

Manual

Prezentarea și prelucrarea statistică a datelor experimentale

**Prof. Dr. Lorentz JÄNTSCHI
Prof. Dr. Ing. Mugur C. BĂLAN**

**AT32/137070/S/1.5/159/POSDRU
Cluj-Napoca, 2014**

FIȘA ATELIERULUI TEMATIC

Universitatea Tehnică din Cluj-Napoca

TEMA ATELIERULUI:

Prezentarea și prelucrarea statistică a datelor experimentale

Numele și prenumele *ETS AT*, inclusiv gradul didactic:

1. *ETS AT 1*: Prof. Dr. Lorentz JĂNTSCHI (responsabil atelier)
2. *ETS AT 2*: Prof. Dr. Mugur C. BĂLAN

Facultate/Departament/Institut de care aparțin:

1. *ETS AT 1*: Fac. IMM / Dep. Fizică & Chimie / Laborator Analiză Instrumentală (responsabil atelier)
2. *ETS AT 2*: Fac. MEC / Dep. Inginerie Mecanică / Laborator Analiză Instrumentală

Locul de desfășurare a atelierului:

Laboratorul de Analiză Instrumentală (acreditat intern UTCN),
B.dul Muncii 103-105 săli C307, C308, C500, C501, C502 tel.
0264401775 & 0264401605, http://lori.academicdirect.org/admin/ia_lorentz.jantschi@gmail.com,
balan.mugur@gmail.com

Încadrarea temei atelierului în domeniile prioritare ale proiectului: Mediu

Cunoștințele pe care le poate dobândi participantul la atelier și aplicabilitatea acestora:

managementul activităților de desfășurare a experimentelor și de colectare a datelor; elementele de bază necesare pentru efectuarea unei analize statistice descriptive și inferențiale; elemente informatice de bază pentru căutări pe Internet și pentru redactarea documentelor; principii de bază în alcătuirea prezentărilor.

Deprinderile pe care le poate dobândi participantul la atelier ca urmare a activităților practice prevăzute:

planificarea colectării și înregistrării datelor; utilizarea programelor de birotică; utilizarea programelor de analiză statistică; alegerea și interpretarea modelelor statistice.

Atelierul tematic contribuie la formarea sau consolidarea următoarelor competențe:

utilizarea programelor informatice pentru ușurarea activităților de cercetare și transfer tehnologic

Aparatura specializată care se folosește în cadrul atelierului:

Nr.crt.	Denumirea echipamentului* (aparat, stand sau instalație)	Firma producătoare	Anul fabricației	Caracteristici principale	Starea actuală a echipamentului
1	Stații meteo	Vantage	2008	achiziție online a 40 de parametrii de mediu	Funcțional (2 buc) Nefuncțional (1 buc)
2	Calculatoare	IBM & HP	2008	dual/quad/eight core	Funcționale (8 buc)
3	Licențe software	Adobe, Statsoft, SPSS Spartan	2008	utilizabile în cadrul laboratorului	Funcționale
4	Biospectrofotometru	PortaLIBS	2008	analiză elementală semicantitativă	Funcțional
5	Fluorometru	Sepadin	2008	analize chimice în soluție	Funcțional
6	Servere	VIA Embedded	2012	gestiune baze de date	Funcționale (3 buc)

Programul AT (4 zile a câte 8 ore):

Interval orar	Conținutul activității	ETS AT care efectuează activitatea**
ZIUA 1: 22 Septembrie 2014		
8-8.50	Înregistrarea participanților, prezentarea tematicii atelierului, înmânarea documentației de lucru	ETS AT 1 (1 oră) ETS AT 2 (1 oră)
9-10.50	Prelegere tema 1 Înregistrarea tabelată a datelor (1 oră) Redactarea documentelor și prezentărilor (1 oră)	ETS AT 2 (2 ore)
11-12	Vizitarea laboratorului în care se desfășoară activitățile AT	ETS AT 2 (2 ore)
16-20	Lucrarea experimentală 1: Utilizarea programului de calcul tabelar Excel pentru înregistrarea datelor și exemple (1 oră) Utilizarea programului de redactare Word și exemple (2 ore) Utilizarea programului de prezentare PowerPoint și exemple (2 ore) Utilizarea programului de calcul tabelar Excel pentru analiza preliminară a datelor (1 oră)	ETS AT 2 (4 ore)
ZIUA 2: 23 Septembrie 2014		
8-9.50	Prelegere tema 2 Principii în colectarea și înregistrarea datelor (1 oră) Principii în analiza statistică a datelor (1 oră)	ETS AT 1 (2 ore)
10 -12	Prezentarea facilităților de cercetare ale Laboratorului de Analiză Instrumentală și exemple de utilizare ale acestora (2 ore)	ETS AT 2 (2 ore)
16-20	Lucrarea experimentală 2: Prelucrarea datelor folosind Excel (4 ore)	ETS AT 1 (4 ore)
ZIUA 3: 24 Septembrie 2014		
8-9.50	Prelegere tema 3 Modele de analiză statistică descriptivă a datelor (1 oră) Modele de analiză statistică inferențială a datelor (1 oră)	ETS AT 1 (2 ore)
10 -12	Utilizarea sistemelor de achiziție a datelor de mediu (2 ore)	ETS AT 2 (2 ore)
16-20	Lucrarea experimentală 3: Prelucrarea datelor folosind programe specializate de statistică (4 ore)	ETS AT 1 (4 ore)
ZIUA 4: 25 Septembrie 2014		
8-9.50	Prelegere tema 4 Principii de urmat în interpretarea statistică a datelor (1 oră) Principii de urmat în prezentarea rezultatelor cercetării (1 oră)	ETS AT 1 (2 ore)
10 -12	Interogarea sistemelor de monitorizare a mediului (2 ore)	ETS AT 2 (2 ore)
16-17.50	Elaborarea raportului colectiv al doctoranzilor, discuții finale	ETS AT 1 (1 oră) ETS AT 2 (1 oră)

Prefață

Activitatea de cercetare, prin natura sa, presupune dobândirea unor temeinice cunoștințe și abilități de proiectare a experimentelor pornind de la intrările (observabilele) și ieșirile (informațiile) pentru care experimentul este menit, de colectare planificată și organizată a rezultatelor derulării experimentului în șirul de repetiții al acestuia și, nu în ultimul rând, de prelucrare statistică și interpretare fenomenologică a întregului volum de date furnizat de experiment. În aceste activități de proiectare, colectare și prelucrare a datelor experimentale cercetătorul trebuie să implice tehnica de calcul și trebuie să facă apel la cunoștințele sale de a opera corect și eficient cu programele informatice care îi facilitează desfășurarea activităților și obținerea rezultatelor.

În acest sens, atelierul tematic 'Prezentarea și prelucrarea statistică a datelor experimentale' (adresat doctoranzilor grupului țintă al proiectului PARTING) își propune să ofere prin intermediul activităților de tip prelegere cunoștințele iar prin intermediul activităților experimentale abilitățile de a opera cu programe specializate ce servesc în colectarea și prelucrarea datelor, și care în același timp pentru a fi utilizate nu necesită cunoștințe sau abilități care să îl îndepărteze pe cercetător de domeniul său propriu de cercetare.

În prima parte a atelierului tematic (în prima zi de activități) se face o prezentare a facilităților de bază pe care le oferă pachetul Office de la Microsoft în ceea ce privește înregistrarea sistematică și prelucrarea primară a datelor (programul MS Excel), colectarea și structurarea documentației (partea introductivă, descrierea materialului și metodei cercetării, realizarea schemelor de lucru) aferente cercetării (programul MS Word), organizarea și validarea conținutului științific (motoare de căutare în Internet), realizarea prezentărilor rezultatelor (programul MS PowerPoint) și a versiunilor portabile ale documentelor (programele Adobe). Prezentarea facilităților de bază ale operării cu aceste programe este urmată de transferul de cunoștințe și abilități de utilizare eficientă a acestora, punând accent pe acele elemente proprii acestor programe care permit utilizatorului să economisească timp de lucru și respectiv să obțină rezultate de calitate de pe urma utilizării lor.

A doua parte a atelierului tematic (a doua zi de activități) este dedicată fixării bunelor practici în colectarea, înregistrarea și analiza datelor, oferind în acest sens facilitățile Laboratorului de Analiză Instrumentală și accesul la datele colectate în cadrul acestui laborator, participanții având prilejul de a-și fixa abilitățile în cadrul laboratorului prin colectare și prelucrare de date.

A treia parte a atelierului tematic (în a 3-a zi de activități) este dedicată transferului de cunoștințe și abilități în ceea ce privește analiza statistică descriptivă și inferențială, participanții având prilejul să-și fixeze aceste abilități folosind datele colectate cu ajutorul sistemelor de achiziție a datelor de mediu existente în cadrul Laboratorului de Analiză Instrumentală. Pe lângă operarea în MS Excel, participanții vor avea prilejul să efectueze prelucrări grafice și statistice în încă alte trei programe specializate: DataPlot (NIST-USA) - reprezentări și prelucrări preliminare ale datelor, SlideWrite (TU Eindhoven) - obținerea și analiza modelelor liniare și neliniare, Statistica (Statsoft) - program specializat de analiză statistică.

Ultima parte a atelierului tematic (în a 4-a zi de activități) este dedicată transferului de cunoștințe și abilități în ceea ce privește extragerea semnificației statistice a modelelor și a parametrilor acestora din analiza de model desfășurată cu ajutorul programelor de analiză statistică, și pe baza acestor semnificații statistice construirea interpretărilor fenomenologice care însoțesc analiza datelor experimentale.

Atelierul tematic este planificat a se desfășura pe baza discuțiilor libere conduse de coordonatori folosind date experimentale obținute în cadrul Laboratorului de Analiză Instrumentală și respectiv la solicitarea participanților folosind date furnizate de aceștia. Este vizată în cadrul laboratorului parcurgerea tuturor etapelor de tratare a datelor experimentale, de la colectare, la prelucrare preliminară, prezentare preliminară, analiză statistică, interpretare statistică, până la prezentarea rezultatelor și concluziilor ce decurg.

Cuprins

Mărimi și măsurarea lor

Analiza dimensională

Probabilități și statistică

Prelucrarea textului și a imaginilor

Calcul tabelar

Analiza statistică a efectelor multiplicative

Proiectarea experimentelor factoriale și interpretarea statistică

Resurse online pentru corelații în contingente ordonate și respectiv corelate

Mărimi și măsurarea lor

O **mărime** este rezultatul unei măsurători efectuate asupra unei observabile cu scopul de a colecta valoarea unei proprietăți [1]. Se poate imagina **spațiul de observare** ca având o structură de arbore (a se vedea *Structura spațiului de observare*) care exprimă relațiile de apartenență dintre observabile în care la bază se află Universul (ca întreg spațiul de observare) iar la suprafață (aproape de noi în calitate de observatori) se află compoziții chimice - ca formă de reprezentare a materiei cu compoziție (de atomi) și relații (între aceștia) bine definite.

Structură	Proprietate
- Univers	Întreg spațiul de observare
- Energie Radiantă	Viteza comparabilă cu cea a luminii
+ Radiații ca β , γ	Diferențiate prin intermediul proprietăților
- Materie	Întreg spațiul de observare nerelativistic
- Corp	Viteza mult mai mică decât a luminii
- Ansamblu de materiale	Compoziție (chimică) variabilă și discontinuă
- Materiale	Compoziție (chimică) variabilă și continuă
- Amestec de substanțe	Compoziție (chimică) bine definită
+ Substanța eterogenă	Compoziție (chimică) variabilă
- Soluție	Stare de agregare bine definită
+ Aliaj	Amestec de metale în stare lichidă sau solidă
- Substanța omogenă	Compoziție (chimică) constantă
+ Compus chimic	Structură chimică bine definită și unică

Structura spațiului de observare

Procesul de observare este o activitate de colectare a cunoștințelor cu ajutorul simțurilor sau instrumentelor. Se presupune existența unui observator și a unei observabile. Procesul de observare transferă o formă abstractă a cunoașterii de la observabilă la observator (ca de exemplu, sub formă de numere sau imagini).

Măsurarea cuprinde două operațiuni serializate: observarea și înregistrarea rezultatelor observației. Măsurarea depinde de natura obiectului observat (material) sau fenomene (imateriale), de metoda de măsurare și de modul de înregistrare a rezultatelor observării. Măsurarea presupune identificarea anterioară a elementului sau a elementelor care fac obiectul investigației și rezultatul măsurării este o proprietate a elementului observat. O serie de măsurători presupune existența unei colecții de elemente distincte - **mulțime** - în care ordinea poate să nu fie relevantă. Mulțimea vidă (\emptyset) este mulțimea cu nici un element în ea.

Proprietatea (ca urmare a unei serii de observații) înregistrată cu exact una din exact două valori numite nefavorabil (și scris ca F sau 0) și favorabil (și scrise ca T sau 1), respectiv, dă o **valoare de adevăr**. Mulțimea de valori de adevăr ($\{0,1\}$ sau $\{T, F\}$) este o mulțime în care elementele sunt ordonate în mod convențional ($0 < 1$, $F < T$). Negația logică (!) este operațiunea (informațională) ce schimbă valoarea de adevăr, în timp ce identitatea logică (\equiv) lasă valoarea de adevăr neschimbată și se exprimă faptul că rezultatul unei operații de măsurare pe două elemente este același. Folosind proprietatea 'valoare de adevăr' pe elementele unei mulțimi conceptul de **submulțime** este raționalizat. Apartenența este o proprietate a unui element de a fi (\in) sau a nu fi (\notin) într-o mulțime.

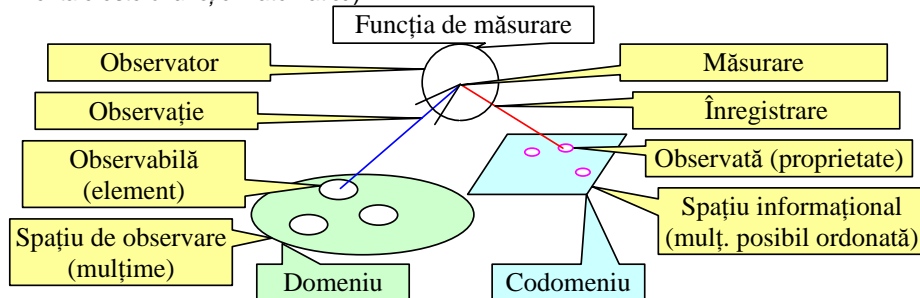
Asocierea totală scrisă ca $S1 \times S2$ și definită prin $S1 \times S2 = \{(e1, e2) \mid E1 \in S1, S2 \in e2\}$ este produsul cartezian al mulțimilor $S1$ și $S2$, iar o submulțime a $S1 \times S2$ este numită relație binară. Dacă $S1=S2$ relațiile sunt numite endo-relații. Câteva proprietăți speciale ale (endo)relațiilor (binare) și exemple de (endo)relații (binare) sunt:

- ÷ Reflexive (RE): $\forall a, (a,a) \in RE$; exemple: $=, \subseteq, |, \leq$ (vezi mai jos);
- ÷ Co-reflexive (CR): $\forall a,b, (a,b) \in CR$ then $a \equiv b$; exemple: $=$ (vezi mai jos);
- ÷ Cvasi-reflexive (QR): $(a,b) \in QR$ atunci $(a,a), (b,b) \in QR$; exemple: \lim (vezi mai jos);
- ÷ Reflexive (IR): $\forall a, (a,a) \notin IR$; exemple: $\neq, \perp, <$ (vezi mai jos);

¹ Lorentz JĂNTSCHI, 2013. Prezentarea și procesarea datelor experimentale. Cluj-Napoca: UTPress.

- ÷ Simetrice (SY): $(a,b) \in SY$ atunci $(b,a) \in SY$; exemple: =, CD, CM (vezi mai jos);
- ÷ Anti-simetrice (NS): $(a,b), (b,a) \in NS$ atunci $a \neq b$; exemple: \leq (vezi mai jos);
- ÷ Asimetrice (AS): $(a,b) \in AS$ atunci $(b,a) \notin AS$; exemple: IH, $<$ (vezi mai jos);
- ÷ Transitive (TS): $(a,b), (b,c) \in TS$ atunci $(a,c) \in TS$; exemple: =, \leq , $<$, \subseteq , $|$, \Rightarrow , IH (vezi mai jos);
- ÷ Totale (TL): $\forall a,b: (a,b) \in TL$ sau $(b,a) \in TL$; exemple: \leq (vezi mai jos);
- ÷ Tri-hotome (TC): exact una dintre $(a,b) \in TL, (b,a) \in TL, a \neq b$; exemple: $<$ (vezi mai jos);
- ÷ Euclidiene (ED): $(a,b), (a,c) \in ED$ atunci $(b,c) \in ED$; exemple: = (vezi mai jos);
- ÷ Seriale (SE): $\exists b: (a,b) \in SE$; exemple: \leq (vezi mai jos);
- ÷ Unicitate (UQ): $(a,b), (a,c) \in UQ$ atunci $b=c$; exemple: $f(\cdot)$ (vezi mai jos);
- ÷ Echivalență (EQ): atunci RE, SY, TS; exemple: =, \sim , \equiv , CM, CD, $||$ (vezi mai jos);
- ÷ Ordine parțială (PO): atunci RE, NS, TS; exemple: $|$ (vezi mai jos);
- ÷ Ordine totală (TO): atunci PO, TL; exemple: Alfabet, \leq (vezi mai jos);
- ÷ Bine ordonate (WO): atunci TO, SE;
- ÷ Co-prime (\perp): cel mai mare divizor comun este 1;
- ÷ Adevărul vacuos (VT): 'daca A atunci B' când A = Fals;
- ÷ Egal (=): atunci RE, CR, SY, NS, TS, ED, EQ;
- ÷ Mai mic sau egal (\leq): atunci RE, NS, TS, TL, SE, PO, TO;
- ÷ Mai mic ($<$): atunci IR, NS, AS, TS, TC, SE;
- ÷ Submulțime (\subseteq): RE, NS, TS, SE, PO;
- ÷ Diferit (\neq): IR, SI;
- ÷ Distanță euclidiană (DI): RE, SI, TS, ED, SE, EQ;
- ÷ Moștenire (IH): AS, TS;
- ÷ Congruență modulo n (CM): EQ;
- ÷ Congruență div n (CD): EQ;
- ÷ Limita unei serii (lim): RE, QR;
- ÷ Funcție matematică ($f(\cdot)$): SE, UQ;
- ÷ Funcție injectivă (inj): $a \neq b$ atunci $f(a) \neq f(b)$;
- ÷ Funcție surjectivă (srj): $\exists x: b=f(a)$;
- ÷ Funcție bijectivă (bij): INJ, SRJ

Similaritatea între conceptul de **funcție matematică** și **funcția de măsurare** este evidentă când analizăm proprietățile relațiilor care se stabilesc între mulțimea observabilelor și mulțimea valorilor asociate din spațiul informațional. Ca și în cazul funcțiilor matematice, atunci când sunt efectuate măsurători experimentale sunt asigurate două proprietăți între elemente observate și proprietățile lor înregistrate. Și anume, pentru toate elementele observate avem înregistrări ale proprietăților lor atunci când facem măsurători - fiind asigurată *serializarea* (SE). O măsură ne oferă (într-un anumit moment de timp și spațiu), o piesă informațională (o înregistrare) și unicitatea (UQ) fiind asigurată de asemenea. Nici o altă proprietate cunoscută (matematică) a relațiilor nu este, în general, valabil pentru funcții matematice și nici pentru funcția de măsurare, așa încât putem spune că ceea ce funcția de măsurare face prin intermediul informațiilor este expresia unei funcții matematice (vezi *Colectarea datelor experimentale este o funcție matematică*).



Colectarea datelor experimentale este o funcție matematică

Pentru o mulțime finită S o **funcție de numerotare** poate fi definită iterativ după cum urmează: $S_0=S$; $S_1=S\setminus\{s_1\}$; ...; $S_i=S\setminus\{s_i\}$; ...; (etc.). Funcția $f(i)=s_i$ este o funcție de numerotare pe mulțimea S , și arată că orice mulțime finită este numărabilă. Alegerea elementelor s_1, \dots, s_i, \dots din mulțimea S este instrumentul specific de măsurare. Este implicită observației, înregistrării, și are ca efect construirea unei submulțimi reunind elementele rămase. S-a arătat mai sus că conceptul de funcție matematică este legat de conceptul de măsurare. Mai departe, funcția de numerotare este instrumentul specific, cu care se face ordonarea în spațiul informațional. Mai mult, în cazul în care mulțimea S conține n elemente (desigur, ar trebui numărate mai întâi), atunci există exact $n!$ posibilități de a enumera elementele sale prin intermediul funcției de numerotare. În afară de numerotarea implicită, funcția de măsurare aduce în spațiul informațional valoarea unei proprietăți observate.

Pentru două (presupus) finite mulțimi A (spațiul nostru de observare) și B (spațiul nostru informațional) sunt exact $|B|^{|A|}$ posibilități de a defini (construi) funcții matematice $f:A\rightarrow B$ (să ne amintim, posibilități de măsurare) care asociază elementele din A cu elementele din B . Pentru o observație cu 0 și 1 ($|B|=2$) asupra unei mulțimi cu n elemente ($|A|=n$) avem un rezultat al numărării ($|A|=n$), un rezultat al posibilităților de enumerare ($|A|=n!$) și un rezultat al posibilităților de observare ($|B|^{|A|}=2^n$). Se poate verifica imediat că $n < n!$. Pentru $n > 3$ și mai mult, $n < 2^n < n!$ pentru $n > 4$. Chiar mai mult decât atât, pentru $n \rightarrow \infty$ $n < 2^n < n!$, adică $\lim_{n \rightarrow \infty} (n/2^n) = \lim_{n \rightarrow \infty} (2^n/n!) = 0$. Dacă o observație cu 0 și 1 este 'cel mai simplu tip de observație' atunci o observație ce înregistrează în spațiul informațional numere reale este cel mai complex tip de observație.

Presupunând că rezultatul observației este un număr real, putem folosi o pereche formată dintr-un bit (0 sau 1) consemnând semnul și un număr real pozitiv pentru a echivala conținutul din spațiul informațional (numărul real cu semn). Mai mult, se poate construi o funcție matematică bijectivă (care aduce o corespondență 1:1) între orice număr real pozitiv $[0, \infty)$ și un număr real din intervalul $[0, 1)$: $f:[0, 1) \rightarrow [0, \infty)$, $f(x) = 1 + 1/(1-x)$. Verificarea că este o funcție bijectivă pentru domeniul de definiție se poate face verificând că $f'(x) = 1/(x-1)^2 > 0$. Rezultă deci că o codificare formată dintr-un semn (un bit) și o succesiune de 0 și 1 (reprezentarea în baza 2 a oricărui număr real subunitar) reflectă în totalitate orice număr real. Trecând însă din nou la limită dimensiunea spațiului observațional ($n \rightarrow \infty$) "puterea" reprezentării prin numere reale ($f:A \rightarrow \mathbb{R}$) este de aceeași cardinalitate cu cea a reprezentării cu numere întregi ($f:A \rightarrow \mathbb{N}$) sau binar ($f:A \rightarrow \{0,1\}$), 2^{\aleph_0} , unde \aleph_0 este cardinalitatea mulțimii numerelor naturale. Acest simplu fapt ne arată că chiar dacă se mărește calitatea reprezentării prin numere reale, rezoluția reprezentării este în continuare insuficientă pentru a egala calitatea enumerării ($\aleph_0!$).

O primă consecință imediată a calității reprezentării din spațiul informațional este existența degenerării. Degenerarea este reprezentarea prin intermediul aceleiași valori a rezultatului observației asupra a două elemente distincte (diferite). Această degenerare este uneori un avantaj (când se pun în evidență similitudinile între proprietățile a două elemente) alături de dezavantaj (când măsurarea care a avut ca scop evidențierea diferențelor între cele două elemente a eșuat în a-și atinge scopul).

O a doua consecință imediată a calității reprezentării din spațiul informațional este că dacă degenerarea nu poate fi evitată prin funcția de măsurare, încă poate fi diminuată prin scala de măsurare. Ar trebui remarcat faptul că nu toate scalele de măsurare induc relații de ordine în spațiul informațional. Exemple naturale sunt grupa de sânge și aminoacizi care constituie codul genetic, și anume sunt situații când codificarea din spațiul informațional nu exprimă o relație de ordine (naturale) între valorile măsurate.

Fie o mulțime cu două elemente ($C=\{a,b\}$) și forțăm ipoteza că ordinea nu este relevantă între ele. Mulțimea submulțimilor lui C este $S_C = \{\emptyset, \{a\}, \{b\}, \{a,b\}\}$. Un ordinea naturală în mulțimea S_C este definită prin cardinalitatea submulțimii. Cardinalitatea ca relație de ordine nu este strictă, pentru că există două submulțimi cu același număr de elemente: $0 = |\{\emptyset\}| < |\{a\}| = 1 = |\{b\}| < |\{a,b\}| = 2$. S-ar putea întreba: "Ce tip de scală de măsură definește cardinalitatea?" - Pentru a oferi un răspuns util trebuie să ne întoarcem la măsurare și noi ar trebui să întrebăm mai întâi: "Ce caracteristici se doresc a fi evaluate?". În cazul în care răspunsul la a doua întrebare este numărul de elemente în subgrupul observat, atunci cardinalitatea este bine definită a fi **cantitativă** - fiind dotată cu o relație de ordine. În cazul în care diferențierea între submulțimile lui C este scopul dorit, atunci cardinalitatea submulțimii nu este suficientă. S-ar putea construi în continuare un alt experiment menit să diferențieze submulțimile pentru care apare degenerarea (în cazul de mai sus pentru submulțimile cu un element) și o nouă funcție de

măsurare ar da răspunsul la întrebarea "Submulțimea conține elementul 'a'?" (complementar cu răspunsul la întrebarea "Submulțimea conține elementul b?"). Aceasta este o măsurătoare tipic **calitativă** - căutăm potriviri.

O altă consecință derivată din căutarea după submulțimile unei mulțimi este că **scala de măsură** care se intenționează a se aplica ar trebui să fie de cel puțin verificată din punct de vedere al consistenței cu scopul propus. Mai mult, chiar și atunci când nu există relații de ordine, pot exista alte relații (cum ar fi complementul logic $\{a\}=\{a,b\}\setminus\{b\}$ în mulțimea submulțimilor mulțimii $\{a,b\}$), care aduce în spațiul informațional faptul că nu întotdeauna rezultatele măsurătorilor sunt independente unul față de altul. Mergând mai departe, tabelul de mai jos clasifică după complexitate (definită de către operațiile permise între valorile înregistrate) scalele de măsură (a se vedea *Scale de măsură*).

