	13.023
13.9	13.5109
14	13.8685
14.1	14.09026
14.2	14.17198
14.3	14.11084
14.4	13.90547
14.5	13.55598
14.6	13.06395
14.7	12.43248
14.8	11.66613
14.9	10.77093
15	9.754318


Microsoft Excel Help

Show All

CORREL

See Also

Returns the correlation coefficient of the array1 and array2 cell ranges. Use the correlation coefficient to determine the relationship between two properties. For example, you can examine the relationship between a location's average temperature and the use of air conditioners.

Syntax

CORREL(array1,array2)

Array1 is a cell range of values.

Array2 is a second cell range of values.

Remarks

- If an array or reference argument contains text, logical values, or empty cells, those values are ignored; however, cells with the value zero are included.
- If array1 and array2 have a different number of data points, CORREL returns the #N/A error value.
- If either array1 or array2 is empty, or if s (the standard deviation) of their values equals zero, CORREL returns the #DIV/0! error value.
- The equation for the correlation coefficient is:

$$Correl(X,Y) = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2 \sum (y-\bar{y})^2}}$$

where x and y are the sample means AVERAGE(array1) and AVERAGE (array2).

Example

The example may be easier to understand if you copy it to a blank worksheet.

How?

	A	B
1	Data1	Data2
2	3	9
3	2	7

Varianta a)		**0.775901**
Varianta b)		**0.775901**
Varianta c)		**0.775901**
Corelatia (X,Y)		
	Medie X	Medie Y
	13.8	**10.03771**

Valori identice!

	Suma C	Suma D	Suma E			
	57.6555	13	424.7427			
	Numarator	Numitor				
	57.6555	74.30784				
	A	**B**	**C**	**D**	**E**	
	-1.2	-9.61406	11.53687	1.44	92.43017	
	-1.1	-8.34566	9.180231	1.21	69.65011	
	-1	-7.07439	7.074386	1	50.04693	
	-0.9	-5.81329	5.231962	0.81	33.79435	
	-0.8	-4.57554	3.660432	0.64	20.93556	
	-0.7	-3.37425	2.361972	0.49	11.38554	
	-0.6	-2.22234	1.333405	0.36	4.938799	
	-0.5	-1.13243	0.566217	0.25	1.282406	
	-0.4	-0.11667	0.04667	0.16	0.013613	
	-0.3	0.813378	-0.24401	0.09	0.661584	
	-0.2	1.646889	-0.32938	0.04	2.712245	
	-0.1	2.373869	-0.23739	0.01	5.635252	
	0	2.985289	0	0	8.91195	
	0.1	3.473193	0.347319	0.01	12.06307	
	0.2	3.830792	0.766158	0.04	14.67496	
	0.3	4.052551	1.215765	0.09	16.42317	
	0.4	4.134267	1.653707	0.16	17.09216	
	0.5	4.073128	2.036564	0.25	16.59037	
	0.6	3.867761	2.320656	0.36	14.95957	
	0.7	3.518267	2.462787	0.49	12.3782	
	0.8	3.02624	2.420992	0.64	9.158127	
	0.9	2.394767	2.15529	0.81	5.734909	
	1	1.628419	1.628419	1	2.651749	
	1.1	0.733221	0.806543	1.21	0.537613	
	1.2	-0.28339	-0.34007	1.44	0.080312	

$$A = X - \overline{X}\ ; B = Y - \overline{Y}\ ; C = A * B\ ; D = A^2\ ; E = B^2$$

In cazul a) se apeleaza functia **CORREL(Array1,Array2)**, unde **Array1**, **Array2** sunt, respectiv, zonele care conţin valorile celor două variabile (trebuie să aibă, evident, acelaşi număr de valori), adica X si Y. Mai jos este fereastra oferita prin apelul functiei CORREL. Se va indica, pe rand fiecare argument in parte: X si Y. Rezultatul obtinut este urmatorul: r_{XY} = **0.775901**.

In cazul b) trebui sa se realizeze calculul direct, adica este nevoie sa se utilizeze 5 vectori A, B, C, D , E definiti tinand seama de expresia dion formula coeficientului de corelatie r_{XY} . Deasupra tabelului de mai sus in care se calculeaza cei 5 vectori se calculeaza valorile intermediare din structura expresiei coeficientului de corelatie si se va obtine acelasi rezultat r_{XY} = **0.775901**.

$$r_{XY} = \frac{\sum \overset{A}{(x_i - \overline{X})}\overset{B}{(y_i - \overline{Y})}}{\sqrt{\left(\sum (x_i - \overline{X})^2\right)\left(\sum (y_i - \overline{Y})^2\right)}}$$

C=A*B, C=A^2, D=B^2

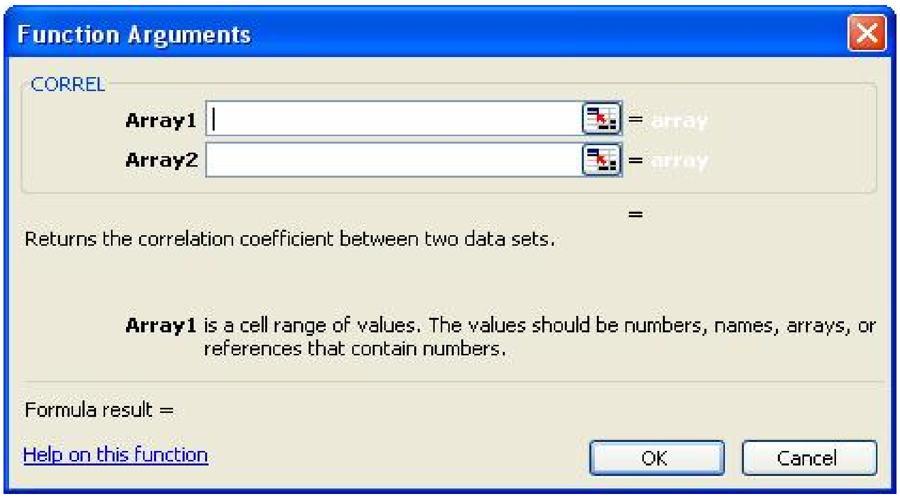


Figura 29. Fereasta oferita de functia CORREL

Cazul c). Calculul coeficientul de corelație al celor doi vectori de date se poate exprima si folosind formula de mai jos:

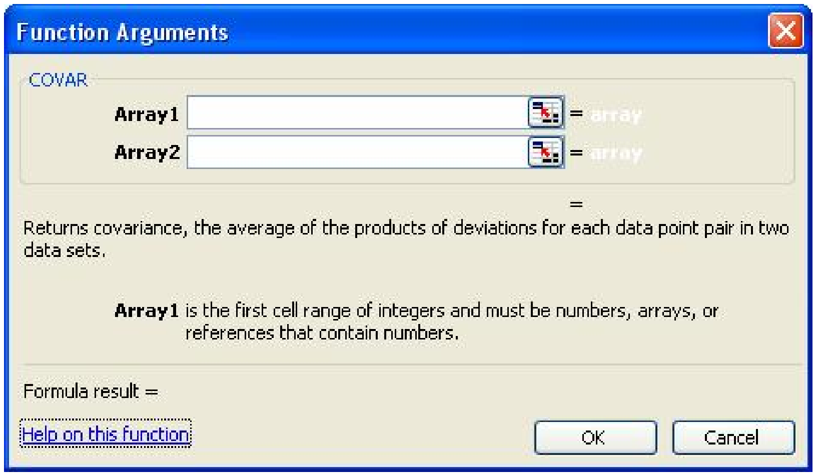
$$r_{XY} = \frac{Cov(X,Y)}{S_X S_Y},$$

unde *Cov(X,Y)* este *covarianta* celor doi vectori X si Y, iar S_X , S_Y sunt abaterile standard pentru X, respectiv Y. Avem: $S_X = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}}$ si $S_Y = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \overline{y})^2}{n}}$ ..

