PREFATA

Lucrarea de fata reprezinta o continuare a cartii “Statistica Aplicata in Farmacie si Studii Clinice” aparuta in Editura Universitara Carol Davila in anul 2007 si sintetizeaza o parte din experienta a doi autori, amandoi in acelasi timp si farmacisti si matematicieni, in aplicarea metodelor statistice in cadrul cercetarilor de dezvoltare de noi medicamente, productia si controlul de calitate a medicamentului, precum si in evaluarea clinica a medicamentului.

Specificul volumului de fata este ca toate aplicatiile prezentate sunt rezolvabile cu “hartie si creion”, ceea ce permite o mult mai buna intelegere a “fenomenului matematic” decat o preluare “de-a gata” a rezultatelor unor softuri mai mult sau mai putin specializate. Numai dupa o astfel de “ucenicie” este recomandat, pentru aplicatii in care volumul de date este cu mult mai mare, a se trece la utilizarea softurilor (utilizare care va face subiectul unui al doilea volum). In cele mai multe cazuri, ca si in « diagnosticul cu calculatorul », rezultatele se obtin rapid si sunt pertinente. In cazurile foarte complexe sunt insa posibile erori foarte grave, care sunt foarte bine ascunse de softuri, desi vizibile usor la o abordare pornind de la regulile fundamentale .

Cartea se adreseaza studentilor de la farmacie , masteranzilor in biostatistica si cursantilor de la ciclul de invatamant de doctorat, in special cei ce pregatesc doctorate in farmacie sau in disciplinele medicale preclinice. Ea face parte dintr-o suita de lucrari de biofarmacie, farmacocinetica, statistica generala si biostatistica , proiectarea studiilor clinice etc, in pregatire in grupul de cercetatori din care fac parte autorii, care vin in plus sa umple si goluri in doua domenii noi in literatura de specialitate in cercetarea medicamentului, definite in Platforma Industriala Europeana Innovative Medicine Initiative : “in vitro – in vivo correlation” si “safety science”.

In forma actuala ea a fost « experimentata » de autori pe multe serii de studenti , masteranzi si doctoranzi si concluzia este ca acestia pot intelege, cu ceva efort, principiile de baza si aplicarea lor.

In final, autorii isi exprima credinta ca, in epoca calculului “industrial”, cu ajutorul calculatorului, arta, matematica si gandirea creatoare in general, trec tot prin varful creionului. Mai departe, acest mod de gandire permite o ascensiune spre mai inalt inclusiv prin unelte sofisticate create tot de om.

Sperand ca lucrul mainilor lor nu va fi trecut cu vederea

Autorii

Bucuresti, 2008

Constantin Mircioiu Roxana Colette Sandulovici

APLICATII NUMERICE
DE
STATISTICA IN FARMACIE
SI IN
STUDIILE CLINICE

Editura Universitara Carol Davila
Bucuresti, 2008

Prof. dr. farm., mat. Constantin Mircioiu
Farm., mat. Roxana Colette Sandulovici

APLICATII NUMERICE
DE
STATISTICA IN FARMACIE
SI IN
STUDIILE CLINICE

pentru

cursul de biostatistica
Facultatea de Farmacie, Universitatea de Medicina si Farmacie
“Carol Davila”, Bucuresti

cursul de biostatistica doctoranzi
Universitatea de Medicina si Farmacie “Carol Davila”, Bucuresti

cursul de biofarmacie si farmacocinetica
Masteratul de Biostatistica
Facultatea de Matematica, Universitatea Bucuresti

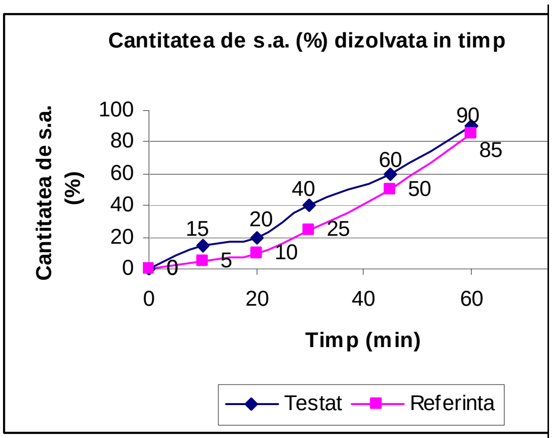
Editura Universitara Carol Davila
Bucuresti, 2008

CUPRINS

Metrici de dizolvare

Consideram ca doua produse sub forma de tablete contin aceeasi substanta activa, in acceasi concentratie si ca am obtinut urmatoarele rezultate in ceea ce priveste cedarea:

Timp (min.)	Cantitatea eliberata %	
	Testat T_i	Referinta R_i
0	0	0
10	15	5
20	20	10
30	40	25
45	60	50
60	90	85



Sa se calculeze "distantele" f_1, f_2, ρ, ρ^w, δ si δ_s intre cele doua profile de dizolvare.

Rezolvare:

Calculam:

T_i	R_i	$\lvert R_i - T_i \rvert$	$(R_i - T_i)^2$
0	0	0	0
15	5	10	100
20	10	10	100
40	25	15	225
60	50	10	100
90	85	5	25

$$f_1 = \frac{\sum \left| \overline{X}_{Ri} - \overline{X}_{Ti} \right|}{\sum \overline{X}_{Ri}}$$

Deci $f_1 = \dfrac{0+10+10+15+10+5}{0+5+10+25+50+85} = \dfrac{40}{175} \approx 0.23$

Factorul

$$f_2 = 50 * \lg \frac{100}{\sqrt{1 + \frac{\sum_{i=1}^{n} \left(\overline{X}_{Ri} - \overline{X}_{Ti} \right)^2}{n}}}$$

$$f_2 = 50 * \lg \frac{100}{\sqrt{1 + \frac{100+100+225+100+25}{5}}} = 50 * \lg \frac{100}{\sqrt{1+\frac{550}{5}}} = 50 * \lg \frac{100}{\sqrt{111}} = 50 * 0.98 = 48.87$$

Facem observatia ca in cazul factorului f_2 este stabilita prin recomandari ale autoritatilor de reglementare, o bariera intre similaritate si non – similaritate, intre curbe si anume valoarea de 50.

Un factor mai mare de 50 justifica decizia de similaritate.

Se admite ca profilele de dizolvare sunt similare si in cazul in care $f_2 \langle 50$, cu conditia ca amandoua produsele sa elibereze peste 85% din substanta activa in primele 15 minute.

Distantele " δ **" si "** δ_s **"** sunt: $\delta = \frac{2\sum |R_i - T_i|}{\sum |R_i + T_i|}$ si $\delta_s = \frac{4\sum |R_i - T_i|^2}{\sum |R_i + T_i|}$

T_i	R_i	$\lvert R_i - T_i \rvert$	$(R_i - T_i)^2$	$R_i + T_i$
0	0	0	0	0
15	5	10	100	20
20	10	10	100	30
40	25	15	225	65
60	50	10	100	110
90	85	5	25	175

$$\delta = 2 * \frac{10+10+15+10+5}{20+30+65+110+175} = 2 * \frac{50}{400} = 2 * \frac{1}{8} = \frac{1}{4} = 0.25$$

$$\delta_s = 4 * \frac{100+100+225+100+25}{20+30+65+110+175} = 4 * \frac{550}{400} = \frac{550}{100} = 5.50$$

Distantele " ρ **" si "** ρ^w **"(**weighted**,** fiecare punct i fiind ponderat cu $\frac{R_i + T_i}{\sum (R_j + T_j)}$ **)**

$$\rho = \frac{1}{n} \sum \max\left(\frac{R_i}{T_i}, \frac{T_i}{R_i}\right) \text{ si } \rho^w = \frac{1}{n} \frac{\sum (R_i + T_i) \max\left(\frac{R_i}{T_i}, \frac{T_i}{R_i}\right)}{\sum (R_i + T_i)}$$

T_i	R_i	$\max\left(\frac{R_i}{T_i}, \frac{T_i}{R_i}\right)$	$R_i + T_i$
0	0		0
15	5	3	20
20	10	2	30
40	25	40/25 = 1.6	65
60	50	60/50 = 1.2	110
90	85	90/85 = 1.06	175

$$\rho = \frac{1}{5} * (3 + 2 + 1.6 + 1.2 + 1.06) = \frac{1}{5} * 8.66 = 1.73$$

$$\rho^{w} = \frac{1}{5} * \frac{15*3 + 20*2 + 40*1.6 + 60*1.2 + 90*1.06}{20 + 30 + 65 + 110 + 175} = \frac{1}{5} * \frac{380.4}{400} = 0.19$$

Pentru normele de tipul δ si ρ nu exista reglementari care sa stabileasca limitele intre similaritate si non – similaritate.

Norma de siguranta d_c provenita din norma cubica $d_c(R,T) = \max(R_i, T_i)$ va da in acest caz:

$d_c(R,T) = 15$

deci curbele nu difera in nici un punct cu mai mult de 15%

1.1. CÂMPURI DE PROBABILITATE

In teoria probabilităţilor fiecărui rezultat posibil al unui experiment aleator, rezultat considerat ca eveniment, i se asociază o masura numerică, numită "probabilitatea" evenimentului respectiv. Această valoare este o caracteristică obiectivă a evenimentului în condiţiile experimentului dat.

Evenimentele pot fi simple, în sensul că nu se pot descompune mai departe in alte evenimente, sau compuse din alte evenimente ce se petrec mai mult sau mai putin simultan. În acest context putem considera două operaţii între evenimente.

Deoarece evenimentele asociate unui experiment aleator sunt parti ale unui eveniment total E este posibil sa consideram ca operatiile cu evenimente sunt operatii cu multimi.

Scriem $A \cap B$ (eveniment intersectie) şi înţelegem prin aceasta un eveniment care constă în producerea evenimentelor A şi B, simultan. Scriem $A \cup B$ (eveniment reuniune) pentru cazul când se produce cel puţin unul din cele două evenimente. Nerealizarea evenimentului A este un eveniment ce se numeste eveniment opus sau contrar lui A si se noteaza $\overline{A}$ sau CA.

Pentru operatiile cu evenimente functioneaza aceleasi proprietati ale lor din teoria multimilor:

1.	$A \cup B = B \cup A$	$A \cap B = B \cap A$
2.	$A \cup (B \cup C) = (A \cup B) \cup C$	$A \cap (B \cap C) = (A \cap B) \cap C$
3.	$A \cup \Phi = A$	$A \cap \Phi = \Phi$
4.	$A \cup A = A$	$A \cap A = A$
5.	$A \cup E = E$	$A \cap E = A$
6.	$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
7.	$A \cup \overline{A} = E$	$A \cap \overline{A} = \Phi$
8.	$\overline{E} = \Phi$	$\overline{\Phi} = E$
9.	$\overline{A \cup B} = \overline{A} \cap \overline{B}$	$\overline{A \cap B} = \overline{A} \cup \overline{B}$
10.	$A - B = A \cap \overline{B}$	

Daca avem o familie nevida de evenimente $\{A_i\}_{i \in I}$, unde I este o familie de indici cel mult numarabila, vom putea extinde operatiile de reuniune si de intersectie astfel:

1.	$\overline{\bigcup_{i \in I} A_i} = \bigcap_{i \in I} \overline{A_i}$	$\overline{\bigcap_{i \in I} A_i} = \bigcup_{i \in I} \overline{A_i}$
2.	$\left(\bigcup_{i \in I} A_i\right) \cap B = \bigcup_{i \in I} (A_i \cap B)$	$\left(\bigcap_{i \in I} A_i\right) \cup B = \bigcap_{i \in I} (A_i \cup B)$

CAMPURI DE PROBABILITATE

Spunem ca doua evenimente A si B sunt incompatibile daca nu se pot realiza simultan:

$$A \cap B = \Phi$$

si spunem ca sunt independente daca realizarile lor nu se influenteaza reciproc.

Un exemplu de evenimente incompatibile si independente este cel al aruncarii cu banul deoarece nu este posibil ca la aceeasi aruncare sa apara amandoua fetele si, de asemenea, aparitia unei fete (stema) nu influenteaza aparitia celeilate fete (banul).

Exemplul clasic de câmp de probabilitate finit îl constituie evenimentele ce pot apărea atunci când, dintr-o urnă în care se află bile albe şi negre se extrag n bile. Dacă proporția bilelor albe în urnă este p, şi deci a celor negre este q = 1 - p, probabilitatea evenimentului A, ca din n bile extrase, k să fie albe, este:

$$P(A) = C_n^k p^k q^{n-k}$$

De exemplu, evenimentul ca din trei bile extrase, două să fie albe - a - şi una să fie neagră - n- se poate descompune în felul următor :

A = (a a n) U (a n a) U (n a a)

şi

P(A) = P(a a n) + P(a n a) + P(n a a) = $p^2q + p^2q + p^2q = 3\ p^2q = C_3^2\, p^2q^{3-2}$

Definitie:

Fie E multimea finita a evenimentelor posibile la efectuarea unui experiment si $\wp(E)$ multimea partilor lui E.

Fie $K \subseteq \wp(E)$ o multime nevida de parti ale lui E. Ea se numeste *corp de evenimente* , daca verifica urmatoarele axiome:

1. $\forall A \in K \quad avem \quad \overline{A} \in K$
2. $\forall A, B \in K \quad avem \quad A \cup B \in K$

Exercitiu:

Daca $K \subseteq \wp(E)$ este un corp de evenimente, verificati urmatoarele proprietati:

a. $\Phi \in K \quad si \quad E \in K$

b. $\forall A_i \in K, i = \overline{1,n} \quad \Rightarrow \quad \bigcup_{i=1}^{n} A_i \in K \quad si \quad \bigcap_{i=1}^{n} A_i \in K$

c. $A, B \in K \Rightarrow A - B \in K$

Solutie:

a. Deoarece $K \neq \phi$, exista cel putin o multime $A \subseteq K$. Rezulta ca $\overline{A} \in K$, deci $A \cup \overline{A} \in K \Rightarrow E \in K$ si $\overline{E} \in K$ ceea ce inseamna ca $\Phi \in K$

b. Daca $A_i \in K, \forall i = \overline{1,n}$, atunci prin inductie completa se obtine ca $\bigcup_{i=1}^{n} A_i \in K$

Deoarece $A_i \in K, \forall i = \overline{1,n}$, avem $\overline{A_i} \in K, \forall i = \overline{1,n}$ si $\bigcup_{i=1}^{n} \overline{A_i} \in K$.

Dar $\bigcup_{i=1}^{n} \overline{A_i} = \overline{\bigcup_{i=1}^{n} A_i} \in K$ ceea ce implica $\overline{\overline{\bigcup_{i=1}^{n} A_i}} = \bigcap_{i=1}^{n} A_i \in K$

c. $A, B \in K \Rightarrow A, \overline{B} \in K \Rightarrow A \cap \overline{B} \in K \Rightarrow A - B \in K$

Definitie:

Fie E o mulțime şi K o familie nevidă de părți ale lui E, K $\subset \wp$(E) cu proprietățile:

1. A$\in$ K $\Rightarrow$CA$\in$ K
2. $(A_i)_{i\in N} \subset$ K$\Rightarrow \bigcup_1^\infty A_i \in$ K
3. E$\in$ K

Deci, este închisă la operațiile de complementare şi reuniune.

Se spune, în acest caz, că familia K, împreună cu operațiile menționate, formează un *corp bolerian*.

Definitie:

Un element $A \in K$ se numeste eveniment compus daca exista doua evenimente $B, D \in K$, $B \neq \Phi$, $D \neq \Phi$, $B \neq A$, $D \neq A$ astfel incat $A = B \cup D$. Un eveniment $A \neq \Phi$ ce nu este compus se numeste eveniment elementar.

Definitie:

Fiind dat un spațiu măsurabil (E, K). O funcție P: $K \to [0,1]$ cu proprietățile:

a) P – măsură şi
b) P(E)=1

se numeşte probabilitate.

Deci, probabilitatea ar fi o măsură “normată”.

CAMPURI DE PROBABILITATE

Definitie:

Se numeşte *măsură* orice funcţie pozitivă definită pe corpul mulţimilor măsurabile, $\mu : K \to R_+$, "aditivă" pe orice familie $(A_i)_{i \in I}$ numărabilă de mulţimi măsurabile disjuncte:

$$\forall n, \forall m, A_n \cap A_m = \Phi \Rightarrow \mu\left(\bigcup_1^\infty A_n\right) = \sum_1^\infty \mu(A_n)$$

Exercitii:

Fie $A \quad si \quad B \in K$. Verificati urmatoarele proprietati:

a. $\forall A \in K \quad avem \quad P(CA) = 1 - P(A)$

b. $P(\Phi) = 0$

c. Daca $A \subset B$, atunci $P(A) \le P(B)$

d. $\forall A \in K \quad avem \quad 0 \le P(A) \le 1$

e. $P(B - A) = P(B) - P(A \cap B)$

f. $P(B - A) = P(B) - P(A)$, daca $A \subset B$

g. $P(B \cup A) = P(A) + P(B) - P(A \cap B)$ daca $A \cap B \ne \Phi$

h. $P\left(\bigcap_{i=1}^{n} A_i\right) \ge 1 - \sum_{i=1}^{n} P(CA_i)$ (inegalitatea lui Boole)

Solutie:

a. $\forall A \in K \quad avem \quad A \cup \overline{A} = E \quad si \quad A \cap \overline{A} = \Phi$

$$P\left(A \cup \overline{A}\right) = P(A) + P\left(\overline{A}\right) = P(E) = 1 \Rightarrow P(CA) = 1 - P(A)$$

b. $\Phi = CE \Rightarrow P(\Phi) = P(CE) = 1 - P(E) = 1 - 1 = 0$

c. $A \subset B \Rightarrow B = A \cup (B \cap CA)$ dar $A \cap (B \cap CA) = \Phi$, deci

$$P(B) = P(A) + P(B \cap CA) \ge P(A)$$

d. $\forall A \in K \quad avem \quad \Phi \subseteq A \subseteq E \Rightarrow P(\Phi) \le P(A) \le P(E) \Rightarrow q.e.d.$

e. $B = B \cap E = B \cap (A \cup CA) = (B \cap A) \cup (B \cap CA)$

Dar $(B \cap A) \cup (B \cap CA) = \Phi$, deci

$P(B) = P(B \cap A) + P(B \cap CA) = P(B \cap A) + P(B - A) \Rightarrow q.e.d.$

f.

$A \subset B \Rightarrow A \cap B = A \Rightarrow P(B) = P(B - A) = P(B) - P(B \cap A) = P(B) - P(A)$

g. Deoarece $A \cap B \ne \Phi$ putem scrie:

$A = A \cap E = A \cap (B \cup CB) = (A \cap B) \cup (A \cap CB)$

$B = B \cap E = B \cap (A \cup CA) = (B \cap A) \cup (B \cap CA)$

Deci, $A \cup B = (A \cap B) \cup (A \cap CB) \cup (B \cap CA)$

Cum evenimentele $A \cap B$, $A \cap CB$, $B \cap CA$ sunt incompatibile doua cate doua obtinem:

$P(A \cup B) = P(A \cap B) + P(A \cap CB) + P(B \cap CA)$

Dar $P(A \cap CB) = P(A) - P(B)$ si $P(B \cap CA) = P(B) - P(A)$ de unde obtinem

$$P(B \cup A) = P(A) + P(B) - P(A \cap B)$$

h. Deoarece $C\left(\bigcap_i A_i\right) = \bigcup_i CA_i$ vom aplica probabilitatea obtinandu-se

$$P\left(C\left(\bigcap_i A_i\right)\right) = P\left(\bigcup_i CA_i\right) \Rightarrow 1 - P\left(\bigcap_i A_i\right) = P\left(\bigcup_i CA_i\right) \leq \sum_i P(CA_i) \Rightarrow$$

$$\Rightarrow P\left(\bigcap_i A_i\right) \geq 1 - \sum_i P(CA_i)$$

Definitie:Probabilitatea condiționată

Fie B un eveniment a cărei probabilitate este diferită de 0. Probabilitatea unui eveniment A, reprezintă proporția în care ne aşteptăm să se realizeze A în cadrul tuturor evenimentelor câmpului de probabilitate la care aparține A

Probabilitatea lui A se mai poate analiza însă şi în contextul în care ştim că s-a produs anterior evenimentul B. Probabilitatea evenimentului A condiționată de B se notează, în acest caz, cu: P(A/B) sau $P_B(A)$.

Dacă s-a constatat experimental o frecvență de apariție k_A şi, respectiv k_B, pentru A şi B, *frecvența relativă de apariție a lui A*, când deja a apărut B, va fi:

$$\frac{k_{AB}}{k_B} = \frac{\frac{k_{AB}}{n}}{\frac{k_B}{n}} \cong \frac{P(A \cap B)}{P(B)}$$

În acest context apare naturală definiția probabilității evenimentului A, condiționată de B, prin formula:

$$P_B(A) = \frac{P(A \cap B)}{P(B)}$$

Exercitiu:

Demonstrati ca raportul de mai sus verifica axiomele probabilitatilor:

a. $P_B(A) \geq 0$

b. $P_B(E) = 1$

c. $P_B(A \cup D) = P_B(A) + P_B(D)$, daca $A \cap D = \Phi$

Demonstratie:

a. Deoarece $P(A) \rangle 0$ si $P(A \cap B) \geq 0$ obtinem inegalitatea ceruta.

b. Conform definitiei avem:

$$P_B(E) = \frac{P(E \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

c.

$$P_B(A \cup D) = \frac{P[(A \cup D) \cap B]}{P(B)} = \frac{P[(B \cap A) \cup (D \cap B)]}{P(B)} =$$

$$= \frac{P(B \cap A) + P(D \cap B)}{P(B)} = \frac{P(B \cap A)}{P(B)} + \frac{P(D \cap B)}{P(B)} =$$

$$= P_B(A) + P_B(D)$$

Teorema probabilității cauzelor

Probabilitatea producerii oricărui eveniment X, este egală cu suma probabilităților de producere a lui X, condiționate de evenimentele complete ale sistemului $(A_i)_{i = 1, n}$ şi

$$P_X(A_j) = \frac{P(A_j) P_{A_j}(X)}{\sum P(A_i) P_{A_i}(X)}$$

CAMPURI DE PROBABILITATE

Exercitii:

1. Masa, rezistenta si inaltimea sunt caracteristici independente ale unui comprimat. Probabilitatea ca un comprimat sa nu corespunda din aceste puncte de vedere sunt: 0,03; 0,05 si 0,02. Care este probabilitatea ca tableta sa corespunda in raport cu cele trei caracteristici?