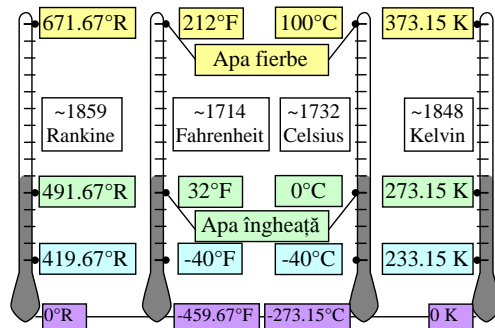
Scală	Tip	Operații	Structură	Statistici	Exemple
Binomială	Logic	"=", "!"	Algebră Booleană	Moda, Fisher Exact	Dead/Alive Fețele unei monezi
(multi) Nomi(n)ală	Discret	"="	Mulțime standard	Moda, Chi squared	Sistemul de grupe de sânge ABO Clasificarea organismelor vii
Ordinală	Discret	"=", "<"	Algebră comutativă	Mediana, Rangul	Numărul de atomi în molecule
Interval	Continuu	"≤", "-"	Spațiu afin (uni-dimensional)	Media, StDev, Corelația, Regresia, ANOVA	Scala de temperatură
Raport	Continuu	"≤", "-", "×"	Spațiu vectorial (uni-dimensional)	GeoMean, HarMean, CV, Logaritm	Dulceața relativă la sucroză pH Scala distanțelor Scala timpului Scala energiei

Scale de măsură

O scală de măsurare este **nominală** dacă între valorile sale o relație de ordine nu poate fi definită. De obicei scala nominală de măsurare este destinată să fie utilizată pentru măsuri calitative. Scala **binară** (sau binomială) este cu doar două valori posibile (între care există o relație de ordine), cum ar fi: $\{Da, Nu\}$, $\{Viu, Mort\}$, $\{Vivo, Vitro\}$, $\{prezent, absent\}$, $\{alcan\ saturated, alt\ tip\ de\ compus\}$, $\{număr\ întreg, număr\ neîntreg\}$. Scala nominală cu mai mult de două valori posibile este numită multinomială. Scala multinomială de măsurare are un număr finit de valori posibile și independent de numărul lor, operează relația de complementaritate. Astfel, pentru $\{0, A, B, AB\}$ grupe sanguine o valoare diferită de oricare dintre cele trei, sigur este cea de a patra. O serie finită de valori poate fi considerată o scală **ordinală** dacă între valorile lor posibile se poate defini o relație de ordine (naturală). Dacă presupunem că "Absent" < "Prezent", "Fals" < "Adevărat", "0" < "1", "Negativ" < "Nenegativ", "Nepozitiv" < "Pozitiv", atunci toate aceste scale de măsură sunt ordinale. Mai mult, un exemplu de scală ordinală cu trei valori este: "Negativ" < "Zero" < "Pozitiv". Un alt lucru important cu privire la scalele ordinale este că nu sunt necesare, cu o cardinalitate finită. Dar este necesară existența unei relații de ordine definită prin "Succesorul unui element" (al unei valori) și complementul acesteia "Predecesorul unui element" (al unei valori). În scala **interval** distanța (sau diferența) între valorile posibile are un sens. De exemplu diferența între 30° și 40° pe scala de temperatură are aceeași semnificație cu diferența între 70° și 80°. Intervalul între două valori este interpretabil (are un sens fizic). Acesta este motivul pentru care are sens calcularea valorii medii a unei variabile de tip interval, ceea ce însă nu are sens pentru valorile unei scale ordinale. În același timp (vezi *Termometrul cu mercur și scale de temperatură*) cum ar fi 80° nu este de două ori mai fierbinte decât 40° (așa cum 2m sunt de 2 ori mai mulți decât 1m), pentru scalele interval raportul dintre două valori nu are nici un sens.

În cele din urmă, pe scalele de tip **raport**, valorile 0 și/sau 1 au întotdeauna o semnificație. Ipoteza este că cea mai mică valoare observabilă este 0. Rezultă prin urmare faptul că dacă două valori sunt luate pe o scală raport, putem calcula raportul lor, și, de asemenea, această măsură posedă o scală de măsurare de tip raport.

În cele din urmă, trebuie să remarcăm că mărimile măsurate pe o scală raport sunt **aditive**, în timp ce mărimile măsurate pe celelalte scale sunt **neaditive**.



Termometrul cu mercur și scale de temperatură

Trebuie menționat faptul că scala de măsurare nu dă precizia de măsurare, și din acest punct de vedere este ilustrativ exemplul țiței (v. *Precizie și exactitate*).



Precizie și exactitate

De scala pe care a fost măsurată proprietatea depinde și modul în care datele se pot prelucra și interpreta. Așa cum s-a ilustrat (v. *Precizie și exactitate*) **precizia** și **exactitatea** unei măsurători sunt la fel de importante ca și valoarea măsurată înseși. Acesta este motivul pentru care se obișnuiește să se exprime valoarea unei măsurători împreună cu precizia sa de măsurare.

Desigur că există mai multe modalități de a exprima precizia unei măsurători. De exemplu, o măsurătoare cu un aparat de măsură nu poate depăși precizia pentru care aparatul a fost construit. Din punctul de vedere al scalei de măsură, o variabilă (din spațiul informațional) care numără moleculele dintr-un spațiu (fizic) dat este "la fel" de tip raport ca și o variabilă ce măsoară temperatura mediului (fizic) în care aceste molecule sunt localizate, chiar dacă rezultatul acestor două operații de măsurare nu are aceeași precizie sau precizie comparabilă (sau nu pare a avea).

Este de dorit în mod evident ca scala de măsurare să încorporeze cât mai multe caracteristici ale variabilei măsurate, precum și să posede numai caracteristicile găsite în variabilă măsurată, pentru că, în caz contrar, scala de măsurare devine o sursă de eroare.

Indiferent de nivelul de structură (moleculară, atomică, subatomică) la care ne referim numărul de particule (compuși chimici, atomi, electroni) la nivel macroscopic (observabil cu ochiul liber sau cu instrumente de mărire) cuprins într-un spațiu de volum definit este imens. Din acest motiv pentru a face referire la macrocantități este nevoie de o unitate de măsură corespunzătoare. Aceasta este molul.

Molul este cantitatea de particule (molecule, ioni, atomi, electroni, altele asemenea sau grupuri ale acestora) al căror tip trebuie specificat și al căror număr este egal cu numărul de atomi de carbon existenți în 0.012 kg (12g) din izotopul acestuia ^{12}C .

Astfel, cantitatea de particule (impropriu spus "cantitate de substanță") se poate exprima prin intermediul numărului de particule (N) sau prin intermediul numărului de moli (n) iar între aceste două modalități de exprimare există relația: $n = N/N_A$, în care N_A este numărul lui Avogadro și exprimă valoarea aproximativă a numărului de atomi de carbon existenți în 0.012 kg (12g) din izotopul acestuia ^{12}C : $N_A = 6.02214 \cdot 10^{23} \text{ mol}^{-1}$. Prin intermediul cantității de substanță o serie de proprietăți observate au caracter intensiv și extensiv: $X_m = X/n$, în care X nominalizează oricare **proprietate extensivă** (care depinde de cantitatea de substanță) iar X_m nominalizează **proprietatea intensivă** corespondentă (care nu mai depinde de cantitatea de substanță).

Energia ca atribut al unei substanțe și consecință a structurii sale atomice, moleculare sau agregate este o mărime intensivă în timp ce energia specifică este corespondentul intensiv al energiei. Similar, energia liberă - eliberată sau absorbită într-un proces este o mărime extensivă în timp ce potențialul chimic este mărimea intensivă asociată. Capacitatea calorică este cantitatea de căldură ce produce schimbarea temperaturii cu 1K și este o mărime extensivă, și capacitatea calorică specifică este

mărimii intensivă asociată.

Masa este proprietate extensivă (M), iar masa molară (M_m) este proprietate intensivă. Volumul (V) este o proprietate extensivă în timp ce volumul molar (V_m) este o proprietate intensivă. Concentrația (molară, molală, procentuală) este o mărime intensivă:

$$n = M/M_m = V/V_m; c_M = n/V_S; c_m = n/m_S; c_{\%m} = m_d/m_S; c_{\%V} = V_d/V_S$$

Se poate remarca că concentrația molară variază cu temperatura, deoarece volumul variază cu temperatura, în timp ce molaritatea este o mărime independentă de temperatură. Se numește o soluție diluată, o soluție ce conține cel mult $10^{-2} \text{ mol} \cdot \text{l}^{-1}$ de solut. În soluțiile diluate ionii de solut sunt separați de cel puțin 10 molecule de solvent. O altă mărime frecvent utilizată la amestecuri este fracția molară x_j (a componentului j) din amestecul cu J ($j \in J$) componenți: $x_j = n_j / \sum_{i \in J} n_i$.

Se poate demonstra că fracția molară este o mărime intensivă. Astfel, fie un amestec P cu compoziția exprimată prin raportul numărului de molecule din fiecare component j în amestec $\alpha_1 : \alpha_2 : \dots : \alpha_J$ (cum ar fi pentru $C_2O_4H_2$, $\alpha_1 : \alpha_2 : \alpha_3 = 2:4:2 = 1:2:1$), și numărul de moli n . Din cele $N = n \cdot N_A$ molecule ale amestecului, pentru a respecta proporția, numărul de molecule din componentul j este $N_j = N \cdot \alpha_j / \sum_{i \in J} \alpha_i$. Frația molară a amestecului este:

$$x_j = n_j / \sum_{i \in J} n_i = (N_j / N_A) / \sum_{i \in J} (N_i / N_A) = (N_j / \sum_{i \in J} N_i) = (N \cdot \alpha_j / \sum_{i \in J} \alpha_i) / \sum_{i \in J} (N \cdot \alpha_i / \sum_{i \in J} \alpha_i) = (\alpha_j) / \sum_{i \in J} \alpha_i$$

Expresia rezultată nu depinde decât de compoziție și nu depinde de numărul de moli sau molecule implicate așa că este o mărime intensivă. Similar, densitatea este o mărime intensivă. În cazul unui amestec cu J componenți:

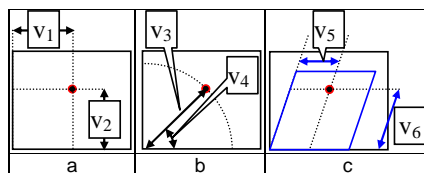
$$\rho = (\sum_{i \in J} m_i) / (\sum_{i \in J} V_i) = (\sum_{i \in J} n_i \cdot M_i) / (\sum_{i \in J} V_i) = (\sum_{i \in J} n \cdot x_i \cdot M_i) / (\sum_{i \in J} V_i) = (\sum_{i \in J} x_i \cdot M_i) / (\sum_{i \in J} V_i / n) = (\sum_{i \in J} x_i \cdot M_i) / V_m$$

În formula de mai sus intervin numai mărimi intensive (x_i , M_i și V_m) și astfel definește o mărime intensivă. Presiunea, într-un sistem în echilibru, este o mărime intensivă atâta timp cât valoarea acesteia în sistem este egală cu valoarea acesteia în orice parte a acestuia. Temperatura, într-un sistem în echilibru, este o mărime intensivă atâta timp cât valoarea acesteia în sistem este egală cu valoarea acesteia în orice parte a acestuia.

În final trebuie făcută remarca că conceptul de mărime intensivă referă macrocantități și își pierde sensul la nivel microscopic. Luând doar temperatura ca exemplu, în spațiu este de câteva grade Kelvin, în timp ce obiectele care se deplasează (cum ar fi o rachetă sau un meteorit) pot ajunge la temperaturi de câteva mii de grade Kelvin, așa cum rezultă din teoria cinetico-moleculară.

Analiza dimensională

În geometrie, pentru a identifica în mod unic poziția unui punct în plan avem nevoie de un reper. Față de acest reper, punctul de probă are două grade de libertate, ceea ce ne arată că poziția sa este în mod unic determinată de fixarea valorilor pentru două proprietăți geometrice ale acestuia. Alegerea reperului nu schimbă condiționarea anterioară. În figura de mai jos (*v. Repere în plan*), reperul poate fi de exemplu unul din colțurile dreptunghiului. Cea mai evidentă modalitate este de a fixa valorile proiecțiilor din punctul de probă pe laturile dreptunghiului (fig. a) însă nu este necesar ca ambele valori să provină din același tip de măsurătoare (în fig. b o măsurătoare este de distanță, alta este de unghi) și nici nu este necesar să fie ortogonale (în fig. c oricare din laturile paralelogramului are o proiecție nenulă pe laturile învecinate).



Repere în plan

Exemplul *Repere în plan* ne arată că modalitățile de a exprima prin valori proprietățile observate pot fi extrem de diferite. În același timp ne arată că ar putea exista, în fiecare caz în parte, o soluție de tipul celei ilustrate în fig. a, în care variabilele asociate celor două valori să fie ortogonale. Chiar însă și în acest caz, nimic nu ne oprește să considerăm că și cazul ilustrat de fig. b este de asemenea ortogonal, deci putem avea chiar mai multe soluții. În toate cazurile însă, am remarcat necesitatea impunerii a două valori fixe, ceea ce ne arată că dimensionalitatea sistemului este 2.

Problema poate fi extinsă la mărimi și relații altele decât geometrice. În general în științe operăm

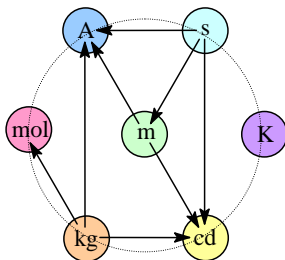
cu mărimi și unități de măsură, și o **analiză dimensională** poate fi condusă dacă se alege un set redus de mărimi și unități pe baza cărora să se poată exprima toate celelalte. Așa cum s-a arătat mai sus, problema nu are o singură soluție, însă unele soluții sunt preferate.

Analiza dimensională asupra mărimilor și relațiilor poate servi la verificarea corectitudinii definiirii acestora. În analiza dimensională a mărimilor și relațiilor fizice, se preferă ca mărimi de bază lungimea, masa, timpul și sarcina electrică. Chiar și aici însă, o remarcă poate fi făcută, și anume că folosind constanta vitezei luminii în vid, se poate exprima o relație de dependență între lungime și timp, iar în ceea ce privește sarcina electrică, definiția sa utilizează lungimea, masa, și timpul.

Un element important al analizei dimensionale îl reprezintă alegerea setului de mărimi de bază. De exemplu în cinematică un set de mărimi de bază este (masa, distanța, timpul) sau (M, L, T). Pe baza acestuia se poate exprima dimensionalitatea vitezei ($L \cdot T^{-1}$), accelerației ($L \cdot T^{-2}$), impulsului ($M \cdot L \cdot T^{-1}$) și a forței ($M \cdot L \cdot T^{-2}$). Setul de mărimi (distanța, viteza, timpul) nu e un set de bază pentru cinematică pentru că masa nu se poate exprima pe baza acestuia și de asemenea cele trei nu sunt independente ($V=L \cdot T^{-1}$).

Cu ajutorul analizei dimensionale se poate stabili corectitudinea unei ecuații din punctul de vedere al omogenității dimensionale și în cazul în care sunt implicate o serie de mărimi diferite într-o expresie simplă se poate chiar stabili forma ecuației din analiza dimensională.

Un sistem de unități de măsură este un set de unități care poate fi folosit pentru a specifica orice care poate fi măsurat. Sistemele actuale de unități de măsură, MKS (metru, kilogram, secundă), CGS (centimetru, gram, secundă), FPS (picior, livră, secundă) diferă unul de celălalt doar prin alegerea unităților de măsură, mărimile măsurate fiind aceleași (distanță, masă, timp). La acestea trei se adaugă și temperatura (măsurată în Kelvin, grade Celsius sau Fahrenheit), însă cum definiția actuală a distanței este făcută prin intermediul constantei vitezei luminii în vid ($d = c \cdot t$, c - constanta vitezei luminii în vid, v. *Constante universale*) practic tot prin intermediul a trei unități fundamentale (temperatura, timpul, masa) sunt construite și definițiile celorlalte de bază (v. *Relații între definițiile mărimilor sistemului fundamental de unități de măsură*).



Relații între definițiile mărimilor sistemului fundamental de unități de măsură

Revenind asupra mărimilor de bază, sistemul internațional de unități recunoaște un număr de 7 ca fiind mărimi de bază:

Mărimă			Unitate	
Nume	Dimensiune	Simbol	Nume	Simbol
masa	M	m	kilogram	kg
cantitatea de substanță	N	n	mol	mol
temperatura termodinamică	Θ	T	kelvin	K
timpul	T	t	secundă	s
lungimea	L	l, x, r	metru	m
curentul electric	I	I, i	amper	A
intensitatea luminoasă	J	I _v	candela	cd

Evaluarea numerică

O **funcție transcendențială** este o funcție care nu satisface o ecuație polinomială ai cărei coeficienți sunt ei înșiși polinoame, în contrast cu o funcție algebrică, care satisface o asemenea ecuație. Cu alte cuvinte, o funcție transcendențială este o funcție care "transcende" algebra, în sensul că aceasta nu poate fi exprimată în termenii unei secvențe finite de operații algebrice de adunare, înmulțire și extragerea rădăcinii. Exemple de funcții transcedentale sunt funcția exponențială, logaritmul și funcțiile

trigonometrice. O **ecuație transcendențială** este o ecuație care conține o funcție transcendențială. Un **număr algebric** este un număr o rădăcină a unui polinom (nenu) într-o variabilă cu coeficienți raționali. Numerele care nu sunt algebrice sunt **numere transcendentale**. Exemplele includ pe π și e .

Inversa funcției $f(w)=w \cdot e^w$ se numește **funcția W a lui Lambert**. Derivata funcției f , $f'(w) = e^w(1+w)$ ne arată că $w=-1$ este un punct de extrem pentru funcția f (de minim, $f(-1)=-e^{-1}$) așa încât față de acest punct de extrem de o parte și de alta a acestuia există exact două soluții (w_1 și w_2) pentru ecuația $y=w \cdot e^w$. Din acest motiv inversa funcției f , funcția W este fie multivalorică (redând pentru un y două valori, w_1 și w_2) fie se restrânge domeniul de definiție al funcției astfel încât să devină (atât f cât și W) funcție bijectivă. Funcția Lambert este o sursă de numere transcendentale, ceea ce înseamnă că soluțiile w asociate ecuației $y=w \cdot e^w$ unde y este un număr algebric exceptând pe $y=0$ sunt toate transcendentale.

Funcția W a lui Lambert este de importanță teoretică, întrucât multe ecuații implicând exponențiale se 'rezolvă' oferind soluțiile (indirecte) ca expresii implicând soluțiile funcției W .

Exemplul 1. Pentru $a, b > 0$ să se rezolve ecuația: $a^x = b \cdot t$. *Rezolvare.* $a^x = b \cdot t \rightarrow 1 = b \cdot t / a^x \rightarrow 1 = b \cdot t \cdot e^{-x \ln(a)} \rightarrow 1/b = t \cdot e^{-x \ln(a)} \rightarrow -\ln(a)/b = -t \cdot \ln(a) \cdot e^{-x \ln(a)} \rightarrow$ Dacă se obține w ca o soluție a ecuației $-\ln(a)/b = w \cdot e^{-w}$ atunci $t = -w / \ln(a)$, care se exprimă formal astfel: $t = -W(-\ln(a)/b) / \ln(a)$.

Exemplul 2. Mai general, pentru $a \neq 0$ și $q > 0$, să se rezolve ecuația: $q^{a \cdot x + b} = c \cdot x + d$. *Indicație.* Se substituie $-t = a \cdot x + a \cdot d/c$ și $R = t \cdot q^t = -(a/c) \cdot q^{b-a \cdot d/c}$.

Exemplul 3. Să se rezolve ecuația $x^x = a$. *Rezolvare.* $x^x = a \rightarrow x \cdot \ln(x) = \ln(a) \rightarrow e^{\ln(x)} \cdot \ln(x) = \ln(a) \rightarrow$ Dacă se obține w ca o soluție a ecuației $\ln(a) = w \cdot e^w$ atunci $x = e^w$, care se exprimă formal astfel: $x = e^{W(\ln(a))}$. *Remarcă.* Dacă $y = w \cdot e^w$, sau scris parametrizat $y = w(y) \cdot e^{w(y)}$ pentru $y = \ln(z)$ atunci și $\ln(z) = w(\ln(z)) \cdot e^{w(\ln(z))}$ are loc, și $e^{w(\ln(z))} = \ln(z) / w(\ln(z))$ astfel încât soluția ecuației $x^x = a$ se mai poate scrie ca: $x = e^{W(\ln(a))} = \ln(a) / W(\ln(a))$.

Alte exemple de utilizare a funcției W a lui Lambert pentru a obține soluțiile ecuațiilor transcendentale includ ecuațiile în forma $x \cdot \log_b(x) = a$ (cu soluția $x = e^{W(a \cdot \ln(b))}$), valorile explicate ale curentului într-un circuit în care sunt în serie o diodă și o rezistență (relația între curent și tensiune în cazul diodei este dat de o lege exponențială [2] în timp ce la rezistor este o dependență liniară), încetătorul atomic Zeeman [3], curgerile și depunerile granulelor și suspensiilor, curgerea fluidelor vâscoase [4], și imagistica creierului [5]. În toate aceste cazuri, ca și în altele de asemeni, când este posibilă obținerea unei soluții analitice sau, în cazul soluțiilor oferite prin intermediul funcției W a lui Lambert, obținerea unor soluții pseudo-analitice, aceste soluții analitice sunt de preferat în favoarea celor obținute prin **evaluare numerică**.

Un caz aparte de ecuații sunt cele în forma $y = a_n x^n + \dots + a_1 x + a_0$, numite **ecuații polinomiale**. Din fericire, astăzi există un algoritm care găsește toate rădăcinile oricărui polinom de orice grad [6], algoritm disponibil online [7] și care este implementat în majoritatea programelor profesionale care necesită rezolvarea de astfel de ecuații. Precizia în care acest algoritm rezolvă ecuațiile polinomiale este limitată numai de precizia de calcul a implementării. De exemplu o implementare C/C++ poate da o precizie de 10^{-14} , o implementare în FreePascal o precizie de 10^{-19} în timp ce o implementare în Fortran poate da o precizie de 10^{-23} .

Pentru ecuațiile la care literatura de specialitate nu ne pune la dispoziție metode deterministe care să ne conducă spre soluție, singura modalitate de abordare rămasă este de a căuta soluțiile. Avantajul pe care îl are utilizarea calculatorului este imens: stocarea expresiei ecuației de rezolvat în calculator oferă practic posibilitatea evaluării rezultatului funcției asociate în orice moment. Fie astfel ecuația de rezolvat $y=f(x)$, în care y este o valoare care trebuie precizată de fiecare dată când facem apel la o rezolvare numerică. Se construiește funcția asociată $g(x) = f(x) - y$. Se poate evalua funcția $g(x)$ în orice

² Carl WAGNER, 1931. On the theory of the rectifier effect (In German). Physikalische Zeitschrift 32:641-645.

³ William D. PHILLIPS, Harold J. METCALF, 1982. Laser Deceleration of an Atomic Beam. Physical Review Letters 48(9):596-599.

⁴ R.M. Corless, G.H. Gonnet, D.E.G. Hare, D.J. Jeffrey, D.E. Knuth, 1996. On the Lambert W function. Advances in Computational Mathematics 5(1): 329-359.

⁵ Roberto C. SOTERO, Iturria-Medina YASSER, 2011. From blood oxygenation level dependent (BOLD) signals to brain temperature maps. Bulletin of Mathematical Biology 73(11):2731-2747.

⁶ M.A. Jenkins, 1975. Algorithm 493: Zeros of a Real Polynomial [C2]. ACM Transactions on Mathematical Software 1(2):178-189.

⁷ <http://www.netlib.org/toms/493>

valoare a domeniului său de definiție prin simpla sa implementare într-o rutină de calcul.

Rezolvarea numerică totdeauna presupune **cunoașterea domeniului** în care se află soluția (sau soluțiile) ecuației. Dacă nu avem nici o indicație asupra domeniului, o bună idee este să se încerce **reprezentarea grafică** a dependenței într-un program specializat (cum este MathCad). Odată identificat domeniul (fie acesta $[a,b]$) în care se află soluția sau soluțiile, următorul pas necesar este separarea soluțiilor, și anume gășirea unui subdomeniu (fie acesta $[c,d]$, așc, d**sb**) în care există o singură soluție. Practic nu există metode standard care să ne ducă la acest subdomeniu, însă din nou o reprezentare grafică ne poate ajuta.

Condiția ca **într-un subdomeniu să existe cel puțin o soluție** este ca $g(c) \cdot g(d) < 0$. Condiția $g(c) \cdot g(d) < 0$ este **suficientă**, în sensul în care dacă $g(c) \cdot g(d) < 0$ atunci există cel puțin o soluție $g(z)=0$ în intervalul $[c,d]$ însă **nu este și necesară**, existența unui număr par de soluții în domeniul $[c,d]$ făcând ca $g(c) \cdot g(d) > 0$ și totuși să existe soluții în acest interval.

Odată identificat un domeniu $[c,d]$ pentru care $g(c) \cdot g(d) < 0$, căutarea soluției ecuației $g(z)=0$ se poate face prin **căutări succesive**, și totdeauna un astfel de algoritm va produce cel puțin o soluție.

Cea simplă metodă de căutare succesivă este prin înjumătățirea intervalului (v. **Căutarea soluției $g(z)=0$ când $g(c) \cdot g(d) < 0$**).

```

Let a0=c; b0=d;
Repeat
  If(a0-b0<eps)then stop;//solution is any of a0 and b0
  If(b0-a0<eps)then stop; //solution is any of a0 and b0
  c0=(a0+b0)/2;
  if(g(c0)<eps)then stop;//solution is c0
  If(g(a0)*g(c0)<0)then b0=c0; else a0=c0;
Until(false);

```

Căutarea soluției $g(z)=0$ când $g(c) \cdot g(d) < 0$

Așa cum se observă în figura de mai sus (v. **Căutarea soluției $g(z)=0$ când $g(c) \cdot g(d) < 0$**) căutarea succesivă a soluției $g(z)=0$ impune definirea unei toleranțe (eps) a cărei valoare minimă este stabilită în funcție de **cerințele problemei de rezolvat** (o precizie de 4 cifre semnificative este foarte rar depășită de instrumentația de analiză fizică și chimică) și de **capacitatea procesorului matematic** al platformei de calcul (s-a arătat mai sus că limita de precizie poate merge până la 23 de cifre semnificative fără a face apel la calculul matematic în precizie arbitrară).

Din fericire, pentru **sistemele de ecuații liniare** sunt elaborate metode, unele dintre ele foarte eficiente, de rezolvare exactă. Diferențele între o metodă și o altă metodă în acest caz o reprezintă precizia cu care este oferită soluția. Pentru a putea rezolva un sistem de ecuații liniare, sunt necesare un număr mare de operații de înmulțire, împărțire, adunare și scădere, iar cea mai eficientă rezolvare are ordinul de complexitate polinomial, fiind de ordinul puterii a 3-a a numărului de ecuații, $O(n^3)$. Astfel, pentru un număr foarte mare de ecuații, sunt implicate și un număr mare de operații (pentru $n = 1000$, $n^3 = 10^9$), ceea ce face ca precizia maximă posibilă să scadă și ea (folosind o implementare C/C++, din precizia de 14 cifre semnificative, scăzând cele 9 datorate operațiilor aritmetice, ajungem la pragul de 5 cifre semnificative ca precizie maximă a evaluării). Morala este că pentru cazurile în care se operează cu sisteme de ecuații foarte mari, utilizarea celor mai precise metode de rezolvare este obligatorie.

În continuare (v. **Procedura Gauss-Jordan pentru regresia liniară multiplă**) este redată procedura Gauss-Jordan pentru rezolvarea sistemelor de ecuații liniare, aplicată pentru modele de regresie liniară, când soluția este furnizată simultan cu calculul de varianță [8].

Pentru ecuații neliniare cu expresii complicate, și cu atât mai mult pentru ecuații transcendentale, cel puțin o soluție poate rezulta prin aproximații succesive. Fie ecuația $f(x)=0$ pentru care căutăm o rădăcină x . O **metodă iterativă** totdeauna pornește de la o valoare inițială x_0 pe care căutăm să o îmbunătățim astfel încât la final să obținem o bună aproximație $|f(x_0)| < \epsilon$ unde ϵ este o valoare arbitrară impusă de cerințele problemei de rezolvat.