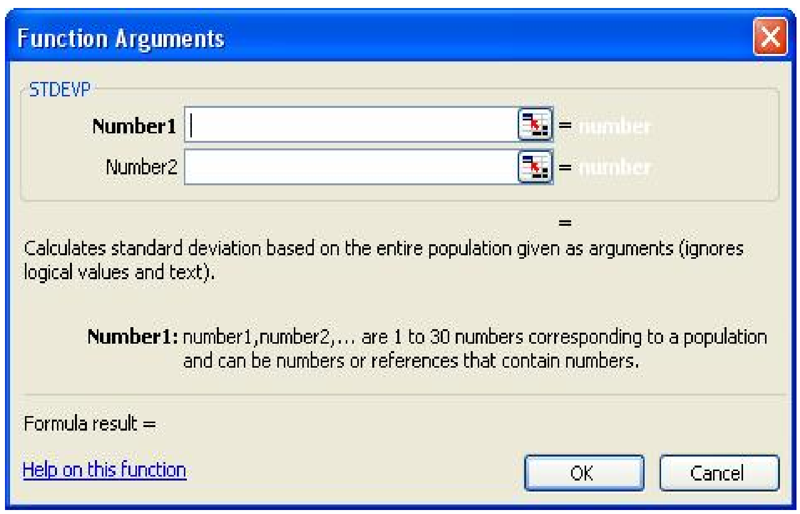
Covarianța (*Covariance*)

Coeficientul de covarianța este o măsură a asocierii liniare dintre două variabile X si Y, $Cov(X,Y) = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{n}$, unde $\overline{x}$ şi $\overline{y}$ reprezintă mediile vectorilor X şi Y. Calculul covarianței folosind funcția statistică din Excel, se face prin apelul functiei

COVAR(Array1,Array2), unde **Array1**, **Array2** sunt, respectiv, zonele care conţin valorile celor două variabile (trebuie să aibă, evident, acelaşi număr de valori), adica X si Y.

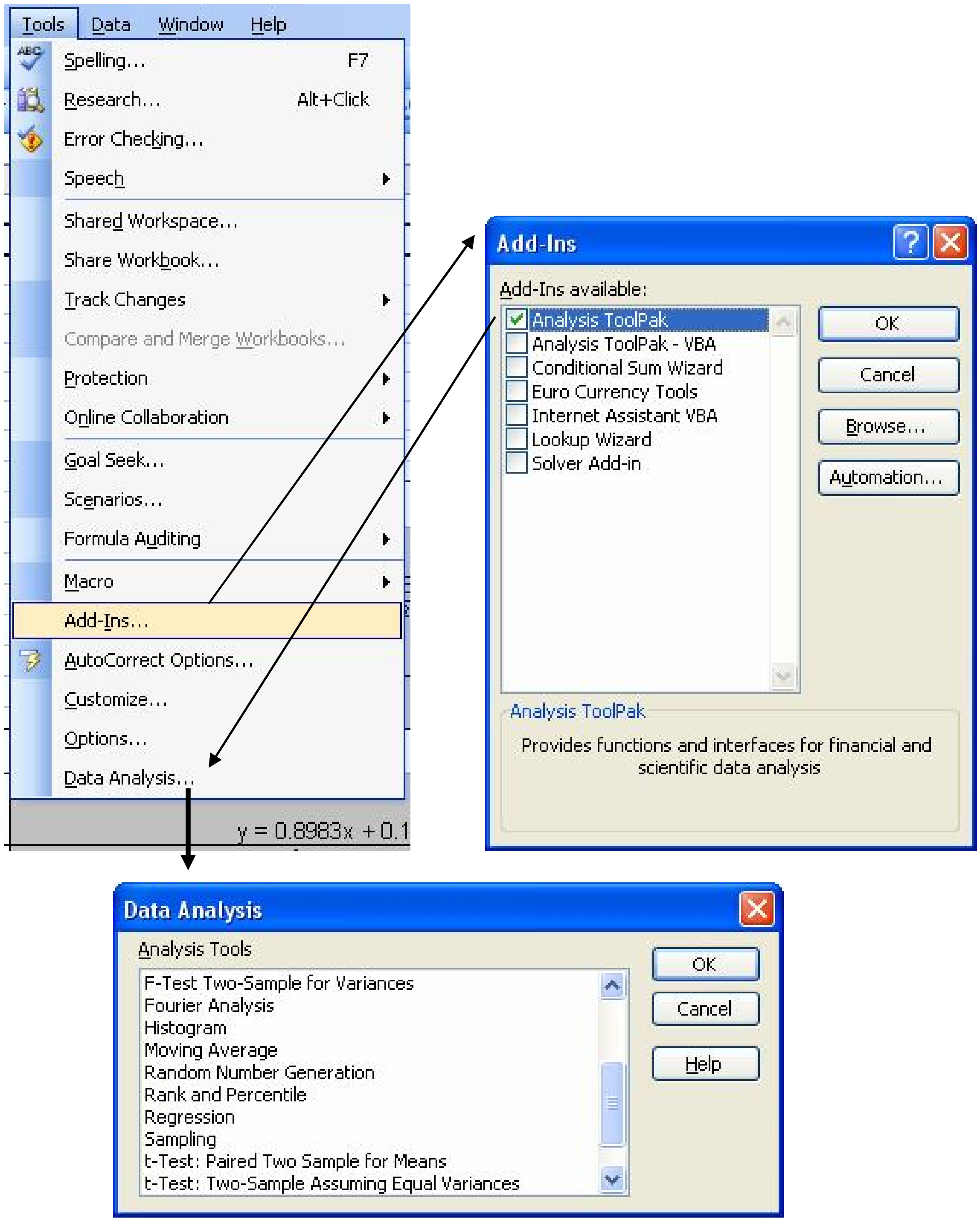


Pentru calculul abaterilor standard S_X , S_Y se apeleaza functia **STDEVP**(**number1**, number2, ...), number1, number2, ... are 1 to 30 number arguments corresponding to a population. You can also use a single array or a reference to an array instead of arguments separated by commas.



In acest fel, si in cazul c) se va obtine acelasi rezultat r_{XY} = **0.775901**.