Solutie:

Fie E_1, E_2, E_3 evenimentele care se realizeaza cand produsul corespunde in raport cu fiecare dintre caracteristici.

$$P(E_1) = 1 - P(CE_1) = 0.97$$

$$P(E_2) = 0.95$$

$$P(E_3) = 0.98$$

Daca CE_1 , CE_2 , CE_3 sunt independente si E_1, E_2, E_3 sunt independente. Asadar:

$$P(E_1 \cap E_2 \cap E_3) = 0.97 * 0.95 * 0.98 = 0.9031$$

2. Cu datele din problema precedenta sa se calculeze probabilitatea ca produsul sa nu corespunda.

Solutie:

$$P(CE_1) = 0.03$$

$$P(CE_2) = 0.05$$

$$P(CE_3) = 0.02$$

Se utilizeaza relatia

$$P(A \cup B \cup D) = P(A) + P(B) + P(D) - P(A \cap B) - P(A \cup D) - P(B \cap D) +$$
$$+ P(A \cap B \cap D)$$

si se obtine

$$P(CE_1 \cup CE_2 \cup CE_3) = 0.03 + 0.05 + 0.02 - 0.03 * 0.05 - 0.03 * 0.02 -$$
$$- 0.05 * 0.02 + 0.03 * 0.05 * 0.02 = 0.0969$$

3. In 5% comprimate rezistenta este necorespunzatoare din cauza nerespectarii formulei de fabricatie, iar 10% din cauza reglajului incorect al masinii de comprimat. Care este probabilitatea ca rezistenta comprimatului sa fie buna?

Solutie:

Fie A(B) evenimentul care se realizeaza cand rezistenta nu corespunde din cauza formulei de fabricatie.

P(A)=0.05 si P(B)=0.10

Se calculeaza $P(CA \cap CB) = P[C(A \cup B)] = 1 - P(A \cup B)$

Dar $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.05 + 0.10 - 0.05 * 0.10 = 0.145$
Deci, $P(CA \cap CB) = 0.855$

4. O capsula este considerata corespunzatoare standardului daca indeplineste conditiile A_1, A_2, A_3, A_4. Datele statistice arata ca 90% dintre capsule indeplinesc conditia A_1, 80% indeplinesc conditia A_2, 85% indeplinesc conditia A_3 si 95% indeplinesc conditia A_4. Care este probabilitatea minima ca o capsula sa corespunda standardului?
Solutie
Se aplica inegalitatea lui Boole
$P(\cap A_i) \geq 1 - \sum P(CA_i)$
$P(CA_i) = 1 - P(A_i)$
Deci $P(\cap A_i) \geq \sum_{i=1}^{n} P(A_i) - (n-1)$
Obtinem $P(A_1 \cap A_2 \cap A_3 \cap A_4) \geq P(A_1) + P(A_2) + P(A_3) + P(A_4) - 3 = 0.5$
Deci probabilitatea ca o capsula sa fie corespunzatoare este cuprinsa intre 0.5 si 1.

5. O instalatie este deservita de trei pompe cu functionare independenta a caror probabilitate de defectare este 0.1 ; 0.15 si 0.25. Instalatia trebuie oprita daca se defecteaza prima pompa sau pompele 2 si 3 simultan. Daca se defecteaza numai una dintre pompele 2 sau 3 instalatia poate functiona. Care este probabilitatea ca instalatia sa functioneze?
Solutie
Fie A_1, A_2, A_3 evenimentele care corespund functionarii pompelor 1, 2 si respectiv 3 si fie A evenimentul cere se realizeaza cand instalatia functioneaza.
$$A = (A_1 \cap A_2 \cap A_3) \cup (A_1 \cap CA_2 \cap A_3) \cup (A_1 \cap A_2 \cap CA_3)$$
$$P(A) = P(A_1 \cap A_2 \cap A_3) + P(A_1 \cap CA_2 \cap A_3) + P(A_1 \cap A_2 \cap CA_3) =$$
$$= P(A_1)P(A_2)P(A_3) + P(A_1)P(CA_2)P(A_3) + P(A_1)P(A_2)P(CA_3) =$$
$$= 0.9 * 0.85 * 0.75 + 0.9 * 0.15 * 0.75 + 0.9 * 0.85 * 0.25 = 0.866$$

6. Un produs este prelucrat in doua etape A si B. Dupa etapa A piesele sunt controlate, iar cele necorespunzatoare sunt reprelucrate. Experientele arata ca piesele corespund dupa etapa A in proportie de 97%, iar dupa etapa B in proportie de 95%. Care este probabilitatea ca o piesa sa corespunda?

Solutie

Conform definitiei probabilitatilor conditionate avem: $P_A(B) = \dfrac{P(B \cap A)}{P(A)}$

$\Rightarrow P(A \cap B) = P(A) * P_A(B) = 0.97 * 0.95 = 0.9215$

7. Se considera doua recipiente de reactiv B_1 si B_2. In recipientul B_1 se afla pastile de KOH, iar in recipientul B_2 pastile de KOH si de NaOH in numar egal. O pastila scoasa la intamplare din unul din recipienti se dovedeste a fi KOH. Care este probabilitatea ca aceasta pastila sa provina din B_1?

Solutie

Fie A evenimentul ca pastila sa fie de KOH.

Se aplica relatia lui Bayes:

$$P_A(B_1) = \frac{P(B_1)P_{B_1}(A)}{P(B_1)P_{B_1}(A) + P(B_2)P_{B_2}(A)}$$

$$P(B_1) = P(B_2) = \frac{1}{2}$$

$$P_{B_1}(A) = 1$$

$$P_{B_2}(A) = \frac{1}{2}$$

Deci $P_A(B_1) = \dfrac{2}{3}$

8. Se considera cinci loturi de comprimate cu structurile:

a. Doua loturi cu 60% comprimate corespunzatoare;

b. Doua loturi cu 55% comprimate corespunzatoare;

c. Un lot cu 70% comprimate corespunzatoare.

Loturile constau din acelasi numar de piese. Se face controlul unui comprimat luat la intamplare.

a) Care este probabilitatea ca acest comprimat sa fie necorespunzator?

b) Daca se presupune ca acest comprimat este necorespunzator care este probabilitatea ca acesta sa provina dintr-un lot de tipul 2?

Solutie

a) Se noteaza cu B evenimentul de a controla un comprimat necorespunzator si cu A_1, A_2, A_3 evenimentele care constau din efectuarea controlului unui comprimat din loturile 1, 2 sau respectiv 3.

$P(B) = \sum_i P(A_i) P_{A_i}(B)$

$P(A_1) = \frac{2}{5}$, $P_{A_1}(B) = 0.4$

$P(A_2) = \frac{2}{5}$, $P_{A_2}(B) = 0.45$

$P(A_3) = \frac{1}{5}$, $P_{A_3}(B) = 0.3$

Deci, $P(B) = 0.4 * \frac{2}{5} + 0.45 * \frac{2}{5} + 0.3 * \frac{1}{5} = \frac{1}{5}$

b) $P_B(A_2) = \frac{P(A_2) P_{A_2}(B)}{P(B)} = \frac{0.45 * \frac{2}{5}}{\frac{2}{5}} = 0.45$

9. Zece loturi dintre care trei de tipul A_1, cinci de tipul A_2 si doua de tipul A_3 trec verificarile la care sunt supuse in proportie de 90%, 75% si respectiv 85%.

a) Care este probabilitatea ca un lot ales la intamplare sa fie corespunzator.

b) Care este probabilitatea ca un lot corespunzator sa fie de tipul A_1.

Solutie

$P(A_1) = 0.3$, $P(A_2) = 0.5$, $P(A_3) = 0.2$

Se noteaza cu B evenimentul care constata in faptul ca lotul ales trece verificarile.

$P(B / A_1) = 0.9$, $P(B / A_2) = 0.75$, $P(B / A_3) = 0.85$

$P(B) = \sum_i P(A_i) P(B / A_i) = 0.815$

$P(A_1 / B) = \frac{P(A_1) P(B / A_1)}{P(B)} = 0.33$

1.2. VARIABILE ALEATOARE

Definiții:

a) Se numeşte *variabilă aleatoare* (întâmplătoare sau statistică) o funcție reală f definită pe mulțimea $\mathcal{K}$ a evenimentelor, cu proprietatea că, oricare ar fi numărul real a, mulțimea $x \in \mathcal{K}$ pentru care $f(x) \leq a$ este un eveniment din $\mathcal{K}$.

În termeni de teoria măsurii, o variabilă aleatoare este o funcție

$f : (E, \mathcal{K}, P) \rightarrow (R, B)$, măsurabilă.

Practic vorbind avem definită probabilitatea ca variabila să aibă valori mai mici decât orice număr dat a.

b) O variabilă aleatoare se numeşte *variabilă aleatoare simplă* dacă ia un număr finit de valori: $f : E \rightarrow R$, $f(E)$ finită şi

$$P(f (x) = x_i) = P(f^{-1}(x_i)) = p_i$$

c) Doua *variabile aleatoare sunt independente*, daca iau valori independente una de cealaltă:

$$P\big((f(x) = x_i) \cap (g(y) = y_j)\big) = P(f(x) = x_i) * P\big(g(y) = y_j\big), \ \forall x_i, y_j$$

Exemplu:

Fie X o variabila aleatoare discreta (v.a.) avand tabelul de distributie:

$$X : \begin{pmatrix} x_1 & x_2 \ldots & x_n \\ p_1 & p_2 \ldots & p_n \end{pmatrix} \text{ unde } \begin{cases} p_i \geq 0, \quad i = \overline{1,n} \\ \sum_{i=1}^{n} p_i = 1 \end{cases}$$

atunci functia de repartitie corespunzatoare va fi:

$$F(x) = \begin{cases} 0, \quad daca \quad x \leq x_1 \\ p_1, \quad daca \quad x_1 < x \leq x_2 \\ p_1 + p_2, \quad daca \quad x_2 < x \leq x_3 \\ . \\ \sum_{i=1}^{k-1} p_i, \quad daca \quad x_{k-1} < x \leq x_k \\ . \\ 1, \quad daca \quad x > x_n \end{cases}$$

Definitie:

Fie $F(x)$ functia de repartitie a unei v.a. X. Daca exista o functie integrabila $f(x)$ astfel incat

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

atunci X se numeste *variabila aleatoare continua*, iar $f(x)$ se numeste *densitatea de probabilitate* sau *densitatea de repartitie* a lui X

Caracteristici ale variabilelor aleatoare

Media:

Se numeşte *valoare medie* (sau speranță matematică) a unei valori aleatoare f, numărul

M(f) = $\sum x_i p_i$, atunci când ξ este o variabilă aleatoare simplă şi, respectiv

M(f) = $\int_{-\infty}^{+\infty} x\rho(x)dx$, atunci când ξ este o variabilă aleatoare continuă, cu densitatea de probabilitate ρ.

Momentul de ordin k al unei variabile aleatoare reprezinta generalizarea notiunii de medie:

$M_k(f) = \sum x_i^k p_i$, atunci când ξ este o variabilă aleatoare simplă

şi respectiv,

$M_k(f) = \int_{-\infty}^{+\infty} x^k \rho(x)dx$, atunci când ξ este o variabilă aleatoare continuă.

Momentul centrat de ordin k al unei variabile aleatoare f este momentul de ordinul k al abaterii sale față de medie.

$$M_k^c(f) = \sum (x_i - \mu_f)^k p_i$$

şi respectiv, $\mu_k^c = \int_{-\infty}^{+\infty} [x - M(f)]^k \rho(x)dx$,în cazul unei variabile aleatoare continue.

Dispersia variabilei aleatoare X de notează D(X) sau σ^2 şi este, în particular, momentul centrat de ordinul doi.

$$D(X) = \sigma^2 = M[(X-M(X))^2] = \int_{-\infty}^{+\infty}\left(x - M(X)\right)^2 \rho(x)dx$$

şi respectiv

$\sigma^2 = M[(X-M(X))^2] = \sum\left(x_i - \mu_X\right)^2 p_i$, atunci când variabila aleatoare este discretă.

Rădăcina pătrată a dispersiei, σ, se numeşte ***abaterea medie pătratică*** a variabilei X, iar s_x abaterea standard.

Exercitiu:

Verificati urmatoarele proprietati ale mediei: dacă f şi g sunt independente, atunci avem:

a) M(af) = aM(f)
b) M(f+g) = M(f) + M(g)
c) M(fg) = M(f)M(g)

Solutie:

Fie variabilele independente:

$$f : \begin{pmatrix} f_1 & f_2 \cdots & f_n \\ p_1 & p_2 \cdots & p_n \end{pmatrix} \text{ unde } \begin{cases} p_i \geq 0, & i = \overline{1,n} \\ \sum_{i=1}^{n} p_i = 1 \end{cases} \quad \text{si}$$

$$g : \begin{pmatrix} g_1 & g_2 \cdots & g_m \\ q_1 & q_2 \cdots & q_m \end{pmatrix} \text{ unde } \begin{cases} q_j \geq 0, & j = \overline{1,m} \\ \sum_{j=1}^{m} q_j = 1 \end{cases}$$

a) $M(af) = \sum_i af_i p_i = a\sum_i f_i p_i = aM(f)$

b) $M(f+g) = \sum_{i,j}\left(f_i + g_j\right)p\left[\left(f = f_i\right)\cap\left(g = g_j\right)\right] =$

$$= \sum_{i,j} (f_i + g_j)[p(f = f_i) * p(g = g_j)] = \sum_{i,j} (f_i + g_j) p_i q_j =$$

$$= \sum_{i,j} (f_i p_i q_j + g_j p_i q_j) = \sum_{i,j} f_i p_i q_j + \sum_{i,j} g_j p_i q_j =$$

$$= \sum_i \left(\sum_j f_i p_i q_j \right) + \sum_j \left(\sum_i g_j p_i q_j \right) =$$

$$= \sum_i f_i p_i \sum_j q_j + \sum_j g_j q_j \sum_i p_i = \sum_i f_i p_i + \sum_j g_j q_j = M(f) + M(g)$$

c) $M(f * g) = \sum_{i,j} (f_i * g_j) p[(f = f_i) \cap (g = g_j)] =$

$$= \sum_{i,j} (f_i * g_j)[p(f = f_i) * p(g = g_j)] = \sum_{i,j} (f_i * g_j) p_i q_j =$$

$$= \sum_{i,j} f_i g_j p_i q_j = \sum_i \sum_j f_i g_j p_i q_j = \sum_i f_i p_i \left(\sum_j g_j q_j \right) = \sum_i f_i p_i M(g) =$$

$$= M(g) \sum_i f_i p_i = M(g) M(f)$$

Exercitiu:
Verificati urmatoarele proprietati ale dispersiei:

a) Pentru orice variabilă aleatoare X şi orice constante a şi b
$D(aX+b) = a^2 D(X)$

b) Dacă X, Y sunt două variabile aleatoare independente
$D(X+Y) = D(X) + D(Y)$

c) Între dispersie, valoarea medie şi momentul de ordinul doi există relația:

$$D(X) = M(X^2) - (M(X))^2$$

Solutie:

a) $D(X) = \sum_i (x_i - \mu)^2 p_i$

$$D(aX + b) = \sum_i ((ax_i + b) - (a\mu + b))^2 p_i = \sum_i (ax_i + b - a\mu - b)^2 p_i =$$

$$= \sum_i (ax_i - a\mu)^2 p_i = \sum_i (a(x_i - \mu))^2 p_i = \sum_i a^2 (x_i - \mu)^2 p_i =$$

$$= a^2 \sum_i (x_i - \mu)^2 p_i = a^2 D(X)$$

b)

$$D(X+Y)=\sum_{i,j}\left(\left(x_i+y_j\right)-\left(\mu_X+\mu_Y\right)\right)^2 p_i q_j =$$

$$=\sum_{i,j}\left(\left(x_i-\mu_X\right)+\left(y_j-\mu_Y\right)\right)^2 p_i q_j =$$

$$=\sum_{i,j}\left(x_i-\mu_X\right)^2 p_i q_j+\sum_{i,j}\left(y_j-\mu_Y\right)^2 p_i q_j-2\sum_{i,j}\left(x_i-\mu_X\right)\left(y_j-\mu_Y\right)p_i q_j =$$

$$=\sum_i\sum_j\left(x_i-\mu_X\right)^2 p_i q_j+\sum_j\sum_i\left(y_j-\mu_Y\right)^2 p_i q_j-2\sum_i\sum_j\left(x_i-\mu_X\right)\left(y_j-\mu_Y\right)p_i q_j =$$

$$=\sum_i\left(x_i-\mu_X\right)^2 p_i\sum_j q_j+\sum_j\left(y_j-\mu_Y\right)^2 q_j\sum_i p_i-2\sum_i\left(x_i-\mu_X\right)p_i\sum_j\left(y_j-\mu_Y\right)q_j =$$

$$=\sum_i\left(x_i-\mu_X\right)^2 p_i+\sum_j\left(y_j-\mu_Y\right)^2 q_j-2*0=D(X)+D(Y)$$

c) D(X) = $\sum\left(x_i-\mu_X\right)^2 p_i = \sum x_i^2 p_i$ -2$\sum x_i\mu_X p_i+\sum\mu_X^2 p_i=$

$$=\sum_i x_i^{\,2} p_i-2\mu_X\sum_i x_i p_i+\mu_X^{\,2}\sum_i p_i=\sum_i x_i^{\,2} p_i-2\mu_X*\mu_X+\mu_X^{\,2}*1=$$

$$=\sum_i x_i^{\,2} p_i-\mu_X^{\,2}=M\left(X^2\right)-\left(M(X)\right)^2$$

Deci, M(X^2) - 2 $\mu_X^2+\mu_X^2$ = M(X^2) – (M(X))^2

Exercitii:

1. Fie urmatoarea variabila aleatoare:

$$X:\begin{pmatrix} 1 & 2 & 3 & 4 \\ 0,2 & 0,1 & 0,3 & 0,4 \end{pmatrix}$$

Sa se determine functia sa de repartitie.

Solutie:

$$F(x)=\begin{cases} 0, \quad daca \quad x\le 1 \\ 0,2 \quad daca \quad 1\langle x\le 2 \\ 0,2+0,1=0,3 \quad daca \quad 2\langle x\le 3 \\ . \\ 0,2+0,1+0,3=0,6 \quad daca \quad 3\langle x\le 4 \\ . \\ 1, \quad daca \quad x\rangle 4 \end{cases}$$

VARIABILE ALEATOARE

2. La o farmacie a fost inregistrat numarul de antibiotice cerut zilnic pe o perioada de 50 de zile, obtinandu-se urmatoarele valori:

Cererea zilnica	3	4	5	6	7	8
Numarul de zile	3	7	12	14	10	4

a. Sa se reprezinte tabelul de distributie a variabilei aleatoare reprezentand cererea zilnica de antibiotice;

b. Sa se determine functia de repartitie corespunzatoare

c. Care este probabilitatea ca numarul cererilor sa fie cuprins intre 4 si 7, putand lua valoarea 4 sau 7;

d. Care este probabilitatea ca cererea de antibiotice sa fie mai mare de 6;

Solutie:

a. $X:\begin{pmatrix} 3 & 4 & 5 & 6 & 7 & 8 \\ \frac{3}{50} & \frac{7}{50} & \frac{12}{50} & \frac{14}{50} & \frac{10}{50} & \frac{4}{50} \end{pmatrix}$ deci

$$X:\begin{pmatrix} 3 & 4 & 5 & 6 & 7 & 8 \\ 0{,}06 & 0{,}14 & 0{,}24 & 0{,}28 & 0{,}20 & 0{,}08 \end{pmatrix}$$

b. Din definitia functiei de repartitie avem $F(x)=P(X\langle x)$, deci:

$$F(x)=\begin{cases} 0, \quad daca \quad x\le 3 \\ 0{,}06 \quad daca \quad 3\langle x\le 4 \\ 0{,}06+0{,}14=0{,}20 \quad daca \quad 4\langle x\le 5 \\ 0{,}06+0{,}14+0{,}24=0{,}44 \quad daca \quad 5\langle x\le 6 \\ 0{,}06+0{,}14+0{,}24+0{,}28=0{,}72 \quad daca \quad 6\langle x\le 7 \\ 0{,}06+0{,}14+0{,}24+0{,}28+0{,}20=0{,}92 \quad daca \quad 7\langle x\le 8 \\ 1 \quad daca \quad x\rangle 8 \end{cases}$$

c.

$$P(4\le X\le 7)=P(4\le X\langle 7)+P(X=7)=F(7)-F(4)+P(X=7)=$$
$$=0{,}72-0{,}06+0{,}20=0{,}86$$

d.

$$P(X\rangle 6)=1-P(X\le 6)=1-[P(X\langle 6)+P(X=6)]=1-F(6)-P(X=6)=$$
$$=1-0{,}44-0{,}28=0{,}28$$

3. Sa se determine $x \in R$ astfel incat variabila aleatoare X sa aiba repartitia:

$$X : \begin{pmatrix} a & b \\ x^2 + x & \frac{x}{4} + \frac{1}{8} \end{pmatrix}$$

Solutie:

Conditiile impuse sunt: $\begin{cases} x^2 + x \in [0,1] \\ \frac{x}{4} + \frac{1}{8} \in [0,1] \\ x^2 + x + \frac{x}{4} + \frac{1}{8} = 1 \end{cases} \Rightarrow x = \frac{1}{2}$

4. Sa se calculeze media si dispersia pentru urmatoarea variabila aleatoare $X : \begin{pmatrix} 1 & 2 \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}$