⁸ William H. PRESS, Saul A. TEUKOLSKY, William T. VETTERLING, Brian P. FLANNERY, 1992. Gauss-Jordan Elimination (§2.1) and Gaussian Elimination with Backsubstitution (§2.2) In: Numerical Recipes in FORTRAN: The Art of Scientific Computing (2nd ed.) Cambridge: Cambridge University Press p. 27-32 (§2.1) and 33-34 (§2.2).

$m_+ = m+1$	$y \sim \hat{y} = a_0 \cdot 1 + a_1 \cdot x_1 + \dots + a_m \cdot x_m$ pentru $(y_i, x_{1,i}, \dots, x_{m,i})_{1 \leq i \leq n}$								
i(linii)\j(coloane)	-1	0	1	...	m	m+1	m+2	...	2m+1
0	$\sum y_i$	n	$\sum x_{1,i}$...	$\sum x_{m,i}$	1	0	...	0
1	$\sum y_i x_{1,i}$	$\sum x_{1,i}$	$\sum x_{1,i} x_{1,i}$...	$\sum x_{1,i} x_{m,i}$	0	1	...	0
...
m	$\sum y_i x_{m,i}$	$\sum x_{m,i}$	$\sum x_{1,i} x_{m,i}$...	$\sum x_{m,i} x_{m,i}$	0	0	...	1
Inițial	S_B		S_A			I_{m+1}			
Operații elementare pe linii	↓		↓			↓			
Final	$A = S_A^{-1} S_B$		I_{m+1}			S_A^{-1}			
0	a_0	1	0	...	0	$(S_A^{-1})_{0,0}$
1	a_1	0	1	...	0		$(S_A^{-1})_{1,1}$
...
m	a_m	0	0	...	1	$(S_A^{-1})_{m,m}$
$m_+ = m$	$y \sim \hat{y} = a_1 \cdot x_1 + \dots + a_m \cdot x_m$ pentru $(y_i, x_{1,i}, \dots, x_{m,i})_{1 \leq i \leq n}$								
i(linii)\j(coloane)	-1	0	1	...	m	m+1	m+2	...	2m+1
0									
1	$\sum y_i x_{1,i}$		$\sum x_{1,i} x_{1,i}$...	$\sum x_{1,i} x_{m,i}$		1	...	0
...
m	$\sum y_i x_{m,i}$		$\sum x_{1,i} x_{m,i}$...	$\sum x_{m,i} x_{m,i}$		0	...	1
Inițial	S_B		S_A			I_m			
Operații elementare pe linii	↓		↓			↓			
Final	$A = S_A^{-1} S_B$		I_m			S_A^{-1}			
0									
1	a_1		1	...	0		$(S_A^{-1})_{1,1}$
...
m	a_m		0	...	1		$(S_A^{-1})_{m,m}$
Variante coeficienți	$\epsilon_j = \hat{y}_j - y_j = a_0 + a_1 x_{1,j} + \dots + a_m x_{m,j} - y_j$					$s^2(a_i) = (S_A^{-1})_{ii} \cdot \sum \epsilon_j^2 / (n - m_+)$			
Semnificație coeficienți	$t_i = (a_i \neq 0) = a_i / s(a_i)$					$p_i = p(a_i \neq 0) = 2 \cdot \text{CDF}_F(- t_i , n - m_+)$			
Semnificația modelului	$r = \begin{cases} \frac{n \cdot \sum y_j \hat{y}_j - \sum y_j \cdot \sum \hat{y}_j}{\sqrt{(n \cdot \sum y_j^2 - \sum y_j \cdot \sum y_j) \cdot (n \cdot \sum \hat{y}_j^2 - \sum \hat{y}_j \cdot \sum \hat{y}_j)}} & , m_+ = m+1 \\ \frac{(\sum y_j \hat{y}_j)^2}{\sum y_j^2 \cdot \sum \hat{y}_j^2} & , m_+ = m \end{cases}$					$F(r) = r^2 \cdot (n - m_+) / (1 - r^2) / m, p(r \neq 0) = \text{CDF}_F(F(r), m, n - m_+)$			

Procedura Gauss-Jordan pentru regresia liniară multiplă

Ecuția $f(x)=0$ se poate transforma astfel încât să se obțină o ecuație echivalentă cu aceasta în forma $x=g(x)$. Succesul metodei iterative depinde de modalitatea de exprimare a ecuației echivalente.

Exemplu. Fie $f(x)=x^2-2x-3$. În tabelul următor (v. *Aplicarea metodelor iterative*) sunt ilustrate mai multe modalități de exprimare a ecuației echivalente.

Ecuție echivalentă	$x = x^2 - x - 3$	$x = (2 \cdot x + 3) / x$	$x = \pm \sqrt{2 \cdot x + 3}$	$x = \pm \sqrt{2 \cdot x + 3}$
Ecuție iterativă	$x_{i+1} = x_i^2 - x_i - 3$	$x_{i+1} = (2 \cdot x_i + 3) / x_i$	$x_{i+1} = \sqrt{2 \cdot x_i + 3}$	$x_{i+1} = -\sqrt{2 \cdot x_i + 3}$
i=0		1	1	9
i=1		5	-3	3.872983
i=2		2.6	9	2.178524
i=3		3.153846	69	1.164924
i=4		2.95122	4689	#NUM!
i=5		3.016529	21982029	#NUM!

Aplicarea metodelor iterative

Așa cum se observă din tabelul anterior (v. *Aplicarea metodelor iterative*) nu orice alegere a metodei iterative conduce la convergența către soluție ($f(x)=0$).

Metoda aproximațiilor succesive poate fi o soluție comodă în anumite cazuri când se cere rezolvarea de **sisteme de ecuații neliniare**. Un exemplu tipic de aplicare este pentru maximizarea acordului între observație și model. Fie exemplul ilustrat mai jos (v. *Observații la contingența de doi factori multiplicativi*).

Factor "A" \ Factor "B"	Nivel b ₁	Nivel b ₂	Nivel b ₃	Nivel b ₄
Nivel a ₁	Obs ₁₁	Obs ₁₂	Obs ₁₃	Obs ₁₄
Nivel a ₂	Obs ₂₁	Obs ₂₂	Obs ₂₃	Obs ₂₄
Nivel a ₃	Obs ₃₁	Obs ₃₂	Obs ₃₃	Obs ₃₄
Nivel a ₄	Obs ₄₁	Obs ₄₂	Obs ₄₃	Obs ₄₄

Observații la contingența de doi factori multiplicativi

În cazul observațiilor la contingența de doi factori multiplicativi (v. *Observații la contingența de doi factori multiplicativi*) ipoteza este că $Obs_{ij} \sim a_i \cdot b_j$ și valoarea așteptată în urma observației este $Exp_{ij} = a_i \cdot b_j$. Datorită unor factori aleatori necunoscuți, observația e afectată de erori, astfel încât aproape niciodată $Exp_{ij} = Obs_{ij}$. Dificultatea însă constă în faptul că nu se cunosc valorile nivelelor factorilor ($a_1, \dots, a_4; b_1, \dots, b_4$) și nici tipul erorii (eroare proporțională cu valoarea observată, eroare în magnitudine uniformă, etc.). Orice tentativă de a rezolva analitic sistemul (prin identificarea valorilor nivelelor) e sortită eșecului, deoarece sistemul are cel puțin o nedeterminare (putând fi astfel rezolvat doar cel mult parametric).

Avem însă posibilitatea să obținem un set de valori inițiale pentru valorile așteptate cu formula:

$$Exp_{i,j} = \sum_{k=1}^r Obs_{i,k} \cdot \sum_{k=1}^c Obs_{k,j} / \sum_{i=1}^r \sum_{j=1}^c Obs_{i,j}$$

În același cadru al presupunerii naturale al efectului multiplicativ al celor doi factori asupra observabilei Obs din punct de vedere matematic se pot formula trei presupuneri cu privire la eroarea pătratică $(Obs_{ij} - Exp_{ij})^2$ produsă de observație:

- ÷ măsurătoarea este afectată de erori absolute întâmplătoare;
- ÷ măsurătoarea este afectată de erori relative întâmplătoare;
- ÷ măsurătoarea este afectată de erori întâmplătoare pe o scară intermediară între erori absolute și erori relative;

Prima dintre ipoteze (erori absolute întâmplătoare) conduce din punct de vedere matematic la minimizarea varianței între model și observație (S^2), a doua dintre ipoteze conduce la minimizarea pătratului coeficientului de variație (CV^2) iar o soluție (una din mai multe soluții posibile) la cea de-a treia dintre ipoteze (X^2) o reprezintă minimizarea statisticii X^2 (v. *Minimizarea diferitelor tipuri de erori pentru acordul între observație și model*).

S^2	CV^2	X^2
$\sum_{i=1}^r \sum_{j=1}^c (Obs_{i,j} - a_i b_j)^2 = \min.$	$\sum_{i=1}^r \sum_{j=1}^c \frac{(Obs_{i,j} - a_i b_j)^2}{(a_i b_j)^2} = \min.$	$\sum_{i=1}^r \sum_{j=1}^c \frac{(Obs_{i,j} - a_i b_j)^2}{a_i b_j} = \min.$

Minimizarea diferitelor tipuri de erori pentru acordul între observație și model

În relațiile de mai sus (v. *Minimizarea diferitelor tipuri de erori pentru acordul între observație și model*) apar exprimați cei doi factori ("A" și "B") a căror independență se verifică prin intermediul efectului multiplicativ ($a_i, 1 \leq i \leq r$ reprezintă contribuția primului factor la valoarea așteptată $E_{i,j}$ iar $b_j, 1 \leq j \leq c$ reprezintă contribuția celui de-al doilea factor la valoarea așteptată $E_{i,j}$ și expresia valorii așteptate $E_{i,j}$ este dată, așa cum presupunerea naturală a fost făcută de produsul celor două contribuții: $E_{i,j} = a_i \cdot b_j$).

Minimizarea cantităților date de relațiile de mai sus în scopul determinării contribuțiilor factorilor A ($A = (a_i)_{1 \leq i \leq r}$) și B ($B = (b_j)_{1 \leq j \leq c}$) se face pe aceleași cale, dată generic de relația:

$$\left(\frac{\partial \cdot (a_i, b_j)}{\partial a_i} = 0 \right)_{1 \leq i \leq r} ; \left(\frac{\partial \cdot (a_i, b_j)}{\partial b_j} = 0 \right)_{1 \leq j \leq c}$$

unde expresia de derivat $(\partial \cdot (a_i, b_j))$ este una din expresiile date de S^2, CV^2 și X^2 .

În urma calculului (derivare) se poate obține că condițiile de minimizare a diferitelor tipuri de erori pentru acordul între observație sunt echivalente cu ecuațiile următoare (v. *Ecuații în minimizarea diferitelor tipuri de erori pentru acordul între observație și model*).

S^2	CV^2	X^2	Domeniu
$a_i = \sum_{j=1}^c b_j O_{i,j} / \sum_{j=1}^c b_j^2$	$a_i = \sum_{j=1}^c \frac{O_{i,j}^2}{b_j^2} / \sum_{j=1}^c \frac{O_{i,j}}{b_j}$	$a_i^2 = \sum_{j=1}^c \frac{O_{i,j}^2}{b_j} / \sum_{j=1}^c b_j$	$i = 1..r$
$b_j = \sum_{i=1}^r a_i O_{i,j} / \sum_{i=1}^r a_i^2$	$b_j = \sum_{i=1}^r \frac{O_{i,j}^2}{a_i^2} / \sum_{i=1}^r \frac{O_{i,j}}{a_i}$	$b_j^2 = \sum_{i=1}^r \frac{O_{i,j}^2}{a_i} / \sum_{i=1}^r a_i$	$j = 1..c$

Ecuatii în minimizarea diferitelor tipuri de erori pentru agrement în între observație și model

Se poate de asemenea arăta matematic că relațiile de mai sus admit o infinitate de soluții și că familiile de soluții ale relațiilor se află în vecinătatea familiei de soluții date valorile inițiale ale valorilor așteptate ($Exp_{i,j}=a_i \cdot b_j$):

$$a_i \cdot b_j = \sum_{k=1}^r O_{i,k} \sum_{k=1}^c O_{k,j} / \sum_{i=1}^r \sum_{j=1}^c O_{i,j}$$

Calea directă de rezolvare a ecuațiilor de minimizare fără a face apel la ecuația de aproximare este ineficientă. De exemplu pentru $r=2$, $c=3$ substituțiile în relația pentru S^2 duc la:

$$\left(\frac{a_2}{a_1}\right)^2 + \frac{(Obs_{1,1}^2 + Obs_{1,2}^2 + Obs_{1,3}^2) - (Obs_{2,1}^2 + Obs_{2,2}^2 + Obs_{2,3}^2)}{(Obs_{1,1}Obs_{2,1} + Obs_{1,2}Obs_{2,2} + Obs_{1,3}Obs_{2,3})} \left(\frac{a_2}{a_1}\right) - 1 = 0$$

care este rezolvabilă în (a_2/a_1) care dovedește că există o infinitate de soluții (pentru orice valoare nenulă a lui a_1 există o valoare a_2 care să verifice ecuația și gradul ecuației este dat de $\min(r,c)$). Ecuațiile ce se obțin pe calea substituției directe devin din ce în ce mai complicate cu creșterea lui r și c și cu coborârea dinspre relația (S^2) către relația (X^2).

Calea indirectă de rezolvare a ecuațiilor de minimizare este prin aproximații succesive făcând apel la soluția aproximativă oferită de estimarea inițială. Astfel, se folosește relația de estimare inițială pentru a obține prima aproximație (aproximația inițială) a soluției după care în fiecare succesiune de aproximații se înlocuiesc vechile valori ale aproximației în partea dreaptă a relațiilor pentru a obține noile aproximații.

În acest caz metoda aproximațiilor succesive **converge rapid către soluția optimă** (v. [9])

Probabilități și statistică

Frecvent operăm cu date provenite din măsurători repetate ale aceluiași fenomen. Aproape niciodată însă, două măsurători independente asupra aceleiași observabile, produce același rezultat. Din acest punct de vedere, odată încheiată observația fenomenului fizic, se pune problema prelucrării și interpretării informației din spațiul informațional și un alt tip de experiment trebuie derulat aici: **experimentul statistic**.

Într-un experiment statistic operăm cu variabile aleatoare - o **variabilă aleatoare** fiind asociată unei măsurabile ale cărei valori sunt colectate în spațiul informațional. Variabila aleatoare are o **serie de valori** care corespund rezultatelor măsurătorilor înregistrate. Totdeauna în spațiul informațional vom poseda serii finite de valori în timp ce în spațiul fizic s-ar putea repeta experimentul de un număr infinit de ori. Din acest motiv, în statistică se diferențiază **eșantionul** - seria de valori colectată în spațiul informațional, de **populație** - întreaga mulțime infinită de valori care ar putea fi adusă din spațiul fizic printr-un proces continuu de măsurare. Din punctul de vedere al reprezentării prin valori în spațiul informațional, avem **variabile aleatoare discrete**, ale căror valori aparțin unei mulțimi finite (înzestrate sau nu cu o relație de ordine) sau infinite dar atunci înzestrate cu o relație de ordine și numărabile (cum este mulțimea numerelor naturale, sau a celor întregi, sau a celor raționale) și **variabile aleatoare continue**, ale căror valori aparțin mulțimii numerelor reale (de puterea continuului) - și asta chiar dacă, în spațiul informațional le stocăm tot folosind o mulțime de valori discrete și ordonate (folosind o anumită precizie de exprimare a numerelor). O altă clasificare a tipului variabilelor aleatoare este în **cantitative** - dacă scala de măsură a fenomenului observat permite operația de adădire ("+") și **calitative** - dacă nu e permisă (nu are sens).

Așa cum s-a anticipat la începutul secțiunii anterioare, o problemă la care statistica are o soluție de exprimare a rezultatului este cu privire la rezultatul urmând o serie de **măsurători repetate**. În tabelul

următor (v. *Măsurile statistice pentru caracterizarea variabilelor cantitative*) sunt prezentate măsurile statistice ce rezultă din aplicarea pe populații și respectiv pe eșantioane, pentru cazul variabilelor cantitative.

Măsură	Referă	Expresie	Interpretare
Suma valorilor	Un șir de numere	$\Sigma(\cdot)$	-
Numărul de valori		$ \cdot $	-
Valoarea medie		$E(\cdot) = \Sigma(\cdot)/ \cdot $	Valoarea așteptată
Moment central de ordin k, k>1		$E_k(\cdot) = E((X-E(X))^k)$	-
Media caracteristicii X	O populație	$\mu = \mu(X) = E(X)$	Tendința centrală
Media observabilei Y	Un eșantion	$m = m(Y) = E(Y)$	
Estimatorul mediei caracteristicii X	O populație	$M(Y) = m(Y)$	
Varianța caracteristicii X	O populație	$Var(X) = E((X-\mu)^2)$	Împrăștierea
Deviația standard a caracteristicii X		$\sigma = \sigma(X) = \sqrt{Var(X)}$	Dispersia
Varianța observabilei Y	Un eșantion	$var = var(Y) = E((Y-E(Y))^2)$	Împrăștierea
Deviația standard a observabilei Y		$s = s(Y) = \sqrt{Var(Y)}$	Dispersia
Estimatorul varianței caracteristicii X	O populație	$VAR(Y) = var(Y) \cdot Y /(Y -1)$	Împrăștierea
Estimatorul deviației standard a caracteristicii X		$S = S(Y) = s(Y) \cdot \sqrt{ Y }/\sqrt{ Y -1}$	Dispersia

Măsurile statistice pentru caracterizarea variabilelor cantitative

Un caz foarte frecvent în măsurătorile repetate este când valorile se distribuie 'normal' (vezi mai jos **distribuția Gauss**), și din acest punct de vedere, este esențială caracterizarea depărtării de la normalitate (v. *Statistici pentru caracterizarea depărtării de normalitate a variabilelor cantitative*).

Symbol și măsură	Referă	Expresie	Mărimi ce intervin
γ_1 , Asimetria caracteristicii X	O populație	$\gamma_1 = \mu_3/\mu_2^{3/2}$	$\mu_k = E_k(X), k>1$
β_2 , Boltirea caracteristicii X		$\beta_2 = \mu_4/\mu_2^2$	
γ_2 , Excesul de boltire al caracteristicii X		$\gamma_2 = \beta_2 - 3$	
g_1 , Asimetria observabilei Y	Un eșantion	$g_1 = m_3/m_2^{3/2}$	$m_k = E_k(Y), k>1$
b_2 , Boltirea observabilei Y		$b_2 = m_4/m_2^2$	
g_2 , Excesul de boltire al observabilei Y		$g_2 = b_2 - 3$	
Estimatorul asimetriei caracteristicii X	O populație	$G_1 = \frac{\sqrt{n_Y(n_Y-1)}}{(n_Y-2)} M_3/M_2^{3/2}$	$n_Y = Y $
Estimatorul boltirii caracteristicii X		$B_2 = \frac{(n_Y-1)(n_Y+1)}{(n_Y-2)(n_Y-3)} M_4/M_2^2$	$M_k = \frac{n_Y}{n_Y-1} E_k(Y),$
Estimatorul excesului de boltire a caracteristicii X		$G_2 = B_2 - 3 \cdot \frac{(n_Y-1)^2}{(n_Y-2)(n_Y-3)}$	$k>1$

Statistici pentru caracterizarea depărtării de normalitate a variabilelor cantitative

Mărimă și notație	Valoare
Media mediei, $\mu_{\bar{Y}}$	$\mu_{\bar{Y}} = \mu(m(Y)) = \mu(X)$
Varianța mediei, $\sigma_{\bar{Y}}^2$	$\sigma_{\bar{Y}}^2 = \sigma^2(m(Y)) = \sigma^2(X)/n_Y$
Media varianței, $\mu(s^2)$	$\mu(s^2) = \mu(s^2(Y)) = \sigma^2(X)(n_Y-1)/n_Y$
Varianța varianței, $\sigma^2(s^2)$	$\sigma^2(s^2) = \sigma^2(s^2(Y)) = \frac{(n_Y-1)^2}{n_Y^3} \mu_4(X) - \frac{(n_Y-1)(n_Y-3)}{n_Y^3} \mu_2^2(X)$

Medii și varianțe ale mediei și varianței observabilei Y ce rezultă din distribuția de eșantionare din populația cu caracteristica X

Mărime și notație	Aproximare
Media mediei, $\mu_{\bar{Y}}$	$\mu_{\bar{Y}} \cong m(Y)$
Varianța mediei, $\sigma_{\bar{Y}}^2$	$\sigma_{\bar{Y}}^2 \cong s^2(Y)/(n_Y - 1)$
Media varianței, $\mu(s^2)$	$\mu(s^2) \cong s^2(Y)$
Varianța varianței, $\sigma^2(s^2)$	$\sigma^2(s^2) \cong \frac{(n_Y - 1)}{n_Y^2} m_4(Y) - \frac{(n_Y - 3)}{n_Y(n_Y - 1)} m_2^2(Y)$

Valori pentru medie, varianța mediei și varianței observabilei Y în ipotezele teoremei limită centrale

Un alt caz foarte important de distribuție este **distribuția de eșantionare**. Extragerea repetată de eșantioane (de volum dat) dintr-o populație face ca valorile obținute să urmeze o distribuție, numită distribuția de eșantionare. Mai sus (v. Medii și varianțe ale mediei și varianței observabilei Y ce rezultă din distribuția de eșantionare din populația cu caracteristica X) au fost prezentate rezultatele care se obțin pentru varianța mărimilor statistice prin extragerea repetată de eșantioane dintr-o populație.

Când valorile parametrilor statistici ai populației nu sunt cunoscute, dar se poate face presupunerea că distribuția populației se comportă suficient de bine [10], aceștia pot fi aproximați cu ajutorul estimatorilor acestora. Formulele de calcul ale mediei și varianței pentru medie și varianță sunt redată în Valori pentru medie, varianța mediei și varianței observabilei Y în ipotezele teoremei limită centrale.

Dacă se pot asuma ipoteze cu privire la distribuția caracteristicii X în populație, atunci se pot obține formule de calcul pentru parametrii statistici (ai populației) și respectiv estimatorii parametrilor statistici ai populației din măsurătorile (statistice) efectuate asupra eșantionului.

Pentru a **măsura agrementul între observație și model** avem la dispoziție o serie de **statistici**.

Fie $Y = (Y_1, \dots, Y_n)$ și $X = (X_1, \dots, X_n)$ serie de observații pereche și obiectivul să fie găsirea unei funcții $f(x, a_1, \dots, a_m)$, pentru care $Y = f(X)$ este cea mai bună soluție posibilă a aproximare $\hat{Y} \sim Y$. Atingerea acestui obiectiv presupune găsirea expresiei funcției f și a valorilor parametrilor a_1, \dots, a_m . Sub ipoteza de acord între observație și model, expresia funcției f se presupune a fi cunoscută (sau cel puțin ar trebui, atunci când se desfășoară o căutare după un anumit set de expresii alternative).

Astfel, a rămas obținerea valorilor parametrilor a_1, \dots, a_m . Pentru a avea o soluție unică pentru valorile parametrilor a_1, \dots, a_m , cel puțin se cere ca $m \leq n$ să fie asigurată. O serie de alternative sunt disponibile pentru aproximarea $Y \sim f(X)$, iar cele considerate cele mai importante sunt exemplificate în continuare.

Minimizarea erorii de acord (minimizarea dezacordului) este o modalitate de estimare a parametrilor necunoscuți ai unei distribuții (presupus cunoscute). În această ipoteză o serie de alternative sunt disponibile (ecuația de mai jos, cu diferite alegeri pentru p și q):

¹⁰ Teorema Limită Centrală

÷ Cronologia contribuțiilor majore:

- Abraham DE MOIVRE. 1733. Approximatio ad Summam Terminorum Binomii $(a+b)^n$ in Seriem expansi. In: The Doctrine of Chance: or The Method of Calculating the Probability of Events in Play (Abraham DE MOIVRE). W. Pearfon 1738: 235-243.
- Joseph L. LAGRANGE. 1776. Mémoire sur l'utilité de la méthode de prendre le milieu entre les résultats de plusieurs observations; dans lequel on examine les avantages de cette méthode par le calcul des probabilités; et où l'on résoud différents problèmes relat ifs à cette matière. Miscellanea Taurinensia 5:167-232.
- Pierre S. LAPLACE. 1812. Théorie Analytique des Probabilités. Courcier, 465 p.
- Aleksandr M. LIAPUNOV. 1901. Nouvelle forme du théoreme sur la limite des probabilités. Mémoires de l'Académie Impériale des Sciences de St. Pétersbourg 12(5):1-24.

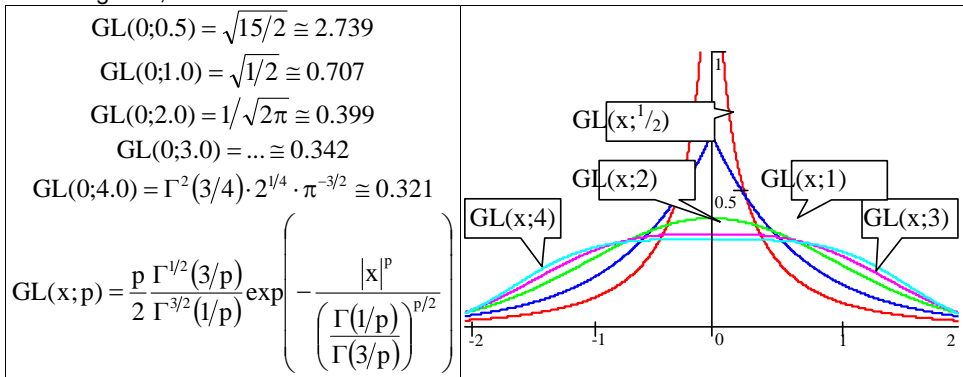
÷ Enunțul teoremei (fie $(X_n)_{n \geq 1}$ variabile independente și $\exists \delta > 0$ a.î. $\mu_{2+\delta}(X_n) < \infty$):

- dacă
$$\lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n \mu_{2+\delta}(X_k)}{\left(\sum_{k=1}^n \sigma_k^2\right)^{(2+\delta)/2}} = 0$$
 atunci
$$\frac{\sum_{i=1}^n (X_n - \mu_1(X_n))}{\sqrt{\sum_{k=1}^n \sigma_k^2}} \xrightarrow{n \rightarrow \infty} N(0,1)$$

$$S(p, q) = \sum_{i=1}^n |Y_i - f(X_i)|^p / f^q(X_i) = \min, \cdot q = 0, 1, p/2, p$$

În cazul în care seria Y reprezintă frecvența (nenulă) a observațiilor distincte X, atunci f(x) ar trebui să fie o funcție pozitivă de asemenea, și modul numărătorului nu mai este necesar.

Distribuția Gauss [11] a valorilor termenilor sumei sunt traduse în p = 2, și distribuția Laplace [12] apare când p = 1 (vezi și [13]). Funcția densitate de probabilitate (PDF) a unor reprezentanți ai familiei care conține distribuțiile Gauss și Laplace standard (μ = 0 și σ = 1), sunt exemplificate în fig. 11. Minimizarea erorii de acord pentru valori p diferite oferă soluții diferite pentru parametrii, și cum se poate observa în figura 1, sunt asociate cu diferite forme de eroare.



Distribuția Gauss-Laplace

Două cazuri particulare sunt de obicei utilizate pentru a estima parametrii necunoscuți ai unei distribuție atunci când p = 2 (vezi [14] și [15]).