Pentru diverse probleme complexe ce necesita anumite *calcule statistice*, trebuie sa se cunoasca si sa se inteleaga semnificatia termenilor si calculelor statistice corespunzatoare si apoi sa se utilizeze *instrumentele statistice* (*Analysis ToolPak*, *Analysis ToolPak* – VBA, *Solver Add-in*, etc.) oferite de programul Excel. Acest lucru este valabil si in cazul problemelor ce necesita rezolvarea ecuatiilor si a sistemelor. Trebuie sa se utilizeze meniul **Tools** → *Add-Ins* (va aparea submeniul ***Data Analysis*** in meniul **Tools**):



Despre programul Microsof Office Excel (versiunea 2007- 2010)

In comparatie cu versiuenea **Microsoft Office Excel versiunea 2003-2007** ce ofera pentru o foaie de calcul (sheet) dimensiune **65536R** x **256 C** (linii si coloane: se actioneaza simultan tastele <**CTRL**>+ <↓>, respectiv <**CTRL**>+ <→>) si extensia pentru fisierul output (rigistru, agenda – work) este data de **.xls**, noua versiune 2007-2010 ofera pentru o foaie de calcul (sheet) cu dimensiunea mult mai mare **1048576R x 16384C** si extensia sub forma. **.xlsx**. Referitor la formatul acestei extensii, trebuie sa facem observatia ca in practica, un utilizator care lucreaza cu *versiunea veche* Excel 2003-2007 si deschide un fisier cu acest format, trebuie sa se asigure ca in *versiunea noua* Excel 2007-2010 este neaparat necesar sa se salveze pentru versiunea Excel 2003-2007.

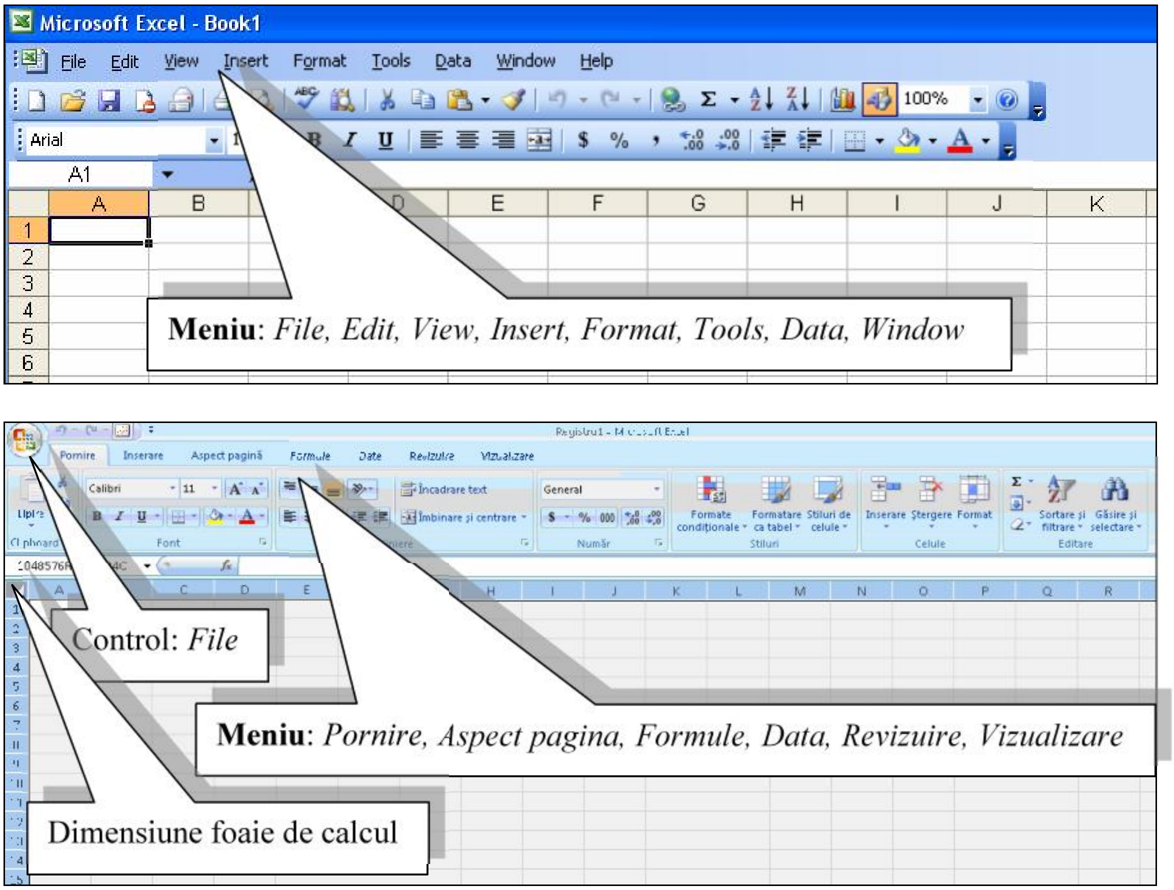
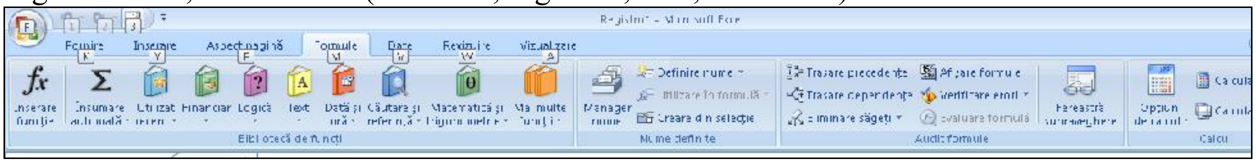


Figura 30. Meniurile principale pentru versiunile Excel 3003-2007 si 2007-2010

MeniulPORNIRE

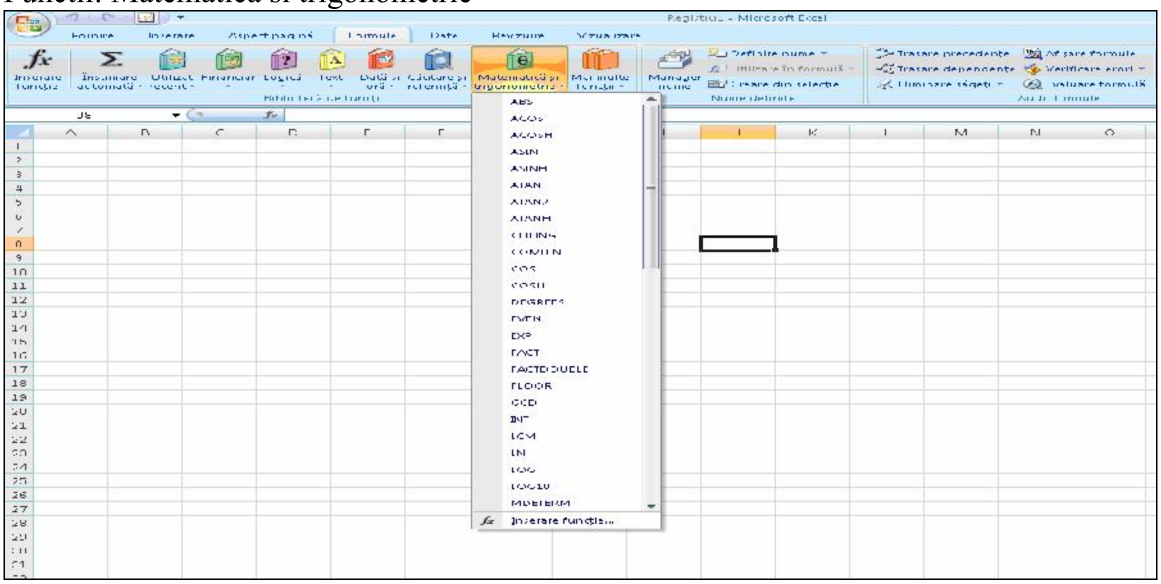
Meniul INSERARE

Meniul FORMULE: Financiar, Logica, Text, Date, Cautare si referinte., Matematica si trigonometrie , Alte functii (Statistica, Inginerie, Cub, Informatii)