Solutie:

$$M(X) = \sum_i x_i p_i = 1 * \frac{1}{3} + 2 * \frac{2}{3} = \frac{5}{3} = 1.66$$

$$D(X) = \sum_i x_i^2 p_i = 1^2 * \frac{1}{3} + 2^2 * \frac{2}{3} = \frac{9}{3} = 3$$

5. Sa se calculeze suma urmatoarelor variabile aleatoare:

$X : \begin{pmatrix} 1 & 2 \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}$ si $Y : \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$

Solutie:

$$X + Y : \begin{pmatrix} 1 & 2 & 3 \\ p_1 & p_2 & p_3 \end{pmatrix}$$

$$p_1 = P(X + Y = 1) = P((X = 1) \cap (Y = 0)) = P(X = 1) * P(Y = 0) =$$
$$= \frac{1}{3} * \frac{1}{2} = \frac{1}{6}$$

$$p_2 = P(X + Y = 2) = P[((X = 1) \cap (Y = 1)) \cup ((X = 2) \cap (Y = 1))] =$$
$$= P(X = 1) * P(Y = 1) + P(X = 2) * P(Y = 0) = \frac{1}{3} * \frac{1}{2} + \frac{2}{3} * \frac{1}{2} = \frac{3}{6} = \frac{1}{2}$$

$p_3 = P(X+Y=3) = P((X=2)\cap(Y=1)) = P(X=2)*P(Y=1) =$

$= \frac{2}{3}*\frac{1}{2} = \frac{2}{6} = \frac{1}{3}$

Deci, $X+Y : \begin{pmatrix} 1 & 2 & 3 \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3} \end{pmatrix}$

6. Cunoscandu-se urmatoarea variabila aleatoare $X : \begin{pmatrix} 1 & 2 \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}$ sa se calculeze $D(3X+2)$

Solutie:

Vom calcula $\mu(x) = \sum_i x_i p_i = 1*\frac{1}{3} + 2*\frac{2}{3} = \frac{5}{3}$

$D(3X+2) = 3^2 D(X) = 9*D(X) = 9*\sum_i (x_i-\mu)^2 p_i =$

$= 9*\left(\left(1-\frac{5}{3}\right)^2 *\frac{1}{3} + \left(2-\frac{5}{3}\right)^2 *\frac{2}{3}\right) =$

$= 9*\left(\frac{4}{9}*\frac{1}{3} + \frac{1}{9}*\frac{2}{3}\right) = \frac{6}{3} = 2$

7. Sa se calculeze media si dispersia pentru suma urmatoarelor variabile aleatoare:

$X : \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$ si $Y : \begin{pmatrix} 0 & 1 & 2 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{pmatrix}$

Solutie:

Vom calcula media si dispersia pentru cele doua variabile aleatoare:

$M(X) = \sum_i x_i p_i = 0*\frac{1}{2} + 1*\frac{1}{2} = \frac{1}{2}$ si

$D(X) = \sum_i (x_i-\mu)^2 p_i = \left(0-\frac{1}{2}\right)^2 *\frac{1}{2} + \left(1-\frac{1}{2}\right)^2 *\frac{1}{2} = \frac{1}{4}*\frac{1}{2} + \frac{1}{4}*\frac{1}{2} = \frac{1}{4}$

Analog, $M(Y) = \sum_i y_i q_i = 0*\frac{1}{4} + 1*\frac{1}{4} + 2*\frac{1}{2} = \frac{5}{4}$ si

$$D(Y)=\sum_i (y_i-\mu)^2 q_i=\left(0-\frac{5}{4}\right)^2*\frac{1}{4}+\left(1-\frac{5}{4}\right)^2*\frac{1}{4}+\left(2-\frac{5}{4}\right)^2*\frac{1}{2}=$$

$$=\frac{25}{16}*\frac{1}{4}+\frac{1}{16}*\frac{1}{4}+\frac{9}{16}*\frac{1}{2}=\frac{44}{32}=\frac{11}{8}$$

$$M(X+Y)=M(X)+M(Y)=\frac{1}{2}+\frac{5}{4}=\frac{7}{4}$$

$$D(X+Y)=D(X)+D(Y)=\frac{5}{4}+\frac{11}{8}=\frac{21}{8}$$

8. Un transport de recipienți cu materie primă sunt controlați astfel: se analizează conținutul unui recipient ales la întâmplare şi se stabileşte dacă poate fi acceptat sau nu. Se cercetează 6 recipienți. Dacă recipientul controlat la extracția de rang k=1,2,3,4,5 nu corespunde întregul transport este respins şi se opreşte analiza.

a) Care este repartiția variabilei aleatoare care dă numărul de recipienți analizați? Se ştie că probabilitatea ca un anumit recipient să corespundă este de 2/3.

b) Să se calculeze media acestei variabile.

Rezolvare:

Variabila aleatoare X ia valorile:

X=1 dacă primul recipient este respins $\Rightarrow$

P(X=1)=P(primul recipient este respins)=$\frac{1}{3}$

X=2 dacă primul recipient corespunde, dar al doilea recipient este respins $\Rightarrow$ P(X=2)=P(primul recipient corespunde şi al doilea recipient este respins)=P(primul recipient corespunde)*P(al doilea recipient este respins)=$\frac{2}{3}*\frac{1}{3}=\frac{2}{3^2}$

X=3 dacă primele două recipiente corespund, dar al treilea recipient este respins

$\Rightarrow$ P(X=3)=P(primul recipient corespunde, al doilea recipient corespunde şi al treilea recipient este respins)=P(primul recipient corespunde)*P(al doilea recipient corespunde)*P(al treilea recipient este respins)=$\frac{2}{3}*\frac{2}{3}*\frac{1}{3}=\frac{2^2}{3^3}$

.
.
.

X=6 dacă toate recipientele corespund $\Rightarrow$

P(X=6)=P(primele 5 recipientele corespund)=$\left(\frac{2}{3}\right)^5$

Deci, repartiția de probabilitatea este:

$$X: \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{3} & \frac{2}{3^2} & \frac{2^2}{3^3} & \frac{2^3}{3^4} & \frac{2^4}{3^5} & \frac{2^5}{3^5} \end{pmatrix}$$

Media acestei variabile este:

$$M(X) = 1*\frac{1}{3} + 2*\frac{2}{3^2} + 3*\frac{2^2}{3^3} + 4*\frac{2^3}{3^4} + 5*\frac{2^4}{3^5} + 6*\frac{2^5}{3^5}$$

9. Fie q probabilitatea ca o reacție de policondensare să se producă şi $p = 1 - q$ probabilitatea ca reacția să se întrerupă. Să se calculeze repartiția, media şi dispersia variabilei aleatoare care dă gradul de policondensare.

Solutie:

Probabilitatea formării unui polimer cu gradul de condensare n, care să conțină n-1 legături formate prin policondensare este q^{n-1}, înmulțită cu probabilitatea de întrerupere $1 - q = p$. Deci repartiția variabilei aleatoare este:

$$X: \begin{pmatrix} 1 & 2 & 3 & .. & n & . \\ p & pq & pq^2 & .. & pq^{n-1} & . \end{pmatrix}$$

Se observă că $\sum_{n=1} pq^{n-1} = \frac{p}{q-1} = 1$

$$E(X) = p + 2pq + 3pq^2 + ... + npq^{n-1} + ... = p(1 + 2q + 3q^2 + ... + nq^{n-1} + ...)$$

Deoarece din $q + q^2 + ... + q^n + ... = \frac{q}{1-q}$ prin derivare se obtine $1 + 2q + 3q^2 + ... + nq^{n-1} = \frac{1}{(1-q)^2}$ rezulta ca

$$E(X) = \frac{p}{(1-q)^2} = \frac{1}{p}$$

$$D(X) = E(X^2) - E^2(X)$$

$$E(X^2) = p + 2^2 pq + 3^2 pq^2 + ... + n^2 pq^{n-1} + ... = p(1 + 2^2 q + 3^2 q^2 + ... + n^2 q^{n-1} + ...)$$

Dar $q+2q^2+3q^3+...+nq^n = \dfrac{q}{(1-q)^2}$

Prin derivare rezulta:

$E(X^2)= p(1+2^2q+3^2q^2+...+n^2q^{n-1}+...)=\dfrac{(1+q)p}{(1-q)^3}$

$(E(X))^2 = \dfrac{1}{p^2}$

Rezulta $D(X)=\dfrac{p(1+q)}{(1-q)^3}-\dfrac{1}{p^2}=\dfrac{q}{p^2}$

Abaterea medie patratica este $\sigma(X)=\dfrac{\sqrt{q}}{p}$

10. Un lot de comprimate este supus controlului pe flux in ceea ce priveste greutatea, rezistenta si inaltimea. Probabilitatea ca un comprimat sa corespunda la fiecare incercare este de 0,9. Experienta este intrerupta daca tableta nu corespunde la o anumita incercare. Sa se scrie repartitia variabilei aleatoare X care reprezinta numarul de comprimate testate.

Solutie:

Se noteaza cu $q=0{,}9$ probabilitatea ca tableta sa reziste si cu $p=1-q=0{,}1$ probabilitatea ca ea sa nu reziste. Probabilitatea ca tableta sa fie supusa la n incercari este $q^{n-1}p$ deoarece ea trebuie sa reziste la n-1 incercari si sa reziste la incercarea n.

Asadar:

$X:\begin{pmatrix} k \\ (0.9)^{k-1}*(0.1)\end{pmatrix}$ k=1,2,....

11. Sa se calculeze media si dispersia variabilei cu repartitia

X: $\begin{pmatrix} 0 & 1 & & q & & n \\ q^n & C_n^1pq^{n-1} & ...... & C_n^kp^kq^{n-k} & .... & p^n \end{pmatrix}$

Solutie:

$E(X)=\sum_{k=0}^{n} kC_n^kp^kq^{n-k}$

Se porneste de la relatia: $\left(pt+q\right)^n = \sum_{k=0}^{n} C_n^k p^k t^k q^{n-k}$. Prin derivare dupa t rezulta

$$np\left(pt+q\right)^{n-1} = \sum_{k=0}^{n} kC_n^k p^k t^{k-1} q^{n-k}$$

Se considera $t=1$ si se obtine:

$$np\left(p+q\right)^{n-1} = \sum_{k=0}^{n} kC_n^k p^k q^{n-k} = E\left(X\right) = np$$

Derivand din nou dupa t obtinem:

$$np\left(pt+q\right)^{n-1} + n\left(n-1\right)p^2t^2\left(pt+q\right)^{n-2} = \sum_{k=0}^{n} k^2 C_n^k p^k t^{k-1} q^{n-k}$$

Se considera $t=1$ si se obtine:

$$np + n\left(n-1\right)p^2 = \sum_{k=0}^{n} k^2 C_n^k p^k q^{n-k} = E\left(X^2\right)$$

$$D\left(X\right) = E\left(X^2\right) - \left(E\left(X\right)\right)^2 = npq$$

12. Sa se calculeze media si dispersia variabilei cu repartitia

$$X: \begin{pmatrix} 0 & 1 & & k & \\ e^{-\lambda} & \dfrac{\lambda e^{-\lambda}}{1!} & \ldots\ldots & \dfrac{\lambda^k e^{-\lambda}}{k!} & \ldots\ldots \end{pmatrix}$$

Solutie: $E\left(X\right) = \sum_{k=0}^{\infty} k\dfrac{\lambda^k e^{-\lambda}}{k!} = \lambda e^{-\lambda}\sum_{k=1}^{\infty}\dfrac{\lambda^{k-1}}{\left(k-1\right)!} = \lambda e^{-\lambda}e^{\lambda} = \lambda$

$$E\left(X^2\right) = \sum_{k=0}^{\infty} k^2\frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\left(k^2-k+k\right)\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}k\left(k-1\right)\frac{\lambda^k}{k!} + e^{-\lambda}\sum_{k=0}^{\infty}k\frac{\lambda^k}{k!} =$$

$$= \lambda^2 e^{-\lambda}\sum_{k=2}^{\infty}\frac{\lambda^{k-2}}{\left(k-2\right)!} + \lambda = \lambda^2 e^{\lambda}e^{-\lambda} + \lambda = \lambda^2 + \lambda$$

$$D\left(X\right) = E\left(X^2\right) - \left(E\left(X\right)\right)^2 = \lambda$$

1.3. DISTRIBUTII DE PROBABILITATE

Distributia binomiala

Distribuția binomială apare la descrierea evenimentelor asociate extracțiilor dintr-o urnă cu bile albe şi bile negre.

Distribuția variabilei aleatoare "numărul k de bile albe din n bile extrase" se poate reprezenta şi sub formă matricială:

$$X = \begin{pmatrix} 0 & 1 & & k & & n \\ C_n^0 p^0 q^n & C_n^1 p^1 q^{n-1} & \cdots & C_n^k p^k q^{n-k} & \cdots & C_n^k p^n q^0 \end{pmatrix}$$

Probabilitatea evenimentului este $p_k = C_n^k p^k q^{n-k}$

Proprietati:

Media şi dispersia unei variabile aleatoare repartizate binomial sunt $M = np$ si $D = npq$

<u>Repartiția binomială apare întotdeauna atunci când un experiment cu numai două răspunsuri posibile se repetă de n ori</u>.

Distributia Poisson

Un caz particular al distributiei binomiale îl prezintă experimentele care se repetă de un număr foarte mare de ori, iar evenimentul în a cărui apariție suntem interesați are o probabilitate foarte mică, categorisit uzual ca "eveniment rar".

Distribuția Poisson se obține la limită, când $n \to \infty$, $p \to 0$, dar np rămâne constant, $np = \lambda$.

Deci, distribuția Poisson este dată de matricea

$$X = \begin{pmatrix} 0 & 1 & & k & & n \\ e^{-\lambda} & \frac{\lambda}{1!} e^{-\lambda} & \ldots & \frac{\lambda^k}{k!} e^{-\lambda} & \ldots & \frac{\lambda^n}{n!} e^{-\lambda} \end{pmatrix}$$

Solutie:

Considerăm deci că $np = \lambda$ şi trecem la limită după n

$$\lim_{n\to\infty} C_n^k p^k q^{n-k} = \lim_{n\to\infty} \frac{n(n-1)\ldots(n-k+1)}{k!} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k} =$$

$$= \frac{1}{k!} * \lim_{n\to\infty} \frac{n(n-1)\ldots(n-k+1)}{n^k} \lambda^k \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

dar $\lim_{n\to\infty}\dfrac{n(n-1)\ldots(n-k+1)}{n^k}=1$ şi

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n-k}=\lim_{n\to\infty}\left[\left(1-\frac{\lambda}{n}\right)^{-\frac{n}{\lambda}}\right]^{\frac{n-k}{n}(-\lambda)}=e^{-\lambda}\text{ şi deci,}$$

$$\lim_{n\to\infty}C_n^k p^k q^{n-k}=\frac{\lambda^k}{k!}e^{-\lambda}$$

Ceea ce arata ca: $X=\begin{pmatrix} 0 & 1 & k & n \\ e^{-\lambda} & \frac{\lambda}{1!}e^{-\lambda}\ldots & \frac{\lambda^k}{k!}e^{-\lambda}\ldots & \frac{\lambda^n}{n!}e^{-\lambda}\end{pmatrix}$

Proprietati:

Media şi dispersia unei variabile aleatoare distribuite Poisson sunt $M(X)=\lambda$ si $D(X)=\lambda$

Solutie:

Vom calcula, după definiție, media şi dispersia unei variabile aleatoare distribuite Poisson şi vom tine cont că

$\sum_{k\ge 0}\frac{\lambda^k}{k!}=e^{\lambda}$, $\sum_{k\ge 0}k\frac{\lambda^k}{k!}=\lambda e^{\lambda}$, $\sum_{k\ge 2}k(k-1)\frac{\lambda^k}{k!}=\lambda^2 e^{\lambda}$, $\sum_{k\ge 1}k\frac{\lambda^k}{k!}=\lambda e^{\lambda}$ obtinandu-se astfel:

$$M(X)=\sum_{k\ge 0}k\frac{\lambda^k e^{-\lambda}}{k!}=e^{-\lambda}\sum_{k\ge 1}\frac{\lambda^k}{(k-1)!}=e^{-\lambda}\lambda\sum_{k\ge 1}\frac{\lambda^{k-1}}{(k-1)!}=e^{-\lambda}\lambda e^{\lambda}=\lambda$$

$$D(X)=e^{-\lambda}\sum_{k\ge 0}\frac{(k-\lambda)^2\lambda^k}{k!}=e^{-\lambda}\left(\sum_{k\ge 0}\frac{k^2\lambda^k}{k!}-2\lambda\sum_{k\ge 0}\frac{k\lambda^k}{k!}+\lambda^2\sum_{k\ge 0}\frac{\lambda^k}{k!}\right)=$$

$$=e^{-\lambda}\left(\sum_{k\ge 1}[k(k-1)+k]\frac{\lambda^k}{k!}-\lambda^2 e^{\lambda}\right)=e^{-\lambda}\left[\sum_{k\ge 2}k(k-1)\frac{\lambda^k}{k!}+\sum_{k\ge 1}k\frac{\lambda^k}{k!}\right]-\lambda^2=$$

$$=e^{-\lambda}\left(\lambda^2 e^{\lambda}+\lambda e^{\lambda}\right)-\lambda^2=\lambda$$

Distributia normala

Spunem că o variabilă aleatoare este normal repartizată $N(m,\sigma^2)$, atunci când densitatea sa de probabilitate este data de formula:

$$\rho(x,m,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-m)^2}{2\sigma^2}}$$

Exercitii:

Sa se verifice urmatoarele proprietati ale densitatii de probabilitatea $\rho(x)$:

a) $\int_{-\infty}^{+\infty}\rho(x)dx=P(-\infty\langle f(t)\langle+\infty)=1$

b) $M[X]=m$

c) $D[X]=\sigma^2$

Aplicatie:

Daca X este o variabilă aleatoare normal repartizată $N(m,\sigma^2)$, atunci variabila aleatoare $Z=\dfrac{X-m}{\sigma}$ este normal repartizată $N(0,1)$.

Solutie:

Tinand cont de proprietatile medie obtinem:

$$M[Z]=M\left[\frac{X-m}{\sigma}\right]=\frac{M[X-m]}{\sigma}=\frac{M[X]-m}{\sigma}=\frac{m-m}{\sigma}=0$$

Tinand cont de proprietatile dispersiei obtinem:

$$D[Z]=D\left[\frac{X-m}{\sigma}\right]=\frac{D[X-m]}{\sigma^2}=\frac{D[X]}{\sigma^2}=\frac{\sigma^2}{\sigma^2}=1$$

Variabila Z se numeste variabila aleatoare normal standardizata.

Distributia χ^2 Helmert - Pearson

Se consideră n observații independente x_1, x_2, …, x_n (variabile aleatoare independente) normal distribuite $N(m,\sigma^2)$.

Variabilele standard $u_i=\dfrac{x_i-m}{\sigma}$, $i=\overline{1,n}$ sunt de asemenea independente, iar suma pătratelor lor va avea o distributie ce poate fi determinată.

Se defineşte $X=\sum_1^n u_i^2$.

Distribuția variabilei X rezultate se notează $\chi^2(n)$ şi este diferită pentru fiecare valoare a lui n, iar parametru n se defineşte ca numărul de gradelor de libertate.

Parametrii (media şi dispersia) unei variabile distribuite χ^2sunt:

$M\left[\chi^2(n)\right] = n$ si $D\left[\chi^2(n)\right] = 2n$

Densitatea de probabilitate este dată de funcția

$$f(\chi^2) = \frac{1}{\Gamma\left(\frac{n}{2}\right)2^{\frac{n}{2}}} e^{-\frac{\chi^2}{2}} \left(\chi^2\right)^{\frac{n}{2}-1},$$

unde Γ este funcția Euler de spețа I-a studiată la cursul de matematică şi anume : $\Gamma(\alpha) = \int_0^{+\infty} e^{-t} t^{\alpha-1} dt$.

Repartitia χ^2 se foloseşte foarte mult în statistica matematică în verificarea ipotezelor asupra egalității dispersiilor.

Distributia STUDENT

Analog cu distribuția χ^2 , repartiția t a fost propusă de Student (pseudonimul lui W.S.Gosset, chimist statistician englez), pentru statistica selecțiilor mici şi exprimă deviațiile mediilor de selecție $\bar{x}$, față de media întregii populații μ, măsurate în $\frac{s}{\sqrt{n}}$ (abaterea standard a mediilor de selecție).

Dacă sunt date două variabile aleatoare $Z \in N(0,1)$ si $V \in \chi^2(n)$ independente, se spune că variabila $t = \frac{Z}{\sqrt{\frac{V}{n}}} \in t(n)$ este repartizată Student cu n grade de libertate.

Mărimea t nu depinde decât de numărul gradelor de libertate.

Distribuția de probabilitate a unei variabile aleatoare repartizate Student tinde pentru $n \to \infty$, la distribuția normală $\rho(t) \to \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$

Densitatea de probabilitate este dată de funcția:

$$f(x) = \frac{1}{\sqrt{n\pi}} * \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} * \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$ unde $x \in R$ şi $n \in N$.

Distributia F (Behrens - Fisher – Snedecor) a raportului a două dispersii

Se consideră frecvent în statistică raportul a două dispersii care estimează aceeaşi dispersie generală a unei colectivități. Dintr-o colectivitate generală se extrag două selecții $U \in \chi^2(n_1)$, $V \in \chi^2(n_2)$.

Raportul lor este o variabilă aleatoare repartizată F

$$F = \frac{\frac{U}{n_1}}{\frac{V}{n_2}} \in F(n_1, n_2)$$

Examinând acest raport se observă că el nu conține dispersia colectivității generale σ^2 , de unde rezultă că distribuția acestui raport nu depinde decât de numărul gradelor de libertate n_1 si n_2 ale celor două dispersii.