Utilizarea momentelor este una dintre alternative. Sub ipoteza că Y~f(X) ar trebui să fie chiar mai precisă (a doua ipoteză fiind caracterul aleatoriu al erorii, cu o medie de zero) și aproximarea $\sum X_i^k Y_i \sim \sum X_i^k \cdot f(x_i)$ este folosită pentru k ≥ 0. Metoda momentelor dă greutate maximă la primele momente, astfel, o soluție a₁, ..., a_m a Y~f(X) poate proveni direct din ecuația $\sum X_i^k Y_i = \sum X_i^k \cdot f(x_i)$. Modul cel mai convenabil pentru cazul general este repetarea căutării parametrilor a₁, ..., a_m începând de la anumite valori inițiale):

$$\sum_{i=1}^n X_i^k Y_i \sim \sum_{i=1}^n X_i^k f(X_i), k = 0, 1, \dots$$

Utilizarea momentelor centrale este o altă alternativă. Se întărește aproximarea $\sum X_i Y_i \sim \sum X_i \cdot f(X_i)$ și $\sum (Y_i - \bar{Y})^k \sim \sum (X_i - \bar{X})^k \cdot f(X_i)$ pentru k ≥ 2. Metoda momentelor dă greutate maximă primelor momente centrale, deci o soluție a₁, ..., a_m a Y~f(X) poate proveni din ecuația:

$$\sum_{i=1}^n X_i Y_i \sim \sum_{i=1}^n X_i f(X_i) \text{ și } \sum_{i=1}^n (Y_i - \bar{Y})^k \sim \sum_{i=1}^n (X_i - \bar{X})^k f(X_i), k = 2, 3 \dots$$

O altă alternativă este **folosind statisticile referitoare la populație**. O ușoară modificare a metodei anterioare poate beneficia de disponibilitatea expresiei pentru media, deviația standard,

¹¹ Carl F. GAUSS, 1809. Theoria Motus Corporum Coelestium. Perthes et Besser, Hamburg. Translated, 1857, as Theory of Motion of the Heavenly Bodies Moving about the Sun in Conic Sections, trans. C. H. Davis. Little, Brown; Boston. Reprinted, 1963, Dover, New York.

¹² Pierre S. LAPLACE, 1812. Théorie Analytique des Probabilités. Paris: Courcier.

¹³ Ronald A. FISHER, 1920. A Mathematical Examination of the Methods of Determining the Accuracy of an Observation by the Mean Error, and by the Mean Square Error. Monthly Notices of the Royal Astronomical Society 80:758-770.

¹⁴ Fisher, R. A. and Mackenzie W. A., 1923. Studies in Crop Variation. II. The manurial response of different potato varieties. Journal of Agricultural Science 13:311-320.

¹⁵ R. A. Fisher, 1924. The Conditions Under Which χ^2 Measures the Discrepancy Between Observation and Hypothesis. Journal of the Royal Statistical Society 87:442-450.

asimetria și excesul asimetriei populației în funcție de parametrii distribuției pentru un număr mare de distribuții bine cunoscute.

Estimarea pe baza șansei maxime oferă, teoretic, estimarea aflată în cel mai mic risc de a fi în eroare. Principiul șansei maxime este că o estimare rezonabilă pentru un parametru este cea care maximizează probabilitatea (P în ecuația de mai jos) de observare a datelor experimentale [16], iar setul mai probabil al parametrilor a_1, \dots, a_m va face P maxim (și implicit MLE este maxim).

$$\text{MLE} = \log_2(P) = \sum_{i=1}^n \log_2(f(X_i))$$

Statistica Benford. Testul Benford folosește distribuția Z (normală) pentru a verifica ipoteza că un șir de numere urmează distribuția Benford, frecvențele după care se distribuie o anumită cifră a fiecărui număr din șir. Un șir de numere urmează distribuția Benford dacă probabilitatea de distribuție a unei cifre (d_i) a numerelor ($d=d_0d_1\dots$) reprezentate în baza de numerație b (uzual baza 10) urmează legea (Benford):

$$\begin{aligned} p(d_0) &= \log_b \left(1 + \frac{1}{d_0} \right), d_0 = 1..(b-1); \\ p(d_1) &= \sum_{k=1}^{b-1} \log_b \left(1 + \frac{1}{k \cdot b + d_1} \right), d_1 = 0..(b-1) \\ p(d_2) &= \sum_{j=1}^{b-1} \sum_{k=0}^{b-1} \log_b \left(1 + \frac{1}{j \cdot b^2 + k \cdot b + d_2} \right), d_2 = 0..(b-1) \\ &\dots \end{aligned}$$

Statistica Benford

Ipoteza acestei legi de distribuție este că valorile măsurătorilor rezultate din observație sunt frecvent distribuite logaritmice și astfel logaritmul setului de măsurători este distribuit uniform. Legea de distribuție este numită după fizicianul Frank BENFORD care a formulat-o intuitiv în 1938 [17], dar demonstrația acesteia a fost dată mult mai târziu [18].

Acest rezultat intuitiv de numărare a aparițiilor a fost găsit aplicându-se la o mare varietate de seturi de date incluzând facturile la electricitate, adresele de străzi, prețurile acțiunilor, numerele populației, ratele de deces, lungimile râurilor, constante fizice și matematice și procesele descrise de legi putere (care sunt foarte comune în natură). Este foarte important de știut că rezultatul (odată observat într-o bază de numerație) are loc independent de baza de numerație în care se exprimă numerele, chiar dacă proporțiile de reprezentare se schimbă. De aici, **acest rezultat poate fi folosit pentru a verifica datele în suspiciunea de alterare (mistificare) a acestora prin compararea frecvențelor teoretice cu cele observate pentru prima cifră a acestora.**

Statistica Jarque-Bera. Testul Jarque-Bera [19, 20] calculează și atribuie probabilitatea statistică ca valorile unui eșantion ce provine din populație normal distribuită să își abată simultan asimetria și excesul de boltire de la valorile teoretice corespunzătoare distribuției normale. Statistica Jarque-Bera se calculează cu relația: $JB = n \cdot (g_1^2 + g_2^2/4)/6$, în care g_1 este asimetria, g_2 este excesul de boltire și n este volumul eșantionului. Statistica JB are o distribuție asimptotică către $\chi^2(df=2)$.

g_1 , Asimetria observabilei Y	Un eșantion	$g_1 = m_3/m_2^{3/2}$	$m_k = E_k(Y)$, $k>1$
b_2 , Boltirea observabilei Y		$b_2 = m_4/m_2^2$	
g_2 , Excesul de boltire al observabilei Y		$g_2 = b_2 - 3$	

¹⁶ R. A. Fisher, 1912. On an Absolute Criterion for Fitting Frequency Curves. Messenger of Mathematics 41:155-160.

¹⁷ BENFORD Frank. 1938. The law of anomalous numbers. Proceedings of the American Philosophical Society 78(4):551-572.

¹⁸ HILL Theodore P. 1995. Base invariance implies Benford's Law. Proceedings of the American Mathematical Society 123(3):887-895.

¹⁹ Carlos M JARQUE, Anil K BERA. 1980. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. Econ Lett 6(3):255-259.

²⁰ Carlos M JARQUE, Anil K BERA. 1981. Efficient tests for normality, homoscedasticity and serial independence of regression residuals: Monte Carlo evidence. Econ Lett 7(4):313-318.

Statistica Kolmogorov-Smirnov ([²¹], [²²]) verifică dacă observațiile independente $(x_i)_{1 \leq i \leq n}$ provin dintr-o populație ce urmează legea de distribuție dată de funcția cumulativă de probabilitate $CDF_i(x)$ prin calcularea maximumului diferenței absolute între $CDF_i(x)$ și funcția cumulativă de probabilitate observată $CDF_o(x)$ în toate punctele observației, și (v. [²³]):

$$D_n = \sup_x \left| CDF_i(x) - \frac{1}{n} \sum_{i=1}^n \begin{cases} 1, & x_i \leq x \\ 0, & x_i > x \end{cases} \right|, \text{KS} = \sqrt{n} \cdot D_n = \sqrt{n} \cdot \max_{1 \leq i \leq n} \left\{ CDF_i(x_i) - \frac{i-1}{n}, \frac{i}{n} - CDF_i(x_i) \right\}$$

Probabilitatea asociată valorii KS și respectiv volumului eșantionului n se obține din distribuția statisticii Kolmogorov-Smirnov. În convergență (pentru n suficient de mare) valoarea probabilității asociate statisticii se obține cu ajutorul formulelor:

$$D_n \xrightarrow{n \rightarrow \infty} D, CDF_{KS}(x) = P(x \leq D) = 1 - 2 \sum_{i=1}^{\infty} (-1)^{i-1} e^{-2i^2 x^2}, \alpha_{KS} = 1 - CDF_{KS}(D) = 2 \sum_{i=1}^{\infty} (-1)^{i-1} e^{-2i^2 D^2}$$

Pentru volume de eșantion mici, singura soluție este obținerea probabilității asociate dintr-un experiment de tip Monte-Carlo [²⁴].

Statistica Kuiper V [²⁵] este o variantă modificată a statisticii Kolmogorov-Smirnov:

$$KV = \sqrt{n} \cdot D_n = \sqrt{n} \cdot \left(\max_{1 \leq i \leq n} \left\{ CDF_i(x_i) - \frac{i-1}{n} \right\} + \max_{1 \leq i \leq n} \left\{ \frac{i}{n} - CDF_i(x_i) \right\} \right)$$

Probabilitatea asociată valorii KV și respectiv volumului eșantionului n se obține din distribuția statisticii Kuiper V, și pentru volume de eșantion mici, singura soluție este obținerea probabilității asociate dintr-un experiment de tip Monte-Carlo [²⁶].

Statistica Anderson-Darling AD se calculează asupra observațiilor ordonate crescător $(x_i \leq x_{i+1})$, pentru $1 \leq i < n$):

$$AD = -n - \sum_{i=1}^n \frac{2i-1}{n} (\ln(CDF_i(x_i)) + \ln(1 - CDF_i(x_{n+1})))$$

și testul Anderson-Darling [²⁷] verifică dacă este o evidență statistică ca un eșantion să nu provină dintr-o funcție de probabilitate dată. Pentru eșantioane mari, varianța statisticii A^2 depinde doar de volumul eșantionului ($[^{28}]$, $\pi=3.14159\dots$) și în această ipoteză valoarea probabilității de a observa o depărtare mai mare de A^2 se obține din distribuția student t, unde $v = n-1$ (dacă parametrii distribuției teoretice $CDF_i(x)$ sunt cunoscuți sau au fost determinați prin maximizarea șansei, MLE), sau $v = n-2$ dacă 1 parametru - de obicei media populației - a fost estimată din momente centrale (și așa mai departe):

$$\text{Var}(A^2) \cong \frac{2(\pi^2 - 9)}{3} + \frac{10 - \pi^2}{n}, \alpha_{AD} \cong 2 \cdot CDF_t(-A^2 / \sqrt{\text{Var}(A^2)}, v)$$

Pentru volume de eșantion mici, singura soluție este obținerea probabilității asociate dintr-un experiment de tip Monte-Carlo [²⁹].

Statistica Cramer-Mises se calculează cu formula ($[^{30}]$, [³¹]):

²¹ Andrey KOLMOGOROV, 1941. Confidence limits for an unknown distribution function. Ann Math Stat 12(4): 461-463.

²² Nikolay V SMIRNOV, 1948. Table for estimating the goodness of fit of empirical distributions. Ann Math Stat 19(2): 279-281.

²³ Aryeh DVORETZKY, Jack KIEFER, Jacob WOLFOWITZ, 1956. Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. Ann Math Stat 27(3): 642-669.

²⁴ Lorentz JÄNTSCHI, 2014. Kolmogorov - Smirnov statistic. <http://l.academicdirect.org/Statistics/tests/KS>

²⁵ Nicolaas H KUIPER, 1960. Tests concerning random points on a circle. Proc Koninklijke Nederlandse Akad van Wetenschappen Ser A 63: 38-47.

²⁶ Lorentz JÄNTSCHI, 2014. Kuiper V statistic. <http://l.academicdirect.org/Statistics/tests/KV>

²⁷ Theodore W ANDERSON, Donald A DARLING, 1952. Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes. Ann Math Stat 23(2): 193-212.

²⁸ Fritz W SCHOLZ, Michael A STEPHENS, 1986. K-Sample Anderson-Darling tests of fit, for continuous and discrete cases. Techn Rep. Univ Washington, GN-22: 81.

²⁹ Lorentz JÄNTSCHI, 2014. Anderson - Darling statistic. <http://l.academicdirect.org/Statistics/tests/AD>

³⁰ Harald CRAMÉR, 1928. On the composition of elementary errors. Scandinv Actuar J 1928(1): 13-74.

³¹ Richard von MISES, 1936. Probability, statistics and truth. Wien: Springer, 316 p. (ed. 4, 1972).

$$CM = \frac{1}{12n} + \sum_{i=1}^n \left(\frac{2i-1}{n} - CDF_i(x_i) \right)^2$$

iar valoarea probabilității asociate se poate obține dintr-un experiment de tip Monte-Carlo [32].

Statistica Watson U^2 este o variantă modificată a statisticii Cramer - von Mises [33]:

$$WU = \frac{1}{12n} + \sum_{i=1}^n \left(\frac{2i-1}{n} - CDF_i(x_i) \right)^2 - n \cdot \left(\frac{\sum_{i=1}^n CDF_i(x_i)}{n} - \frac{1}{2} \right)$$

iar valoarea probabilității asociate se poate obține dintr-un experiment de tip Monte-Carlo [34].

Statistica Chi-Square. Marele dezavantaj pe care îl posedă testul Chi-Square [35] este că presupune împărțirea domeniului variabilei aleatoare sau împărțirea domeniului de probabilitate într-un număr arbitrar de intervale și compararea frecvențelor observate cu frecvențele teoretice în aceste intervale, pentru a calcula statistica Chi-Square și respectiv probabilitatea asociată din distribuția Chi-Square:

$$X^2 = \sum_{k=1}^m \frac{(O_k - E_k)^2}{E_k}, \quad \alpha_{CS} = 1 - CDF_{\chi^2}(X^2, v)$$

Calculul valorilor așteptate (E_i) se bazează pe valorile observate, astfel încât creează o dependență care face ca numărul total al gradelor de libertate (m , în suma X^2) să se reducă cu o unitate ($v = m-1$). Dacă unul sau mai mulți parametri ai distribuției teoretice nu au fost cunoscuți și au fost estimați altfel decât prin maximizarea șansei de observare atunci numărul gradelor de libertate se diminuează și mai mult corespunzător cu numărul acestora (v , [36]).

Deoarece în practică se aplică o metodă euristică la calculul valorii X^2 , aceasta are drept consecință apariția erorilor de tipul I (respingerea incorectă a ipotezei de distribuție folosind o statistică X^2 mai mare decât valoarea sa adevărată).

Statistica Fisher Chi-Square. Întrucât dezavantajul major al statisticii Chi-Square este numărul arbitrar de împărțire a intervalului valorilor observate în subintervale, odată eliminat acest dezavantaj suntem în posesia unui instrument foarte puternic de discriminare. Într-adevăr, Fisher propune utilizarea statisticii Chi-Square pentru a calcula probabilitatea de observare a unui fenomen observat în două serii de observații independente [37]. Similar se poate proceda pentru statistici calculate independent [38], singura enigmă rămânând în acest caz expresia de calcul a numărului gradelor de libertate (n_{χ^2}) în funcție de numărul statisticilor calculate independent (n_{st}), numărul gradelor de libertate ale distribuției teoretice testate (n_{td}) și numărul observațiilor experimentale (n_{ob}):

$$FCS = - \sum_{i=1}^{n_{st}} \ln(p_i), \quad \alpha_{FCS} = 1 - CDF_{\chi^2}(FCS, n_{\chi^2}), \quad n_{\chi^2} = n_{\chi^2}(n_{st}, n_{td}, n_{ob})$$

unde, în cel mai pesimist scenariu, numărul gradelor de libertate este $n_{\chi^2} = \min(n_{st}, n_{td}, n_{ob})$.

Măsuri ale tendinței centrale. Fie X un șir de n valori X_1, X_2, \dots, X_n . Pentru calculul valorii medii, următorii indicatori sunt cei mai folosiți:

÷ Media Aritmetică, $AM(X)$, dată de:

$$AM(X) = \sum_{i=1}^n X_i / n$$

³² Lorentz JÄNTSCH, 2014. Cramer - von Mises statistic. <http://academicdirect.ro/Statistics/tests/CM>

³³ Geoffrey S WATSON, 1961. Goodness-of-fit tests on a circle. Biometrika 48(1-2): 109-114.

³⁴ Lorentz JÄNTSCH, 2014. Watson U^2 statistic. <http://academicdirect.ro/Statistics/tests/WU>

³⁵ Karl PEARSON, 1900. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Phil Mag Ser 5 50(302): 157-175.

³⁶ Ronald A FISHER, 1922. On the interpretation of χ^2 from contingency tables, and the calculation of p. J R Stat Soc 85: 87-94.

³⁷ Ronald A FISHER, 1948. Combining independent tests of significance. Am Stat 2(5): 30.

³⁸ Sorana D BOLBOACĂ, Lorentz JÄNTSCH, Adriana F SESTRĂȘ, Radu E SESTRĂȘ, Doru C PAMFIL, 2011. Pearson-Fisher chi-square statistic revisited. Information 2(3): 528-545.

- ÷ Media Geometrică, $GM(X)$, obținută din expresia (de notat că pentru n impar, expresia pentru GM este totdeauna determinată, iar când produsul $\prod X_i$ este negativ și n este par valoarea obținută este complexă):

$$GM(X) = \sqrt[n]{\prod_{i=1}^n X_i} = \exp(AM(\ln(X)))$$

- ÷ Media Armonică, $HM(X)$, dată de:

$$HM(X) = n / \sum_{i=1}^n 1/X_i = 1 / AM(1/X)$$

- ÷ Media lui Euler, $EM(X)$, calculată ca:

$$EM(X) = \sqrt{\sum_{i=1}^n X_i^2 / n} = \sqrt{AM(X^2)}$$

- ÷ Valoarea mediană, $m(X)$, este numărul (π a.î. X_π este șir ordonat):

$$m(X) = \begin{cases} \left(X_{\pi(\frac{n}{2})} + X_{\pi(\frac{n}{2}+1)} \right) / 2, & n \text{ par} \\ X_{\pi(\frac{n+1}{2})}, & n \text{ impar} \end{cases}$$

- ÷ Valorile la modă sunt din mulțimea:

$$\tilde{X} = \{X_i \mid f_i = \sup\{f_j, 1 \leq j \leq n\}, f_j = |\{x \in X \mid x = X_j\}|\}$$

Prelucrarea textului și a imaginilor

În cadrul laboratorului se parcurg următoarele resurse; se efectuează operații de editare, salvare, închidere și deschidere cu acestea; se salvează fișierele rezultat relevante.

Activitate

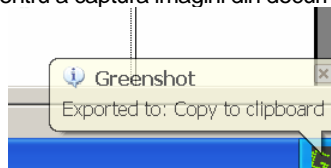
Resursă: GreenShot

Domeniu de interes: capturi de ecran

Se descarcă și se instalează pe calculator de la adresa: <http://getgreenshot.org/>

Facilități: poate captura sub formă de imagine din orice tip de document deschis pe calculator indiferent de protecțiile de securitate ale acestuia.

Operații: Se utilizează programul pentru a captura imagini din documente PDF securizate



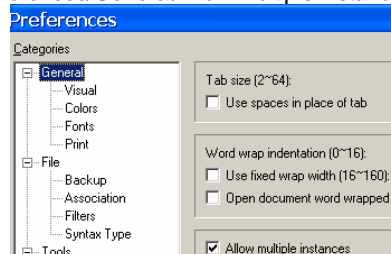
Resursă: Crimson Editor

Domeniu de interes: editare text; copiere din pagini web în mod text neformatat

Se descarcă și se instalează pe calculator de la adresa: <http://www.crimsoneditor.com/>

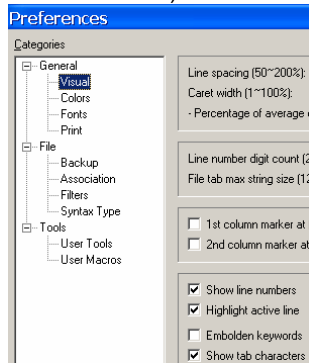
Facilități:

- ÷ Sesiuni multiple: Tools/Preferences/General/Allow multiple instances

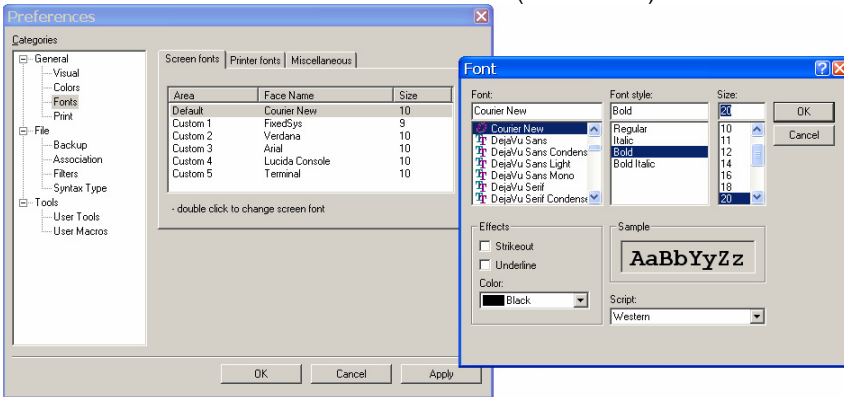


- ÷ Numerotarea liniilor din pagină: Tools/Preferences/General/Visual/Show line numbers

÷ Afișarea Tab-ului (separator de coloană în Excel): -"/Show tab characters

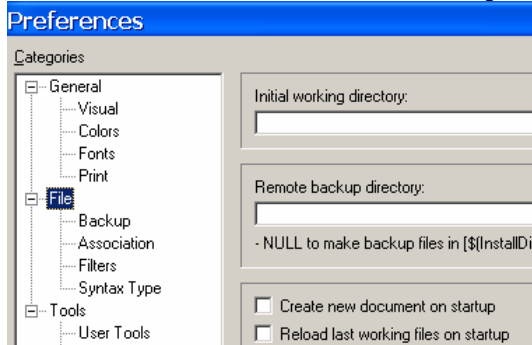


÷ Dimensiunea literelor: Tools/Preferences/General/Fonts/(Screen fonts)/Default/Size → 20



÷ Configurarea inițializării: Tools/Preferences/File/☑Create new document on startup

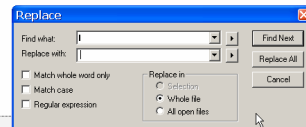
÷ Configurarea inițializării: Tools/Preferences/File/☑Reload last working files ...



Operații:

÷ Înlocuirea automată a spațiilor cu tab-uri: se copiază un tabel de pe o pagină web, se apasă tab-ul, se taie apoi acesta cu foarfecea, se selectează caracterul de înlocuit (unul sau mai multe spații consecutive) și se accesează interfața de înlocuire: Search/Replace/Replace all

1	Id	r2	t0	v0	v1	v2	v3	TP	TO	TM	SP	ST	SD	VP	VT	VD
2	1	0.85	25.6	9.91	10.2	1.84	7.71	1	0	0	1	0	0	1	0	0
3	2	0.66	25.1	9.88	12.4	2.02	9.18	1	0	0	1	0	0	0	0	1
4	3	0.57	24.2	9.95	29.8	1.8	22.6	1	0	0	1	0	0	0	0	1
5	4	0.92	32.6	9.94	8	1.82	6.18	1	0	0	1	0	0	1	0	0
6	5	0.95	24.6	9.94	10.2	1.81	9.16	1	0	0	1	0	0	1	0	1
7	6	0.79	33.5	10.1	12	1.83	6.81	1	0	0	1	0	0	1	0	1



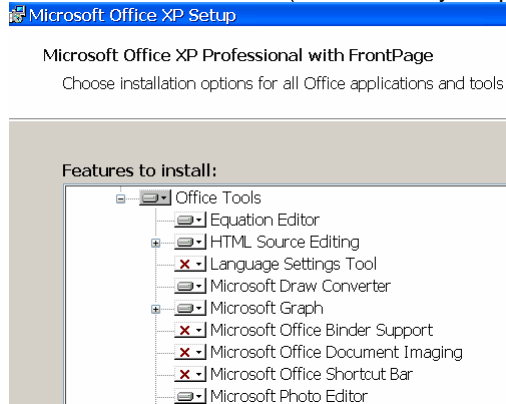
Resursă: MS Word

Domeniu de interes: editor de documente

Versiuni: http://en.wikipedia.org/wiki/Microsoft_Word

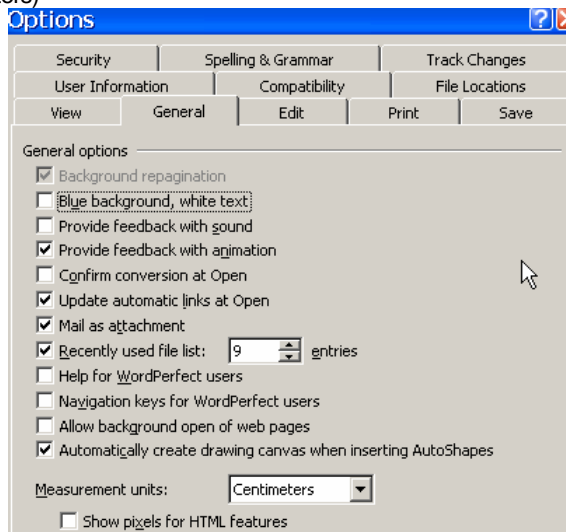
Facilități MS Word XP ("Office XP", v. 10.0 din 2001):

- ÷ Editare de ecuații matematice: nu e activă la instalarea implicită; se activează prin 'instalare' (sau 'reinstalare' pe calea: Control Panel/Add or Remove Programs/Microsoft Office XP/Change/Add or Remove Features/Features to install):
 - Office Tools/Equation Editor (Run all from my computer)
- ÷ Editare de imagini: nu e activă la instalarea implicită; se activează prin 'instalare' (sau 'reinstalare' pe calea: Control Panel/Add or Remove Programs/Microsoft Office XP/Change/Add or Remove Features/Features to install):
 - Office Tools/Microsoft Photo Editor (Run all from my computer)

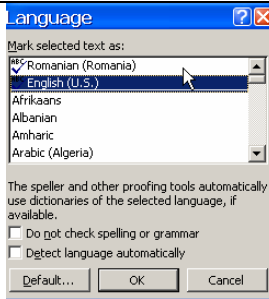


÷ Afișarea marginilor textului în pagină: (MS Word)/Tools/Options.../View/Text boundaries

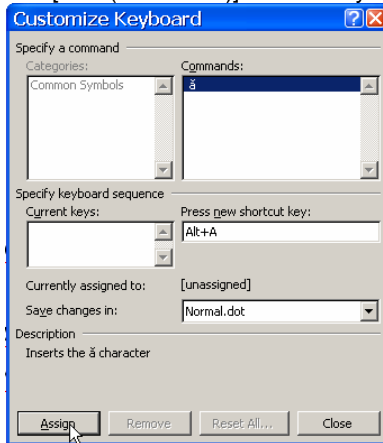
÷ Schimbarea unității de măsură în pagină: (MS Word)/Tools/Options.../General/Mesurement units (Inches↓Centimeters)



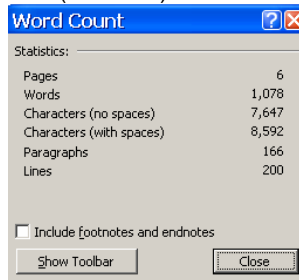
÷ Schimbarea limbii implicite de corectare: (MS Word)/Tools/Language/Set Language .../Romanian (Romania)



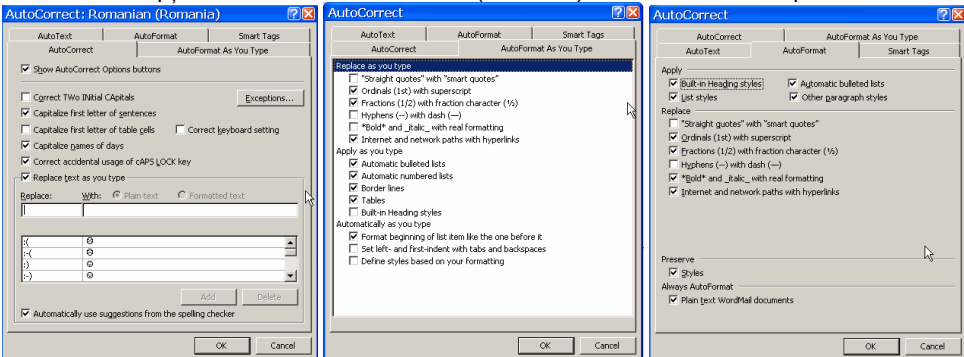
⇒ Instalarea scurtăturilor de taste pentru caractere românești: (MS Word)/Insert/Symbol.../Symbols/(Font:(normal text))Shortcut Key/Press new shortcut key



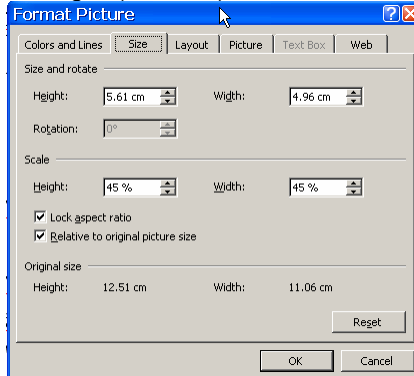
⇒ Numărarea cuvintelor din document: (MS Word)/Tools/Word count...