Meniul DATE

Functii: Matematica si trigonometrie



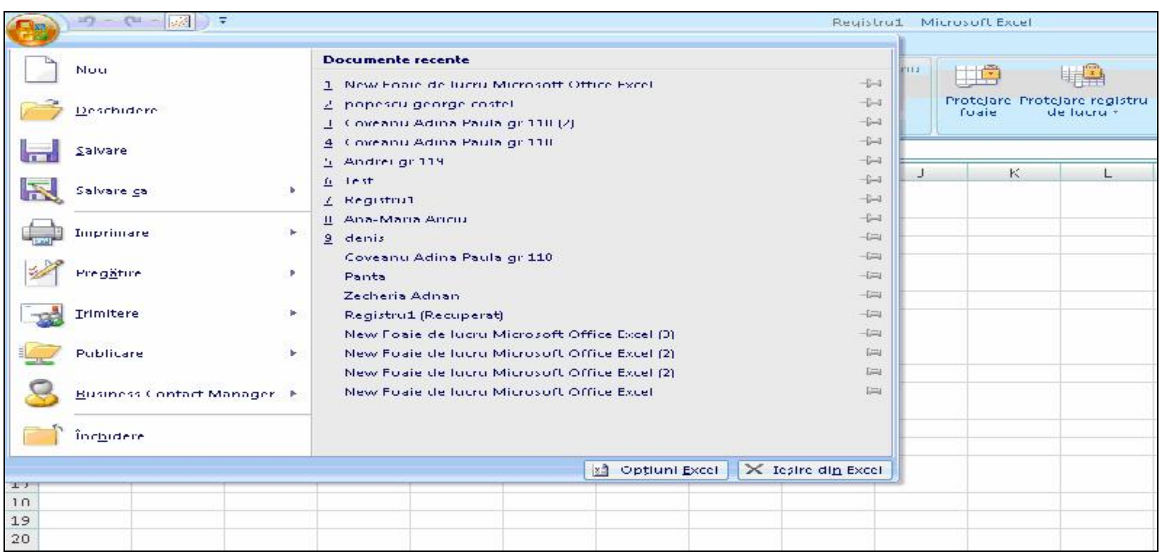


Figura 31. Centrul de Control: File

Regresia liniară (*Regression, Linear Regression*)

Date fiind valorile observate pentru două variabile aleatoare X şi Y, fie acestea (x_i,y_i), $i=1,\ldots,n$, prin *funcție de regresie* se va înțelege acea funcție $Y = f(X)$ care aproximează cel mai bine setul de date observate. De regulă, criteriul ales este dat de *metoda celor mai mici pătrate* (MCMMP), adică acea funcție f pentru care se minimizează suma patratelor erorilor intre valorile masurate si cele estimate (*procedeu de fitare*), adica suma

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

Dacă f este o *funcție liniară*, atunci se obține *regresia liniară*, reprezentată grafic printr-o dreaptă (*dreapta de regresie*). Dreapta de regresie, împreună cu abaterile standard ale variabilelor X şi Y, sau cu coeficientul de corelație, pot constitui o rezumare rezonabilă a distribuției comune a celor două variabile X si Y. Adecvanța modelului liniar este mai bună atunci când diagrama de împrăştiere are formă de elipsă.

Metoda celor mai mici pătrate (MCMMP)

Dependența funcțională a unei variabile aleatoare Y (*dependentă-efect*) față de altă variabilă X (*independentă-cauză*) poate fi studiată empiric, pe cale experimentală, efectuîndu-se o serie de măsurători asupra variabilei Y pentru diferite valori ale variabilei X. Rezultatele se pot prezenta sub formă de tabel sau grafic.
Problema care apare în acest caz este de a găsi *reprezentarea analitică* a dependenței funcționale căutate (*procedeu de fitare*), adică de a alege o expresie (*formulă sau model matematic*) care să descrie rezultatele experimentului printr-un model matematic.
Formula se alege dintr-o mulțime de formule determinate, de exemplu:

- $y = ax + b$ (*dreapta*),
- $y = ax^2 + bx + c$ (*parabola*),
- $y = ae^{bx} + c$ (*exponentiala*),
- $y = a + b \sin(\omega t + \varphi)$ (*sinusoida*).

Pin urmare, problema constă în a determina *parametrii* a, b, c, etc. în timp ce formula (expresia analitică) este cunoscută dinainte, ca urmare a unor considerente teoretice sau după forma prezentării grafice a datelor, în mod empiric.

Să considerăm, cazul general când avem p parametri, si astfel vom nota dependența funcțională prin

$$y = f(x;\, a_0, a_1, \ldots, a_p)$$

Parametri $a_0, a_1,\ldots, a_p$ nu se pot determina exact pe baza valorilor empirice $y_1, y_2,\ldots,y_n$ ale funcției, deoarece acestea din urmă conțin erori aleatoare. Problema reprezintă obținerea unei estimari "*suficient de bune*".

Formularea problemei

Dacă toate măsurătorile valorilor varabilei Y sunt y_1, y_2,...,y_n, atunci estimațiile parametrilor a_0, a_1,..., a_p se determină din condiția ca suma pătratelor abaterilor valorilor măsurate y_k de la cele calculate $f(x_k;\ a_0,\ a_1,...,\ a_n)$ să ia valoarea minimă, adică sa fie minimă expresia

$$S=\sum_{k=1}^{n}[y_k - f(x_k;a_0,a_1,...a_p)]^2$$

.

Consideraţia formulată se păstrează şi în general, pentru determinarea parametrilor unei funcții de mai multe variabile (2, 3, etc.), adica o variabila dependenta (efect) si mai multe variabile independente (cauze). De exemplu, pentru variabila *Z* (efect) ce depinde de două variabile independente (cauze) *X* şi *Y*, adică *Z=f(X,Y)*, estimațiile parametrilor a_0, a_1,..., a_p se determină din condiția ca expresia

$$S = \sum_{k=1}^{n}[z_k - f(x_k, y_k; a_0, a_1,..., a_p)]^2$$

să fie minimă.

Determinarea valorilor parametrilor a_0, a_1,..., a_p, se face prin aplicarea condițiilor de obtinere a *valorii minime* in derivatele partiale ale funcției *S* considerată în variabilele a_0, a_1,..., a_p , adică funcția cu p variabile $S(a_0,\ a_1,...,\ a_p)$. Obținerea acestor valori înseamnă rezolvarea sistemului de *p* ecuații cu *p* necunoscute.

$$\frac{\partial S}{\partial a_0} = 0\,,\ \frac{\partial S}{\partial a_1} = 0\,,\ldots,\ \frac{\partial S}{\partial ap} = 0\,.$$

Dreapta de regresie

În cazul modelului liniar (cel mai simplu) se studiază numai două variabile *X* (cauza), *Y*(efect) şi se doreşte găsirea dependenței *Y = f(X)*, unde *f(x)* = ax + b este o dependenta liniara (*functie de gradul I*) cu *p=2* parametri *a* si *b*.