Densitatea de probabilitate este dată de funcția:

$$f(x) = \frac{\Gamma\left(\frac{n_1 + n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right) * \Gamma\left(\frac{n_2}{2}\right)} * \left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} * x^{\frac{n_1}{n_2} - 1} * \left(1 + \frac{n_1}{n_2} * x\right)^{-\frac{n_1 + n_2}{2}}$$,când $x \rangle 0$.

Observatie:

Cunoaşterea mediei şi dispersiei unei variabile aleatoare dă o indicație asupra intervalului în care se află valorile variabilei, cu cea mai mare probabilitate. Mai exact cu cât ne îndepărtăm mai mult de valoarea medie, cu atât valorile respective sunt mai puțin probabile ca valori ale variabilei date.

Aplicatii:

1. Probabilitatea ca un comprimat sa se sfarame este de 0.1. Presupunand ca un lot cuprinde 200 comprimate sa se calculeze numarul mediu de comprimate care se sfarama si abaterea medie patratica.

Solutie:

Deoarece p =0.1 este constant se utilizeaza repartitia binomiala

$$E(X) = np = 0.1 * 200 = 20$$

$$D(X) = \sigma^2 = npq = 200 * 0.1 * 0.9 = 18$$

2. Se stie ca 3 muncitori din 2000 fac alergie la medicamente. Care este probabilitatea ca intr-o fabrica cu 1000 de muncitori sa existe 3 muncitori alergici?

Solutie:

Se poate utiliza repartitia binomiala cu p=3/2000=0.0015 si n=1000.

$$P_n^k = C_n^k p^k q^{n-k}$$

$$P_{1000}^3 = C_{1000}^3 \left(\frac{3}{2000}\right)^3 \left(1 - \frac{3}{2000}\right)^{1000-3} = 0.12555$$

Folosind repartitia Poisson se obtine:

$$P^3 = \frac{e^{-1.5}(1.5)^3}{3!} = 0.125 \text{ unde } \lambda = np = 1.5$$

3. Se stie ca probabilitatea ca un lot sa fie corespunzator este de 0.9.

a. Care este probabilitatea ca din 10 loturi sa fie rebutat unul singur?

b. Care este probabilitatea ca din 20 de loturi sa fie rebutat mai mult de 2 loturi?

Solutie:

a) Fie X numarul de loturi rebutate. Se ia p=0.1 si n=10.

$$P(X=1) = C_n^k p^k q^{n-k} = C_{10}^1 * 0{,}1 * 0{,}9^9 = \frac{10!}{1!(10-1)!} * 0{,}1 * 0{,}9^9 = 0{,}9^9 = 0{,}387$$

b)

$$P(X \rangle 2) = 1 - P(X \le 2) = 1 - \left[P(X=0) + P(X=1) + P(X=2)\right] =$$

$$= 1 - \left[C_{10}^0 * 0{,}1^0 * 0{,}9^{10} + C_{10}^1 * 0{,}1 * 0{,}9^9 + C_{10}^2 * 0{,}1^2 * 0{,}9^8\right] =$$

$$= 1 - \left(0{,}9^{10} + 0{,}9^9 + 0{,}45 * 0{,}9^8\right) = 1 - 0{,}9^8 * \left(0{,}9^2 + 0{,}9 + 0{,}45\right) =$$

$$=1-0,9^8*\left(0,9^2+0,9+\frac{0,9}{2}\right)=1-0,9^9*\left(0,9+1+\frac{1}{2}\right)=1-0,9^9*\frac{4,8}{2}=$$

$$=1-0,387*2,4=1-0,929=0,071$$

4. Se stie ca 20% dintre anumite produse sunt defecte. Se aleg la intamplare 8 produse.

a. Care este probabilitatea ca 0.1 sau 2 produse sa fie defecte?

b. Care este probabilitatea ca cel putin doua produse sa fie defecte?

c. Care este probabilitatea ca cel mult doua produse sa fie defecte?

Solutie:

Fie X numarul de produse defecte. Se ia p=0.2 si n=8.

a) $P(X=0.1\,sau\,2)=P(X=0)+P(X=1)+P(X=2)=$

$$=C_8^0*0,2^0*0,8^8+C_8^1*0,2*0,8^7+C_8^2*0,2^2*0,8^6=$$

$$=0,8^8+\frac{8!}{1!(8-1)!}*0,2*0,8^7+\frac{8!}{2!(8-2)!}*0,2^2*0,8^6=$$

$$=0,8^6\left(0,8^2+1,6*0,8+28*0,04\right)=0,8^7\left(0,8+1,6+1,4\right)=3,8*0,8^7=0.798$$

b) $P(X\geq 2)=1-P(X\langle\, 2)=1-P(X\leq 1)=1-[P(X=0)+P(X=1)]=$

$$=1-\left[C_8^0*0,2^0*0,8^8+C_8^1*0,2*0,8^7\right]=1-\left[0,8^8+\frac{8!}{1!(8-1)!}*0,2*0,8^7\right]=$$

$$=1-0,8^7(0,8+1,6)=1-0,8^8*3=1-0,17*3=1-0,51=0,49$$

c) $P(X\leq 2)=P(X=0)+P(X=1)+P(X=2)=0,798$

5. 8% din recipientele cu materie prima sunt rebutate. Care este probabilitatea ca din 20 de recipiente 2 sa fie rebutate.

Solutie:

Folosind repartitia binomiala pentru p=0.08 si n=20 se obtine $P_{20}^2=\frac{20!}{2!18!}(0.08)^2(0.92)^{18}=0.2711$

Folosind repartitia Poisson pentru $\lambda=np=1.6$ se obtine $P^2=\frac{e^{-1.6}(1.6)^2}{2!}=0.258\Rightarrow$ Se observa ca rezultatele sunt apropiate.

6. Se stie ca 10% din productia unei fabrici este alcatuita din produse cu defectiuni. Sa se calculeze probabilitatea ca din 10 produse alese la intamplare (cu punerea inapoi a produsului) k sa fie defecte, k=0,1,2,...,7.
Solutie:
Se poate aplica fie repartitia binomiala, fie repartitia Poisson.

$P_n^k = C_n^k p^k q^{n-k}$, $P^k = \dfrac{\lambda^k e^{-\lambda}}{k!}$ unde n=10; p=0.1; q=0.9 si $\lambda = np = 1$

k	0	1	2	3	4	5	6	7
P_n^k	0.348	0.387	0.193	0.057	0.011	0.001	0.001	0
P^k	0.368	0.368	0.184	0.061	0.015	0.003	0.001	0

7. Probabilitatea de a avea un consum zilnic normal de energie este de 0.75. Fie X numarul de zile lucratoare pe saptamana, in care consumul este normal. Sa se determine :

a. Repartitia variabilei aleatoare X

b. Probabilitatea unui consum normal in cel mai putin 4 zile

c. Media, dispersia si abaterea medie patratica a lui X.

Solutie:
Se aplica repartitia binomiala cu n=6 , p=3/4 si q=1/4

a) $X : \begin{pmatrix} k \\ C_n^k p^k q^{n-k} \end{pmatrix}$

Deci $X : \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{4096} & \frac{18}{4096} & \frac{135}{4096} & \frac{540}{4096} & \frac{1215}{4096} & \frac{1458}{4096} & \frac{729}{4096} \end{pmatrix}$

b) $P(X \geq 4) = P(X = 4) + P(X = 5) + P(X = 6) = 0.83$

c) $E(X) = np = 4.5$

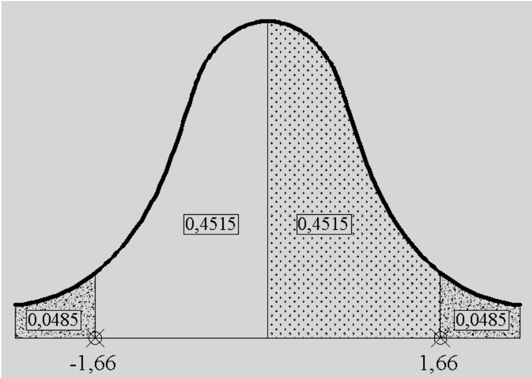
$D(X) = npq = 1.125$

$\sigma(X) = \sqrt{D(X)} = 1.06$

8. Norma prevede pentru caprolactama punctul de solidificare $\geq 67.2^{\circ}$C. Experientele efectuate asupra unui lot arata ca punctul de solidificare este repartizat normal cu media $\mu = 67.7^0 C$ si dispersia $\sigma^2 = 0.09$. Sa se determine procentul de sarje care nu corespund calitatii.
Solutie:

Din ipoteza avem ca variabila aleatoare X (punctul de solidificare) este normal repartizata cu media $\mu = 67.7^0 C$ si dispersia $\sigma^2 = 0.09$ deci,

$$Z = \frac{X - E(X)}{\sqrt{D(X)}} \in N(0,1) \Rightarrow \frac{67{,}2 - 67{,}7}{0{,}3} = \frac{-0{,}5}{0{,}3} = -\frac{5}{3} = -1{,}66$$



Se observa ca:

$$P(X \langle 67.2) = P\left(\frac{X - 67{,}7}{\sqrt{0{,}09}} \langle \frac{67{,}2 - 67{,}7}{\sqrt{0{,}09}} \right) = P(Z \langle -1{,}66) = 0{,}5 - 0{,}4515$$

Deci, $P(X \langle 67.2) = 0.0485$

9. Un echipament pentru producerea apei pentru solutii injectabile este alcatuit din patru sisteme independente. Probabilitatea ca un sistem sa se defecteze in timpul functionarii este de 0.1. Care este repartitia variabilei aleatoare X care da numarul de sisteme ce se pot defecta in timpul functionarii?

Solutie:

X este probabilitatea binomiala in care $n = 4$, $p = 0.1$ si $q = 1 - p = 0.9$

$$X : \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ (0.9)^4 & C_4^1 (0.9)^3 (0.1) & C_4^2 (0.9)^2 (0.1)^2 & C_4^3 (0.9)^1 (0.1)^3 & (0.1)^4 \end{pmatrix}$$

10. Un operator poate controla vizual doua fiole de vaccin pe minut. Pe durata unui schimb (480 minute) se constata ca 40 de fiole de vaccin prezinta defecte de inchidere. Considerand ca numarul de fiole defecte este repartizat Poisson sa se determinde probabilitatea ca din 5 fiole cel putin doua sa fie defecte.

Solutie:

Intr-un minut se constata $\frac{40}{480} = 0.083$ fiole defecte si sunt inspectate vizual 2 fiole. Deci, 5 fiole de vaccin vor fi inspectate in $\frac{5}{2} = 2.5$ minute.

In acest timp se constata $0.083 * 2.5 = 0.2$ fiole defecte. La o repartitie Poisson λ este media, deci $\lambda = 2$

$$P(k) = \frac{(0.2)^k e^{-0.2}}{k!}$$

Probabilitatea ceruta corespunde lui $k = 2, \quad 3, \quad 4, \quad 5$

$$P = P(2) + P(3) + P(4) + P(5) = 0.0175$$

11. Sa se calculeze probabilitatea ca o variabila aleatoare normala sa ia valori intr-un interval de lungime $k\sigma$ de-o parte si de alta a valorii medii. Se ia k=1,2,3.

Solutie:

Se calculeaza

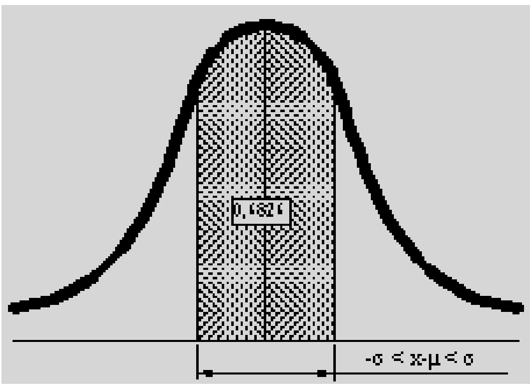
$$P(|X - \mu| \langle \, k\sigma) = P(\mu - k\sigma \,\langle\, X \,\langle\, \mu + k\sigma) = P(-k\sigma \,\langle\, X - \mu \,\langle\, k\sigma) =$$

$$= P\left(-k \,\langle\, \frac{X - \mu}{\sigma} \,\langle\, k\right) = P(-k \,\langle\, Z \,\langle\, k) = \Phi(k) - \Phi(-k) = 2 * P(\, Z \,\langle\, k) = 2 * z_k$$

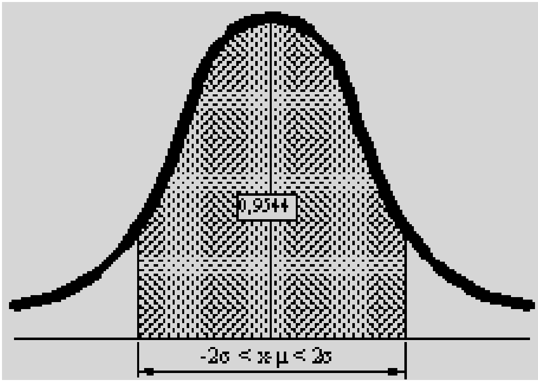
unde $Z = \frac{X - \mu}{\sigma} \in N(0,1)$ si $z_k = P(Z \,\langle\, z_k)$

a) Pentru $k = 1$ avem

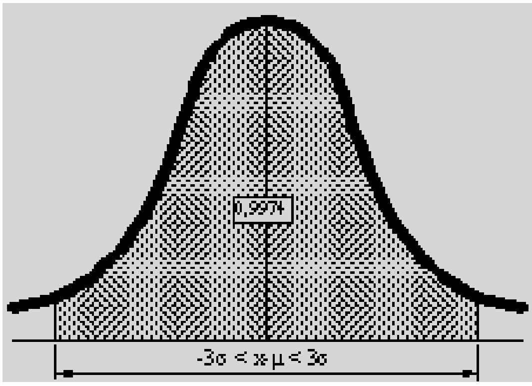
$$P(|X - \mu| \,\langle\, \sigma) = P(-1 \,\langle\, Z \,\langle\, 1) = 2 * z_1 = 2 * 0{,}3413 = 0{,}6826$$



b) Pentru $k = 2$ avem
$$P(|X - \mu| \langle\ 2\sigma) = P(-2 \ \langle\ Z \ \langle\ 2) = 2 * z_2 = 2 * 0{,}4772 = 0{,}9544$$



c) Pentru $k = 3$ avem
$$P(|X - \mu| \langle\ 3\sigma) = P(-3 \ \langle\ Z \ \langle\ 3) = 2 * z_3 = 2 * 0{,}4987 = 0{,}9974$$



Se observa ca probabilitatea ca o variabila aleatoare normala sa ia valori inafara intervalului $(-3\sigma\ , 3\sigma)$ de-o parte si de alta a valorii medii este infima.

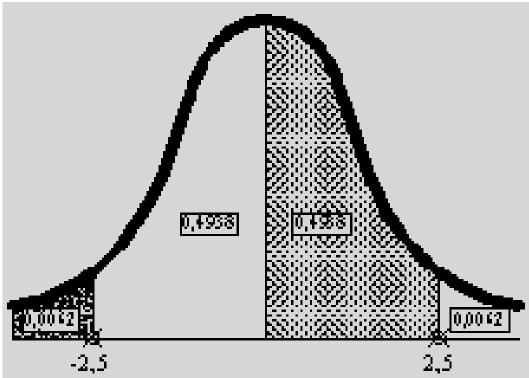
12. O variabila aleatoare este repartizata normal cu media 30 si dispersia $\sigma^2 = 100$. Care este probabilitatea ca variabila aleatoare sa ia valori mai mari decat 5 (mai mici decat -5)?

Solutie:

a)

$$P(X\langle 5)=P\left(\frac{X-\mu}{\sigma}\langle\frac{5-30}{10}\right)=P\left(\frac{X-\mu}{\sigma}\langle -2{,}5\right)=\Phi(-2.5)=$$

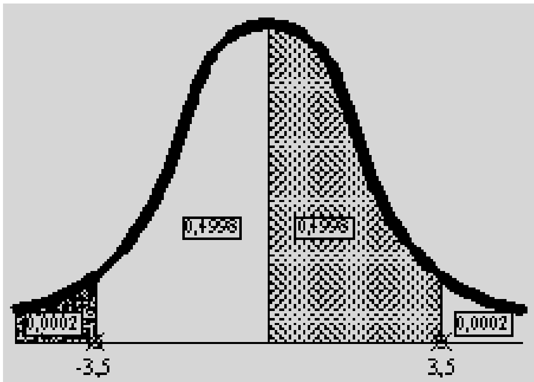
$$=0{,}5-0{,}4938=0.0062$$



b)

$$P(X\ \langle -5)=P\left(\frac{X-\mu}{\sigma}\langle\frac{-5-30}{10}\right)=P\left(\frac{X-\mu}{\sigma}\langle -3{,}5\right)=\Phi(-3.5)=$$

$$=0{,}5-0{,}4998=0.0002$$



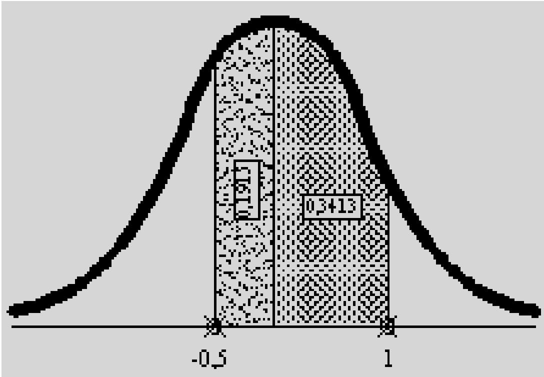
13. O variabila aleatoare are media $\mu=2$ si dispersia $\sigma^2=4$. Sa se calculeze $P(0\le X\langle 3)$ si $P(|X|\le 1)$.

Solutie:

a)

$$P(0\le X\langle 3)=P\left(\frac{0-2}{2}\le\frac{X-2}{2}\langle\frac{3-2}{2}\right)=P(-0{,}5\le Z\ \langle 1)=$$

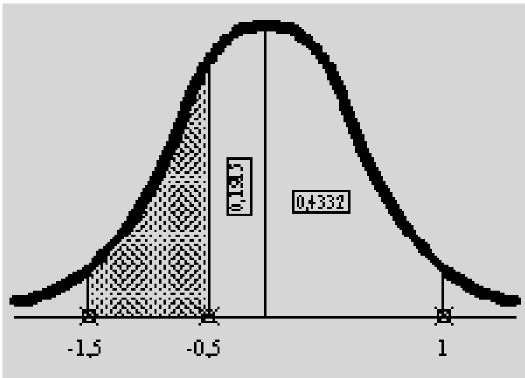
$$= \Phi\left(\frac{0-2}{2}\right) - \Phi\left(\frac{3-2}{2}\right) = 0,1915 + 0,3413 = 0,5328$$



b)

$$P(|X| \leq 1) = P(-1 \langle X \langle 1) = P\left(\frac{-1-2}{2} \leq \frac{X-2}{2} \leq \frac{1-2}{2}\right) = P(-1,5 \leq Z \leq -0,5) =$$

$$= \Phi\left(\frac{1-2}{2}\right) - \Phi\left(\frac{-1-2}{2}\right) = 0,4332 - 0,1915 = 0.2417$$



14. Fie X o variabila aleatoare repartizata normal cu parametrii μ si σ. Sa se calculeze $E(|X-\mu|)$

Solutie:

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

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

$$=-\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{0}ye^{-\frac{y^2}{2^2}}dy+\frac{1}{\sigma\sqrt{2\pi}}\int_{0}^{+\infty}ye^{-\frac{y^2}{2^2}}dy=\frac{2}{\sigma\sqrt{2\pi}}\int_{0}^{+\infty}ye^{-\frac{y^2}{2^2}}dy=$$

$$=-\frac{2\sigma^2}{\sigma\sqrt{2\pi}}e^{-\frac{y^2}{2^2}}\Bigg|_{0}^{+\infty}=\sigma\sqrt{\frac{2}{\pi}}$$

15. Densitatea de repartitie a duratei de functionare a unui sistem de centrifugare este $f(\tau)=\frac{1}{\bar{\tau}}e^{-\frac{\tau}{\bar{\tau}}}$ unde $\bar{\tau}$ este 3 ani. Sa se determine probabilitatea ca intr-o instalatie formata din trei sisteme de centrifugare sa nu se defecteze nici o centrifuga in primii 6 ani de functionare.

Solutie:

Probabilitatea ca durata de functionare a unei singure centrifugi sa fie mai mare de 6 ani este

$$\int_6^{\infty}\frac{1}{\bar{\tau}}e^{-\frac{\tau}{\bar{\tau}}}d\tau=e^{-\frac{6}{\bar{\tau}}}=e^{-2}$$

Deoarece duratele de functionare ale evaporatoarelor sunt variabile independente, probabilitatea cautata va fi $\left(e^{-2}\right)^3=e^{-6}$

16. Intr-un strat fluidizat se afla M=1000 kg de granula. Debitul de material prin strat este de $G_m=4000\ kg/h$.

Sa se determine ce fractiune de particule se afla in strat un timp mai mic (mai mare) decat timpul mediu de stationare.

Solutie:

$\bar{\tau}=\frac{M}{G_m}=\frac{1000}{4000}*60=15\min$ si $f(\tau)=\frac{1}{\bar{\tau}}e^{-\frac{\tau}{\bar{\tau}}}$

$$P\left(\tau\rangle\,\bar{\tau}\right)=\int_{\bar{\tau}}^{\infty}f(\tau)d\tau=\int_{\bar{\tau}}^{\infty}\frac{1}{\bar{\tau}}e^{-\frac{\tau}{\bar{\tau}}}d\tau=-e^{-\frac{\tau}{\bar{\tau}}}\Big|_{\bar{\tau}}^{\tau}=e^{-1}$$

$$P\left(\tau\langle\,\bar{\tau}\right)=\int_0^{\bar{\tau}}f(\tau)d\tau=1-e^{-1}$$

INTERVALE DE INCREDERE

Problema estimării intervalelor pentru un parametru θ se reduce la găsirea unui interval de încredere (θ_L, θ_U) cu un coeficient de încredere $1-\alpha$ astfel încât

$$P(\theta_L \langle \theta \langle \theta_U) = 1-\alpha .$$

În stabilirea intervalelor se utilizează caracteristicile numerice cuantile. Se numesc *cuantile de ordin* β valoarea x_β a variabilei aleatoare x pentru care $F(x_\beta) = P(x \langle x_\beta) = \beta$ adică valoarea variabilei aleatoare care are la stânga ei aria β sub curba densității de probabilitate.