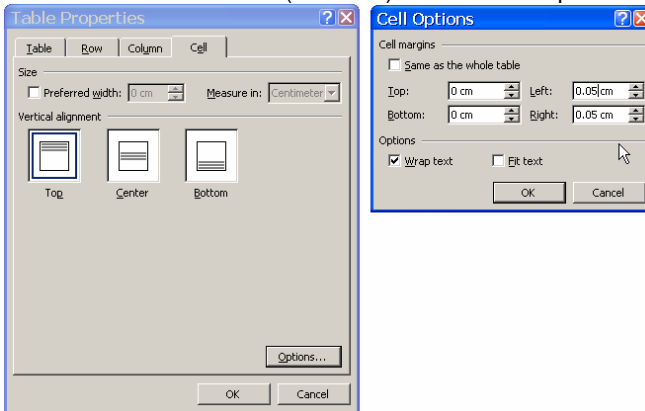
⇒ Schimbarea opțiunilor de corectare automată: (MS Word)/Tools/AutoCorrect Options...



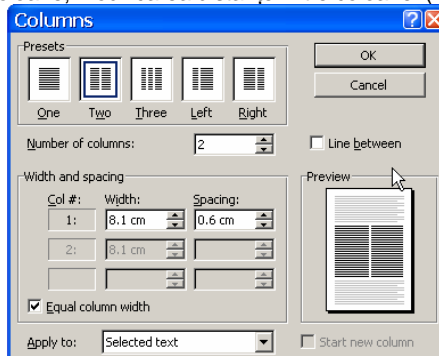
÷ Modificarea dimensiunii unei imagini: (MS Word)/Format/Picture.../Size



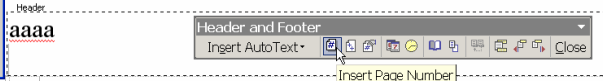
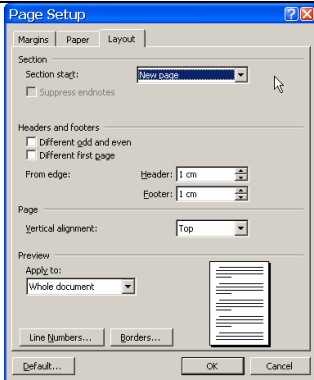
÷ Modificarea formatului celulelor unui tabel: (MS Word)/Table/Table Properties/Cell/Cell Options



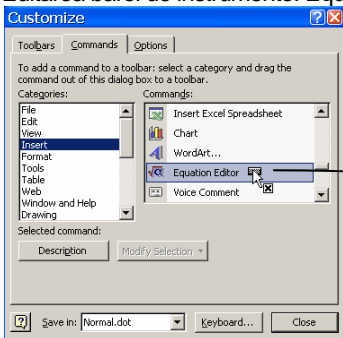
÷ Scrierea pe mai multe coloane; modificarea distanței între coloane: (MS Word)/Format/Columns...



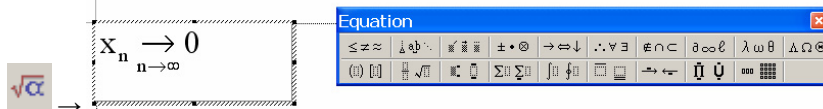
÷ Editarea capului și piciorului paginii: (MS Word)/File/Page Setup.../Layout & (MS Word)/View/Header and Footer



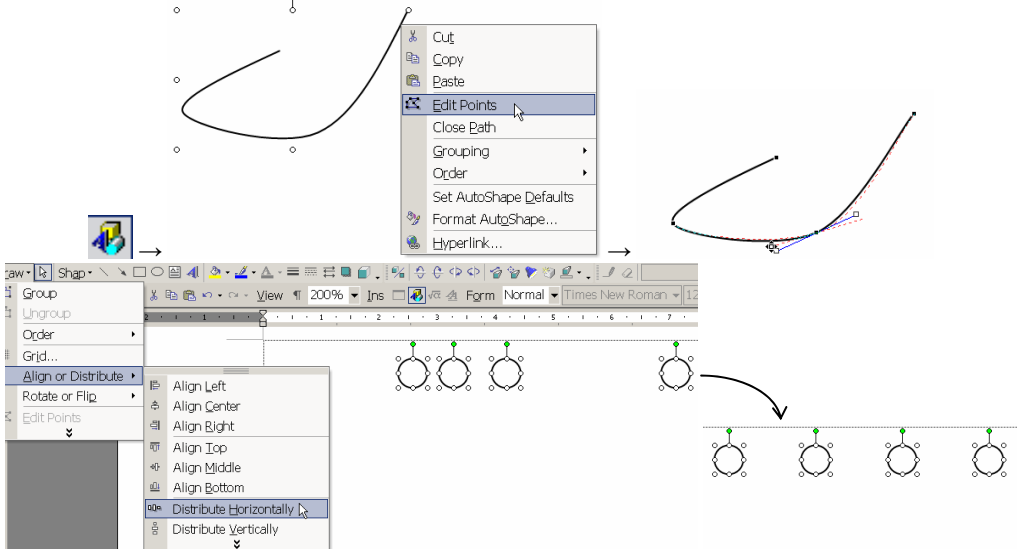
✦ Editarea barei de instrumente: Equation Editor: (MS Word)/Tools/Customize.../



✦ Utilizarea editorului de ecuații:

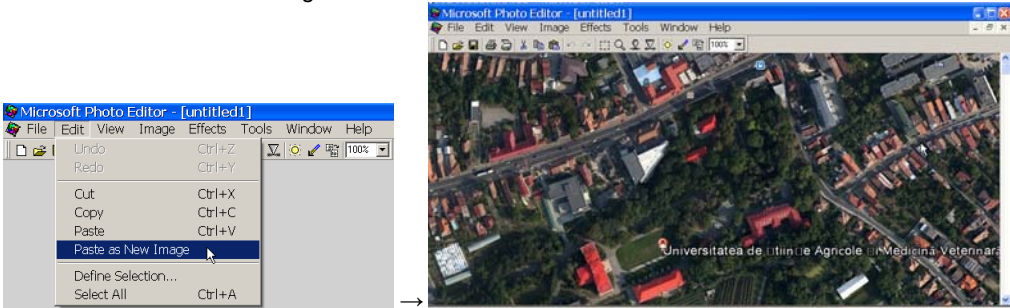


✦ Utilizarea editorului de desene:

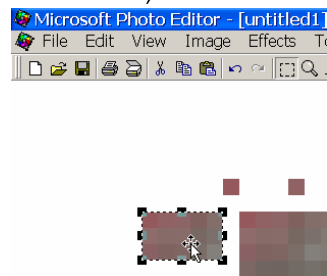
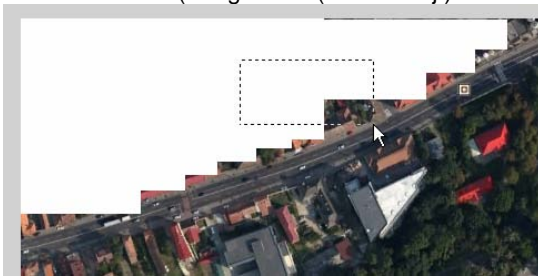


Resursă: MS Photo Editor

÷ Utilizarea editorului de imagini: Microsoft Office Tools ► Microsoft Photo Editor



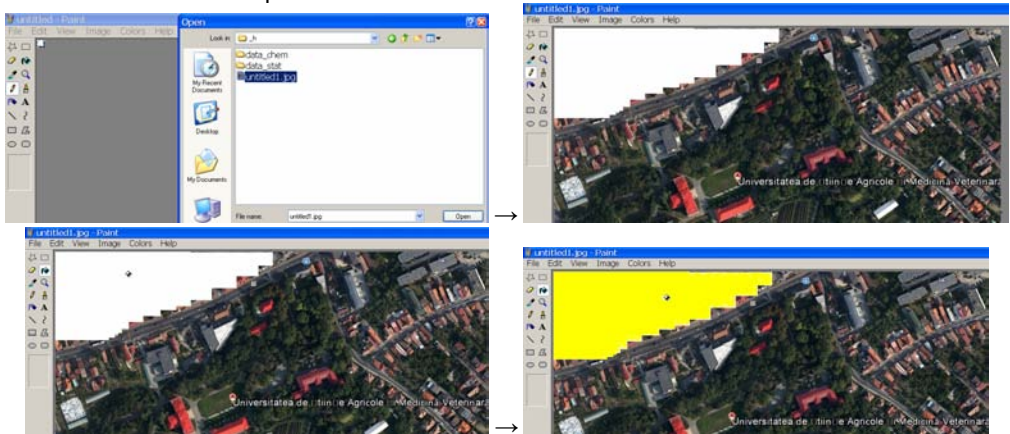
(Google Earth("usamv cluj") → Greenshot → PhotoEditor)



(decuparea imaginii - selecție și tăiere) și (editarea la nivel de pixel)

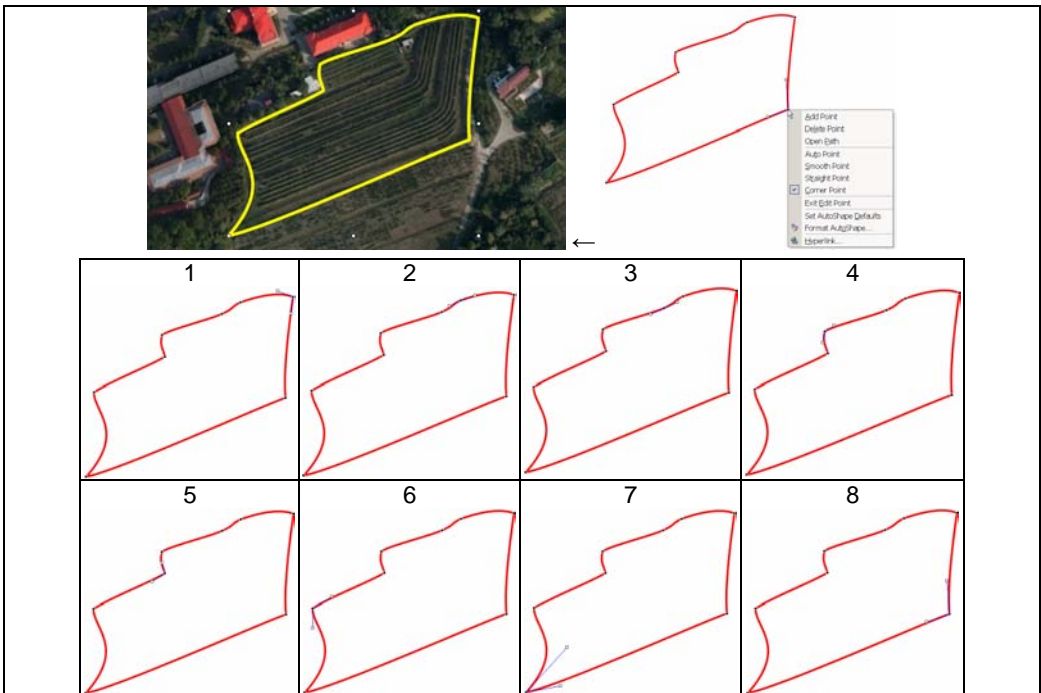
Resursă: MS Paint

÷ Utilizarea editorului de poze: Accesories ► Paint



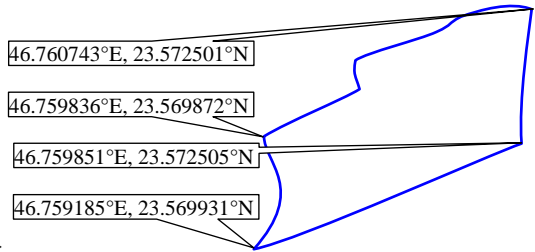
Resursă: MS Draw (MS Word Drawing)

÷ Delimitarea unui teren pe o imagine: în MS Word Drawing



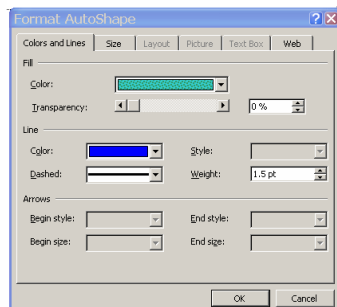
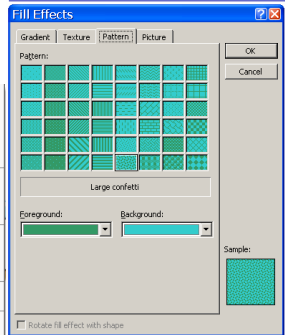
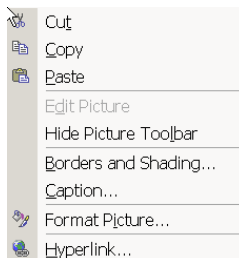
Resursă: Google Earth

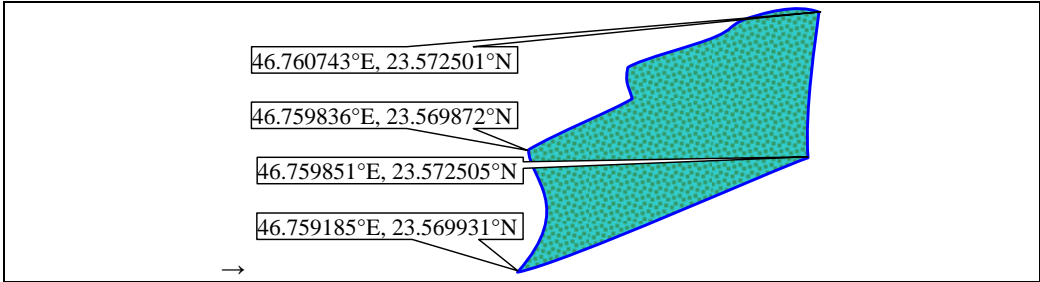
÷ Delimitarea coordonatelor locației din Google Earth



Resursă: MS Word

÷ Colorarea conținutului delimitat de un contur (MS Word)





Calcul tabelar

Se parcurg în cadrul laboratorului următoarele resurse; se efectuează operații de editare, salvare, închidere și deschidere cu acestea; se salvează fișierele rezultat relevante.

Nr | Activitate

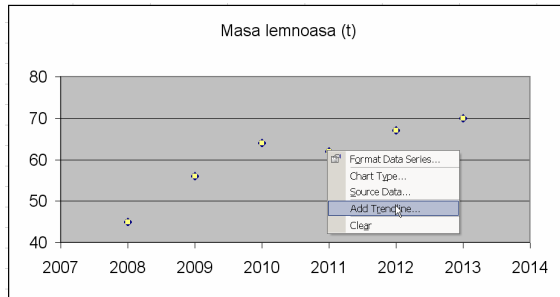
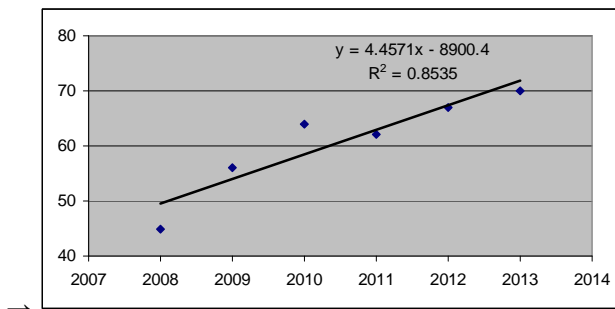
1 Resursă: MS Excel
 Domeniu de interes: calcul tabelar
 Versiuni: http://en.wikipedia.org/wiki/Microsoft_Excel
 Facilități MS Word XP ("Office XP", v. 10.0 din 2001):

- ÷ Module adiționale pentru calcule (inexistente fără 'activare'): (MS Excel)/Tools/Add-Ins...

('Analysis ToolPak' și 'Solver Add-in')

- ÷ Efectuarea unei **analize de tendință**:

	A	B
1	An	Masa lemnoasa (t)
2	2008	45
3	2009	56
4	2010	64
5	2011	62
6	2012	67
7	2013	70

÷ **Analiza modelului de regresie:** Tools/Data Analysis.../Regression/

	A	B	C	D	E	F	G
1	An	Masa lemnoasa (t)					
2	2008		45				
3	2009		56				
4	2010		64				
5	2011		62				
6	2012		67				
7	2013		70				
8							
9							
10							
11							

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.9238
R Square	0.8535
Adjusted R Square	0.8169
Standard Error	3.8625
Observations	6

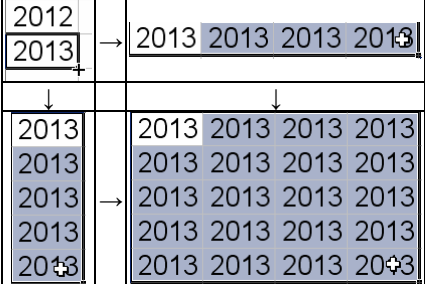
ANOVA					
	df	SS	MS	F	Significance F
Regression	1	347.6571	347.6571	23.3029	0.00848
Residual	4	59.6762	14.9190		
Total	5	407.3333			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-8900.42	1856.33	-4.79	0.0087	-14054.43	-3746.40
An	4.4571	0.9233	4.8273	0.0085	1.8936	7.0207

- ÷ Modelul de regresie este valid când:
- "Significance F" < 0.05 - e acceptată ipoteza unei asocieri liniare între "An" și "Masa lemnoasă"
 - P-value("Intercept") < 0.05 - e acceptată valoarea "Intercept" ca fiind o valoare nenulă cel mai probabil dată de calculul al cărui rezultat este redat în Coefficients("Intercept"); dacă P-value("Intercept") ≥ 0.05 atunci trebuie să se caute un model liniar fără acest coeficient ("") la "Constant is zero" în fereastra "Regression" și respectiv la "Set intercept =" în "Add Trendline"
 - P-value("An") < 0.05 - este acceptată variabila "An" ca fiind o variabilă nenulă în contribuția sa la varianța observată pentru "Masa lemnoasă"; dacă P-value("An") ≥ 0.05 atunci trebuie să se caute un model liniar ce nu include această variabilă - situație posibilă numai când modelul este 'liniar multiplu' și variabila dependentă de interes (în cazul de mai sus 'masa lemnoasă' este modelată în funcție de mai mult de o caracteristică independentă (în cazul de mai sus a fost considerată o singură caracteristică independentă, 'An')

Operații în MS Excel:

÷ Extinderea unei valori de la o celulă la celule adiacente:



÷ Extinderea unei formule la celulele adiacente:

	A		A		A	B	C	D	
1	2012	1	2012	→	1	2012	2013	2014	2015
2	=A1+1	2	2013		2	2013	2014	2015	2016

	A
1	2012
2	2013

	A
1	2012
2	2013
3	2014
4	2015
5	2016

÷ Fixarea coloanei

	A		A		A	B	C	D	
1	2012	1	2012	→	1	2012	2013	2014	2015
2	=\$A1+1	2	2013		2	2013	2013	2013	2013

	A
1	2012
2	2013

	A
1	2012
2	2013
3	2014
4	2015
5	2016

÷ Fixarea liniei

	A		A		A	B	C	D	
1	2012	1	2012	→	1	2012	2013	2014	2015
2	=\$A1+1	2	2013		2	2013	2014	2015	2016

	A
1	2012
2	2013

	A
1	2012
2	2013
3	2013
4	2013
5	2013

÷ Fixarea liniei și a coloanei:

	A		A
1	2012	1	2012
2	=A\$1+1	2	2013

→

	A	B	C	D
1	2012	2013	2014	2015
2	2013	2013	2013	2013

	A
1	2012
2	2013

↓

	A
1	2012
2	2013
3	2013
4	2013
5	2013

	A	B	C	D
1	2012	2013	2014	2015
2	2013	2013	2013	2013
3	2013	2013	2013	2013
4	2013	2013	2013	2013
5	2013	2013	2013	2013

Analiza statistică a efectelor multiplicative

Calculul valorilor așteptate ($E_{i,j}$), a valorilor χ^2 și a probabilităților asociate se face din formulele [39]:

$$E_{i,j} = \frac{\left(\sum_{j=1}^n O_{i,j}\right) \cdot \left(\sum_{i=1}^m O_{i,j}\right)}{\left(\sum_{i=1}^m \sum_{j=1}^n O_{i,j}\right)}, \quad \chi^2_{i,j} = \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}}, \quad p_{total} = p_{\chi^2} \left(\sum_{i=1}^m \sum_{j=1}^n X_{i,j}^2, (m-1) \cdot (n-1)\right),$$

$$p_{linia "i"} = p_{\chi^2} \left(\sum_{j=1}^n X_{i,j}^2, (n-1)\right), \quad p_{coloana "j"} = p_{\chi^2} \left(\sum_{i=1}^m X_{i,j}^2, (m-1)\right)$$

÷ Date observate (O_{ij} , Kg/m2/an lemn):

↓sol/tratament→	udare	fertilizareN	fertilizareP
argilos	140	160	80
cernoziom	260	205	185
nisipos	120	128	100

→

"O"	b1	b2	b3
a1	140	160	80
a2	260	205	185
a3	120	128	100

÷ Sume valori observate ($\sum_i O_{i,j}$, $\sum_j O_{i,j}$, $\sum_{i,j} O_{i,j}$):

	A	B	C	D	E
1 "O"	b1	b2	b3	sum	
2 a1		140	160	80	=SUM(B2:D2)
3 a2		260	205	185	↓
4 a3		120	128	100	...
5 sum	=SUM(B2:B4)	→

→

	A	B	C	D	E
1 "O"	b1	b2	b3	sum	
2 a1	140	160	80	380	
3 a2	260	205	185	650	
4 a3	120	128	100	348	
5 sum	520	493	365	1378	

÷ Valori așteptate ($E_{i,j}$):

A..F	G	H	I	J
1 ...	"E" b1	b2	b3	
2 ...	a1	=E2*B\$5/\$E\$5	→	...
3 ...	a2		↓	...
4 ...	a3	

→

A..F	G	H	I	J
1 ...	"E"	b1	b2	b3
2 ...	a1	143.4	136.0	100.7
3 ...	a2	245.3	232.5	172.2
4 ...	a3	131.3	124.5	92.2

÷ Erori pătratice ($X^2_{i,j}$):

³⁹ Ronald A. FISHER, 1923. Studies in Crop Variation. II. The Manurial Response of Different Potato Varieties. Journal of Agricultural Science 13:311-320.

A..K	L	M	N	O	A...K	L	M	N	O
1 ...	"X2" b1		b2 b3		1 ...	"X2"	b1	b2	b3
2 ...	a1	=(B2-H2)^2/H2	→ ...	→	2 ...	a1	0.08	4.25	4.24
3 ...	a2		↓	3 ...	a2	0.88	3.26	0.96
4 ...	a3		4 ...	a3	0.98	0.10	0.66

÷ Sume pătratice ($\sum_i X_{ij}^2$, $\sum_j X_{ij}^2$, $\sum_i X_{ij}^2$):

A..K	L	M	N	O	P	A...K	L	M	N	O	P
1 ...	"X2" b1		b2 b3	sum		1 ...	"X2"	b1	b2	b3	sum
2 ...	a1		0.08 4.25 4.24	=SUM(M2:O2)		2 ...	a1	0.08	4.25	4.24	8.57
3 ...	a2		0.88 3.26 0.96	↓		3 ...	a2	0.88	3.26	0.96	5.10
4 ...	a3		0.98 0.10 0.66	...		4 ...	a3	0.98	0.10	0.66	1.74
5 ...	sum	=SUM(M2:M4)	→		5 ...	sum	1.94	7.62	5.86	15.4

÷ Probabilități de asociere a factorilor (de efect multiplicativ al factorilor, $O_{ij} \sim a_i \cdot b_j$):

A..K	L	M	N	O	P	Q
1 ...	"X2" b1		b2 b3	sum	pX2	
2 ...	a1		0.08 4.25 4.24	8.57	=chidist(p2,count(m2:o2)-1)	
3 ...	a2		0.88 3.26 0.96	5.10	↓	
4 ...	a3		0.98 0.10 0.66	1.74	...	
5 ...	sum		1.94 7.62 5.86	15.4		
6 ...	pX2	=chidist(m5,count(m2:m4)-1)	→ ...		=chidist(p5,(count(m6:o6)-1)*(count(q2:q4)-1))	

A..K	L	M	N	O	P	Q
1 ...	"X2" b1	b2	b3	sum	pX2	
2 ...	a1	0.08	4.25	4.24	8.57	0.014
3 ...	a2	0.88	3.26	0.96	5.10	0.078
4 ...	a3	0.98	0.10	0.66	1.74	0.419
5 ...	sum	1.94	7.62	5.86	15.4	
6 ...	pX2	0.379	0.022	0.053		0.004

÷ Interpretare test:

Singurii factori asupra cărora nu se poate respinge ipoteza de efect multiplicativ sunt pentru b1 ("udarea", $p = 37.9\%$) și b3 ("fertilizarea cu P", $p = 5.3\%$) diferitelor tipuri de soluri și respectiv aplicarea diferitelor tratamente asupra a2 ("cernoziom", $p = 7.8\%$) și a3 ("nisipos", $p = 41.9\%$). Asupra ansamblului general de factori "tratament" vs. "tip sol" este însă respinsă ipoteza de efect multiplicativ al celor 2 factori ($p = 0.4\%$, "distinct semnificativ statistic", $p < 1\%$) asupra producției de lemn (în $\text{Kg}/\text{m}^2/\text{an}$) pe baza observațiilor experimentale (aici au fost considerate valori ipotetice) colectate din teren.

Proiectarea experimentelor factoriale și interpretarea statistică

Experimentele factoriale vizează identificarea cantitativă a influenței unuia sau mai multor factori a căror prezență și manifestare este cunoscută în prezența unei posibile serii de alți factori a căror prezență și manifestare este necunoscută însă este întâmplătoare.