În urma celor *n* probe (*masuratori, observatii*) se cunosc datele $(x_i\ ,y_i)$, i=1,..., *n* şi trebuie să se determine coeficienții *a* şi *b* astfel încât suma

$$\mathrm{S} = \sum_{i=1}^{n}\left[y_i - (ax_i + b)\right]^2$$

să fie minimă. Condițiile de obținere a parametrilor a şi b sunt:

$$\begin{cases} \dfrac{\partial S}{\partial a} = 0 \\ \dfrac{\partial S}{\partial b} = 0 \end{cases}$$

, ceea ce conduce la sistemul de 2 ecuatii cu 2 necunoscute:

$$\begin{cases} 2\sum_{i=1}^{n}\left[y_i - (ax_i + b)\right](-x_i) = 0 \\ 2\sum_{i=1}^{n}\left[y_i - (ax_i + b)\right] = 0 \end{cases} \rightarrow \begin{cases} -2\sum_{i=1}^{n} x_i y_i + 2\sum_{i=1}^{n} ax_i^2 + 2\sum_{i=1}^{n} bx_i = 0 \\ 2\sum_{i=1}^{n} y_i - 2\sum_{i=1}^{n} ax_i - 2\sum_{i=1}^{n} b = 0 \end{cases}$$

Se notează: $\sum_{i=1}^{n} x_i y_i = S_{xy}$ $\sum_{i=1}^{n} x_i^2 = S_{xx}$ $\sum_{i=1}^{n} x_i = S_x$ $\sum_{i=1}^{n} y_i = S_y$ si sistemul de ecuații devine:

$$\begin{cases} S_{xy} - aS_{xx} - bS_x = 0 \\ S_y - aS_x - nb = 0 \end{cases}$$

. Se obțin urmatoarele expresii pentru cei doi parametri a si b:

$$a = \frac{S_x S_y - nS_{xy}}{(S_x)^2 - nS_{xx}} \text{ şi } b = \frac{1}{n}\left(S_y - aS_x\right)$$

Cei doi parametri ai funcției model *f(x)* = ax + b reprezintă:

- *a* - *panta dreptei de regresie*, adică *a=tg(α)*, unde α este unghiul dintre graficul funcției *f* si axa OX (absciselor);
- *b* - valoarea pe axa OX unde graficul funcției *f* intersectează axa OY (ordonatelor).

Trebuie să facem observația că indiferent de gradul de împrăştiere al punctelor, întotdeauna se poate găsi o *dreaptă de regresie*, dar în cazul unei dispersii mari aceasta devine inutilă. De aceea un studiu preliminar al distribuției punctelor (norul de puncte) se impune cu necesitate.
Calitatea unei drepte de regresie poate fi analizată după *coeficientul de determinare* R^2 (*R-squared value on chart, pătratul coeficientului de corelație multiplă*) ce are valori in intervalul [0,1] si se calculează cu relația:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} [y_i - f(x_i)]^2}{\sum_{i=1}^{n} [E(f(x)) - f(x_i)]^2}, \text{ unde } E(f(x)) = \frac{1}{n}\sum_{i=1}^{n} f(x_i).$$

O valoare 1 pentru acest coeficient are semnificația că funcția model f explică întreaga variabilitate (dependentă) a lui y, iar valoarea 0 că nu există nici o relație liniară între variabila Y şi variabila X. O valoare de 0.5 a lui R^2 poate fi interpretată în sensul că aproximativ 50% din variația variabilei Y poate fi determinata de către variabila independentă X.

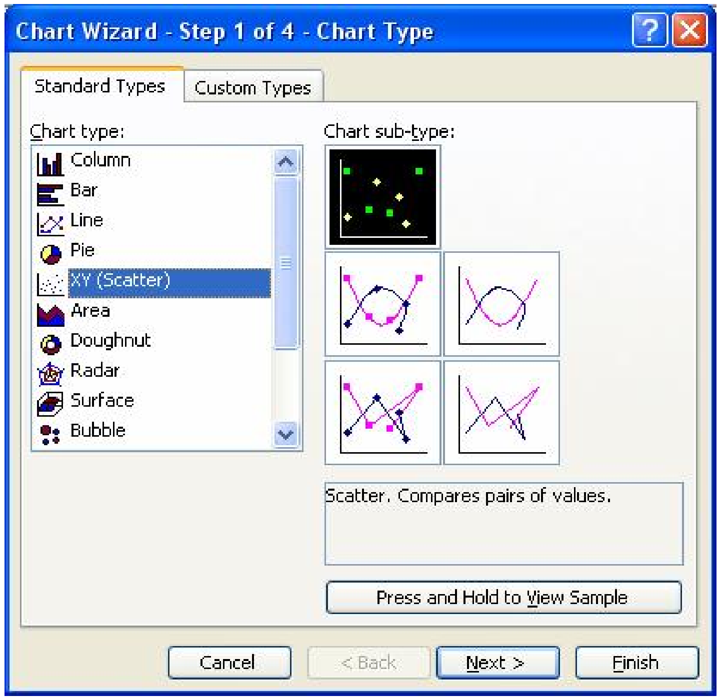
EXEMPLE

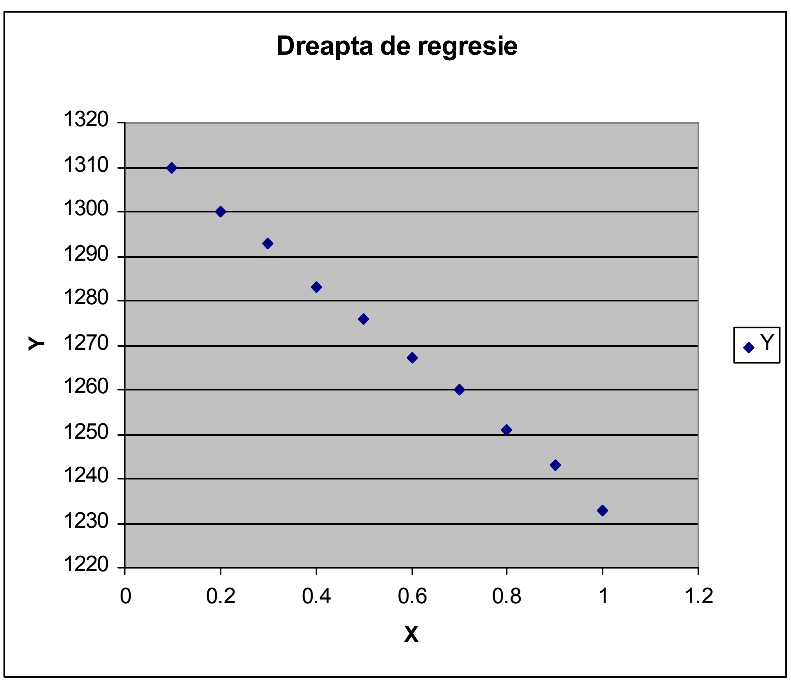
Exemplul 1.

Folosind programul Excel sa se determine *drepta de regresie* pentru doua variabile X si Y (de exemplu, in cadrul unui proces electric: *variabila intensitate* I(mA) si *variabila Tensiune* U(mV) ce depinde deaceasta) si sa se obtina *calitatea aproximarii* (fitarii) prin calculul *coeficientul de determinare* R^2.