Pentru a estima un interval se alege $1-\alpha$, se citesc din tabelele cuantilele, de exemplu $x_{1-\frac{\alpha}{2}}$ şi $x_{\frac{\alpha}{2}}$ şi se stabileste intervalul pornind de la formula distributiei de probabilitate.

Estimarea intervalelor de încredere pentru medii

a) <u>Cazul când se cunoaste dispersia.</u>

Se consideră o populație repartizată normal $N(\mu, \sigma^2)$. Dacă se cunoaşte dispersia se foloseste faptul că $Z = \dfrac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ este repartizată $N(0,1)$. Se notează cu z_α cuantila de ordinul α pentru repartiția $N(0,1)$. Evident

$$P\left(z_{\frac{\alpha}{2}} \langle z \langle z_{1-\frac{\alpha}{2}}\right) = P\left(z_{\frac{\alpha}{2}} \langle \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \langle z_{1-\frac{\alpha}{2}}\right) =$$

$$= P\left(z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}} \langle \overline{X} - \mu \langle z_{1-\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}\right) =$$

$$= P\left(-\overline{X} + z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}} \langle -\mu \langle -\overline{X} + z_{1-\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}\right) =$$

$$= P\left(-\overline{X} - z_{1-\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}} \langle -\mu \ \langle -\overline{X} + z_{1-\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}} \right) =$$

$$= P\left(\overline{X} - z_{1-\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}} \langle \ \mu \ \langle \ \overline{X} + z_{1-\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}} \right) = 1 - \alpha$$

Aşadar intervalul căutat este

$$(\theta_L, \theta_U) = \left(\overline{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \overline{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right)$$

Mărimea $E = z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ poartă numele de eroare.

b) <u>Cazul când dispersia este necunoscută</u>

Dacă nu se cunoaste dispersia în estimarea intervalelor se utilizează dispersia de selecție care este un estimator nedeplasat al dispersiei deoarece $E(s^2) = \sigma^2$

Se consideră $x_1, x_2, ..., x_n$ o selecție dintr-o populație de tipul $N(\mu, \sigma^2)$.

Mărimea $T = \dfrac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}}$ este repartizată $T(n-1)$ şi, ca urmare

$$P\left(t_{n-1,\frac{\alpha}{2}} \ \langle \ T \ \langle \ t_{n-1,1-\frac{\alpha}{2}} \right) = P\left(t_{n-1,\frac{\alpha}{2}} \ \langle \ \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}} \ \langle \ t_{n-1,1-\frac{\alpha}{2}} \right) =$$

$$= P\left(t_{n-1,\frac{\alpha}{2}} * \frac{s}{\sqrt{n}} \ \langle \ \overline{X} - \mu \langle \ t_{n-1,1-\frac{\alpha}{2}} * \frac{s}{\sqrt{n}} \right) =$$

$$= P\left(-\overline{X} + t_{n-1,\frac{\alpha}{2}} * \frac{s}{\sqrt{n}} \ \langle -\mu \langle -\overline{X} + t_{n-1,1-\frac{\alpha}{2}} * \frac{s}{\sqrt{n}} \right) =$$

$$= P\left(-\overline{X} - t_{n-1,1-\frac{\alpha}{2}} * \frac{s}{\sqrt{n}} \ \langle -\mu \langle -\overline{X} + t_{n-1,1-\frac{\alpha}{2}} * \frac{s}{\sqrt{n}} \right) =$$

$$= P\left(\overline{X} - t_{n-1,1-\frac{\alpha}{2}} * \frac{s}{\sqrt{n}} \ \langle \ \mu \langle \ \overline{X} + t_{n-1,1-\frac{\alpha}{2}} * \frac{s}{\sqrt{n}} \right) = 1 - \alpha$$

Deoarece repartitia Student este simetrică față de origine $t_{n-1,1-\frac{\alpha}{2}} = -t_{n-1,1-\frac{\alpha}{2}}$

Ca urmare intervalul căutat este

$$(\theta_L, \theta_U) = \left(\overline{X} - t_{n-1,1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}, \overline{X} + t_{n-1,1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \right)$$

În acest caz eroarea este $E = t_{n-1,1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$

Dacă numărul de experiențe este $n \rangle 30$, se poate folosi aproximația

$$t_{n-1,1-\frac{\alpha}{2}} = z_{1-\frac{\alpha}{2}}$$

Estimarea intervalului de încredere $1-\alpha$ pentru diferența a două medii

Se consideră două selecții din populații normal repartizate $N(\mu_1, \sigma_1^2)$ şi $N(\mu_2, \sigma_2^2)$.

a) <u>Cazul dispersiilor σ_1^2, σ_2^2 cunoscute.</u>

Fie o selecție aleatoare $x_{11}, x_{12}, ..., x_{1n_1}$ din populația $N(\mu_1, \sigma_1^2)$ şi o selecție $x_{21}, x_{22}, ..., x_{2n_2}$ dintr-o populație $N(\mu_2, \sigma_2^2)$. Estimatorii nedeplasați ai mediilor μ_1 şi μ_2 sunt: $\overline{X_1} = \frac{\sum_1^{n_1} x_{1i}}{n_1}$ si $\overline{X_2} = \frac{\sum_1^{n_2} x_{2i}}{n_2}$

Considerând variabila aleatoare $\overline{X_1} - \overline{X_2}$, ea este normal repartizata avand media $M(\overline{X_1} - \overline{X_2}) = \mu_1 - \mu_2$ şi dispersia $D(\overline{X_1} - \overline{X_2}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$ ținand cont că x_{1i} şi x_{2i} sunt independente.

Variabila aleatoare $z = \frac{(\overline{X_1} - \overline{X_2}) - (\mu_1 - \mu_2)}{\sqrt{D(\overline{X_1} - \overline{X_2})}} = \frac{(\overline{X_1} - \overline{X_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ este repartizată N(0,1).

Deci, aplicand rationamentul anterior obtinem:

$$\left(\overline{X_1}-\overline{X_2}\right)-z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \langle\, \mu_1-\mu_2 \,\langle \left(\overline{X_1}-\overline{X_2}\right)+z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

Aşadar, intervalul de estimație pentru diferența mediilor este

$$(\Theta_1,\Theta_2)=\left(\left(\overline{X_1}-\overline{X_2}\right)-z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}},\ \left(\overline{X_1}-\overline{X_2}\right)+z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right)$$

În acest caz, eroarea este $E=z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$.

b) <u>Dispersii necunoscute dar presupuse egale</u>

În cazul în care nu cunoaştem dispersiile dar ştim că sunt egale $\sigma_1^2=\sigma_2^2=\sigma^2$ utilizăm dispersia ponderată de selecție

$$s_p^2=\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}=\frac{\sum_1^{n_1}\left(x_{1i}-\overline{X_1}\right)^2+\sum_1^{n_2}\left(x_{2i}-\overline{X_2}\right)^2}{n_1+n_2-2}$$

ca un estimator nedeplasat pentru σ^2, variabila aleatoare

$$T=\frac{\left(\overline{X_1}-\overline{X_2}\right)-(\mu_1-\mu_2)}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

fiind repartizată $T(n_1+n_2-2)$

Deoarece repartiția Student este simetrică $t_{n_1+n_2-2,\frac{\alpha}{2}}=-t_{n_1+n_2-2,1-\frac{\alpha}{2}}$ rezultă că

$$\overline{X_1}-\overline{X_2}-t_{n_1+n_2-2,1-\frac{\alpha}{2}}s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\langle \mu_1-\mu_2 \langle \overline{X_1}-\overline{X_2}-t_{n_1+n_2-2,1-\frac{\alpha}{2}}s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

Deci, intervalul de incredere este:

$$(\Theta_1,\Theta_2)=\left(\overline{X_1}-\overline{X_2}-t_{n_1+n_2-2,1-\frac{\alpha}{2}}s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}},\ \overline{X_1}-\overline{X_2}+t_{n+n_2-2,1-\frac{\alpha}{2_1}}s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right)$$

cu eroarea $E=t_{n_1+n_2-2,1-\frac{\alpha}{2}}s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$.

Estimarea intervalelor de încredere pentru dispersie

Considerăm o selecție de volum n dintr-o populație normală $N(\mu,\sigma^2)$.

Variabila aleatoare $v=\frac{(n-1)s^2}{\sigma^2}$ este repartizată $\chi^2(n-1)$ şi ca urmare

$$\chi^2_{n-1,\frac{\alpha}{2}} \langle (n-1)\frac{s^2}{\sigma^2} \langle \chi^2_{n-1,1-\frac{\alpha}{2}} \text{ si } \frac{(n-1)s^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}} \langle \sigma^2 \langle \frac{(n-1)s^2}{\chi^2_{n-1,\frac{\alpha}{2}}}.$$

Estimarea intervalului de încredere pentru raportul a două dispersii

Se consideră selecția aleatoare $x_{11}, x_{12},..., x_{1n_1}$ dintr-o populație $N(\mu_{,}, \sigma_1^2)$ şi o selecție $x_{21}, x_{22},..., x_{2n_2}$ dintr-o populație $N(\mu_2, \sigma_2^2)$.

Raportul $F=\frac{\frac{s_1^2}{\sigma_1^2}}{\frac{s_2^2}{\sigma_2^2}}$ este repartizat $F(n_1-1, n_2-1)$ şi deci intervalul de estimație pentru raportul dispersiilor este:

$$(\Theta_L, \Theta_U)=\left(\frac{s_2^2}{s_1^2} f_{n_1-1,n_2-1,\frac{\alpha}{2}}, \frac{s_2^2}{s_1^2} f_{n_1-1,n_2-1,1-\frac{\alpha}{2}}\right)$$

TEOREMA LIMITA CENTRALA

Aplicatii:

1. Sa se calculeze un interval de incredere pentru media determinarilor colorimetrice exprimate in molaritati 10^{-4}: 1.22; 1.23; 1.18; 1.29, daca se cunoaste dispersia pentru o observatie individuala $\sigma^2 = 16*10^{-10} M^2$. Se alege $1-\alpha = 0.95$

Solutie

Se utilizeaza faptul ca $Z = \dfrac{\overline{X}-\mu}{\dfrac{\sigma}{\sqrt{n}}}$ este repartizat $N(0,1)$

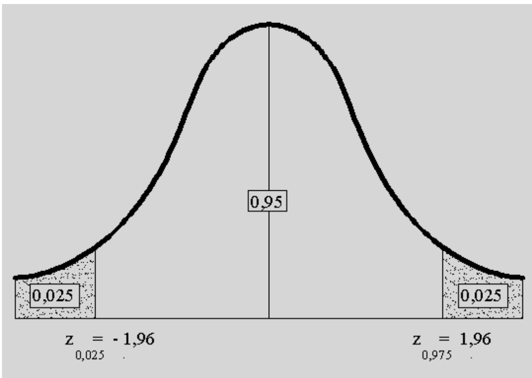
$$z_{\frac{\alpha}{2}} \langle \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \langle z_{1-\frac{\alpha}{2}} \Rightarrow -z_{1-\frac{\alpha}{2}} \langle \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \langle z_{1-\frac{\alpha}{2}} \Rightarrow -z_{1-\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}} \langle \overline{X}-\mu \langle z_{1-\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}$$

$$\Rightarrow -\overline{X} - z_{1-\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}} \langle -\mu \langle -\overline{X} + z_{1-\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}} \Rightarrow$$

$$\Rightarrow \overline{X} - z_{1-\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}} \langle \mu \langle \overline{X} + z_{1-\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}$$

Dar avem: $\overline{X} = \dfrac{\sum x_i}{n} = \dfrac{(1,22+1,23+1,18+1,29)*10^{-4}}{4} = 1.23*10^{-4} M$;

$z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96$; $n = 4$



Deci intervalul de incredere pentru media μ este:

$$1,23*10^{-4} - 1,96*\frac{7*10^{-5}}{2} \langle \mu \langle 1,23*10^{-4} + 1,96*\frac{7*10^{-5}}{2}$$

$$\Rightarrow 0,838*10^{-4} M \langle \mu \langle 1,622*10^{-4} M$$

TEOREMA LIMITA CENTRALA

2. La receptionarea unei substante in recipienti care trebuie sa aiba 40 kg se efectueaza un control cantarind prin sondaj 4 recipiente. Se obtin greutatile: 39.75; 40.25; 39.50; 39.50. Sa se determine un interval de incredere pentru greutatea medie cu un coeficient de incredere $1-\alpha = 0.95$ daca se presupune ca greutatile sunt distribuite normal.

Solutie

Deoarece nu se cunoaste dispersia σ^2 se utilizeaza faptul ca $T = \dfrac{\overline{X} - \mu}{\frac{s_X}{\sqrt{n}}}$ este repartizat $T(n-1)$ unde:

$$\overline{X} = \frac{\sum x_i}{n} = \frac{(39{,}75 + 40{,}25 + 39{,}50 + 39{,}50)}{4} = 39{,}75$$

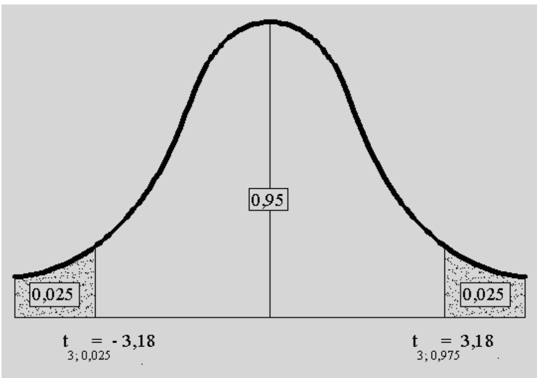
$$s_X^2 = \frac{1}{n-1}\sum\left(x_i - \overline{X}\right)^2 =$$

$$= \frac{(39{,}75-39{,}75)^2 + (40{,}25-39{,}75)^2 + (39{,}50-39{,}75)^2 + (39{,}50-39{,}75)^2}{4-1} =$$

$$= \frac{0 + (0{,}50)^2 + (-0{,}25)^2 + (-0{,}25)^2}{3} = \frac{1}{3}\left[\left(\frac{1}{2}\right)^2 + \left(-\frac{1}{4}\right)^2 + \left(-\frac{1}{4}\right)^2\right] =$$

$$= \frac{1}{3}\left(\frac{1}{4} + \frac{1}{16} + \frac{1}{16}\right) = \frac{1}{3} * \frac{3}{8} = \frac{1}{8} \Rightarrow s_X = \frac{1}{2\sqrt{2}}$$

$n = 4$; $t_{3;0.975} = 3{,}18$



Intervalul de incredere este:

$$39{,}75 - 3{,}18 * \frac{1}{2*2\sqrt{2}} \langle\, \mu \,\langle\, 39{,}75 + 3{,}18 * \frac{1}{2*2\sqrt{2}}$$

Eroarea este $E = 3{,}18 * \frac{1}{2*2\sqrt{2}} \cong 0{,}56$, deci $39{,}19 \langle\, \mu \,\langle\, 40{,}31$

3. 7 containere au greutatile 9.8; 10.2; 10.4; 9.8; 10.0; 10.2; 9.6. Sa se gaseasca un interval de incredere cu $1-\alpha = 0.95$ pentru media greutatii presupunand ca greutatea este distribuita normal.

Solutie

Vom utiliza variabila aleatoare $T = \dfrac{\overline{X} - \mu}{\frac{s_X}{\sqrt{n}}}$ repartizata $T(n-1)$

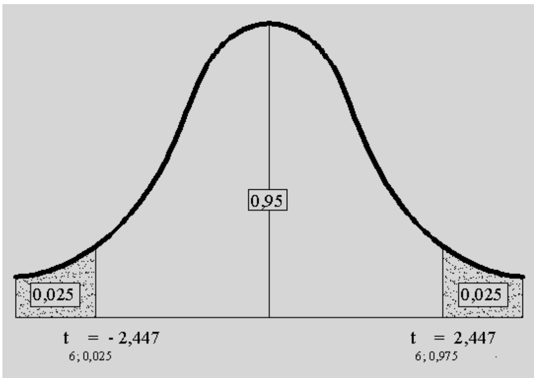
$$\overline{X} = \frac{\sum x_i}{n} = \frac{9{,}8 + 10{,}2 + 10{,}4 + 9{,}8 + 10 + 10{,}2 + 9{,}6}{7} = 10.0\,;$$

$$s_X^2 = \frac{\sum \left(x_i - \overline{X}\right)^2}{n-1} =$$

$$\frac{(9{,}8-10)^2 + (10{,}2-10)^2 + (10{,}4-10)^2 + (9{,}8-10)^2 + (10{,}2-10)^2 + (9{,}6-10)^2}{7-1} =$$

$$= \frac{1}{6}\left[\left(\frac{2}{10}\right)^2 + \left(\frac{2}{10}\right)^2 + \left(\frac{4}{10}\right)^2 + \left(-\frac{2}{10}\right)^2 + \left(\frac{2}{10}\right)^2 + \left(-\frac{4}{10}\right)^2\right] \cong 0{,}09$$

$t_{6;0.975} = 2.447$



Intervalul de incredere: $10 - 2.447 * \frac{0.3}{\sqrt{7}} \langle\, \mu \,\langle\, 10 + 2.447 * \frac{0.3}{\sqrt{7}}$

Deci $\left(\Theta_L, \Theta_U\right) = (9.74\,; 10.26)$

TEOREMA LIMITA CENTRALA

4. La determinarea continutului de penicilina a unui numar de 2500 fiole s-a gasit greutatea medie $\overline{X} = 126.18\, mg$, $s = 4.05\, mg$. Sa se calculeze un interval de incredere pentru medie. Se alege $1-\alpha = 0.99$

Solutie:

Deoarece $n = 2500 \rangle 30$, se poate folosi aproximația $t_{n-1,1-\frac{\alpha}{2}} = z_{1-\frac{\alpha}{2}}$

si se va lua $z_{0.995} = 2.58$

Avem $126.18 - 2.58 * \frac{4.05}{\sqrt{2500}} \langle \mu \langle 126.18 + 2.58 * \frac{4.05}{\sqrt{2500}}$, deci intervalul de incredere este $125,97 \langle \mu \langle 126,39$

5. Se efectueaza 8 titrari volumetrice si se obtin rezultatele: 76.48; 76.43; 77.20; 76.45; 76.25; 76.48; 76.48; 76.60 cm^3. Sa se calculeze un interval de incredere pentru media masuratorilor. Se ia $1-\alpha = 0.95$

Solutie

$$\overline{X} = \frac{\sum x_i}{n} = \frac{76,48 + 76,48 + 77,20 + 76,45 + 76,25 + 76,48 + 76,48 + 76,60}{8} =$$

$$= 76.546$$

$$s_X^2 = \frac{\sum \left(x_i - \overline{X}\right)^2}{n-1} = 0.0790$$

$$t_{n-1;1-\frac{\alpha}{2}} = t_{7;0.975} = 2.36$$

Avem $t_{n-1;\frac{\alpha}{2}} \langle \frac{\overline{X} - \mu}{s_X} \langle t_{n-1;1-\frac{\alpha}{2}}$, deci,

$\overline{X} - t_{n-1;1-\frac{\alpha}{2}} * s_X \langle \mu \langle \overline{X} + t_{n-1;1-\frac{\alpha}{2}} * s_X$ adica $76.31 \langle \mu \langle 76.79$

6. Determinarile succesive efectuate in doua vase deschise care contin HCl au dat normalitatile:

N_1	N_2
15.75	15.58
15.64	15.49
15.92	15.72

Se stie ca dispersia concentratiilor este $\sigma^2 = 0,016$.

TEOREMA LIMITA CENTRALA

Sa se construiasca un interval de incredere cu coeficientul $1-\alpha = 0.95$ pentru diferenta $\mu_1 - \mu_2$ unde μ_1 si μ_2 sunt valorile teoretice ale concentratiilor medii.

Solutie

Se utilizeaza variabila $z = \dfrac{\overline{X_1} - \overline{X_2} - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

$$\overline{X_1} = \frac{\sum x_{1i}}{n} = \frac{15,75 + 15,64 + 15,92}{3} = 15,77$$

$$\overline{X_2} = \frac{\sum x_{2i}}{n} = \frac{15,58 + 15,49 + 15,72}{3} = 15,60$$

$$z_{0.975} = 1.96$$

Avem, $z_{\frac{\alpha}{2}} \langle Z \langle z_{1-\frac{\alpha}{2}}$,deci $-z_{1-\frac{\alpha}{2}} \langle \dfrac{\overline{X_1} - \overline{X_2} - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_1^2}{n_2}}} \langle z_{1-\frac{\alpha}{2}}$ ceea ce inseamna ca: $\overline{X_1} - \overline{X_2} - err \langle \mu_1 - \mu_2 \langle \overline{X_1} - \overline{X_2} + err$

unde eroarea este:

$$err = z_{1-\frac{\alpha}{2}} * \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = 1,96 * \sqrt{\frac{0,016}{3} + \frac{0,016}{3}} = 1,96 * 0,10 = 0,20$$

Obtinem astfel: $15,77 - 15,60 - 0,20 \langle \mu_1 - \mu_2 \langle 15,77 - 15,60 + 0,20$

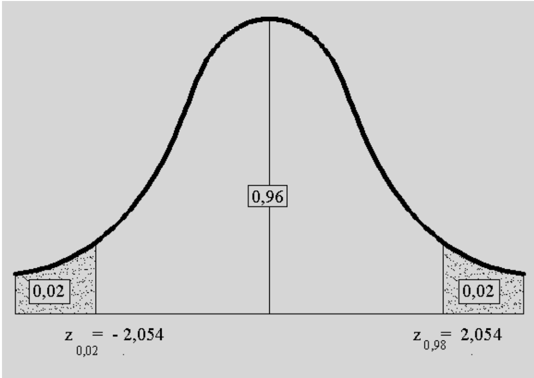
$-0,03 \langle \mu_1 - \mu_2 \langle 16,57$

7. Au fost examinate 75 esantioane de substanta de tipul 1 cu procentul de substanta activa 8.2 si abaterea medie patratica 0.8 si 50 esantioane de substanta de tipul 2 cu procentul de substanta activa 7.6 si abaterea medie patratica 0.6. Sa se gaseasca un interval cu coeficientul de incredere $1-\alpha = 0.96$ pentru diferenta $\mu_1 - \mu_2$ a continuturilor medii de substanta activa.