În acest sens se încadrează:

- ÷ O serie de experimente desfășurate în condiții naturale (unde factorii întâmplători ce pot influența rezultatele experimentelor sunt temperatura, lumina, precipitațiile), pentru care exemplul reprezentativ este influența tratamentului și/sau a soiului la creșterea plantelor în anumite condiții naturale (definite de o locație geografică);
- ÷ O serie de experimente desfășurate în condiții similare celor naturale (unde factorii întâmplători ce pot influența rezultatele experimentelor pot fi din nou temperatura, lumina și umiditatea dar și factorul uman sau echipamentele de lucru implicate), pentru care exemplul reprezentativ este influența tratamentului și/sau a tipului de material prelucrat într-o hală în care sunt amplasate mașini de prelucrare (cum ar fi prin așchiere);
- ÷ O serie de experimente care în ciuda faptului că sunt desfășurate în condiții de laborator, prin natura lor folosesc elemente naturale (sol sau material biologic), cum sunt experimentele pe culturi microbiologice ce vizează identificarea unor antibiotice care oferă cel mai bun efect de tratament;
- ÷ În fapt în orice șir de experimente ce vizează identificarea unui set de valori pentru unul, doi, sau

mai mulți factori când se știe că factorii (dacă sunt cel puțin doi) influențează multiplicativ rezultatul observației;

În continuare pentru a păstra rigurozitatea raționamentului și consecvența terminologiei, se va purta discuția având ca reper experimentele desfășurate în condiții naturale.

Fenomenele naturale (precum creșterea și dezvoltarea materialului vegetal) sunt influențate de o serie de **factori**. Asupra unora dintre acești factori putem exercita control aceștia fiind numiți **factori controlabili**. Este implicit aici că pentru un factor controlabil avem la dispoziție o serie de valori pe care le putem fixa acestor factori.

În categoria factorilor controlabili pot intra, în cazul unui studiu ce implică cea mai bună alegere dintr-o serie de alternative în ceea ce privește materialul (propagat) și/sau tratamentul (aplicat): alegerea materialului sau alegerea tratamentului - aceste alegeri sunt variabilele factor (controlabil) în condițiile existente (într-o anumită locație geografică caracterizată de o serie de valori medii pentru temperatură și precipitații), de asemenea putând fi legată de efectuarea unei serii de alte intervenții (alte tratamente) înainte sau pe parcursul propagării materialului biologic.

În acest exemplu (precum și în celelalte situații similare), fiecare dintre aceste intervenții poate fi privită ca un factor (fiecare **eșantion** de **material** supus aceluiași **tratament** va fi definit de un șir de valori, unde fiecare valoare atribuită fiecărui factor controlabil al seriei corespunde unui 'tratament'). În același timp și o serie de tratamente poate fi privită ca un factor (și în acest caz fiecare **eșantion** de material supus aceleiași succesiuni de tratamente, urmând aceeași cronologie va fi caracterizat de o valoare atribuită factorului controlabil 'seria de tratamente'). Asupra unora dintre acești factori nu putem exercita control cu costuri rezonabile și pentru aceștia experimentele care se vor desfășura pot doar să evalueze magnitudinea influenței acestor **factori incontrolabili** în valoarea unei caracteristici urmărite.

Un **experiment** vizează identificarea valorilor unor factori controlabili în care se obțin cele mai bune rezultate (rezultate definite de o funcție de măsurare) în propagarea (sau tratamentul) materialului. Întrucât întotdeauna există și o serie de factori incontrolabili (și variabilitatea naturală este doar unul dintre aceștia) este foarte important să se asigure condițiile ca cel puțin acei factori incontrolabili care pot avea o contribuție majoră să aibă o **influență omogenă**. Un exemplu de astfel de factor incontrolabil care poate avea o contribuție majoră este compoziția solului sau materialului, care poate varia longitudinal și transversal. Acesta este motivul pentru care într-un experiment care vizează identificarea celui mai bun material biologic de propagat, având de ales din două sau mai multe alternative, pentru a neutraliza efectele de compoziție (cu latitudinea și longitudinea sau cu lungimea și lățimea) se vor realiza mai multe **eșantioane**, urmând o **schemă de repetiții**. Un **eșantion** este o unitate din câmpul (sau materialul) pe care se desfășoară experimentul care este supusă aceleiași intervenții (care poate referi aplicarea unui tratament, unei succesiuni de tratamente, sau utilizarea unui anume soi sau unui anume cuțit pentru tăiere) și asupra căreia se realizează o măsurătoare (sau mai multe măsurători) și rezultatele sunt înregistrate distinct, păstrând evidența de proveniență a măsurătorii. Un **experiment factorial** este un experiment proiectat și desfășurat pe **eșantioane** având definiți unul sau mai mulți factori controlabili (fie F_1, \dots, F_n acești factori). Pentru fiecare dintre acești factori există controlul asupra unui șir de valori (fie pentru F_1 valorile $v_{1,1}, v_{1,2}, \dots, v_{1,m_1}, \dots$, pentru F_n valorile $v_{3,1}, v_{3,2}, \dots, v_{n,m_n}$). O **variantă** este o alegere unică de valori pentru factori. Pentru F_1, \dots, F_n factori o variantă este oricare șir de valori $(v_{1,i_1}, \dots, v_{n,i_n})$ unde $1 \leq i_1 \leq m_1, \dots, 1 \leq i_n$. Pentru cei F_n factori există exact $m_1 \cdot \dots \cdot m_n$ variante. O **repetiție** este un șir sau o matrice de **eșantioane** care cuprinde o singură dată fiecare variantă. Experimentul factorial este format din mai multe repetiții, urmând o **schemă de repetiții** cum sunt blocuri randomizate, pătrate latine, dreptunghiuri latine sau grilaje balansate. Oricare însă ar fi această schemă de repetiții, ea vizează totdeauna neutralizarea efectelor longitudinale și transversale, așa cum este exemplificat în continuare.

Fie propagarea unui material biologic într-un câmp pentru care se vizează alegerea celei mai potrivite din două opțiuni posibile (fie aceste opțiuni A și B). Pentru acestea se realizează un amplasament (în câmp) în forma redată mai jos. Fie un efect al latitudinii definit de factorul multiplicativ x și un efect al longitudinii definit de factorul multiplicativ y . Pentru această configurație de opțiuni (A și B) și factori (x și y) este urmată o analiză de variație (v. **Schemă de repetiție vs. efecte longitudinale și transversale**).

÷ Amplasament în câmp

A	B
B	A

÷ Efecte de latitudine (x) și longitudine (y)

1	x
y	xy

÷ Efecte observate și sume (pe linii, pe coloane, total)

1	x	1+x
y	xy	y+xy
1+y	x+xy	1+x+y+xy

÷ Efecte așteptate (v. calculul valorilor așteptate ($E_{i,j}$))

$\frac{(1+x)(1+y)}{(1+x+y)} = 1$	$\frac{(1+x)x(1+y)}{(1+x+y)} = x$
$\frac{y(1+x)(1+y)}{(1+x+y)} = y$	$\frac{y(1+x)x(1+y)}{(1+x+y)} = xy$

÷ Efectul latitudinii (x) și longitudinii (y) asupra culturilor amplasate în câmp: efectul opțiunii "A" este estimat în caracteristica măsurată (Z) sub forma valorii "a" (de exemplu "a" kg/ha masă lemnoasă pentru cultura soiului "A"); efectul opțiunii "B" este estimat în caracteristica măsurată (Z) sub forma valorii "b" (în același exemplu atunci "b" sunt kg/ha masă lemnoasă pentru cultura soiului "B"); ambele opțiuni (A și B) sunt afectate (efecte multiplicative) de latitudine (x) și longitudine (y) astfel încât pentru amplasamentul din câmp dat, în funcție de poziția parcelelor, efectele factorilor latitudine (x) și longitudine (y) în caracteristica măsurată Z sunt:

a	xb
yb	xya

÷ Efecte estimate și sume (pe linii, pe coloane, total):

a	xb	a+xb
yb	xya	yb+xya
a+yb	xb+xya	a+xb+yb+xya

÷ Efecte așteptate (v. calculul valorilor așteptate ($E_{i,j}$))

$\frac{(a+xb)(a+yb)}{(a+xb+yb+xya)}$	$\frac{(a+xb)x(b+ya)}{(a+xb+yb+xya)}$
$\frac{y(b+xa)(a+yb)}{(a+xb+yb+xya)}$	$\frac{y(b+xa)x(b+ya)}{(a+xb+yb+xya)}$

÷ Diferențe între efecte estimate și așteptate:

$a - \frac{(a+xb)(a+yb)}{(a+xb+yb+xya)} = D_{1,1}$	$xb - \frac{(a+xb)x(b+ya)}{(a+xb+yb+xya)} = D_{1,2}$
$yb - \frac{y(b+xa)(a+yb)}{(a+xb+yb+xya)} = D_{2,1}$	$xya - \frac{y(b+xa)x(b+ya)}{(a+xb+yb+xya)} = D_{2,2}$

÷ Evaluarea diferențelor:

$D_{1,1} = \dots = \frac{xy(aa-bb)}{a+xb+yb+xya}$	$D_{1,2} = \dots = \frac{xy(bb-aa)}{a+xb+yb+xya}$
$D_{2,1} = \dots = \frac{xy(bb-aa)}{a+xb+yb+xya}$	$D_{2,2} = \dots = \frac{xy(aa-bb)}{a+xb+yb+xya}$

Schemă de repetiție vs. efecte longitudinale și transversale

Așa cum se observă (v. **Schemă de repetiție vs. efecte longitudinale și transversale**) diferențele între valorile așteptate și cele estimate (a fi observate) depind în egală măsură de efectul latitudinii (x) și longitudinii (y) cumulated diferențe pătrăcite între opțiuni (a și b).

Schema ilustrată mai sus (v. **Schemă de repetiție vs. efecte longitudinale și transversale**) ilustrează rolul proiectării experimentului în modalitatea de prelucrare și interpretare a rezultatelor obținute în urma experimentelor factoriale în câmp.

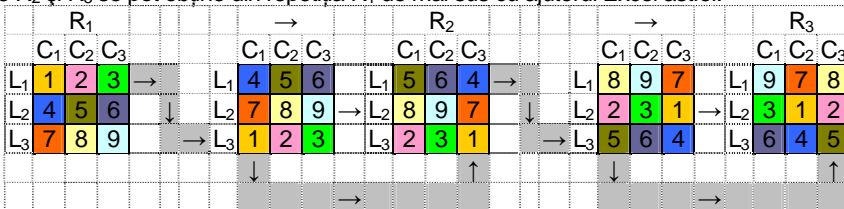
Fie un experiment factorial în câmp în care se urmărește influența factorului "variantă" (varietate, soi, tratament) cu privire la "producție" (în tone/hectar) folosind o schemă de repetiție cu trei repetiții, în care prima repetiție este redată mai jos:

R ₁	C ₁	C ₂	C ₃
L ₁	1	2	3
L ₂	4	5	6
L ₃	7	8	9

Este ideal ca:

- ÷ Efectul longitudinii (L₁ → L₃) să fie uniformizat asigurând prezența fiecărei cifre pe fiecare poziție: cifra "1" să fie în R₁ pe L₁, în R₂ pe L₂, în R₃ pe L₃; idem pentru cifrele "2".."9";
- ÷ Efectul latitudinii (C₁ → C₃) să fie uniformizat asigurând prezența fiecărei cifre pe fiecare poziție: cifra "1" să fie în R₁ pe L₁, în R₂ pe L₂, în R₃ pe L₃; idem pentru cifrele "2".."9";

Repetițiile R₂ și R₃ se pot obține din repetiția R₁ de mai sus cu ajutorul Excel astfel:



- ÷ Se pomește de la R₁;
- ÷ Se selectează prima linie și se mută sub ultima linie și matricea obținută se ridică toată cu o linie;
- ÷ Se selectează prima coloană și se mută după ultima coloană și matricea obținută se deplasează spre stânga cu o coloană când se obține R₂;
- ÷ Se continuă de la R₂;
- ÷ Se selectează prima linie și se mută sub ultima linie și matricea obținută se ridică toată cu o linie;
- ÷ Se selectează prima coloană și se mută după ultima coloană și matricea obținută se deplasează spre stânga cu o coloană când se obține R₃.

Rezultatul obținut este schema de repetiție, redată mai jos:

Obs	R ₁			R ₂			R ₃		
	C ₁	C ₂	C ₃	C ₁	C ₂	C ₃	C ₁	C ₂	C ₃
L ₁	1	2	3	5	6	4	9	7	8
L ₂	4	5	6	8	9	7	3	1	2
L ₃	7	8	9	2	3	1	6	4	5

- ÷ Verificarea faptului că fiecare variantă este avantajată (sau dezavantajată) în mod egal de poziția geografică a parcelei se verifică numărând pe orizontală și pe verticală suma pozițiilor relative ale fiecărei variante:

Verticală	Orizontală
6 = 1 + 2 + 3 = Σ ₁ (poziții) = ... = Σ ₉ (poziții)	15 = 1 + 6 + 8 = Σ ₁ (poziții) = Σ ₄ (poziții) = Σ ₇ (poziții)
	15 = 2 + 4 + 9 = Σ ₂ (poziții) = Σ ₅ (poziții) = Σ ₈ (poziții)
	15 = 3 + 5 + 7 = Σ ₃ (poziții) = Σ ₆ (poziții) = Σ ₉ (poziții)

Desfășurarea experimentelor în câmp și analiza de varianță se desfășoară în conformitate cu schema de repetiție, în cazul de mai sus cu un factor (varietate, soi, tratament) având 9 valori (numerotate de la 1 la 9), folosind cele 3 repetiții, fiecare repetiție având un aranjament de 3 linii și 3 coloane.

Fie rezultatele experimentale redate în tabelul următor (reprezentat direct în Excel pentru simplificarea calculului):

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Rep	R1			R2			R3			LCRV	Sume	Nr	S ²	S ² /Nr			
2		C1	C2	C3	C1	C2	C3	C1	C2	C3	L1	215.2	9	46311	5145.7			
3	L1	1	2	3	5	6	4	9	7	8	L2	227.2	9	51620	5735.5			
4	L2	4	5	6	8	9	7	3	1	2	L3	215.3	9	46354	5150.5			
5	L3	7	8	9	2	3	1	6	4	5	TL	657.7	27	432569	16021.1			
6		R1			R2			R3			C1	215.2	9	46311	5145.7			
7	Obs	C1	C2	C3	C1	C2	C3	C1	C2	C3	C2	215.1	9	46268	5140.9			
8		C1	C2	C3	C1	C2	C3	C1	C2	C3	C3	227.4	9	51711	5745.6			
9	L1	27.4	28.8	22.7	26.5	19.5	18.7	28.7	19.1	23.8	TC	657.7	27	432569	16021.1			
10	L2	20.5	19.2	21.4	20.2	37.3	24.6	22.1	27.5	34.4	R1	225.5	9	50850	5650.0			
11	L3	24.3	25.4	35.8	24.7	18.7	25.5	20.8	19.6	20.5	R2	215.7	9	46526	5169.6			
12		R1			R2			R3			R3	216.5	9	46872	5208.0			
13	O ²	C1	C2	C3	C1	C2	C3	C1	C2	C3	TR	657.7	27	432569	16021.1			
14		C1	C2	C3	C1	C2	C3	C1	C2	C3	V1	80.4	3	6464	2154.7			
15	L1	751	829	515	702	380	350	702	380	350	V2	87.9	3	7726	2575.5			
16	L2	420	369	458	408	1391	605	408	1391	605	V3	63.5	3	4032	1344.1			
17	L3	590	645	1282	610	350	650	610	350	650	V4	58.8	3	3457	1152.5			
18		R1			R2			R3			V5	66.2	3	4382	1460.8			
19		C1	C2	C3	C1	C2	C3	C1	C2	C3	V6	61.7	3	3807	1269.0			
20		C1	C2	C3	C1	C2	C3	C1	C2	C3	V7	68	3	4624	1541.3			
21		C1	C2	C3	C1	C2	C3	C1	C2	C3	V8	69.4	3	4816	1605.5			
22		C1	C2	C3	C1	C2	C3	C1	C2	C3	V9	101.8	3	10363	3454.4			
23		C1	C2	C3	C1	C2	C3	C1	C2	C3	TV	657.7	27	432569	16021.1			
24		C1	C2	C3	C1	C2	C3	C1	C2	C3	O ²		27	16753.1	620.5			

Formulele de calcul implementate în coloanele O și P se efectuează după cum urmează:

N	O	P
1	Icrv Sume	Nr
2	L1 =sum(b9:d9,f9:h9,j9:l9)	=count(b9:d9,f9:h9,j9:l9)
3	L2 ↓	↓
4	L3 =sum(b11:d11,f11:h11,j11:l11)	=count(b9:d9,f9:h9,j9:l9)
5	TL =sum(o2:o4)	=sum(p2:p4)
6	C1 =sum(b9:b11,f9:f11,j9:j11)	=count(b9:b11,f9:f11,j9:j11)
7	C2 =sum(c9:c11,g9:g11,k9:k11)	=count(c9:c11,g9:g11,k9:k11)
8	C3 =sum(d9:d11,h9:h11,l9:l11)	=count(d9:d11,h9:h11,l9:l11)
9	TC =sum(o6:o8)	=sum(p6:p8)
10	R1 =sum(b9:d11)	=count(b9:d11)
11	R2 =sum(f9:h11)	=count(f9:h11)
12	R3 =sum(j9:l11)	=count(j9:l11)
13	TR =sum(o10:o12)	=sum(p10:p12)
14	V1 =sumif(\$b\$3:\$l\$5,concatenate("=",right(n14,1)),\$b\$9:\$l\$11)	=countif(\$b\$3:\$l\$5,concatenate("=",right(n14,1)))
15	V2 ↓	↓
...
22	V9 =sumif(\$b\$3:\$l\$5,concatenate("=",right(n22,1)),\$b\$9:\$l\$11)	=countif(\$b\$3:\$l\$5,concatenate("=",right(n22,1)))
23	TV =sum(o14:o22)	27
24	O ²	27

Formulele de calcul implementate în coloanele Q, R servesc la calculul varianțelor și sunt implementate după cum urmează:

	Q	R
1	S ²	S ² /Nr
...	=o2 ²	=q2/p2
	↓	↓

	↓	↓
...	=o23 ²	↓
24	=sum(b15:l17)	=q24/p24

Formulele de calcul implementate în coloanele T..Z calculează varianțele și sunt implementate după cum urmează:

	T	U	W	X	Y	Z
1	Analiza de varianță (ANOVA)					
2	Diferență de sume pătratice		Grade de libertate	s ²	F	p_F
3	L	=sum(r2:r4)-r5	=count(r2:r4)-1	=u3/w3	=X3/X\$8	=fdist(y3,w3,w\$8)
4	C	=sum(r6:r8)-r9	=count(r6:r8)-1	=u4/w4	=X4/X\$8	=fdist(y4,w4,w\$8)
5	R	=sum(r10:r12)-r13	=count(r10:r12)-1	=u5/w5	=X5/X\$8	=fdist(y5,w5,w\$8)
6	V	=sum(r14:r22)-r23	=count(r14:r22)-1	=u6/w6	=X6/X\$8	=fdist(y6,w6,w\$8)
7	T	=q24-r23	=p24-1	=u7/w7		
8	E	=u24-sum(u2:u23)	=w7-sum(w3:w6)	=u8/w8		

când rezultă valorile:

	T	U	W	X	Y	Z
1	Analiza de varianță (ANOVA)					
2	Diferență de sume pătratice		Grade de libertate	s ²	F	p_F
3	L	10.6	2	5.29	0.38	69.2%
4	C	11.1	2	5.56	0.40	67.9%
5	R	6.6	2	3.29	0.24	79.3%
6	V	537	2	67.1	4.82	0.8%
7	T	732	26	28.2		
8	E	167	12	13.9		

Rezultatele obținute în tabelul de mai sus arată că singura variabilitate semnificativ statistică ($p_F(V,E) < 5\%$) care este în fapt înalt semnificativ statistică ($p_F(V,E) < 1\%$) se înregistrează în cadrul variantelor aplicate.

În cele ce urmează, se folosește următorul raționament: pentru fiecare varianță (de la V1 la V9) observațiile vin dintr-o distribuție normală (cu număr mic de observații, 3 observații, câte 1 în fiecare repetiție). La această distribuție normală se poate face estimarea valorii medii (cu formulele " $=o14/p14$ " până la " $=o22/p22$ ") și a abaterii standard prin intermediul erorii pătratice medii (MS_E^2 , calculată în celula x8 (" $=u8/w8$ "). De aici încolo, compararea valorilor observate pentru două medii (a două variante) se face folosind testul Student t în modificarea adusă de Welch [40]:

$$t(\bar{x}_1, \bar{x}_2, s_1, s_2) = (\bar{x}_1 - \bar{x}_2) \Big/ \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, \quad v = \left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^2 \Big/ \left(\frac{s_1^4}{n_1^2(n_1-1)} + \frac{s_2^4}{n_2^2(n_2-1)} \right)$$

Astfel,

÷ Pentru compararea mediei unei variante cu o valoare de referință ("r", care este în acest caz valoare adevărată):

$$t(\bar{x}_1, r, s_1, 0) = \frac{\bar{x}_1 - r}{s_1} \cdot \sqrt{n_1}, \quad v = n_1 - 1$$

⁴⁰ Bernard L. WELCH, 1947. The generalization of "Student's" problem when several different population variances are involved. *Biometrika* 34(1-2): 28-35.

÷ Pentru compararea mediei unei variante cu o media experimentului (când $s_1 = s_2 = s$, $n_1 \neq n_2$):

$$t(\bar{x}_1, \bar{x}_2, s) = \frac{\bar{x}_1 - \bar{x}_2}{s} \sqrt{\frac{n_1 n_2}{n_1 + n_2}}, \quad v = \frac{(n_1 + n_2)(n_1 - 1)(n_2 - 1)}{n_2^2(n_2 - 1) + n_1^2(n_1 - 1)}$$

÷ Pentru compararea a 2 variante (V_i și V_j , când ($s_1 = s_2 = s$ și $n_1 = n_2 = n$):

$$t(\bar{x}_1, \bar{x}_2, s_1, s_2) = \frac{\bar{x}_1 - \bar{x}_2}{s} \sqrt{\frac{n}{2}}, \quad v = 2(n-1)$$

Efectuând calculele pentru exemplul dat, se obțin următoarele valori:

	V1	V2	V3	V4	V5	V6	V7	V8	V9
Medie	26.800	29.300	21.167	19.600	22.067	20.567	22.667	23.133	33.933
Eroare (SE)	3.731	3.731	3.731	3.731	3.731	3.731	3.731	3.731	3.731
n	3	3	3	3	3	3	3	3	3

Pentru o producție medie dată de 25 (tone/hectar) probabilitățile ca producțiile ce se vor obține să nu fie semnificativ statistic diferite (mai mari sau mai mici) de această valoare (25 tone/hectar) dacă se cultivă variantele V1..V9 sunt:

	V1	V2	V3	V4	V5	V6	V7	V8	V9
\bar{x}	26.800	29.300	21.167	19.600	22.067	20.567	22.667	23.133	33.933
r	25.000	25.000	25.000	25.000	25.000	25.000	25.000	25.000	25.000
Eroare (SE)	3.731	3.731	3.731	3.731	3.731	3.731	3.731	3.731	3.731
"n"	3	3	3	3	3	3	3	3	3
"t" = $\sqrt{n} \cdot (\bar{x}_1 - r) / s_E$	0.836	1.996	-1.780	-2.507	-1.362	-2.058	-1.083	-0.867	4.147
"v" = "n"-1	2	2	2	2	2	2	2	2	2
=tdist(abs("t"), "n", 1)	24.6%	9.2%	10.9%	6.5%	15.3%	8.8%	19.6%	23.9%	2.7%
probabilități de a fi:	">r"	">r"	"<r"	"<r"	"<r"	"<r"	"<r"	"<r"	">r"

Pentru o producție medie dată de amestecul variantelor probabilitățile ca producțiile ce se vor obține să nu fie semnificativ statistic diferite (mai mari sau mai mici) de această valoare (medie) dacă se cultivă variantele V1..V9 sunt:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s} \sqrt{\frac{n \cdot 9n}{n + 9n}} = \frac{\bar{x}_1 - \bar{x}_2}{s} \sqrt{\frac{9n}{10}} = \frac{\bar{x}_1 - \bar{x}_2}{s} \sqrt{\frac{27}{10}}$$

$$v = \frac{(n + 9n)^2 (n - 1)(9n - 1)}{81n^2 (9n - 1) + n^2 (n - 1)} = \frac{100(n - 1)(9n - 1)}{81(9n - 1) + (n - 1)} = \frac{1300}{527}$$

	V1	V2	V3	V4	V5	V6	V7	V8	V9
\bar{x}	26.800	29.300	21.167	19.600	22.067	20.567	22.667	23.133	33.933
$\bar{\bar{x}}$	24.359	24.359	24.359	24.359	24.359	24.359	24.359	24.359	24.359
Eroare (SE)	3.731	3.731	3.731	3.731	3.731	3.731	3.731	3.731	3.731
"n"	3	3	3	3	3	3	3	3	3
"t" = $\sqrt{9n/10} \cdot (\bar{x}_1 - \bar{\bar{x}}_2) / s$	1.075	2.176	-1.406	-2.096	-1.010	-1.670	-0.745	-0.540	4.217
"v" = $\frac{100(n - 1)(9n - 1)}{81(9n - 1) + (n - 1)}$	2.467	2.467	2.467	2.467	2.467	2.467	2.467	2.467	2.467
=1-pt(abs("t"), "v")	18.8%	6.8%	13.6%	7.3%	20.1%	10.6%	26.0%	31.7%	1.8%
probabilități de a fi:	">r"	">r"	"<r"	"<r"	"<r"	"<r"	"<r"	"<r"	">r"

pt(v1,v2): Funcție MathCad de calcul a probabilității cumulative a distribuției "t"
(Nici Excel nici EasyFit nu calculează corect pentru grade de libertate fracționare)

Compararea variantelor (de la V1 la V9) se poate realiza folosind aceeași modalitate:

$$t(\bar{x}_1, \bar{x}_2, s_1, s_2) = \sqrt{n/2} \cdot (\bar{x}_1 - \bar{x}_2) / s = \sqrt{3/2} \cdot (\bar{x}_1 - \bar{x}_2) / s, \quad v = 2(n-1) = 4$$

Pentru simplitate se pot înscrie datele într-o nouă foaie de calcul, când valorile (pentru medii variante V1..V9, eroare standard experiment s_E , număr de observații n și grade de libertate v) și formulele de calcul sunt după cum urmează:

	A	B	C	D	E..J	K	L	M	N
1	t_m		V4	V6	...	V9		sE	3.731
2			19.600	20.567	...			n	3
3	V4	19.600	=sqrt(\$n\$2/2)*(\$b3-c\$2)/\$n\$1	→	...	=sqrt(\$n\$2/2)*(\$b3-k\$2)/\$n\$1		v	4
4	V6	20.567	↓	↓			
...			
11	V9	33.933	=sqrt(\$n\$2/2)*(\$b11-c\$2)/\$n\$1	→	...	=sqrt(\$n\$2/2)*(\$b11-k\$2)/\$n\$1			
12									
13	p_t		V4	V6	V3	V9			
14			19.600	20.567	21.167				33.933
15	V4	19.600	=tdist(abs(c3),\$n\$3,1)	→	...	=tdist(abs(k3),\$n\$3,1)			
16	V6	20.567	↓	↓			
...			
23	V9	33.933	=tdist(abs(c11),\$n\$3,1)	→	...	=tdist(abs(k11),\$n\$3,1)			

când rezultă valorile:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	t_m		V4	V6	V3	V5	V7	V8	V1	V2	V9		sE	3.731
2			19.600	20.567	21.167	22.067	22.667	23.133	26.800	29.300	33.933		n	3
3	V4	19.600	0.000	-0.967	-1.567	-2.467	-3.067	-3.533	-7.200	-9.700	-14.333		v	4
4	V6	20.567	0.967	0.000	-0.600	-1.500	-2.100	-2.567	-6.233	-8.733	-13.367			
5	V3	21.167	1.567	0.600	0.000	-0.900	-1.500	-1.967	-5.633	-8.133	-12.767			
6	V5	22.067	2.467	1.500	0.900	0.000	-0.600	-1.067	-4.733	-7.233	-11.867			
7	V7	22.667	3.067	2.100	1.500	0.600	0.000	-0.467	-4.133	-6.633	-11.267			
8	V8	23.133	3.533	2.567	1.967	1.067	0.467	0.000	-3.667	-6.167	-10.800			
9	V1	26.800	7.200	6.233	5.633	4.733	4.133	3.667	0.000	-2.500	-7.133			
10	V2	29.300	9.700	8.733	8.133	7.233	6.633	6.167	2.500	0.000	-4.633			
11	V9	33.933	14.333	13.367	12.767	11.867	11.267	10.800	7.133	4.633	0.000			
12														
13	p_t		V4	V6	V3	V5	V7	V8	V1	V2	V9			
14			19.600	20.567	21.167	22.067	22.667	23.133	26.800	29.300	33.933			
15	V4	19.600	0.500	0.383	0.317	0.232	0.186	0.155	0.039	0.017	0.005			
16	V6	20.567	0.383	0.500	0.427	0.324	0.264	0.223	0.055	0.023	0.006			
17	V3	21.167	0.317	0.427	0.500	0.391	0.324	0.277	0.069	0.028	0.007			
18	V5	22.067	0.232	0.324	0.391	0.500	0.427	0.372	0.098	0.038	0.009			
19	V7	22.667	0.186	0.264	0.324	0.427	0.500	0.443	0.123	0.048	0.010			
20	V8	23.133	0.155	0.223	0.277	0.372	0.443	0.500	0.148	0.056	0.012			
21	V1	26.800	0.039	0.055	0.069	0.098	0.123	0.148	0.500	0.229	0.040			
22	V2	29.300	0.017	0.023	0.028	0.038	0.048	0.056	0.229	0.500	0.101			
23	V9	33.933	0.005	0.006	0.007	0.009	0.010	0.012	0.040	0.101	0.500			

Interpretarea testului este după cum urmează:

- ÷ Experimentul în câmp a arătat că în condițiile experimentale date:
- cultura propagată folosind varianta V9 va da o producție distinct semnificativ statistică ($p < 1\%$) mai mare decât culturile propagate folosind oricare dintre variantele V4, V6, V3, V5;
 - cultura propagată folosind varianta V9 va da o producție semnificativ statistică ($p < 5\%$) mai mare decât culturile propagate folosind oricare dintre variantele V4, V6, V3, V5, V7, V8 și V1;
 - cultura propagată folosind varianta V9 și/sau varianta V2 va da o producție semnificativ statistică ($p < 5\%$) mai mare decât culturile propagate folosind oricare dintre variantele V4, V6, V3, V5, V7;
- Astfel, dacă variantele referă propagarea unor varietăți diferite în spiritul asigurării unei diversități de cultură în baza experimentului factorial în câmp desfășurat se recomandă propagarea varietăților V2 și V9 iar dacă variantele referă aplicarea unor tratamente diferite asupra aceluiași material biologic atunci în baza experimentului factorial în câmp desfășurat se recomandă propagarea varietății V9.