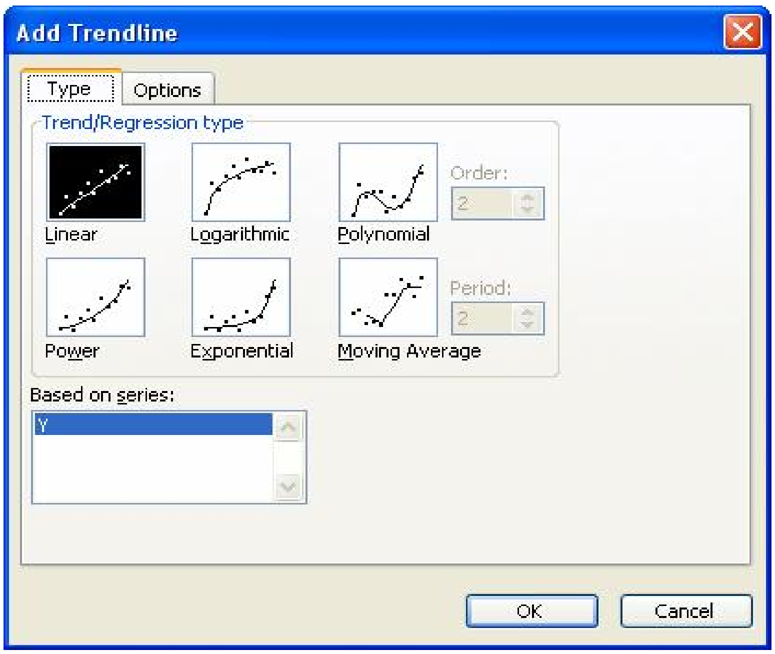
Intr-o foaie de calcul Excel presupunem ca apar valorile masurate pentru variabilele X si Y. Pentru obtinerea dreptei de regresie si a coeficientului de determinare R^2 , trebuie sa se parcurga urmatorii pasi:

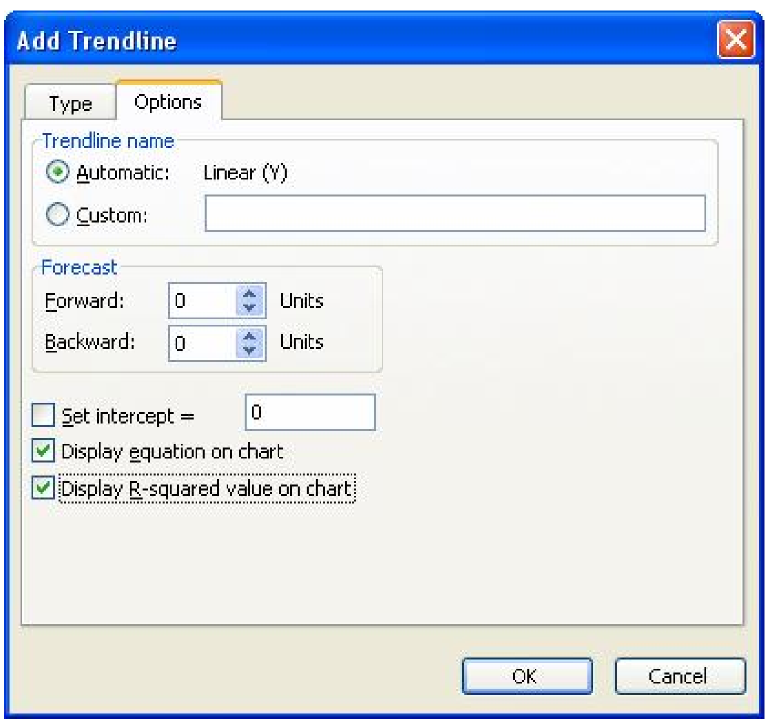
Pasul 1. Reprezentarea norului de puncte (diagrama de imprastiere) pentru variabilele X si Y. Pentru acest lucru trebuie sa se selecteze valorile aflate in cele 2 coloane ale celor 2 variabile, se actioneaza **Insert → Chart** si se alege tipul de grafic **XY (Scatter)** (*Standard Types*), de unde din cele 5 variante de grafice se opteaza pentru prima varianta (*Scatter-Compares pairs of values*); se parcurg etapele pentru a genera graficul respectiv, si care apare mai jos;





Pasul 2. Determinarea si reprezentarea dreptei de regresie. Se selecteaza graficul obtinut la pasul 1 (norul de puncte) si se actioneaza **Chart → Add Trendline** de unde se alege tipul **Linear** (*Standard Types*), Eticheta **Add Trendline Options** este prezentată în figura următoare şi permite definirea altor atribute ale liniei de trend*:* **Display equation on chart** – marcarea boxei de control are efectul trecerii pe grafic a ecuației estimate, **Display R-squared value on chart** – este utilă pentru afişarea coeficientului de determinare R^2 (pătratul coeficientului de corelație multiplă).





In urma aparitiei graficului ce reprezinta dreapta de regresie, se obtin urmatoarele rezultate:

$y = f(x) = -83.636x + 1317.6$, $\boldsymbol{a = -83.636}$, $\boldsymbol{b = 1317.6}$ si $\mathbf{R^2 = 0.999}$.

X	Y
0.1	1310
0.2	1300
0.3	1293
0.4	1283
0.5	1276
0.6	1267
0.7	1260
0.8	1251
0.9	1243
1	1233

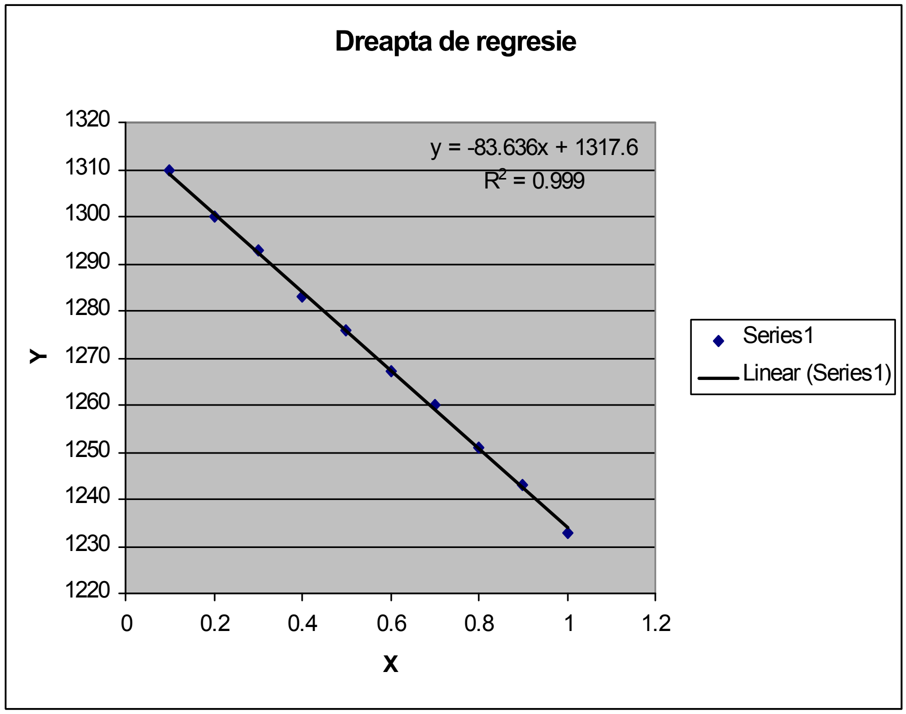


Figura 32. Setul de valori si dreapta de regresie (modelul liniar)