Solutie:

$$\overline{X_1} - \overline{X_2} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \langle \mu_1 - \mu_2 \langle \overline{X_1} - \overline{X_2} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$\overline{X_1} - \overline{X_2} = 8.2 - 7.6 = 0.6 \;;\; z_{1-\frac{\alpha}{2}} = z_{0.98} = 2.054$$



$$0.6-2.054\sqrt{\frac{0.64}{75}+\frac{0.36}{50}} \langle \mu_1-\mu_2 \langle 0.6+2.054\sqrt{\frac{0.64}{75}+\frac{0.36}{50}}$$

$$0.342 \langle \mu_1-\mu_2 \langle 0.858$$

8. Pentru a studia influenta concentratiei de component catalitic asupra reactiei de NO_2 se fac doua grupe de experiente indexate prin 1 si 2 pentru concentratiile 0.5 si respectiv 1%. Se obtin datele (conversiile).

1	5.18	5.52	5.42
2	5.58	5.62	5.82

Sa se construiasca un interval de incredere cu coeficientul 0.95 pentru diferenta $\mu_1-\mu_2$. Se considera ca dispersiile sunt necunoscute dar egale si $1-\alpha=0,95$

Solutie:

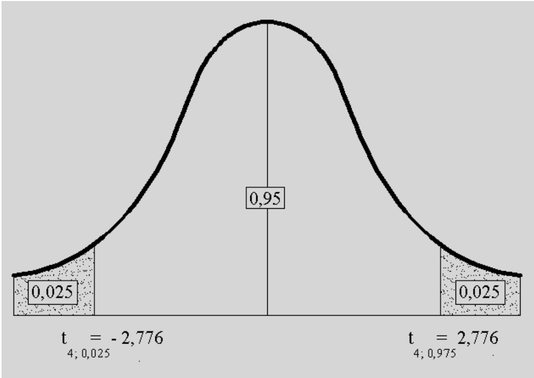
Se utilizeaza variabila $T=\dfrac{\overline{X}_1-\overline{X}_2-\mu_1-\mu_2}{S_p\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}$ distribuita Student cu n_1+n_2-2 grade de libertate.

Avem

$\overline{X}_1=5.37$; $\overline{X}_2=5.67$; $s_1^2=0.08425$; $s_2^2=0.07325$

$$S_p^2=\frac{2s_1^2+2s_2^2}{n_1+n_2-2}=0.07875;\ t_{4;0.975}=2.776$$

TEOREMA LIMITA CENTRALA



Deci

$$5.37-5.67-2.776*\sqrt{0.07875}*\sqrt{\frac{1}{3}+\frac{1}{3}}\ \langle\ \mu_1-\mu_2$$

$$\mu_1-\mu_2\ \langle\ 5.37-5.67+2.776*\sqrt{0.07875}*\sqrt{\frac{1}{3}+\frac{1}{3}}$$

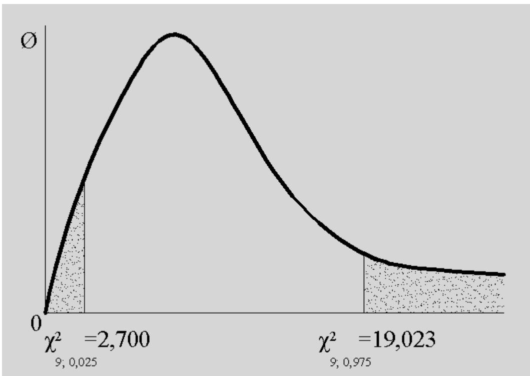
Rezulta $-0.9361\ \langle\ \mu_1-\mu_2\ \langle\ 0.2361$

9. Greutatile unor containere sunt (in kg): 16.4; 16.1; 15.8; 17.0; 16.1; 15.9; 15.8; 16.9; 15.2; 16.0. Sa se gaseasca un interval de incredere cu $1-\alpha=0.95$ pentru dispersia acestora.

Solutie

Se utilizeaza variabila aleatoare $V=\frac{(n-1)s^2}{\sigma^2}$ unde $s^2=0.286$

$$\frac{(n-1)s^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}\ \langle\ \sigma^2\ \langle\ \frac{(n-1)s^2}{\chi^2_{n-1,\frac{\alpha}{2}}}$$

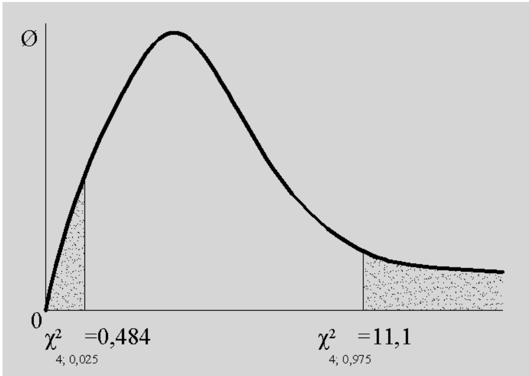


$\chi^2_{9;0.975} = 19.023$; $\chi^2_{9;0.025} = 2.700$

$\frac{9*0.286}{19.023} \langle \sigma^2 \langle \frac{9*0.286}{2.700}$ Rezulta $0.125 \langle \sigma^2 \langle 0.953$

10. Cinci masuratori similare asupra debitului apei reci la un schimbator de caldura au dat rezultatele (kg / s): 5.76; 6.03; 5.84; 5.90; 5.89. Sa se gaseasca un interval de incredere $1-\alpha = 0.95$ pentru dispersia debitelor.
Solutie:

$\overline{X}_1 = 5.88$; $s^2 = 0.00975$;



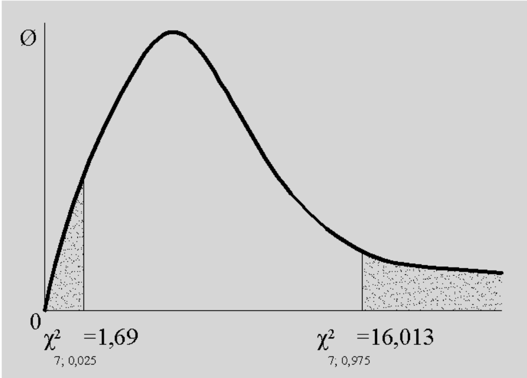
$\chi^2_{4;0.975} = 11.1$; $\chi^2_{4;0.025} = 0.484$

$\frac{4*0.00975}{11.1} \langle \sigma^2 \langle \frac{4*0.00975}{0.484}$ Rezulta $0.00351 \langle \sigma^2 \langle 0.08057$

11. Se efectueaza 8 titrari volumetrice si se obtin rezultatele: 76.48; 76.43; 77.20; 76.45; 76.25; 76.48; 76.48; 76.60 cm^3. Sa se calculeze un interval de incredere pentru dispersia masuratorilor. Se alege $1-\alpha = 0.95$.
Solutie:

$$\frac{(n-1)s^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}} \langle \sigma^2 \langle \frac{(n-1)s^2}{\chi^2_{n-1,\frac{\alpha}{2}}}$$

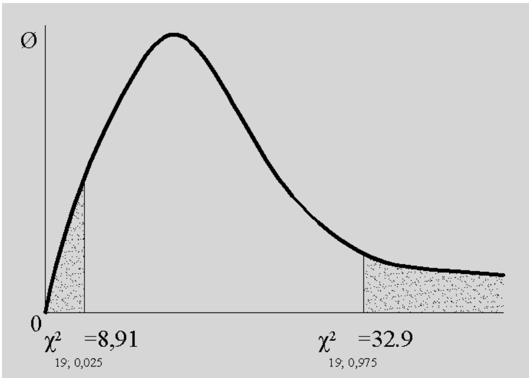
$s^2 = 0.0790$;



$\chi^2_{7;0.975} = 16.013$; $\chi^2_{7;0.025} = 1.690$

$\frac{7*0.0790}{16.013} \langle \sigma^2 \langle \frac{7*0.0790}{1.690}$ Rezulta $0.03452 \langle \sigma^2 \langle 0.3262$

12. S-au facut 20 de analize ale unei materii prime si s-au gasit o dispersie a concentratiei de substanta activa $s^2 = 0.2132$. Sa se calculeze un interval de incredere pentru dispersia masuratorilor. Se alege $1-\alpha = 0.95$.
Solutie:



$\chi^2_{19;0.975} = 32.9$; $\chi^2_{19;0.025} = 8.91$

$\frac{19*0.2132}{32.9} \langle \sigma^2 \langle \frac{19*0.2132}{8.91}$

Rezulta $0.1231 \langle \sigma^2 \langle 0.4545$

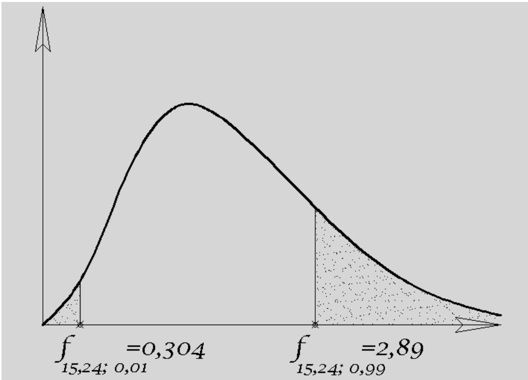
TEOREMA LIMITA CENTRALA

13. S-a efectuat o analiza asupra continutului de substanta activa a doua loturi de materie prima. S-au analizat 25 de probe din primul lot si 16 probe din cel de al doilea lot de materie. S-a obtinut o concentratia medie 82 cu dispersia 64 pentru primul lot si concentratia medie 78 cu dispersia 49 pentru al doilea lot. Sa se gaseasca un interval de incredere cu coeficientul $1-\alpha=0.98$ pentru raportul dispersiilor.

Solutie:

$n_1=25$; $n_2=16$; $s_1^2=64$; $s_2^2=49$

$f_{15.24;0.01}=0.304$; $f_{15.24;0.99}=2.89$



$$\frac{s_1^2 * f_{n_2-1,n_1-1;\frac{\alpha}{2}}}{s_2^2} \langle \frac{\sigma_1^2}{\sigma_1^2} \langle \frac{s_1^2 * f_{n_2-1,n_1-1;1-\frac{\alpha}{2}}}{s_2^2}$$

$$\frac{64}{49} * 0.304 \langle \frac{\sigma_1^2}{\sigma_1^2} \langle \frac{64}{49} * 2.89$$

Rezulta $0.397 \langle \frac{\sigma_1^2}{\sigma_1^2} \langle 3.775$ sau $0.630 \langle \frac{\sigma_1}{\sigma_2} \langle 1.943$

14. Se stie ca 90% din produsele unei intreprinderi sunt corespunzatoare. Sa se gaseasca un interval de estimatie pentru proportia p de produse corespunzatoare. Se alege $1-\alpha=0.95$ si se face o selectie de 100 produse.

Solutie

Se utilizeaza $Z=\frac{\hat{p}-p}{\sqrt{\frac{\hat{p}*\hat{q}}{n}}}$

$\hat{p} = 0.9\,;\ \hat{q} = 0.9\,;\ z_{0.975} = -z_{0.025} = 1.96$

$$0.9 - 1.96 * \sqrt{\frac{0.9 * 0.1}{100}} \ \langle\ p \ \langle\ 0.9 + 1.96 * \sqrt{\frac{0.9 * 0.1}{100}}$$

$0.841 \langle\ p \ \langle\ 0.959$

VERIFICAREA IPOTEZELOR STATISTICE

Ipoteze statistice

Ipotezele statistice sunt ipoteze asupra repartiției unor variabile aleatoare. Ele se referă fie la parametrii repartiției, fie la legea propriu zisa de repartiție. In cele ce urmeaza ne vom referi numai la ipotezele privind parametrii.

Notații conventionale

Ipoteza testată, presupusă adevarată, se numeşte <u>ipoteza nulă</u> şi se notează H_0. Testarea necesită şi formularea unei ipoteze complementare, numită <u>ipoteză alternativă</u> şi notată H_A. Dacă se acceptă H_0, în mod normal se respinge H_A şi invers.

Dacă testul priveşte valoarea unui parametru θ, de exemplu $H_0 : \theta = \theta_0$ şi $H_A : \theta = \theta_1$ se poate întâmpla ca toți ceilalți parametri ce caracterizează distribuțiile să fie cunoscuți şi, după acceptarea uneia din cele două ipoteze, distribuțiile $\rho(x, \theta_0)$ şi $\rho(x, \theta_1)$ devin complet definite. În acest caz, ipotezele sunt numite "simple". Dacă însă ceilalți parametric nu sunt cunoscuți complet, ipotezele se numesc "ipoteze compuse". De exemplu, dacă distribuția este normală şi parametrul cautat este μ , iar dispersia este necunoscută, suntem în cazul unei ipoteze compuse.

Probabilitatea unei decizii gresite

La verificarea ipotezelor se pot comite două feluri de erori:

1. Erorile de tipul 1 constau în respingerea ipotezei H_0 atunci când aceasta este adevărată.
2. Erorile de tipul 2 constau în acceptarea ipotezei H_0 atunci când aceasta este falsă.

Notatii uzuale:

α = P (respinge H_0 / H_0 adevărată) = riscul de a respinge în mod greşit H_0 se numeşte <u>nivel de semnificație</u>

β = P (acceptă H_0 / H_0 falsă) = P (respinge H_A / H_A adevărată) = riscul de a respinge în mod greşit H_A

$\pi = 1 - \beta = P\left(respinge\ H_0 \mid H_0\ \ falsa\right)$ se numeşte <u>puterea testului.</u>

VERIFICAREA IPOTEZELOR STATISTICE

Pentru a verifica o ipoteză se folosesc datele de selecție pentru calcularea unui test statistic. Domeniul de valori ale testului care corespunde respingerii ipotezei H_0 cu probabilitatea α se numeşte <u>regiune critică.</u>

Metodologia de verificare cuprinde în principiu următoarele etape:

1. se presupune, pe baza unor teste anterioare sau pe baza structurii fenomenului studiat, o repartiție pentru populația statistică din care se face selectia;
2. se formulează ipoteza;
3. se calculează valoarea testului ales şi se compară cu limitele de acceptare, respectiv respingere;
4. se acceptă sau se respinge, în funcție de rezultat, ipoteza H_0.

Ipoteze asupra mediei

1. Dispersia cunoscută

Se consideră X o selecție dintr-o populație normală $N\left(\mu,\sigma^2\right)$. Ca urmare a teoremei limită centrală, variabila aleatoare

$$Z=\frac{\overline{X}-E\left(\overline{X}\right)}{\sqrt{D\left(\overline{X}\right)}}=\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$$ este repartizata $N(0,1)$.

Pentru un nivel de semnificatie α , ipotezele şi criteriile de acceptare sau respingere sunt prezentate mai jos:

H_0	H_A	Regiunea critică
$\mu=\mu_0$	$\mu\neq\mu_0$	$z\rangle z_{1-\frac{\alpha}{2}}$
		$z\langle -z_{1-\frac{\alpha}{2}}$
$\mu=\mu_0$	$\mu\rangle\mu_0$	$z\rangle z_{1-\alpha}$
$\mu=\mu_0$	$\mu\langle\mu_0$	$z\langle -z_{1-\alpha}$

2. Dispersia necunoscută

În acest caz se înlocuieşte în formula anterioară σ cu estimația sa S_X şi se ține cont că variabila aleatoare $T=\frac{\overline{X}-\mu}{\frac{S_X}{\sqrt{n}}}$ este repartizată Student cu n-1 grade de libertate.

Dispersia de selectie S_X este de forma urmatoare: $S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2$

Ipoteze asupra diferențelor a două medii

1. Cazul când se cunosc dispersiile

Se consideră două populații normale $N\left(\mu_1, \sigma_1^2\right)$ şi $N\left(\mu_2, \sigma_2^2\right)$, o selecție aleatoare din $x_{11}, x_{12}, ..., x_{1n_1}$ din populația $N\left(\mu_1, \sigma_1^2\right)$ şi o selecție aleatoare $x_{21}, x_{22}, ..., x_{2n_2}$ din populația $N\left(\mu_2, \sigma_2^2\right)$.

Variabila aleatoare

$$z = \frac{\left(\overline{X_1} - X\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{D\left(\overline{X_1} - \overline{X_2}\right)}} = \frac{\left(\overline{X_1} - \overline{X_2}\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

este repartizată N(0,1).

2. Cazul dispersiilor necunoscute, dar presupuse egale

În cazul în care nu cunoaştem dispersiile dar ştim că sunt egale $\sigma_1^2 = \sigma_2^2 = \sigma^2$ utilizăm dispersia ponderată de selecție

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{\sum_1^{n_1}\left(x_{1i} - \overline{X_1}\right)^2 + \sum_1^{n_2}\left(x_{2i} - X\right)^2}{n_1 + n_2 - 2}$$

ca un estimator nedeplasat pentru σ^2 .

După cum s-a arătat anterior, mărimea $T = \frac{\left(\overline{X_1} - \overline{X_2}\right) - \left(\mu_1 - \mu_2\right)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ este repartizată $T\left(n_1 + n_2 - 2\right)$

3.Cazul observațiilor perechi

In cazul când observațiile formează în mod natural perechi considerăm variabila aleatoare $d = X_1 - X_2$.

Un exemplu de observatii perechi este masurarea concentrațiilor în n probe, fiecare din ele cu două metode diferite sau cazul când două medicamente se administrează unui aceluiaşi lot de voluntari, în două perioade diferite.

În cazul în care selecțiile aparțin la aceiaşi populație, media lui d va fi zero: $E(\overline{d}) = 0$.

Când se cunosc dispersiile avem $D(\overline{d}) = \sigma_d^2 = \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}$ şi variabila aleatoare $\frac{\overline{d}}{\sigma_d}$ este repartizată $N(0,1)$.

Când nu se cunosc dispersiile se folosesc dispersiile de selecție şi se ține cont că variabila aleatoare $\frac{\overline{d}}{\frac{s_d}{\sqrt{n}}}$ după cum se poate arăta uşor, este repartizată Student cu n-1 grade de libertate.

Compararea proporțiilor

Dacă vom considera un experiment în care răspunsul este de tip da sau nu, de exemplu vindecare sau nevindecare, supraviețuire sau moarte, etc., numărul de rezultate k de un anumit tip în n repetări ale experimentului este o variabilă aleatoare repartizată binomial.

Variabila aleatoare standardizată $z = \frac{k - E(k)}{\sqrt{D(k)}} = \frac{k - np}{\sqrt{npq}} = \frac{\frac{k}{n} - p}{\sqrt{\frac{pq}{n}}}$ se aproximează ca fiind normal repartizată.

Fie două populații de tip "urna Poisson cu bile albe şi bile negre", cu parametrii (probabilitatea bilei albe) p_1 şi respectiv p_2 . În două selecții din cele două populații, de volum n_1 şi respectiv n_2 presupunem că s-a obținut răspuns "pozitiv" de k_1 şi respectiv k_2 ori.

Estimarea dispersiei

Considerăm o selecție de volum n dintr-o populație normală $N(\mu, \sigma^2)$. Variabila aleatoare $v = \frac{(n-1)s^2}{\sigma^2}$ este repartizată $\chi^2(n-1)$.

Solutie:

$$v = \frac{(n-1)s^2}{\sigma^2} = \frac{\sum_1^n (x_i - \overline{X})^2}{\sigma^2} = \frac{\sum_1^n [(x_i - \mu) - (\overline{X} - \mu)]^2}{\sigma^2} =$$

$$= \frac{\sum_1^n (x_i-\mu)^2 - 2(\overline{X}-\mu)(n\overline{X}-n\mu) + n(\overline{X}-\mu)^2}{\sigma^2} =$$

$$= \frac{\sum_1^n (x_i-\mu)^2 - n(\overline{X}-\mu)}{\sigma^2} = \sum_1^n \left(\frac{x_i-\mu}{\sigma}\right)^2 - \left(\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right)^2$$

Dar $\frac{x_i-\mu}{\sigma}$ este repartizat N(0,1) căci $M\left(\frac{x_i-\mu}{\sigma}\right) = \frac{M(x_i)-\mu}{\sigma} = 0$

şi $D^2\left(\frac{x_i-\mu}{\sigma}\right) = 1$

Deci v este o sumă de n-1 pătrate de variabile de tip N(0,1).

Estimarea raportului a două dispersii

Se consideră selecția aleatoare $x_{11}, x_{12}, ..., x_{1n_1}$ dintr-o populație $N(\mu_1, \sigma_1^2)$ şi o selecție aleatoare $x_{21}, x_{22}, ..., x_{2n_2}$ dintr-o populație $N(\mu_2, \sigma_2^2)$.

Raportul $F = \frac{\frac{s_1^2}{\sigma_1^2}}{\frac{s_2^2}{\sigma_2^2}}$ este repartizat $F(n_1-1, n_2-1)$.

Aplicatii:

1. Se presupune că efectuarea a 4 măsuratori de normalitate conduc la valoarea medie $\overline{X} = 1.2*10^{-3}M$. Se ştie ca dispersia corespunzătoare unei măsuratori este $\sigma^2 = 49*10^{-8}M^2$. Să se verifice ipoteza potrivit căreia valoarea medie a normalității este $\mu_0 = 10^{-3}M$. Se alege $\alpha = 0.05$.