Studii de caz

Analiza factorială. Se dorește să se realizeze un experiment în care să fie colectate (în spațiul informațional) observații cu privire la structura de mai jos, în care pe fiecare din cele 6 nivele pot fi prezenți atomi de carbon, azot și bor.

Câte informații distincte se vor regăsi, maxim, în spațiul informațional, știind că structura este rigidă și fixată într-un ansamblu mai mare (cazul a) și respectiv mobilă (cazul b)?

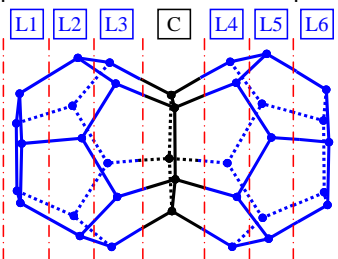
Soluție. Straturi: 6 (L1, L2, L3, L4, L5, L6); Nivele: 3 (B, N, C); Număr total de combinații: 729; Structuri distincte: 378. Exemplu de degenerare a influenței factorilor: Proprietăți din Spartan '10

MolVol: Molecular volume Å³, SurfA: Surface area Å², Ovality:

Ovality dimensionless 1.234; HOMO: Highest Occupied Molecular

Orbital Energy eV; LUMO: Lowest Unoccupied Molecular Orbital

Energy eV; Estimated polarizability: *10⁻³⁰ m³; Electronegativity= -HOMO+LUMO/2; Hardness= -HOMO-LUMO/2



Property	Distinct	Grp.	Zero	Property	Distinct	Grp.	Zero
DipoleT_0	376	352		Lumo-Homo_0	353	375	
DipoleT_1	377	351		Lumo-Homo_1	331	397	
DipoleT_2	377	351		Lumo-Homo_2	372	356	
EnergyHF_0	131	510	87	MolVol_0	175	512	41
EnergyHF_1	139	507	82	MolVol_1	167	530	31
EnergyHF_2	346	382		MolVol_2	372	356	
HOMO_0	337	389	2	Ovality_0	46	516	166
HOMO_1	320	408		Ovality_1	51	590	87
HOMO_2	367	361		Ovality_2	56	645	27
LUMO_0	318	410		Polariz_0	349	379	
LUMO_1	300	428		Polariz_1	316	412	
LUMO_2	367	361		Polariz_2	371	357	
Lumo+Homo_0	355	373		SurfA_0	221	477	30
Lumo+Homo_1	324	404		SurfA_1	202	520	6
Lumo+Homo_2	356	370	2	SurfA_2	368	360	

Precizie și exactitate. Să se analizeze sub aspectul preciziei următoarele măsurători ale vitezei vântului în raport cu afirmația că în acea zi nu s-au înregistrat deplasări de aer.

Moment	Viteza	Direcția	Durata
9 ³⁰	2.8ms ⁻¹	NV	1 min.
10 ³⁰	1.0ms ⁻¹	E	2 min.
11 ³⁰	1.4ms ⁻¹	NE	1 min.
12 ³⁰	3ms ⁻¹	S	1 min.

Ce observație lipsește pentru ca să se susțină afirmația?

Soluție. Se alcătuieste un tabel de forma celui de mai jos:

v	t	dir	dx	dy	sdx	sdv
2.8	60	NV	-118.8	118.8	-118.8	118.8
1	120	E	120.0	0.0	1.2	118.8
1.4	60	NE	59.4	59.4	60.6	178.2
3	60	S	0.0	-180.0	60.6	-1.8

Separare cromatografică maximă. În cromatografie se folosesc o serie de măsurări pentru a caracteriza separarea compuşilor dintr-un amestec:

÷ $RF(i, e) = l(i, e)/l(e)$, în care i este un compus separat, e faza mobilă, $l(i, e)$ distanța de migrare a lui i în e , și $l(e)$ distanța de migrare a lui e și ce conduce la **mulțimea factorilor de retardare (RF)**

÷ $RFO(i, e) = l(\pi(i), e)/l(e)$, unde $\pi(i)$ este o permutare ce transformă lista într-o listă ordonată (crescător sau descrescător) și ce conduce la **șirul ordonat al factorilor de retardare (RFO)**

÷ $nc(e) = \text{Count}(\{i | l(\pi(i+1), e) - l(\pi(i), e) > (w(\pi(i+1), e) + w(\pi(i), e)) / 8)\})$ ce conduce la **numărul de**

componentei observați la o distanță (unul de altul) de cel puțin 1σ (nc)

÷ $mnc = \max\{nc(e) | e \in E\}$, în care E mulțimea tuturor fazelor mobile posibile și ce conduce la *numărul de componente existenți în amestec (mnc)*

÷ $RSM(i, j, e) = 2 \cdot (l(i, e) - l(j, e)) / (w(i, e) + w(j, e))$, unde $w(i, e)$ este lărgimea spotului i iar $w(j, e)$ este lărgimea spotului j , și ce conduce la *matricea rezoluțiilor (RSM)*

÷ $RSO(i, e) = 2 \cdot (l(\pi(i+1), e) - l(\pi(i), e)) / (w(\pi(i+1), e) + w(\pi(i), e))$, unde din nou π este permutarea ce ordonează (crescător) șirul factorilor de retardare și ce conduce la *șirul rezoluțiilor spoturilor adiacente (RSO)*

÷ $RSS(e) = \sum_{i=1}^{nc(e)} RSO(i, e)$ ce conduce la *suma rezoluțiilor (RSS)*

÷ $QN_{eff}(e) = \sqrt{N_{eff}(e)} = 4 \cdot l(e) \cdot nc(e) / \sum_{i=1}^{nc(e)} w(e, i)$ ce conduce la *radical din numărul efectiv de talere (QN_{eff}); ajustând condițiile experimentale date idealul de separare se obține când $4 \cdot RSS(e) \rightarrow QN_{eff}(e)$*

÷ $RSP(e) = 25 \cdot RSS(e) / QN_{eff}(e)$ ce dă *rezoluția împărțită la numărul de talere (RSP); idealul de separare se obține când $RSP(e) \rightarrow 100$*

÷ $RSA(e) = RSS(e) / nc(e)$ ce dă *rezoluția medie a separării (RSA)*

÷ $RRP(e) = \prod_{i=1}^{nc(e)} RSO(i, e) / \sum_{i=1}^{nc(e)} RSO(i, e)$ care este *produsul rezoluțiilor relative (RRP)*

÷ $RSH(e, p) = \left(\sum_{i=1}^{nc(e)} (RSO(i, e))^p / nc(e) \right)^{1/p}$ unde p este o valoare reală arbitrară și ce conduce la *media Hölder a rezoluțiilor (RSH); când $p=2$ $RSH(e, 2)$ este un mai bun descriptor al separării decât RSA*

÷ $RFD(e) = \sqrt{\sum_{i=1}^{nc(e)} (RFO(i+1, e) - RFO(i, e) - 1/mnc)^2} / \sqrt{nc(e) \cdot (nc(e) + 1)}$, unde $1/mnc$ este diferența

teoretică între 2 factori de retardare și ce conduce la un indice de separare medie exprimat ca deviație între ideal ($1/mnc$) și observat ($RFO(i+1, e) - RFO(i, e)$) - *deviația factorilor de retardare (RFD)*

÷ $IEne(e) = mnc^2 - \sum_{i=1}^{mnc} n_i^2$, unde n_i este numărul de spoturi în al i -lea interval echidistant și ce conduce

la *energia informațională asociată separării (IEne)*

÷ $IEnt(e) = \sum_{i=1}^{mnc} n_i \cdot \log(n_i)$, unde n_i este numărul de spoturi în al i -lea interval echidistant și ce conduce

la *entropia informațională asociată separării (IEnt)*

÷ $QF(e) = \min_{1 \leq i, j \leq nc(e)} RSM(i, j, e) = \min_{1 \leq i \leq nc(e)} RSO(i, e)$ care conduce la *factorul de calitate a separării (QF)*

dat de cea mai defavorabilă separare

Să se calculeze și apoi să se analizeze proprietățile (operațiile matematice permise) de acești parametrii.

Soluție. Imaginând procesul de separare drept un proces în care apariția migrării spoturilor este o problemă de durată, din punct de vedere experimental este de interes ce se întâmplă cu valorile acestor parametrii când se detectează apariția unui nou spot (eventual la începutul sau la sfârșitul șirului de spoturi deja existent). Fie e un eluent (același). Vom simplifica notațiile făcând referire mereu la același eluent.

RF. Fie $\{RF(i), 1 \leq i \leq nc\}$ mulțimea factorilor de retardare. Apariția unui nou factor face ca la mulțimea factorilor de retardare să se adauge un element. Din acest punct de vedere, mulțimea factorilor de retardare poate fi văzută ca o mulțime de caracteristici. Asocierea între compusul chimic și factorul de retardare al acestuia e unică (pe calea directă, adică un compus are un singur factor de retardare în

condițiile experimentale date; reciproca însă nu este adevărată, putând exista 2 compuși cu același factor de retardare) și poate indica ulterior absența compusului dintr-un amestec prin absența valorii factorului de retardare din mulțimea valorilor măsurate pentru amestec.

RFO. Șirul ordonat al factorilor de retardare nu este o mulțime standard, fiind echipat și cu o relație de ordine. Din acest punct de vedere, șirul ordonat al factorilor de retardare necesită o operație simplă de adăugare la început sau la sfârșit doar dacă valoarea adăugată este fie mai mică, fie mai mare decât toate valorile deja existente în șir; în caz contrar, adăugarea necesită și identificarea poziției în care trebuie făcută inserarea.

nc. Numărul de componenți observați își modifică valoarea la apariția unui nou spot numai dacă este îndeplinită condiția ca distanța dintre spoturile adiacente să fie cel puțin 1/8 din lărgimea acestora. Din acest punct de vedere, adăugarea unui nou compus în amestec poate avea ca efect unul din următoarele:

- ÷ $nc \rightarrow nc-1$ când noul spot apare între 2 spoturi deja existente, dar apariția acestuia face ca cele 2 spoturi existente inițial împreună cu cel de-al treilea nou spot să nu mai poată fi distinse unul de celălalt;
- ÷ $nc \rightarrow nc$ când noul spot apare în vecinătatea altuia dar suficient de departe de oricare alt spot
- ÷ $nc \rightarrow nc+1$ când noul spot apare suficient de departe de oricare alt spot

RSM. Matricea rezoluțiilor suferă o operațiune de adăugare a unei linii și a unei coloane pentru fiecare nou compus adăugat în amestec. Liniile și coloanele existente inițial în această matrice rămân cu valori neschimbate.

RSO. Șirul rezoluțiilor ordonate nu se poate obține din ordonarea matricei RSM (de exemplu ordonând o linie prin permutare de coloane). Matricea rezoluțiilor ordonate se obține din diferențele între valorile din șirul ordonat al factorilor de retardare (RFO \rightarrow RSO). Din acest punct de vedere, rezultatul obținut prin adăugarea unui nou compus în amestec poate produce apariția sau dispariția unui element în șirul rezoluțiilor ordonate sau să se păstreze numărul acestora (vezi nc) necesitând totodată și schimbarea valorilor adiacente valorii inserate (dacă numărul elementelor rămâne același sau crește cu o unitate).

Metoda adaosului standard. Se consideră o probă sub formă de soluție în care compușii dizolvați sunt cunoscuți dar sunt în cantități necunoscute. De asemenea sunt disponibili acești compuși preparați separat sub formă de soluții apoase. Care este modalitatea de determinare a cantităților necunoscute, știind că identificarea se realizează cu ajutorul unui detector care produce un semnal proporțional cu concentrația (fie aceasta molară) de compus identificat?

Soluție. Fie x și y cantitățile în grame de compuși cunoscuți C_1 și C_2 dizolvați în 100 ml H_2O (în proba necunoscută) și fie a și b cantitățile în grame de compuși cunoscuți dizolvați în 100 ml H_2O (în etaloanele preparate din acești compuși). Fie masele molare M_1 (în g, pentru C_1) și M_2 (în g, pentru C_2).

În acest caz avem la dispoziție 3 soluții: S_U (soluția necunoscută), S_1 (din C_1) și S_2 (din C_2):

Soluție	S_U	S_1	S_2
Volum	100 ml	100 ml	100 ml
Substanță dizolvată	$x gC_1, y gC_2$	$a gC_1$	$b gC_2$

Imaginăm o serie de 3 experimente cu acești compuși în care folosim câte 25 ml din soluția cu cantități dizolvate necunoscute (S_U) și câte un volum (deocamdată neprecizat) din soluțiile etalon. Un calcul simplu dă compoziția amestecurilor:

Experiment	S_U	S_1	S_2	$V(H_2O)$	$m(C_1)$	$m(C_2)$
1	25 ml	V_{11} ml	V_{12} ml	$25+V_{11}+V_{12}$	$25 \cdot x/100+V_{11} \cdot a/100$	$25 \cdot y/100+V_{12} \cdot b/100$
2	25 ml	V_{21} ml	V_{22} ml	$25+V_{21}+V_{22}$	$25 \cdot x/100+V_{21} \cdot a/100$	$25 \cdot y/100+V_{22} \cdot b/100$
3	25 ml	V_{31} ml	V_{32} ml	$25+V_{31}+V_{32}$	$25 \cdot x/100+V_{31} \cdot a/100$	$25 \cdot y/100+V_{32} \cdot b/100$

Detectorul va înregistra semnale. În acest moment avem 2 soluții de implementare: când detecția se realizează după separare, caz în care pentru fiecare compus identificat avem câte un semnal în fiecare experiment (fie R_{11} și R_{12} semnalele din primul experiment și asemeni R_{21} , R_{22} , R_{31} și R_{32}) sau detecția se realizează direct asupra amestecului, caz în care avem câte un singur semnal din fiecare experiment (fie Z_1 semnalul înregistrat în primul experiment corespunzător detecției, Z_2 semnalul înregistrat în al doilea experiment și Z_3 semnalul înregistrat în al treilea experiment) proporționale cu concentrațiile molare, deci este necesar să evaluăm aceste concentrații:

Experiment	$n(C_1)/V(H_2O)$	$n(C_2)/V(H_2O)$
1	$(25 \cdot x / 100 + V_{11} \cdot a / 100) / M_1 / (25 + V_{11} + V_{12})$	$(25 \cdot y / 100 + V_{12} \cdot b / 100) / M_2 / (25 + V_{11} + V_{12})$
2	$(25 \cdot x / 100 + V_{21} \cdot a / 100) / M_1 / (25 + V_{21} + V_{22})$	$(25 \cdot y / 100 + V_{22} \cdot b / 100) / M_2 / (25 + V_{21} + V_{22})$
3	$(25 \cdot x / 100 + V_{31} \cdot a / 100) / M_1 / (25 + V_{31} + V_{32})$	$(25 \cdot y / 100 + V_{32} \cdot b / 100) / M_2 / (25 + V_{31} + V_{32})$
Semnal	Separat pentru C_1	Separat pentru C_2
1	R_{11}	R_{12}
2	R_{21}	R_{22}
3	R_{31}	R_{32}
Semnal	Împreună pentru C_1 și C_2 în amestec	
1	Z_1	
2	Z_2	
3	Z_3	

Se exprimă raportul între concentrație și semnal care pentru un anumit compus este același de la un experiment la altul. Rezolvăm în continuare pe cele două cazuri:

Cazul semnalului separat

$$\div \text{ Pentru } C_1: \frac{x/4 + a \cdot V_{11}/100}{M_1 \cdot (25 + V_{11} + V_{12}) \cdot R_{11}} = \frac{x/4 + a \cdot V_{21}/100}{M_1 \cdot (25 + V_{21} + V_{22}) \cdot R_{21}} = \frac{x/4 + a \cdot V_{31}/100}{M_1 \cdot (25 + V_{31} + V_{32}) \cdot R_{31}}$$

$$\div \text{ Pentru } C_2: \frac{y/4 + b \cdot V_{12}/100}{M_2 \cdot (25 + V_{11} + V_{12}) \cdot R_{12}} = \frac{y/4 + b \cdot V_{22}/100}{M_2 \cdot (25 + V_{21} + V_{22}) \cdot R_{22}} = \frac{y/4 + b \cdot V_{32}/100}{M_2 \cdot (25 + V_{31} + V_{32}) \cdot R_{32}}$$

Simplificând:

$$\div \text{ Pentru } C_1: \frac{x \cdot 25 + a \cdot V_{11}}{(25 + V_{11} + V_{12}) \cdot R_{11}} = \frac{x \cdot 25 + a \cdot V_{21}}{(25 + V_{21} + V_{22}) \cdot R_{21}} = \frac{x \cdot 25 + a \cdot V_{31}}{(25 + V_{31} + V_{32}) \cdot R_{31}}$$

$$\div \text{ Pentru } C_2: \frac{y \cdot 25 + b \cdot V_{12}}{(25 + V_{11} + V_{12}) \cdot R_{12}} = \frac{y \cdot 25 + b \cdot V_{22}}{(25 + V_{21} + V_{22}) \cdot R_{22}} = \frac{y \cdot 25 + b \cdot V_{32}}{(25 + V_{31} + V_{32}) \cdot R_{32}}$$

Pentru fiecare dintre cei doi compuși, avem exact 2 soluții (nu neapărat identice) corespunzătoare celor 2 egalități. În fapt, problema poate fi rezolvată și mai corect minimizând eroarea de observare (temă de casă!).

În continuare, fie aceste două egalități următoarele:

$$x \cdot \frac{25}{(25 + V_{11} + V_{12}) \cdot R_{11}} + \frac{a \cdot V_{11}}{(25 + V_{11} + V_{12}) \cdot R_{11}} = x \cdot \frac{25}{(25 + V_{21} + V_{22}) \cdot R_{21}} + \frac{a \cdot V_{21}}{(25 + V_{21} + V_{22}) \cdot R_{21}}$$

$$x \cdot \frac{25}{(25 + V_{11} + V_{12}) \cdot R_{11}} + \frac{a \cdot V_{11}}{(25 + V_{11} + V_{12}) \cdot R_{11}} = x \cdot \frac{25}{(25 + V_{31} + V_{32}) \cdot R_{31}} + \frac{a \cdot V_{31}}{(25 + V_{31} + V_{32}) \cdot R_{31}}$$

De unde:

$$x = a \cdot \frac{\frac{V_{21}/R_{21}}{25 + V_{21} + V_{22}} - \frac{V_{11}/R_{11}}{25 + V_{11} + V_{12}}}{\frac{25/R_{11}}{25 + V_{11} + V_{12}} - \frac{25/R_{21}}{25 + V_{21} + V_{22}}} \text{ și/sau } x = a \cdot \frac{\frac{V_{31}/R_{31}}{25 + V_{31} + V_{32}} - \frac{V_{11}/R_{11}}{25 + V_{11} + V_{12}}}{\frac{25/R_{11}}{25 + V_{11} + V_{12}} - \frac{25/R_{31}}{25 + V_{31} + V_{32}}}$$

În mod similar rezultă și expresiile pentru y.

Cazul amestecului

În cazul măsurării directe pe amestec, se suprapune semnalul de la primul compus cu semnalul de la al doilea compus, însă nu în mod necesar cu aceeași intensitate:

$$\div \text{ Experiment 1: } \frac{x/4 + a \cdot V_{11}/100}{M_1 \cdot (25 + V_{11} + V_{12})} \cdot \alpha + \frac{y/4 + b \cdot V_{12}/100}{M_2 \cdot (25 + V_{11} + V_{12})} \cdot \beta = Z_1$$

$$\div \text{ Experiment 2: } \frac{x/4 + a \cdot V_{21}/100}{M_1 \cdot (25 + V_{21} + V_{22})} \cdot \alpha + \frac{y/4 + b \cdot V_{22}/100}{M_2 \cdot (25 + V_{21} + V_{22})} \cdot \beta = Z_2$$

$$\div \text{ Experiment 3: } \frac{x/4 + a \cdot V_{31}/100}{M_1 \cdot (25 + V_{31} + V_{32})} \cdot \alpha + \frac{y/4 + b \cdot V_{32}/100}{M_2 \cdot (25 + V_{31} + V_{32})} \cdot \beta = Z_3$$

Ecuțiile de mai sus formează un sistem de 3 ecuații cu 4 necunoscute (x, y, α și β). Este evident că

pentru a rezolva acest sistem este nevoie de încă o ecuație, deci de încă un experiment:

$$\text{Experiment 4: } \frac{x/4 + a \cdot V_{41}/100}{M_1 \cdot (25 + V_{41} + V_{42})} \cdot \alpha + \frac{y/4 + b \cdot V_{42}/100}{M_2 \cdot (25 + V_{41} + V_{42})} \cdot \beta = Z_4$$

Sistemul de ecuații rezultat se poate scrie prin intermediul valorilor cunoscute astfel (unde c_{11}, \dots, c_{45} sunt valori cunoscute):

$$x \cdot \alpha \cdot c_{11} + \alpha \cdot c_{12} + y \cdot \beta \cdot c_{13} + \beta \cdot c_{14} = c_{15}$$

$$x \cdot \alpha \cdot c_{21} + \alpha \cdot c_{22} + y \cdot \beta \cdot c_{23} + \beta \cdot c_{24} = c_{25}$$

$$x \cdot \alpha \cdot c_{31} + \alpha \cdot c_{32} + y \cdot \beta \cdot c_{33} + \beta \cdot c_{34} = c_{35}$$

$$x \cdot \alpha \cdot c_{41} + \alpha \cdot c_{42} + y \cdot \beta \cdot c_{43} + \beta \cdot c_{44} = c_{45}$$

Chiar dacă pare complicat, în esență este un sistem de ecuații ușor de rezolvat. Făcând substituțiile $\gamma = x \cdot \alpha$ și $\delta = y \cdot \beta$ sistemul devine un sistem linear și omogen în necunoscutele α, β, γ și δ :

$$\gamma \cdot c_{11} + \alpha \cdot c_{12} + \delta \cdot c_{13} + \beta \cdot c_{14} = c_{15}$$

$$\gamma \cdot c_{21} + \alpha \cdot c_{22} + \delta \cdot c_{23} + \beta \cdot c_{24} = c_{25}$$

$$\gamma \cdot c_{31} + \alpha \cdot c_{32} + \delta \cdot c_{33} + \beta \cdot c_{34} = c_{35}$$

$$\gamma \cdot c_{41} + \alpha \cdot c_{42} + \delta \cdot c_{43} + \beta \cdot c_{44} = c_{45}$$

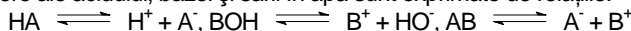
Odată obținută soluția acestui sistem (α, β, γ și δ) valorile necunoscutele x și y sunt imediate: $x = \gamma/\alpha$ și $y = \delta/\beta$.

Calculul pH-ului în funcție de volumul de soluție adăugată. Să se exprime variația pH-ului în titrarea în soluție apoasă (H_2O) a unui acid slab (HA) cu o bază slabă (BOH).

Soluție. Modelul reacției de titrare pornește de la scrierea ecuației reacției chimice a titrării unui acid slab HA cu o bază slabă BOH, pentru care echilibrul este caracterizat de constanta de solubilitate a sării obținute: $HA + BOH \rightleftharpoons AB + H_2O$.

Dacă reacția se desfășoară în apă, trebuie să considerăm influența asupra pH-ului din disocierea moleculelor de apă: $HOH \rightleftharpoons HO^- + H^+$.

Procesele de disociere ale acidului, bazei și sării în apă sunt exprimate de relațiile:



Titarea începe cu adăugarea unei mici cantități de bază în acid. În acest moment sunt prezente în soluție speciile: H^+, HO^-, HA și A^- . Din ecuația de disociere a acidului și apei rezultă:

$$[H^+] \cdot [A^-] = K_a \cdot [HA], [H^+] \cdot [HO^-] = K_w$$

unde $[\cdot]$ este exprimă concentrația molară ($[H^+]$ este concentrația molară), K_a este constanta de aciditate iar K_w este constanta de disociere a apei la temperatura considerată. Dacă se aplică bilanțul de masă pentru acid și sare, rezultă că C_a , concentrația analitică a acidului și respectiv C_s , concentrația analitică a sării sunt date de: $C_a = [HA] + [H^+] - [HO^-]$, $C_s = [A^-] - [H^+] + [HO^-]$

După substituțiile corespunzătoare în ecuațiile de mai sus, se obține o ecuație de gradul 3 a pH-ului (ecuația Brønsted-Lowry [⁴¹], [⁴²]), $[H^+] = x: x^3 + (K_a + C_s) \cdot x^2 - (K_w + C_x \cdot K_a) \cdot x - K_w \cdot K_a = 0$.