Trebuie sa precizam ca programul Excel ofera prin **Trandine** mai multe tipuri de regresii (modele liniare si neliniare):

- *Linear* – modelul liniar (regresia simplă), $y = a + bx$.
- *Polynomial* – modelul polinomial de ordin 2, 3, 4, 5, sau 6, $y = a_0 + a_1x + a_2x^2 + \ldots + a_kx^k$.
- *Logarithmic* – modelul logaritmic, $y = a + b \ln x$.
- *Exponential* – modelul exponenţial, $y = ae^{bx}$
- *Power* – modelul putere, $y = a x^b$.
- *Moving Average* – modelul de tip MA (medii glisante), în care se calculează o serie nouă cu valori obţinute ca medie aritmetică a valorilor din seria iniţială: $y_n = (x_n + x_{n-1} + \ldots + x_{n-k+1})/k$, unde k este ordinul modelului. Este modelul prin care se elimină influenţele pe termen foarte scurt sau scurt. Pentru o alegere corectă se poate utiliza informaţia cunoscută din cercetări anterioare sau cea furnizată vizual de aspectul norului de puncte.

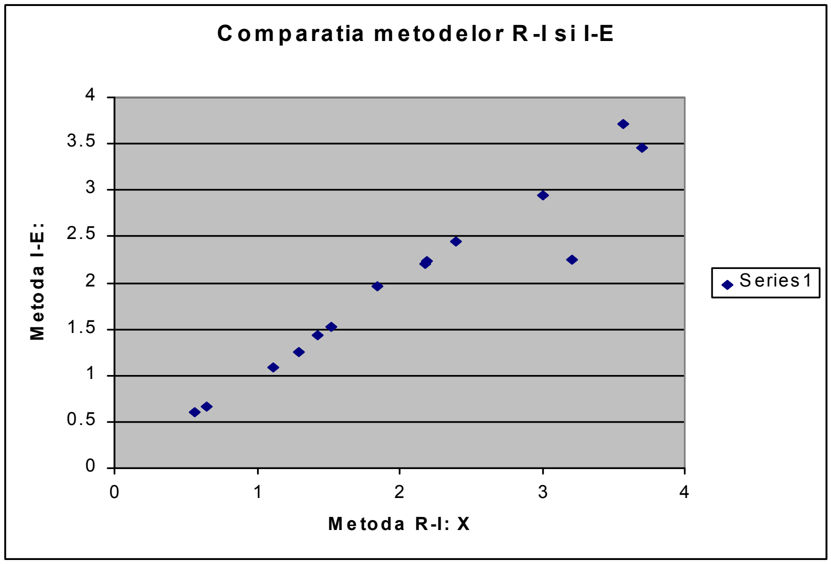
Exemplul 2.

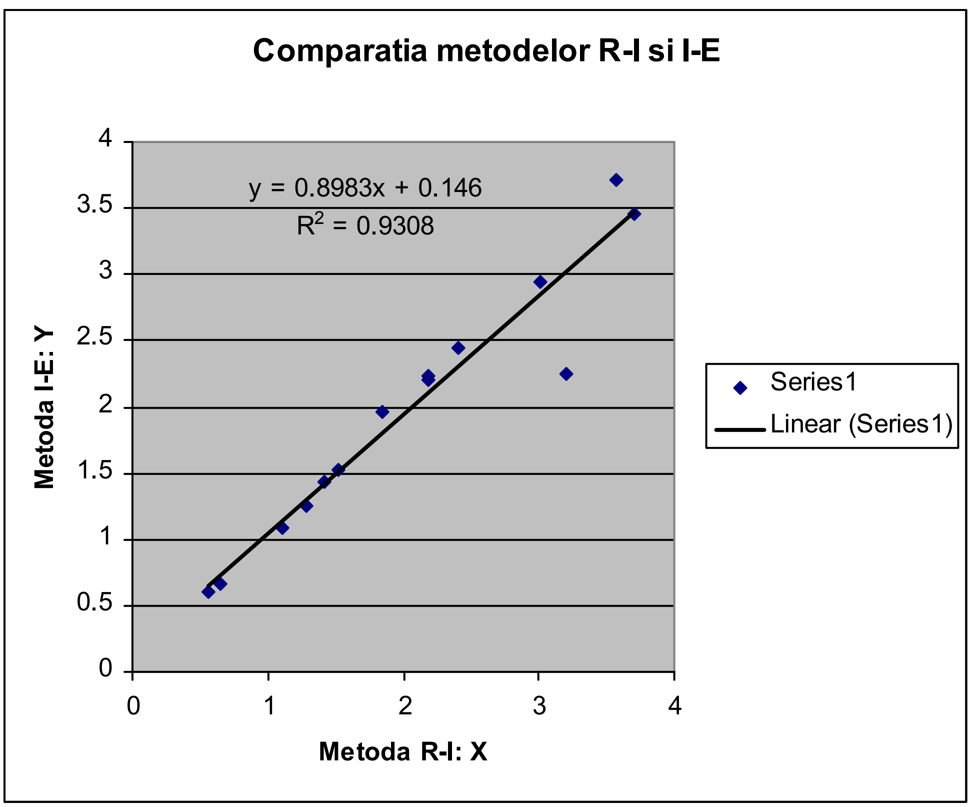
Pentru dozarea unui antibiotic într-un lichid biologic se propun două metode: o metodă *radio-imunologică* (R-I) şi o metodă *imuno-enzimatică* (I-E). Se se realizează testarea comparativă a celor două metode. Datele pentru cele doua metode sunt prezentate în tabelul de mai jos. Coeficientul de corelație intre vectorii R-I (X) şi I-E (Y). Dreapta de regresie şi coeficientul de determinare.

Coeficientul de corelație se obtine apeland functia Excel **CORREL (X,Y) = 0.964795**. In urma aparitiei graficului ce reprezinta dreapta de regresie, se obtin urmatoarele rezultate:

$y = f(x) = 0.8983\,x + 0.146$, $\boldsymbol{a} = \mathbf{0.8983}$, $\boldsymbol{b} = \mathbf{0.146}$ si $\mathbf{R^2 = 0.9308}$.