Soluție:

Se utilizează testul $Z = \frac{\overline{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$

Ipotezele statistice sunt:

$$H_0: \quad \mu = 10^{-3} M \text{ vs.}$$
$$H_A: \quad \mu \neq 10^{-3} M$$

Regiunea critica este $z \rangle z_{1-\frac{\alpha}{2}}$; $z \langle -z_{1-\frac{\alpha}{2}}$

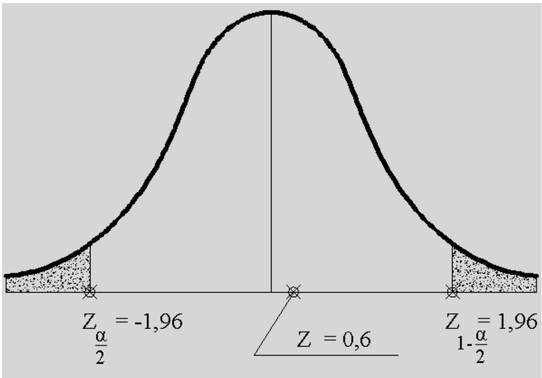
Calculand variabila aleatoare Z se obtine:

$$Z = \frac{1,2*10^{-3} - 10^{-3}}{\frac{7*10^{-4}}{\sqrt{4}}} = \frac{0,2*10^{-3}*2}{7*10^{-4}} = \frac{4}{7} \cong 0,6$$

$z_{1-\frac{\alpha}{2}} = z_{1-\frac{0,05}{2}} = z_{0,975}$

Din tabelul distributiei normale se obtine $z_{0,975} = 1,96$

Deci, regiunea critică este $z \rangle 1.96$ sau $z \langle -1.96$



Cum $-1,96 \langle 0,6 \langle 1,96 \Rightarrow$ z nu apartine regiunii critice, deci se acceptă ipoteza H_0

2. O întreprindere trebuie să livreze recipiente cu material recuperat cu greutatea de 15 kg şi abaterea medie pătratică $\sigma = 0.5\,kg$. Un control efectuat asupra a 49 de piese duce la o valoare medie $\overline{X} = 14.8\,kg$. Să se verifice ipoteza potrivit căreia masa medie este de 15 kg. Se alege $\alpha = 0,01$.

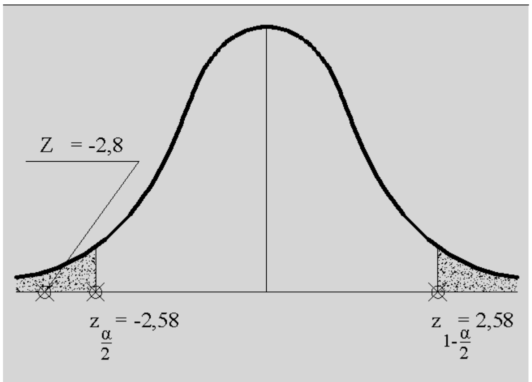
Soluție:

$H_0: \quad \mu = 15$

$H_A: \quad \mu \neq 15$

$$Z = \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{14.8 - 15}{\frac{0.5}{\sqrt{49}}} = \frac{-0,2 * 7}{0,5} = \frac{-1,4}{0,5} = -2,8\,;$$

$z_{1-\frac{\alpha}{2}} = z_{0.995} = 2.58$



Deoarece $-2,8 \langle -2,58$, rezulta ca z este in regiunea critica, deci ipoteza H_0 se respinge.

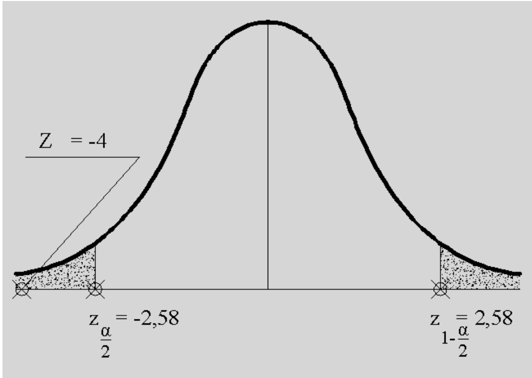
3. Experientele anterioare arata ca greutatea unui comprimat este o variabila aleatoare cu abaterea medie patratica de 15 mg. O selectie de volum 9 ne da o greutate medie $\overline{X} = 400\ mg$. Sa se verifice la un prag de semnificatie $\alpha = 0.01$ ipoteza potrivit careia greutatea medie este de 420 mg.

Soluție:

$H_0: \quad \mu = 420$ vs. $H_A: \quad \mu \neq 420$

$$Z = \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{400 - 420}{\frac{15}{3}} = -4$$

$z_{1-\frac{\alpha}{2}} = z_{0.995} = 2.58$



Regiunea critică este $Z > 2.58$ sau $Z < -2.58$.

Deoarece $-4 < -2.57$ rezulta ca Z este in regiunea critica, deci se respinge H_0

4. O selectie de 16 loturi de coprolactoma cristalizata are continutul de baze volatile $\overline{X} = 0.23$ miliechivalenti / kg. Presupunand ca acest continut este o variabila aleatoare cu abaterea medie patratica 0.07 sa se verifice ipoteza :

$H_0: \quad \mu = 0.20$

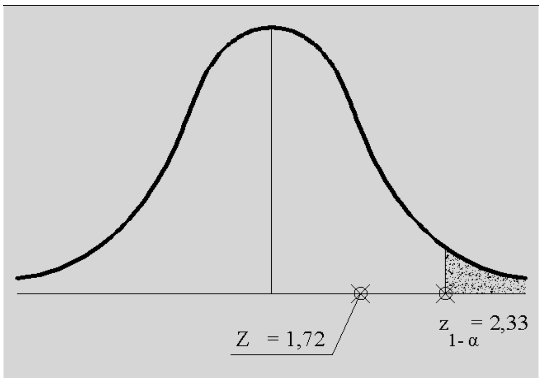
$H_A: \quad \mu > 0.20$

Se alege $1-\alpha = 0.99$.

Soluție :

$$Z = \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{0.23 - 0.20}{\frac{0.07}{\sqrt{16}}} = \frac{0{,}03 * 4}{0{,}07} = \frac{0{,}12}{0{,}07} \cong 1.72\,; \quad z_{0.99} = 2.33$$

Regiunea critică este $Z > z_\alpha$.



Deoarece $1{,}72 \langle\ 2{,}33$, Z nu este in regiunea critica, deci se accepta ipoteza H_0

5. Durata de functionare a unui electrod este o variabila aleatoare cu $\sigma = 200\ h$. O selectie de 25 astfel de electrozi da o durata de functionare de 1380 h. Cu $\alpha = 0.01$ sa se verifice ipoteza:

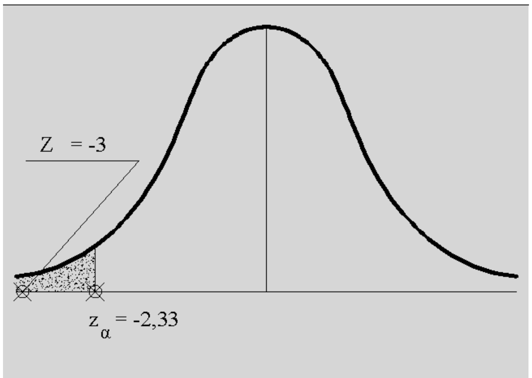
$H_0 :\quad \mu = 1500\ h$

$H_A :\quad \mu \langle\ 1500\ h$

Soluție:

Variabila aleatoare este $Z = \dfrac{\overline{X} - \mu_0}{\dfrac{\sigma}{\sqrt{n}}}$ unde $\overline{X} = 1380\ h$, $\mu_0 = 1500\ h$ si $n = 25$

Deci, $Z = \dfrac{1380 - 1500}{\dfrac{200}{\sqrt{25}}} = \dfrac{-120 * 5}{200} = -3$



Regiunea critică este $Z \langle\ z_\alpha$ unde $z_\alpha = z_{0.01} = -2.33$

Deoarece $-3 \langle -2{,}33$ rezulta ca Z se afla in regiunea critica, deci se respinge ipoteza $H_0 :\quad \mu = 1500\ h$

6. Rezultatele unor cantariri de etaloane sunt urmatoarele: 26,7 ; 26,8 ; 25,8 ; 25,7. Se poate afirma ca media greutatilor este mai mica decat 26.5 g? Se alege $\alpha = 0.05$.

Soluție:

$H_0: \quad \mu = 26.5$

$H_A: \quad \mu \langle 26.5$

Se utilizează variabila $T = \dfrac{\overline{X} - \mu_0}{\dfrac{S}{\sqrt{n}}}$ care este repartizata $T(n-1)$,

regiunea critică fiind $T \langle t_{n-1;\alpha}$,

$$\overline{X} = \frac{\sum x_i}{n} = \frac{26{,}7 + 26{,}8 + 25{,}8 + 25{,}7}{4} = 26{,}25$$

$$S^2 = \frac{1}{n-1}\sum\left(x_i - \overline{X}\right)^2 =$$

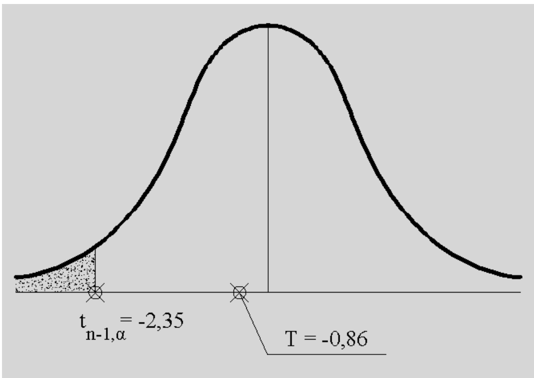
$$= \frac{1}{4-1}\left[(26{,}7 - 26{,}25)^2 + (26{,}8 - 26{,}25)^2 + (25{,}8 - 26{,}25)^2 + (25{,}7 - 26{,}25)^2\right] =$$

$$= \frac{1}{3}\left[0{,}45^2 + 0{,}55^2 + (-0{,}45)^2 + (-0{,}55)^2\right] = \frac{1{,}01}{3} \cong 0{,}34$$

Deci, $S \cong 0{,}58$

$$T = \frac{26{,}25 - 26.5}{\dfrac{0{,}58}{\sqrt{4}}} = \frac{-0{,}25 * 2}{0{,}58} = -\frac{0{,}50}{0{,}58} \cong -0{,}86$$

$$t_{n-1,\alpha} = -t_{n-1;1-\alpha} = -t_{3;0.95} = -2{,}35$$



Deoarece $-2{,}35 \langle -0{,}86$, T nu este situat in regiunea critică deci se accepta ipoteza H_0 adica $\mu \langle 26.5$

7. Douazeci si cinci de determinari ale unei conversii au dat valoarea medie $\overline{X} = 54.76$ si dispersia de selectie $S^2 = 4$. Sa se verifice daca conversia medie este mai mica decat 55. Se alege $\alpha = 0.01$.

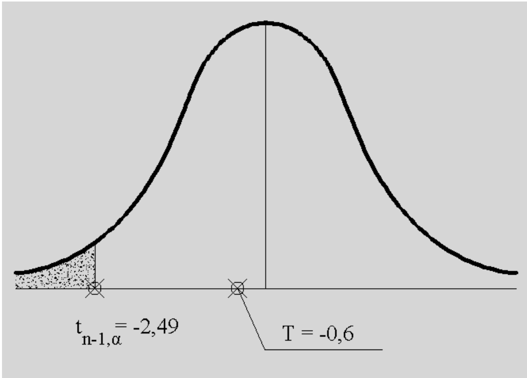
Soluție:

$H_0 : \quad \mu = 55$

$H_A : \quad \mu \langle 55$

$$T = \frac{\overline{X} - \mu_0}{\frac{S}{\sqrt{n}}} = \frac{54{,}76 - 55}{\frac{2}{\sqrt{25}}} = \frac{-0{,}24 * 5}{2} = -\frac{1{,}20}{2} = -0{,}60$$

Regiunea critică este $T \langle t_{n-1;\alpha}$ unde $t_{n-1,\alpha} = -t_{n-1;1-\alpha} = -t_{24;0.99} = -2{,}49$.



Deci T nu apartine regiunii critice ceea ce inseamna ca se poate accepta ipoteza H_0.

8. Se stie ca greutatea media a unor recipienti este $\mu = 6.80\,kg$. O selectie de volum $n = 9$ ne da greutatea medie $\overline{X} = 6.5$ si dispersia de selectie medie $S^2 = 0.25$. Acest rezultat infirma experientele anterioare? Se alege $\alpha = 0.05$.

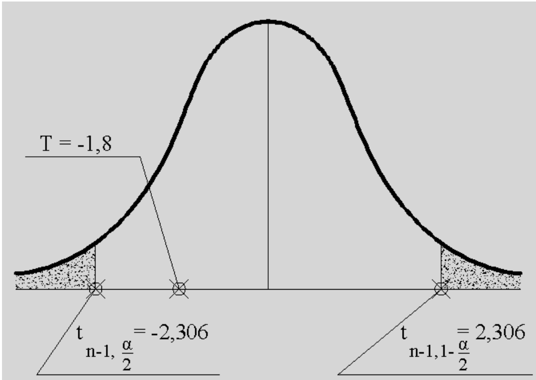
Soluție:

$H_0 : \quad \mu = \mu_0$; $H_A : \quad \mu \neq \mu_0$

$$T = \frac{\overline{X} - \mu_0}{\frac{S}{\sqrt{n}}} = \frac{6.5 - 6.8}{\frac{0.5}{3}} = -1.8$$

$$t_{n-1;1-\frac{\alpha}{2}} = t_{8;0.975} = 2.306$$

Regiunea critică este $T > t_{n-1;1-\frac{\alpha}{2}}$ sau $T < -t_{n-1;1-\frac{\alpha}{2}}$.



Aşadar se acceptă H_0

9. S-au facut 25 determinari asupra continutului de component activ al unui amestec etalon si s-a determinat media $\overline{X} = 34.45$ mg/l si $S = 0.9$. In etalon s-a introdus cantitatea $\mu_0 = 34.00$ mg/l. Valoarea gasita pentru medie este intamplatoare sau este datorata unor erori sistematice de metoda? Se alege $\alpha = 0.05$.

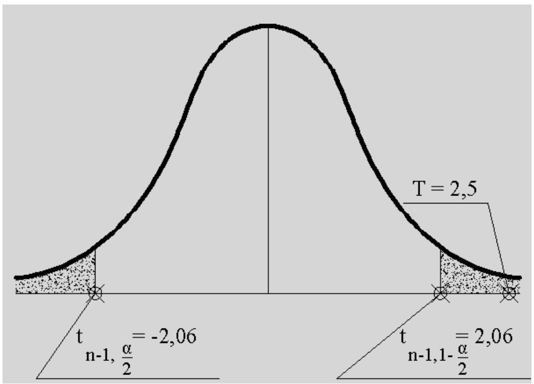
Soluție

Se verifica ipoteza $H_0: \quad \mu = \mu_0$; $H_A: \quad \mu \neq \mu_0$

$$T = \frac{\overline{X} - \mu_0}{\frac{S}{\sqrt{n}}} = \frac{34,45 - 34,00}{\frac{0,9}{\sqrt{25}}} = \frac{0,45 * 5}{0,9} = 2,5$$

Regiunea critică este $T > t_{n-1;1-\frac{\alpha}{2}}$ sau $T < -t_{n-1;1-\frac{\alpha}{2}}$.

$t_{24;0.975} = 2,06$



Deoarece $2,5 \rangle 2,06$, T se gaseste in regiunea critica deci respingem ipoteza H_0. Aşadar valoarea gasita pentru media de selectie este datorata unor erori sistematice de metoda.

10. Patru termometre sunt introduse intr-un mediu cu temperatura fixa 1000°C. Ele dau indicatiile: 986, 1005, 991, 994. Sa se verifice ipoteza potrivit careia abaterile de la valoarea $\mu_0 = 1000\,^\circ C$ sunt datorate experientelor. Se alege $\alpha = 0.05$.

Soluție:

Se verifica ipoteza $H_0: \quad \mu = \mu_0$; $H_A: \quad \mu \neq \mu_0$

$T = \dfrac{\overline{X} - \mu_0}{\dfrac{S}{\sqrt{n}}}$; Regiunea critică este $T \rangle t_{n-1;1-\frac{\alpha}{2}}$ sau $T \langle -t_{n-1;1-\frac{\alpha}{2}}$.

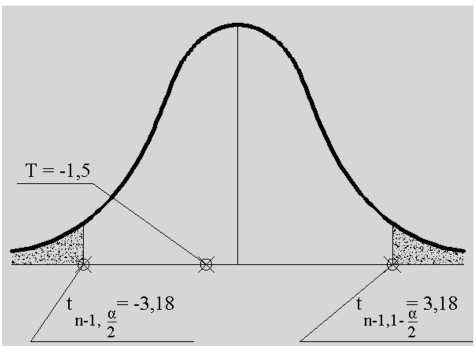
$$\overline{X} = \frac{\sum x_i}{n} = \frac{986 + 1005 + 991 + 994}{4} = 994;$$

$$S^2 = \frac{1}{n-1} \sum \left(x_i - \overline{X}\right)^2$$

$$S^2 = \frac{1}{n-1}\left[(986-994)^2 + (1005-994)^2 + (991-994)^2 + (994-994)^2\right]$$

$$= \frac{1}{3}\left[(-8)^2 + 11^2 + (-3)^2 + 0\right] = \frac{1}{3} * 194 \cong 64,7$$

Deci $T = \dfrac{994 - 1000}{\sqrt{\dfrac{64.4}{4}}} = -\dfrac{6*2}{8,04} \cong -1,5$ unde $t_{3;0.975} = 3,18$



Deoarece $-3{,}18 \langle -1{,}5 \langle 3{,}18$ nu se respinge H_0 adica erorile sunt datorate experientelor.

11. Zece determinari ale procentului de clor dintr-o soluție au condus la $\overline{X} = 0.832\,\%$ si $S = 0.02\,\%$. Daca adevaratul continut este $\mu_0 = 0.9\,\%$ sa se verifice ipoteza

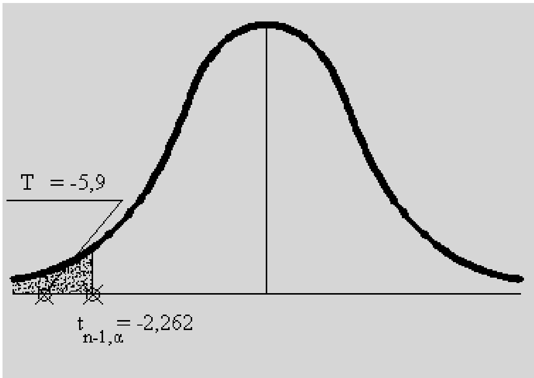
$H_0 :\quad \mu = \mu_0$; $H_A :\quad \mu \langle \mu_0$

Se alege $\alpha = 0.05$.

Soluție :

$$T = \frac{\overline{X} - \mu_0}{\frac{S}{\sqrt{n}}} = \frac{0.832 - 0.9}{\frac{0.02}{\sqrt{10}}} = -5.9$$

Zona critică este $T \langle t_{n-1,\alpha}$;



$t_{9;0.05} = -2.262$, aşadar se respinge ipoteza H_0

12. Determinarile succesive efectuate in doua vase deschise care contin HCl au dat rezultatele :

I	II
15.75	15.58
15.64	15.49
15.92	15.72

Sa se stabileasca daca cele doua vase au in medie aceeasi compozitie. Se cunoaste distersia $\sigma_1^2 = \sigma_2^2 = 0.16$. Se alege $1 - \alpha = 0.95$

Soluție :

Se verifica $H_0: \quad \mu_1 = \mu_2$; $H_A: \quad \mu_1 \neq \mu_2$

Se utilizează variabila $Z = \dfrac{\overline{X_1} - \overline{X_2} - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$ unde

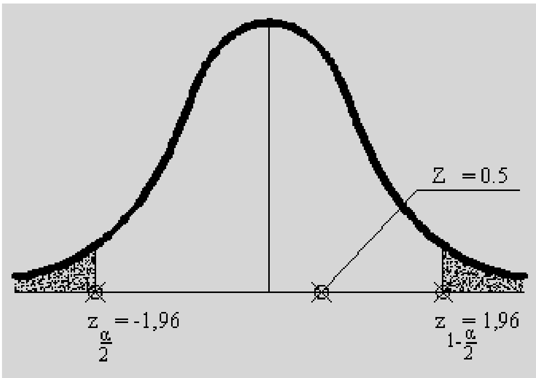
$\overline{X_1} = \dfrac{15,75 + 15,64 + 15,92}{3} = 15.77$ si $\overline{X_2} = \dfrac{15,58 + 15,49 + 15,72}{3} = 15.60$

Conform ipotezei $H_0: \mu_1 = \mu_2$, deci $\mu_1 - \mu_2 = 0$

Deci $Z = \dfrac{15,77 - 15,60 - 0}{\sqrt{\dfrac{0,16}{3} + \dfrac{0,16}{3}}} = \dfrac{0,17}{0,4}\sqrt{\dfrac{3}{2}} \cong 0,5$;

$z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96$

Zona critică este $Z \rangle z_{1-\frac{\alpha}{2}}$ sau $Z \langle -z_{1-\frac{\alpha}{2}}$



Nu se poate respinge H_0 deoarece $-1.96 \langle 0,5 \langle 1.96$

13. Doua cantare automate M_1 si M_2 sunt folosite pentru ambalarea unui produs in pachete de 1000 g. Se stie ca produsele ambalate au masele repartizate normal cu mediile μ_1 si μ_2 si abaterile medii patratice $\sigma_1 = 3\ g$, respectiv $\sigma_2 = 4\ g$. Se cantaresc cate 100 pachete din produsele ambalate de fiecare dintre cantare si se obtin rezultatele $\overline{X_1} = 1007\ g$ si $\overline{X_2} = 1002\ g$.