Ecuația de mai sus admite o soluție unică în intervalul (0,1). Ținând seama că:

$$C_s = C_b \cdot V_x / (V_a + V_x), C_x = (C_a \cdot V_a - C_b \cdot V_x) / (V_a + V_x)$$

unde C_b este concentrația analitică a bazei, V_x este volumul de bază adăugat, C_x concentrația analitică a acidului după adăugare iar V_a este volumul inițial de acid. Substituind în ecuația Brønsted, aceasta poate fi rezolvată numeric (v. <http://academicdirect.org/Education/Training/titration/v1.1/>).

La punctul de echivalență, se pornește modelul de la același punct inițial și se consideră toate echilibrele menționate. La hidroliză mică, $C_s = [A^-] = [B^+]$ așa că $[H^+] = x$:

$$x = \sqrt{K_w \cdot K_a \cdot (K_b + C_s) / K_b \cdot (K_a + C_s)}$$

După punctul de titrare, prin deduceri similare se obține că $[H^+] = x$:

$$x^3 + (K_w/K_b + C_x) \cdot x^2 - (K_w + C_s \cdot K_w/K_b) \cdot x - K_w^2/K_b = 0$$

unde expresiile lui C_x și C_s sunt: $C_x = (C_b \cdot V_x - C_a \cdot V_a) / (V_a + V_x)$, $C_s = C_a \cdot V_a / (V_a + V_x)$.

⁴¹ Bronsted JN, 1923. Some remarks on the concept of acids and bases (In German). Recueil des Travaux Chimiques des Pays-Bas 42(8): 718-728.

⁴² Lowry TM, 1923. The Uniqueness of Hydrogen. Chemistry and Industry 42(3): 43-47.

Modelarea în Excel a procesului de acțiune enzimatică Michaelis-Menten

Mecanismul a fost pentru prima dată observat la invertază ^[43], este caracterizat de ecuația: $S + E \leftrightarrow C \rightarrow P + E$, (S - substrat, E - enzimă, C - complex, P - produs (concentrații: s, e, c, p), se bazează pe ipoteza că enzima (E) nu suferă modificări în ceea ce privește cantitatea sa totală (și astfel nici concentrația) în timp ($e + c = \text{constant}$). Modelul cinetic permite exprimarea evoluției sistemului către echilibru. Presupunem scrierea ecuațiilor de viteză de reacție pentru fiecare reacție elementară și aplicarea principiului conservării numărului de atomi. Soluție:

1. Scrierea reacțiilor elementare; scrierea ecuațiilor de viteză						
(1): $S + E \xrightarrow{k_1} C$, $v_{(1)} = k_1 \cdot s \cdot e$	(2): $C \xrightarrow{k_2} S + E$, $v_{(2)} = k_2 \cdot c$	(3): $C \xrightarrow{k_3} P + E$, $v_{(3)} = k_3 \cdot c$				
2. Aplicarea principiului conservării numărului de atomi						
(S): $\dot{s} = v_{(2)} - v_{(1)}$	(E): $\dot{e} = v_{(2)} + v_{(3)} - v_{(1)}$	(C): $\dot{c} = v_{(1)} - v_{(2)} - v_{(3)}$	(P): $\dot{p} = v_{(3)}$			
3. Presupuneri și notații						
$s(0) = s_0$	$e(0) = e_0$	$c(0) = 0$	$p(0) = 0$	$e = e_0 - c$		
4. Ecuații de rezolvat						
$\dot{s} = k_2 \cdot c - k_1 \cdot s \cdot (e_0 - c)$		$\dot{c} = k_1 \cdot s \cdot (e_0 - c) - (k_2 + k_3) \cdot c$				
5. Diferite abordări						
aproximația "QSSA"	Briggs & Haldane, ^[44]	$\dot{c} = 0 \Rightarrow$	$c = \frac{e_0 \cdot s}{\kappa + s}$; $-\dot{s} = \dot{p} = \frac{k_3 \cdot e_0 \cdot s}{\kappa + s}$; $\kappa = \frac{k_2 + k_3}{k_1}$			
aproximația "EA"	Henri ^[45]	$\dot{s} = 0 \Rightarrow$	$c = \frac{e_0 \cdot s}{\kappa + s}$; $\dot{p} = \frac{k_3 \cdot e_0 \cdot s}{\kappa + s}$; $\kappa = \frac{k_2}{k_1}$			
Cazul general	ecuație implicită \rightarrow nu are soluție analitică!	spațiul fazelor	$\frac{dy}{dx} = b \frac{-x - y - xy}{-x + ay + xy}$			
	substituții în ecuația explicită $t = k_1 \cdot e_0 \cdot \tau$, τ timp inițial	$0 < a < 1$ $b > 0$	$a = \frac{k_2}{k_2 + k_3}$; $b = \frac{k_2 + k_3}{k_1 \cdot e_0}$; $x = \frac{k_1 \cdot s}{k_2 + k_3}$; $y = \frac{c}{e_0}$ $\dot{x} = -x + ay + xy$; $\dot{y} = b(x - y - xy)$			
6. Rezolvare numerică (i=1..n)						
$x_0 = 3$ $y_0 = 0$	$x_{i+1} = x_i + \delta(-x_i + a x_i + x_i y_i)$ $y_{i+1} = y_i + b \delta(x_i - y_i - x_i y_i)$	$\delta = 10^{-2}$ $n = 3000$	$a \in \{\frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}\}$	$b \in \{\frac{250}{25}, \frac{50}{25}, \frac{10}{25}, \frac{4}{25}\}$		
7. Foaie de calcul Excel:						
	A	B	C	D	E	F
1	$x_0 =$	3		i	xi	yi
2	$y_0 =$	0		=0	=B1	=B2
3	$\delta =$	1.0e-2		=D2+1	=E2+\$B\$3*(-E2+\$B\$4*E2+E2*F2)	=F2+\$B\$5*\$B\$3*(E2-F2-E2*F2)
4	a=	0.2	
5	b=	10	

Modelarea mecanismului Lindemann - Hinshelwood al complexului activat

Prima oară comunicat de Lindemann ^[46] și ulterior analizat de Hinshelwood ^[47] furnizează un mecanism de acțiune enzimatică. Se modelează mecanismul $R+R \leftrightarrow R^*+R \rightarrow P$. Soluție:

⁴³ Leonor MICHAELIS, Maud L MENTEN, 1913. Die Kinetik der Invertinwirkung. Biochem Z 49: 333-369.

⁴⁴ George E BRIGGS, John BS HALDANE, 1925. A note on the kinematics of enzyme action. Biochem J 19: 338-339.

⁴⁵ Victor HENRI, 1903. Lois Générales de l'Action des Diastases. Paris: Hermann.

⁴⁶ Frederick A. LINDEMANN, Svante ARRHENIUS, Irving LANGMUIR, N. R. DHAR, J. PERRIN, W. C. McC. LEWIS, 1922. Discussion on "the radiation theory of chemical action". Transactions of the Faraday Society, 17:598-606. doi: 10.1039/TF9221700598

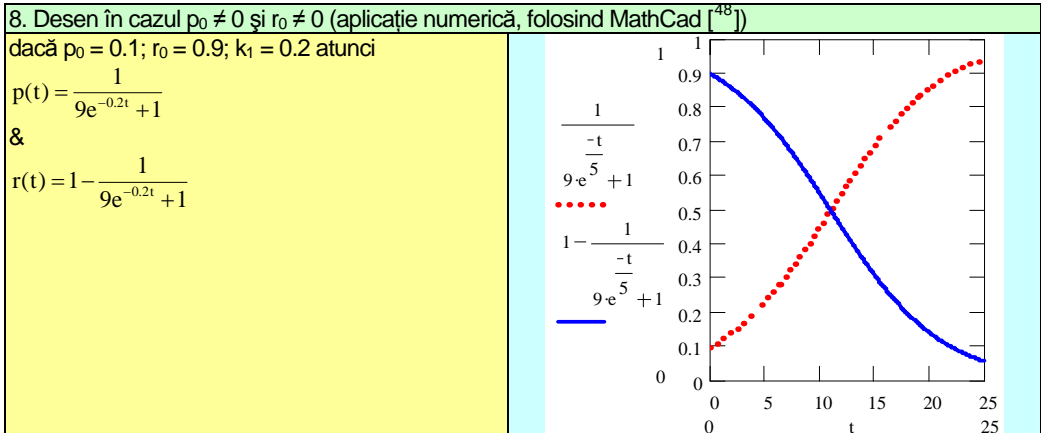
⁴⁷ Cyril N. HINSHELWOOD, 1926. On the Theory of Unimolecular Reaction. Proc. R. Soc. Lond. A 113:230-233. DOI:10.1098/rspa.1926.0149

1. Scrierea reacțiilor elementare; scrierea ecuațiilor de viteză							
(1): $R + R \rightarrow R^* + R, v_{(1)}=k_1[R]^2$		$R^* + R \rightarrow R + R, v_{(2)}=k_2[R][R^*]$		$R^* \rightarrow P, v_{(3)}=k_3[R^*]$			
2. Aplicarea principiului conservării numărului de atomi (necunoscute $[R] = x; [R^*] = y; [P] = z$):							
(R): $\dot{x} = -v_{(1)} + v_{(2)}$		(R*): $\dot{y} = v_{(1)} - v_{(2)} - v_{(3)}$		(P): $\dot{z} = v_{(3)}$			
3. Presupuneri și notații							
$r(0) = r_0$	$r^*(0) = 0$	$p(0) = 0$	$k_1 = a$	$k_2 = b$	$k_3 = c$		
4. Ecuații de rezolvat							
$\dot{x} = -ax^2 + bxy$		$\dot{y} = ax^2 - bxy - cy$		$\dot{z} = cy$			
5. Abordare greșită							
Căutarea unei soluții analitice este fără de succes.							
6. Rezolvare numerică (i=1..n)							
$x_0 = 3$ $y_0 = 0$ $z_0 = 0$	$x_{i+1} = x_i + \delta(-ax_i^2 + bx_i y_i)$ $y_{i+1} = y_i + \delta(ax_i^2 - bx_i y_i - cy_i)$ $z_{i+1} = z_i + \delta cy_i$		$\delta = 10^{-2}$ $n = 3000$	$a = 10^{-1}$ $b = 10^{-2}$ $c = 10^{-3}$			
7. Foaie de calcul Excel:							
	A	B	C	D	E	F	G
1	x0=	1		i	xi	yi	zi
2	y0=	0		=0	=B1	=B2	=B3
3	z0=	0		=D2+1	=E2+(-B\$1*E2^2+ B\$2*E2*F2)*B\$4	=F2+(B\$1*E2^2- B\$2*E2*F2-B\$3*F2)*B\$4	=G2+B\$3*F2*B\$4
4	δ=	1e-2	
5	a=	1e-1	
6	b=	1e-2	
7	c=	1e-3	

Simularea cineticii autocatalizei

Autocataliza este un caz special de cinetică chimică în care reacțiile au loc doar în prezența a atât reactanților cât și a produșilor de reacție și au o ecuație generală de reacție dată de: $R \rightarrow P$, cu $v_{(1)} = k[R] \cdot [P]$. Soluție:

1. Scrierea reacțiilor elementare; scrierea ecuațiilor de viteză					
(1): $R \rightarrow P, v_{(1)} = k[R] \cdot [P]$					
2. Aplicarea principiului conservării numărului de atomi (necunoscute $[R] = x; [R^*] = y; [P] = z$):					
(R): $\dot{r} = -v_{(1)} = -k_1 r p$			(P): $\dot{p} = v_{(1)} = k_1 r p$		
3. Presupuneri și notații					
$[R] = r$	$[P] = p$	$k_1 = a$	$r+p=r_0+p_0=b$	$r(0)=r_0$	$p(0)=p_0$
4. Ecuații de rezolvat					
$\dot{p} = a(b - p)p$			$\dot{r} = -ar(b - r)$		
5. Abordări - există soluție analitică					
$\dot{p} = a(b - p)p \Rightarrow \frac{dp}{p(b - p)} = a dt \Rightarrow \frac{1}{b} \ln \frac{p}{b - p} = at + c \Rightarrow \frac{p}{b - p} = e^{b(k_1 t + c)} \Rightarrow p = \frac{b}{1 + e^{-b(k_1 t + c)}} = \frac{b}{1 + e^{-bk_1 t} e^{-bc}}$					
6. Constantele "b" & "c" - din valorile inițiale ale concentrației (la momentul t = 0).					
$\frac{1}{b} \ln \frac{p(0)}{b - p(0)} = a0 + c \Rightarrow \frac{1}{b} \ln \frac{p_0}{r_0} = c; bc = \ln \frac{p_0}{r_0}; -bc = \ln \frac{r_0}{p_0}; e^{-bc} = \frac{r_0}{p_0}$					
7. Soluția analitică și interpretarea ei					
$p = p(t) = p_0 \frac{r_0 + p_0}{r_0 e^{-(r_0 + p_0) k_1 t} + p_0}$			dacă $p_0 = 0$ atunci $p = 0$ & astfel nu evoluează dacă $r_0 = 0$ atunci $p = p_0$ & astfel nu evoluează		



Resurse online pentru corelații în contingente ordonate și respectiv corelate

Corelații în contingente ordonate. Fie un tabel general de contingență reprezentând contingenta dintre două variabile $(x_i)_{1 \leq i \leq R}$ și $(y_j)_{1 \leq j \leq C}$ (unde $R =$ nr. de rânduri, $C =$ nr. de coloane):

R/C	Y_1	...	Y_j	...	Y_C
X_1	$n_{1,1}$...	$n_{1,j}$...	$n_{1,C}$
...
X_j	$n_{j,1}$...	$n_{j,j}$...	$n_{j,C}$
...
X_R	$n_{R,1}$...	$n_{R,j}$...	$n_{R,C}$
$X_1 < \dots < X_j < \dots < X_R$ $Y_1 < \dots < Y_j < \dots < Y_C$					

Formulele de calcul pentru măsurile de corelație a rangurilor (vezi *Calcularea coeficienților de corelație și a varianței asociate*) necesită definirea unei serii de notații (vezi *Notații pentru calcularea coeficienților de corelație*).

Simbol	Formula	Semnificația
$n_{i.}; n_{.j}; n$	$\sum_j n_{i,j}; \sum_i n_{i,j}; \sum_i \sum_j n_{i,j}$	total pe rând, pe coloană, și general
$R_i; C_j$	i or $R1_i; j$ or $C1_j$	scor pentru rândul i și coloana j
$\bar{R}; \bar{C}$	$\sum_i n_{i,R_i} / n; \sum_j n_{.j,C_j} / n$	scor mediu pe rând și coloană
$R1_i; C1_j$	$\sum_{k < i} n_{k.} + (n_{i.} + 1) / 2; \sum_{l < j} n_{.l} + (n_{.j} + 1) / 2$	scorul rangului pentru rândul i și coloana j
$C_{i,j}; D_{i,j}$	$\sum_{k > i} \sum_{l > j} n_{k,l} + \sum_{k < i} \sum_{l < j} n_{k,l}; \sum_{k > i} \sum_{l < j} n_{k,l} + \sum_{k < i} \sum_{l > j} n_{k,l}$	concordanțe și discordanțe
$C; D$	$\sum_i \sum_j n_{i,j} C_{i,j}; \sum_i \sum_j n_{i,j} D_{i,j}$	număr dublu de concordanțe număr dublu de discordanțe

Notații pentru calcularea coeficienților de corelație

În contextul notațiilor de mai sus (vezi *Notații pentru calcularea coeficienților de corelație*), următorul tabel (vezi *Calcularea coeficienților de corelație și a varianței asociate*) redă în mod unitar formulele de calcul ale măsurilor de corelație a rangurilor și respectiv statisticile asociate acestora.

⁴⁸ MathSoft, 1997. MathCad7 Professional (software). MathSoft. URL: <http://www.mathsoft.com>

Măsura	Expresia
r	$\frac{ss_{rc}}{w}, w = \sqrt{ss_r ss_c}, ss_r = \sum_i \sum_j n_{i,j} (R_i - \bar{R})^2, ss_c = \sum_i \sum_j n_{i,j} (C_j - \bar{C})^2$ $ss_{rc} = \sum_i \sum_j n_{i,j} (R_i - \bar{R})(C_j - \bar{C}), b_{i,j} = (R_i - \bar{R})^2 ss_c + (C_j - \bar{C})^2 ss_r$
Varianța asimptotică	$\frac{1}{w^4} \left(\sum_i \sum_j n_{i,j} \left(w(R_i - \bar{R})(C_j - \bar{C}) - \frac{b_{i,j} ss_{rc}}{2w} \right)^2 \right)$
Varianța pentru $r \neq 0$	$\frac{1}{w^2} \left(\sum_i \sum_j n_{i,j} (R_i - \bar{R})(C_j - \bar{C}) - \frac{ss_{rc}}{n} \right)^2$
ρ_s	$\frac{v}{w}, v = \sum_i \sum_j n_{i,j} R_0 C_0, w = \frac{1}{12} \sqrt{FG},$ $F = n^3 - \sum_i n_{i,\cdot}^3, G = n^3 - \sum_j n_{\cdot,j}^3, R_0 = R_1 - n/2, C_0 = C_1 - n/2$ $\bar{z} = \sum_i \sum_j n_{i,j} z_{i,j}, z_{i,j} = wv_{i,j} - vw_{i,j}, w_{i,j} = \frac{-n}{96w} (Fn_{\cdot,j}^2 + Gn_{i,\cdot}^2)$ $v_{i,j} = n \left(R_0 C_0 + \frac{1}{2} \sum_{i,l} n_{i,l} C_0 \frac{1}{2} \sum_k n_{k,j} R_0 + \sum_{k>i} \sum_{k,l} n_{k,l} C_0 + \sum_{k>l} \sum_{k,l} n_{k,l} R_0 \right)$
Varianța asimptotică	$\frac{1}{n^2 w^4} \sum_i \sum_j n_{i,j} (z_{i,j} - \bar{z})^2$
Varianța pentru $\rho_s \neq 0$	$\frac{1}{n^2 w^2} \sum_i \sum_j n_{i,j} (v_{i,j} - \bar{v})^2, \bar{v} = \sum_i \sum_j n_{i,j} v_{i,j} / n$
t_b	$\frac{C-D}{w}, w = \sqrt{w_r w_c}, w_r = n^2 - \sum_i n_{i,\cdot}^2, w_c = n^2 - \sum_j n_{\cdot,j}^2$ $d_{i,j} = C_{i,j} - D_{i,j}, v_{i,j} = n_{i,\cdot} w_c - n_{\cdot,j} w_r$
Varianța asimptotică	$\frac{1}{w^4} \left(\sum_i \sum_j n_{i,j} (2w d_{i,j} + t_b \cdot v_{i,j})^2 - n^3 t_b^2 (w_r + w_c)^2 \right)$
Varianța pentru $t_b \neq 0$	$\frac{4}{w_r w_c} \left(\sum_i \sum_j n_{i,j} (C_{i,j} - D_{i,j})^2 - \frac{(C-D)^2}{n} \right)$
γ	$\frac{C-D}{C+D}$
Varianța asimptotică	$\frac{16}{(C+D)^4} \sum_i \sum_j n_{i,j} (D \cdot C_{i,j} - C \cdot D_{i,j})^2$
Varianța pentru $\gamma \neq 0$	$\frac{4}{(C+D)^2} \left(\sum_i \sum_j n_{i,j} (C_{i,j} - D_{i,j})^2 - \frac{(C-D)^2}{n} \right)$
d_s	$\frac{C-D}{w_r}, w_r = n^2 - \sum_i n_{i,\cdot}^2, d_{i,j} = C_{i,j} - D_{i,j}$
Varianța asimptotică	$\frac{4}{w_r^4} \left(\sum_i \sum_j n_{i,j} (w_r d_{i,j} - (C-D)(n - n_{i,\cdot}))^2 \right)$
Varianța pentru $d_s \neq 0$	$\frac{4}{w_r^2} \left(\sum_i \sum_j n_{i,j} (C_{i,j} - D_{i,j})^2 - \frac{(C-D)^2}{n} \right)$

Calcularea coeficienților de corelație și a varianței asociate

Interpretarea coeficienților de corelație. În analiza asocierii dintre două variabile se evaluează direcția (pozitivă sau negativă), gradul de asociere (valoarea absolută a coeficientului de corelație) și respectiv forma asocierii:

- ÷ Direcția asocierii: valorile pozitive indică o relație în care creșterea valorilor unei variabile (fie aceasta X) este însoțită de creșterea valorilor celeilalte variabile (atunci aceasta este Y); valorile negative indică o relație în care creșterea valorilor unei variabile (X) este însoțită de scăderea valorilor celeilalte variabile (Y);
- ÷ Valori posibile: $-1 \leq r, \rho, \tau, \gamma \leq +1$. Valoarea -1 indică o relație perfectă inversă între cele două variabile; valoarea $+1$ indică o relație perfectă directă în timp ce valoarea 0 indică absența asocierii dintre cele două variabile;
- ÷ Forma asocierii este liniară pentru analiza condusă cu coeficientul de corelație Pearson sau orice dependență monotonă pentru analiza condusă cu coeficienții de corelație Spearman, Kendall și Goodman-Kruskal.

Interpretarea unui coeficient de corelație se face cu ajutorul probabilității asociate testului statistic adecvat (H_0 (ipoteza nulă): coeficientul de corelație nu este semnificativ diferit de zero vs. H_1 (ipoteza alternativă): coeficientul de corelație este semnificativ diferit de zero; vezi *Testul de semnificație pentru Pearson (r), Spearman (ρ_S), Sommer (d_S), Goodman-Kruskal (γ)*).

$$T(\text{coef} \neq 0) = -\sqrt{\text{coef}^2 / \text{Var}_{\neq 0}(\text{coef})^2}, p_{\text{coef}} = 2 \cdot \text{CDF}_1(T(\text{coef} \neq 0), n-1), \text{coef} \in \{r, \rho_S, \tau_b, \gamma, d_S\}$$

Testul de semnificație pentru Pearson (r), Spearman (ρ_S), Sommer (d_S), Goodman-Kruskal (γ)

Un coeficient de corelație al cărui interval de încredere nu conține valoarea zero este considerat semnificativ statistic. Intervalul de încredere al coeficientului de corelație se calculează în funcție de volumul eșantionului, permițând astfel interpretarea corectă a semnificației puterii de asociere (v. *Intervale de încredere pentru Pearson (r), Spearman (ρ_S), Sommer (d_S), Goodman-Kruskal (γ)*).

$$\text{coef} \pm \text{InvCDF}_1(1-\alpha/2, n-1) \cdot \frac{\sqrt{\text{Var}_{\neq 0}(\text{coef})}}{\sqrt{n}}, \text{coef} \in \{r, \rho_S, \tau_b, \gamma, d_S\}$$

Intervale de încredere (1- α) pentru

Pearson (r), Spearman (ρ_S), Kendall (τ_b), Sommer (d_S), Goodman-Kruskal (γ)

Testarea semnificației coeficientului de corelație se face prin aplicarea testelor statistice. Ipotezele testate sunt: nu există asociere între variabila X și Y (H_0 , ipoteza nulă) vs. există asociere între variabilele X și Y (H_1 , ipoteza alternativă). Dacă testul se aplică la un prag de semnificație α (uzual 5%), vom accepta ipoteza alternativă dacă semnificația (p) va fi mai mică decât α .

O implementare a analizei de corelație, care calculează coeficienții de corelației Pearson (r), Spearman (ρ), Kendall (τ), și semi-cantitativ (r_{SQ} [49]) împreună cu semnificația statistică a acestora se găsește la adresa: http://l.academicdirect.org/Statistics/linear_dependence/.

Analiza corelațiilor corelate se realizează prin aplicarea testelor statistice pentru dependențe (vezi *Teste statistice în corelații corelate*, în care m_2, m_3, m_m numărul de grade de libertate consumate în obținerea ecuațiilor de regresie a lui Y_1 cu Y_2 și respectiv Y_1 cu Y_3 ; dacă acestea diferă pentru m_m se aplică formula de medie dată la *Generalizarea testului Student t*).

Denumire	Formule de calcul a probabilității de a nu exista corelare (2 vs. 3)
Hotelling [50]	$H = \sqrt{\frac{(n-3-m_m)(1+r_{2,3})(r_{1,2}-r_{1,3})^2}{2[1+2 \cdot r_{2,3} \cdot r_{1,2} \cdot r_{1,3} - r_{2,3}^2 - r_{1,2}^2 - r_{1,3}^2]}}$, $p_H = 2 \cdot \text{CDF}_1(-H, n-3-m_m)$
Williams [51]	$\frac{1}{W^2} = \frac{1}{H^2} + \frac{1}{2(n-1)} \cdot \frac{(r_{1,2}+r_{1,3}) \cdot (1-r_{2,3})^3}{(1+r_{2,3}) \cdot (r_{1,2}-r_{1,3})^2}$, $p_W = 2 \cdot \text{CDF}_1(-W, n-3-m_m)$

⁴⁹ Sorana D BOLBOACĂ, Lorentz JÄNTSCHI, 2006. Pearson versus Spearman, Kendall's Tau correlation analysis on structure-activity relationships of biologic active compounds. Leonardo J Sci 9: 179-200.

⁵⁰ Harold HOTELLING, 1931. The generalization of Student's ratio. Ann Math Stat 2(3): 360-378.

⁵¹ Evan J WILLIAMS, 1959. The comparison of regression variables. J Roy Stat Soc Ser B 21(2): 396-399.

Steiger [⁵²] & Fisher [⁵³]	$z(r) = \ln\left(\frac{1+r}{1-r}\right), \quad \rho(z) = \frac{e^{2z} - 1}{e^{2z} + 1}, \quad z_{1,2} = z(r_{1,2}) - 0.5 \cdot r_{1,2} / (n - 1 - m_2),$ $z_{1,3} = z(r_{1,3}) - 0.5 \cdot r_{1,3} / (n - 1 - m_3), \quad z_{2,3} = z(r_{2,3}) - 0.5 \cdot r_{2,3} / (n - 1 - m_m), \quad \rho_1 = \rho(\sqrt{z_{1,2} \cdot z_{1,3}}),$ $\rho_{2,3} = \rho(z_{2,3}), \quad S = \sqrt{(z_{1,2} - z_{1,3})^2 (1 - \rho_1^2)^2 (n - 3 - m_m) / (2 + \rho_{2,3} \cdot (\rho_1^2 \cdot (4 - \rho_{2,3}) - 2) - 3 \cdot \rho_1^2)},$ $p_s = 2 \cdot \text{CDF}_z(- S)$
--	---

Teste statistice în corelații corelate ($r_{1,2}$ vs. $r_{1,3}$)

Analiza corelațiilor corelate utilizând formulele de mai sus se poate realiza utilizând programul online [⁵⁴]: <http://l.academicdirect.org/Statistics/tests/Correlated/>.

⁵² James K. STEIGER, 1980. Tests for comparing elements of a correlation matrix. Psychol Bull.87: 245-251.

⁵³ Ronald A FISHER, 1921. On the "probable error" of a coefficient of correlation deduced from a small sample. Metron 1: 3-32.

⁵⁴ Lorentz JÄTSCHI, 2014. Correlated correlation analysis. <http://l.academicdirect.org/Statistics/tests/Correlated/>