X	Y
0.56	0.60
0.65	0.67
1.11	1.08
1.29	1.25
1.42	1.44
1.52	1.53
1.84	1.96
2.18	2.21
2.19	2.23
2.40	2.44
3.01	2.95
3.21	2.25
3.57	3.71
3.70	3.46

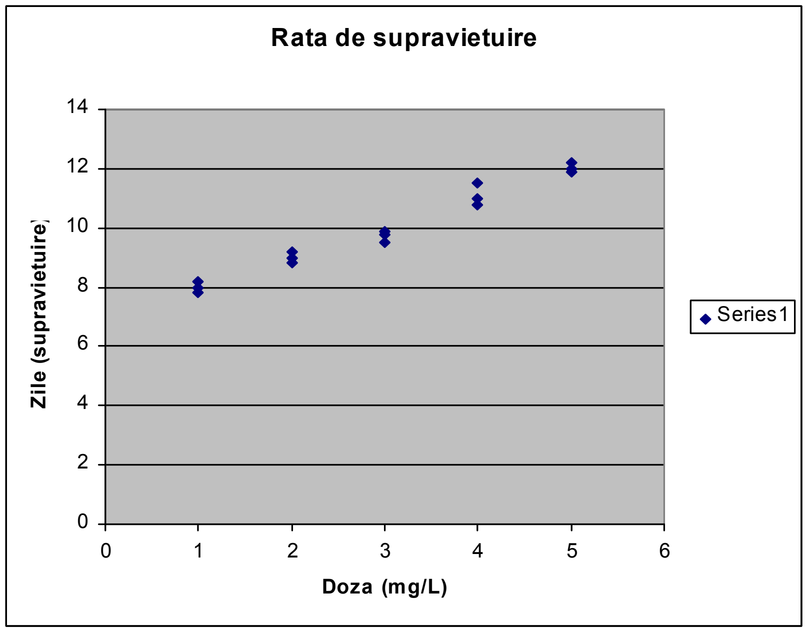


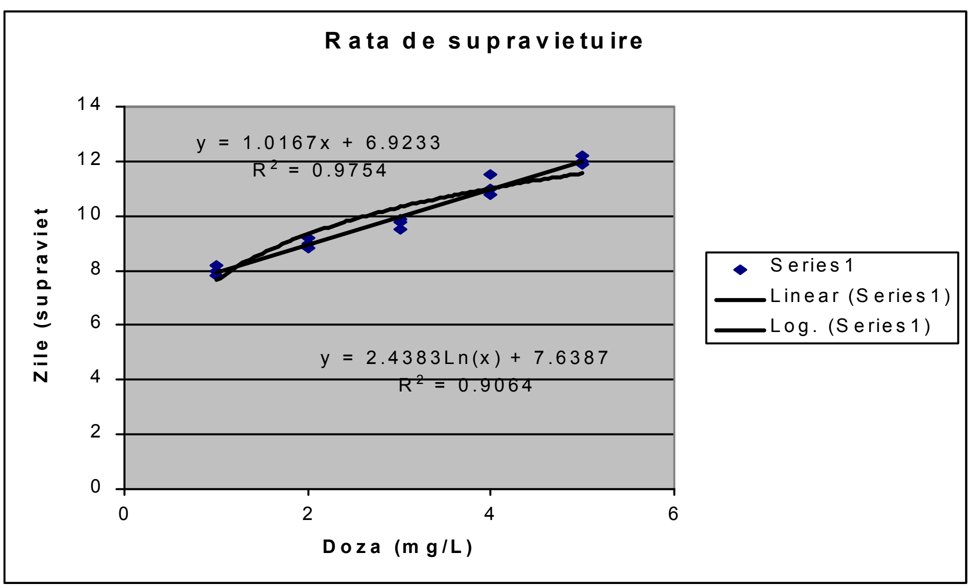


Exemplul 3.
Pentru studierea efectului unei anumite *substanțe medicamentoase* se injectează aleator cu diferite doze 15 şoareci. Se urmăreşte *numărul de zile de supravieţuire* la soareci. Analizând datele, se poate trage concluzia că rata de supravieţuire creşte liniar în funcţie de doza injectată? Sa se studieze reprezentarea norului de puncte si sa se compare *modelul liniar* si *modelul exponential*.

Rezolvare.

Doza(X)	Zile(Y)
1	8
1	7.8
1	8.2
2	8.8
2	9
2	9.2
3	9.8
3	9.5
3	9.9
4	11
4	10.8
4	11.5
5	12
5	12.2
5	11.9





In cazul modelului linear (dreapata de regresie) se obtin: $y = 1.0167\,x + 6.9233$, si $\mathbf{R^2} =$ **0.9308**, iar in cazul neliniar (logaritmic) avem $y = 2.4383\ \mathrm{Ln}(x) + 7.6387$, si $\mathbf{R^2} =$

0.9064. In concluzie, deoarece cazul liniar (dreapta de regresie) ofera $\mathbf{R^2 = 0.9308}$**,** coeficientul de determinare mai mare, acesta este mai bun in aproximare.

Bibliografie

1. Lucian Boiculese - Biostatistica teme , Scoala doctorala, UMF Iasi
2. Lucia Căbulea, Nicoleta Breaz, Interpretarea statistică a informațiilor. Elememnte de data mining şi prognoză, Modul de instruire nr.7, Universitatea "1 decembrie 1918" Alba Iulia,
3. Valentin Clocotici, Dicționar explicativ de statistică (online), Universitatea "A.I.Cuza" Iasi, http://profs.info.uaic.ro/~val/statistica/StatGloss.htm
4. Alexandra Fotis, http://anale-informatica.tibiscus.ro/download/lucrari/1-2-13-Fortis.pdf
5. Leonard Mihaly Cozmuta, Prelucrarea datelor experimentale, http://chimie-biologie.ubm.ro/Cursuri%20on-line/MIHALY%20COZMUTA%20LEONARD/Prelucrarea%20datelor%20experimentale.pdf
6. Mihai Miron, Liliana Miron, Masurari electrice si electronice, Brasov. 2003, http://www.afahc.ro/invatamant/electro/mee.pdf
7. http://ebooks.unibuc.ro/Fizica/Sabina/2.pdf
8. http://www.phys.ubbcluj.ro/~daniel.andreica/pdf/Mec-CURS/01-INTRODUCERE-IN-CALCULUL-ERORILOR.pdf
9. http://www.phys.ubbcluj.ro/~dana.maniu/Laborator%20Optica/Calcul%20de%20erori.pdf
10. http://www.utgjiu.ro/math/mbuneci/book/mn2007/c04.pdf
11. http://cs.upm.ro/~bela.finta/.files/carti/Numerika.pdf
12. Peter J. Steinbach, Two-Dimensional Distributions of Activation Enthalpy and Entropy from Kinetics by the Maximum Entropy Method, Biophysical Journal Volume 70 March 1996, pp. 1521-1528, http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1225079/pdf/biophysj00049-0431.pdf
13. Joseph W. Ochterski, Thermochemistry in Gaussian, http://www.gaussian.com/g_whitepap/thermo.htm
14. Marjorie Olmstead, http://courses.washington.edu/phys431/index.php
15. Kai Oliver Arras, An Introduction To Error Propagation, 1998, http://www.nada.kth.se/~kai-a/papers/arrasTR-9801-R3.pdf
16. Peter Keusch, University of Regensburg, http://www.demochem.de/eyr-e.htm
17. David W. A. Bourne, Pharmacokinetics and Biopharmaceutics, (Java Applets - On line Graphs, JavaScript Calculators Online), http://www.boomer.org/c/p1/
18. David W. A. Bourne, Mathematical modeling of pharmacokinetic data, Technomic Publishing Co., ISBN 1-56676-204-9, 1995
19. M. Vlada, pagina principala, http://www.unibuc.ro/prof/vlada_m/
20. M. Vlada, Birotica.Tehnologii multimedia, Editura Universitatii din Bucuresti, 2004
21. M. Vlada, on-line – http://ebooks.unibuc.ro/informatica/Birotica/
22. M. Vlada, on-line – http://ebooks.unibuc.ro/informatica/eureka/