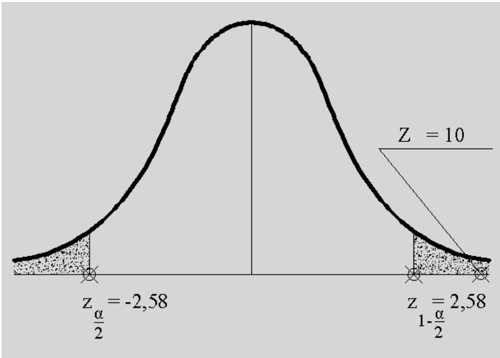
La un prag de semnificatie $\alpha = 0.01$ sa se verifice daca pachetele ambalate au aceeasi masa.

Soluție:

$H_0: \quad \mu_1 = \mu_2$; $H_A: \quad \mu_1 \neq \mu_2$

$$Z = \frac{\overline{X_1} - \overline{X_2} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{1007 - 1002 - 0}{\sqrt{\frac{3^2}{100} + \frac{4^2}{100}}} = \frac{5}{\sqrt{\frac{25}{100}}} = 5 * 2 = 10$$

$z_{0.995} = 2.58$



Deoarece $Z \rangle z_{1-\frac{\alpha}{2}}$ se respinge ipoteza H_0.

14. Se analizeaza doua loturi de materii prime indexate prin 1 si 2. Pentru cele 75 (80) de probe din lotul 1 (2) se obtin concentratiile medii $\overline{X_1} = 44$ (respectiv $\overline{X_2} = 40$). Se stie ca dispersiile sunt $\sigma_1^2 = 100$ si $\sigma_2^2 = 40$. Cu un prag de semnificatie $\alpha = 0.01$ sa se verifice daca diferenta medie de concentratie intre cele doua loturi este mai mare decat 2.

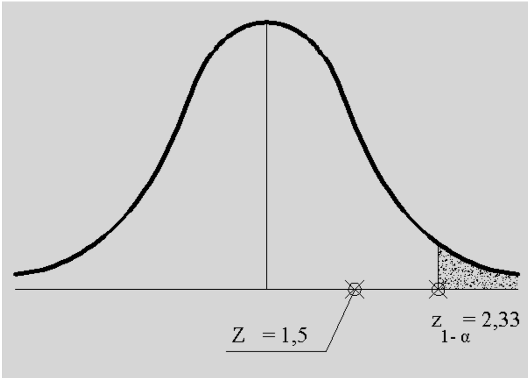
Soluție:

$H_0: \quad \mu_1 - \mu_2 = 2$; $H_A: \quad \mu_1 - \mu_2 \rangle 2$

$$Z = \frac{\overline{X_1} - \overline{X_2} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$n_1 = 75$	$n_2 = 80$
$\overline{X_1} = 44$	$\overline{X_2} = 40$
$\sigma_1^2 = 100$	$\sigma_2^2 = 40$
$\mu_1 - \mu_2 = 2$	

Deci, $Z = \dfrac{\overline{X_1} - \overline{X_2} - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = \dfrac{44 - 40 - 2}{\sqrt{\dfrac{100}{75} + \dfrac{40}{80}}} = \dfrac{2}{\sqrt{\dfrac{4}{3} + \dfrac{1}{2}}} = 2 * \sqrt{\dfrac{6}{11}} \cong 1.5$



Regiunea critică este $Z \rangle z_{1-\alpha}$ unde $z_{1-\alpha} = z_{0.99} = 2.33$, deci se acceptă H_0

15. Pentru a studia efectul concentratiei de catalizator asupra conversiei se fac doua grupe de observatii si se obtin datele:

I	II
5.18	5.58
5.52	5.62
5.42	5.82

Se poate considera ca cele doua tipuri de catalizator duc la aceeasi conversie medie? Se alege $\alpha = 0.05$ si se considera dispersiile egale.

Soluție :

$H_0 : \quad \mu_1 = \mu_2$; $H_A : \quad \mu_1 \neq \mu_2$

$\overline{X_1} = \dfrac{5{,}18 + 5{,}52 + 5{,}42}{3} \cong 5.37$; $\overline{X_2} = \dfrac{5{,}58 + 5{,}62 + 5{,}82}{3} \cong 5.67$;

Avem: $s_1^2 = \dfrac{1}{n_1 - 1} \sum_{i=1}^{n_1} \left(x_{1i} - \overline{X_1}\right)^2$; $s_2^2 = \dfrac{1}{n_2 - 1} \sum_{i=1}^{n_2} \left(x_{2i} - \overline{X_2}\right)^2$ si

$S_p^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$

Deci,

$s_1^2 = \dfrac{1}{3-1}\left[(5{,}18 - 5{,}37)^2 + (5{,}52 - 5{,}37)^2 + (5{,}42 - 5{,}37)^2\right] \cong \dfrac{1}{2} * 0{,}06 = 0{,}03$

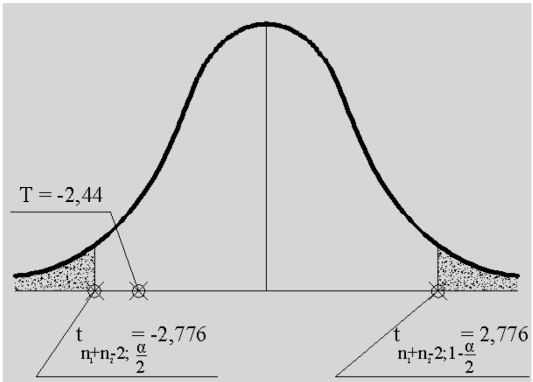
$$s_2^2 = \frac{1}{3-1}\left[(5{,}58-5{,}67)^2 + (5{,}62-5{,}67)^2 + (5{,}82-5{,}67)^2\right] \cong \frac{1}{2} * 0{,}03 = 0{,}015$$

$$S_p^2 = \frac{2s_1^2 + 2s_2^2}{3+3-2} = \frac{2(0{,}03+0{,}015)}{4} = 0.0225,\ \text{deci } S_p = 0{,}15$$

$T = \dfrac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ este distribuita T cu $n_1 + n_2 - 2$ grade de libertate.

$$T = \frac{5{,}37 - 5{,}67 - 0}{0{,}15\sqrt{\frac{1}{3}+\frac{1}{3}}} = -\frac{0{,}3}{0{,}15}\sqrt{\frac{3}{2}} = -2\sqrt{\frac{3}{2}} = -\sqrt{\frac{4*3}{2}} = -\sqrt{6} \cong -2{,}44$$

Regiunea critica este $T \langle\, t_{n_1+n_2-2;\frac{\alpha}{2}}$ si $T \,\rangle\, t_{n_1+n_2-2;1-\frac{\alpha}{2}}$



$t_{4;0.975} = 2.776$

Deoarece $-2.776 \langle -2{,}44 \langle 2.776$ nu se poate respinge H_0.

16. Pentru a compara doua benzine cu cifrele octanice 90 si 98 o anumita cantitate este folosita in 5 automobile pentru incercare. Se masoara distanta parcursa pana la oprire si se obtin valorile

	98	90
$\overline{X}$ (km)	22.7	21.3
S (km)	0.45	0.55

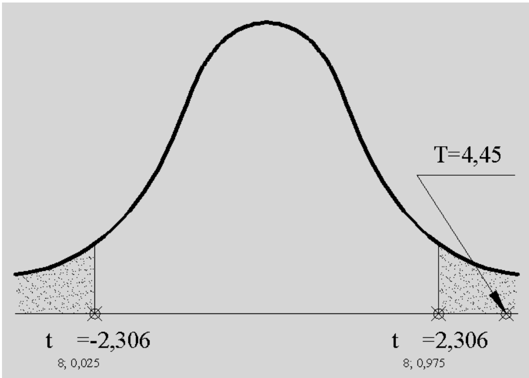
Sa se verifice daca cele doua benzine sunt diferite. Se alege $\alpha = 0.05$

Soluție:

$H_0: \quad \mu_{98} = \mu_{90}$; $H_A: \quad \mu_{98} \neq \mu_{90}$

$$S_p^2 = \frac{2s_1^2 + 2s_2^2}{n_1 + n_2 - 2} = \frac{4*0.45^2 + 4*0.55^2}{5+5-2} = 0.252 \text{ deci } S_p = 0.5$$

$$T = \frac{\overline{X}_{98} - \overline{X}_{90}}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{22.7 - 21.3}{0.5*\sqrt{\frac{2}{3}}} = 4.45;$$



$t_{8;0.975} = 2.306$ deci se respinge H_0

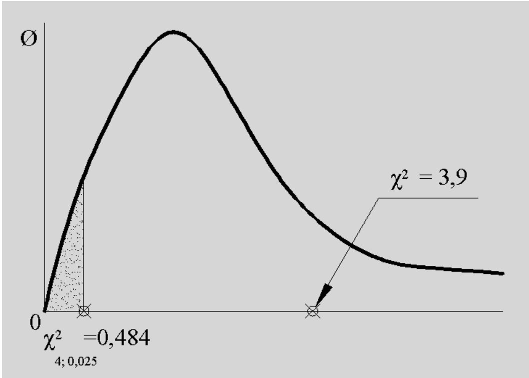
17. Cinci determinari de debit pentru un schimbator de caldura au dat valorile: 5.84; 5.76; 6.03; 5.90; 5.87 kg / s. Se poate presupune ca dispersia pentru aceste masuratori este mai mica decat 0.01? se alege $\alpha = 0.025$

Soluție :

$H_0: \quad \sigma^2 = 0.01$; vs. $H_A: \quad \sigma^2 < 0.01$

Se utilizează variabila $\chi^2 = (n-1)\frac{S^2}{\sigma_0^2}$; $S^2 = 0.00975$;

$\chi^2 = \frac{4*0.00975}{0.01} = 3.9$ Regiunea critică este data de $\chi^2 < \chi^2_{n-1,\alpha}$



Deoarece $\chi^2_{4;0.025} = 0.484$ nu se respinge H_0 si nu putem considera $\sigma^2 \langle\ 0.01$

18. Doua echipe de experimentatori au efectuat cate 13 observatii asupra unor temperaturi de reactie. S-au obtinut rezultatele:

$\overline{X_1} = 310.8423\,^{\circ}$C, $s_1^2 = 1.1867$, respectiv

$\overline{X_2} = 310.5246\,^{\circ}$C, $s_2^2 = 1.5757$

Exista diferente semnificative intre rezultatele obtinute? Se alege $\alpha = 0.02$.

Soluție:

Inainte de a aplica testul T se aplica testul F pentru a ne asigura de egalitatea dispersiilor.

$H_0: \ \sigma_1^2 = \sigma_2^2$; vs. $H_A: \ \sigma_1^2 \neq \sigma_2^2$

Se utilizează variabila $F = \dfrac{s_1^2}{s_2^2} = \dfrac{1.5757}{1.1867} = 1.3$

Se alege $\alpha = 0.02$; $f_{12,12;0.99} = 4.16$; $f_{12,12;0.01} = 0.241$

Deoarece $0.241 \langle 1.3 \langle 4.16$ se acceptă ipoteza H_0.

In continuare se aplica testul T

$$T = \frac{\overline{X_1} - \overline{X_2}}{S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} = \frac{0.3117}{0.46095} = 0.689;$$

$S_p^2 = \dfrac{12*1.1867 + 12*1.5757}{24} = 1.3812$; $t_{24;0.025} = -2.064$; $t_{24;0.975} = 2.064$

Deoarece $-2.064 \langle 0.689 \langle 2.064$ nu se respinge ipoteza H_0, deci intre cele doua echipe nu exista deosebiri semnificative.

19. Doua pompe cu debitul nominal de 100 l/min au functionat cu debitele:

1	97.8	98.9	101.2	98.8	102	99	99.1	100.8	100.9	100.5
2	97.2	100.5	98.2	98.3	97.5	99.9	97.9	96.8	97.4	97.2

Sa se verifice daca sunt caracterizate de aceeasi dispersie. Se alege $\alpha = 0.05$

Soluție

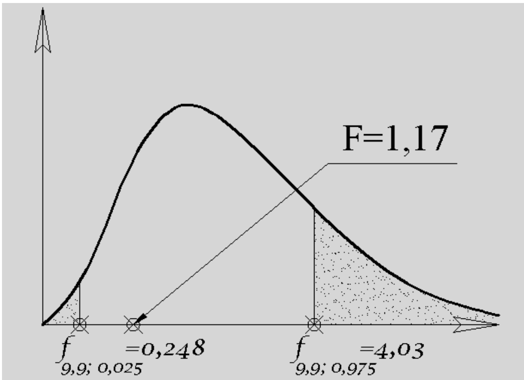
Se aplica testul F:

$H_0: \quad \sigma_1^2 = \sigma_2^2; \; H_A: \quad \sigma_1^2 \neq \sigma_2^2$

$\overline{X_1} = 99.9, \; s_1^2 = 1.69,$

$\overline{X_2} = 98.1, \; s_2^2 = 1.44$

$$F = \frac{s_1^2}{s_2^2} = \frac{1.69}{1.44} = 1.17; \; f_{9,9;0.975} = 4.03; \; f_{9,9;0.025} = 0.248$$



Deoarece $0.248 \langle 1.17 \langle 4.03$ nu se respinge ipoteza H_0.

20. Doua clase de experiente dau rezultatele:

$\overline{X_1} = 1.219, \; s_1^2 = 0.28, \; n_1 = 16$

$\overline{X_2} = 1.179, \; s_2^2 = 0.143, \; n_2 = 15$

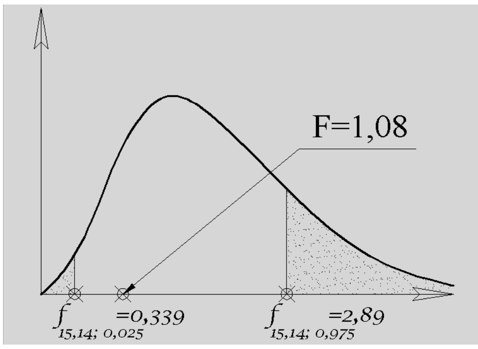
Se poate trage concluzia ca ambele experiente duc la acelasi rezultat? Se alege $\alpha = 0.05$

Soluție:

Se aplica testul F si pe aceasta baza testul T.

$H_0: \quad \sigma_1^2 = \sigma_2^2; \; H_A: \quad \sigma_1^2 \neq \sigma_2^2$

$$F = \frac{s_1^2}{s_2^2} = \frac{0.208}{0.143} = 1.08\,;\ f_{15,14;0.975} = 2.89\,;\ f_{15,14;0.025} = 0.339$$



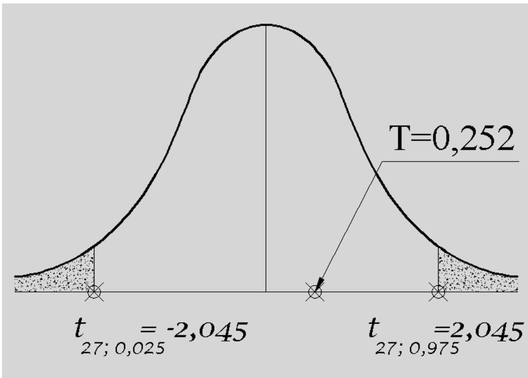
Deoarece $0.339 \langle 1.08 \langle 2.89$ nu se respinge ipoteza H_0.

In continuare se aplica testul T

$H_0 :\quad \mu_1 = \mu_2$; vs. $H_A :\quad \mu_1 \neq \mu_2$

$$S_p^2 = \frac{15 * 0.208 + 14 * 0.193}{29} = 0.201$$

$$T = \frac{\overline{X_1} - \overline{X_2}}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{1.219 - 1.179}{0.447 * 0.35} = 0.252\,;\ t_{27;0.975} = 2.045$$



Deoarece $-2.045 \langle 0.252 \langle 2.045$ se acceptă ipoteza $H_0 :\quad \mu_1 = \mu_2$.

21. S-au efectuat doua serii de cate 25 de experiente obtinandu-se abaterile standard $s_1 = 0.023$ si $s_2 = 0.019$. Sa se compare dispersiile celor doua serii de experiente.

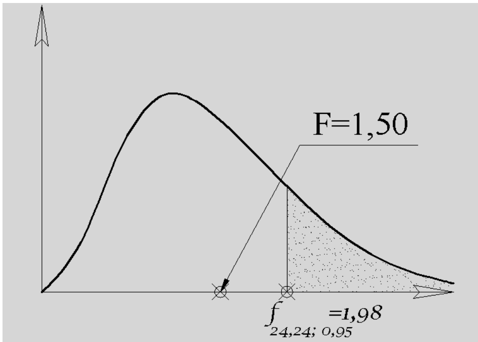
Soluție

Se aplica testul F

$H_0: \quad \sigma_1^2 = \sigma_2^2;\ H_A: \quad \sigma_1^2 \neq \sigma_2^2$

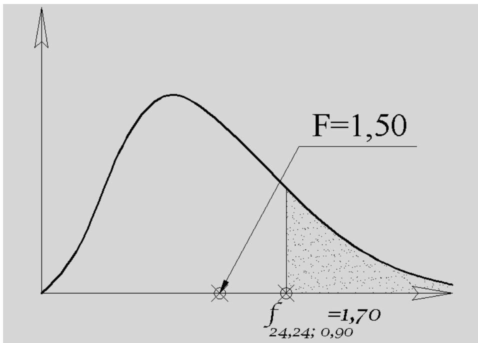
$$F = \frac{s_1^2}{s_2^2} = \frac{0.023^2}{0.019^2} = 1.5$$

Se alege $\alpha = 0.05$ si $f_{24,24;0.95} = 1.98$,



deci nu se poate respinge ipoteza H_0.

Pentru $\alpha = 0.10$ si $f_{24,24;0.90} = 1.70$ concluzia este aceeasi.



TESTE NEPARAMETRICE

Testele independente de distributie, numite şi teste de rang, înlocuiesc valorile variabilei cantitative observate cu rangurile lor. Testele neparametrice sunt valabile şi pentru variabile normal distribuite, dar sunt mai puțin eficiente, pentru acelaşi prag de semnificație fiind necesare eşantioane mai mari decât pentru testele parametrice.

Aplicarea lor este posibila atunci cand variabilele aleatoare sunt continue si independente.

Să considerăm o populație finită de N elemente, la care asociem numerele $x_1, x_2, ..., x_N$. Dacă presupunem că toate elementele au aceeaşi probabilitate $\frac{1}{N}$, atunci media şi dispersia populației sunt urmatoarele:

$$\mu = E(X) = \frac{1}{N}\sum_1^N x_i \text{ şi } \sigma^2 = D(X) = \frac{N-1}{N^2}\sum_1^N x_i^2 - \frac{2}{N^2}\sum_{i\neq j} x_i x_j$$

Solutie:

a) $\mu = E(X) = \sum_1^N x_i p_i = \frac{1}{N}\sum_1^N x_i$

b)

$$\sigma^2 = D(X) = E(X^2) - (E(X))^2 = \sum_1^N x_i^2 p_i - \left(\sum_1^N x_i p_i\right)^2 =$$

$$= \frac{1}{N}\sum_1^N x_i^2 - \frac{1}{N^2}\left(\sum_1^N x_i\right)^2 =$$

$$= \left(\frac{1}{N} - \frac{1}{N^2}\right)\sum_1^N x_i^2 - \frac{2}{N^2}\sum_{i\neq j} x_i x_j = \frac{N-1}{N^2}\sum_1^N x_i^2 - \frac{2}{N^2}\sum_{i\neq j} x_i x_j$$

Multimea tuturor selecțiilor posibile de mărimea n din populație va include:

$(x_1, x_2, ..., x_{n-1}, x_n)$

$(x_1, x_2, ..., x_{n-1}, x_{n+1})$

.

.

.

$(x_{N-n+1}, x_{N-n+2}, ..., x_N)$

Aceste probe sunt formate prin alegerea a n elemente din N. Există C_N^n căi de a alege o astfel de probă. Presupunem că fiecare probă are aceeaşi probabilitate de a fi selectată, $\frac{1}{C_N^n}$.

Să considerăm media selecției j: $\overline{X_j} = \frac{1}{n}\sum_{i=1}^{n} x_{ji}$ şi să considerăm variabila aleatoare $\overline{X} = \left(\overline{X_j}\right)_{j=1,C_N^n}$

Proprietati:

1. Vom demonstra ca *<u>media mediei probei este egală cu media populației</u>*, ceea ce inseamna ca $E\left(\overline{X}\right) = \mu$

Solutie:

$$E\left(\overline{X}\right) = \sum_{j=1}^{C_N^n} \overline{X_j}\, p_j = \frac{1}{C_N^n}\sum_{j=1}^{C_N^n} \overline{X_j} =$$

$$= \frac{1}{C_N^n}\begin{bmatrix} \frac{1}{n}\left(x_1 + x_2 + ... + x_{n-1} + x_n\right) + \frac{1}{n}\left(x_1 + x_2 + ... + x_{n-1} + x_{n+1}\right) + ... + \\ + \frac{1}{n}\left(x_{N-n+1} + x_{N-n+2} + ... + x_N\right) \end{bmatrix}$$

Probele care conțin x_1 se obțin prin selectarea a (n-1) alte elemente din populația disponibilă de (N-1) elemente şi, aceasta se poate face în C_{N-1}^{n-1} moduri. Vor fi deci C_{N-1}^{n-1} probe conținând x_1 şi la fel se aplică pentru fiecare x_i.

$$C_N^n = \frac{N!}{n!(N-n)!} = \frac{N}{n}\frac{(N-1)!}{(n-1)!(N-n)!} = \frac{N}{n}C_{N-1}^{n-1} \Rightarrow \frac{C_{N-1}^{n-1}}{C_N^n} = \frac{n}{N}$$

În consecință

$$E\left(\overline{X}\right) = \frac{1}{C_N^n}\left(\frac{1}{n}C_{N-1}^{n-1}\sum_1^N x_i\right) = \frac{1}{n} * \frac{C_{N-1}^{n-1}}{C_N^n}\sum_1^N x_i = \frac{1}{n} * \frac{n}{N}\sum_1^N x_i = \frac{\sum_1^N x_i}{N} = \mu$$

ceea ce aveam de demonstrat.

2. Similar se poate arata ca dispersia variabilei aleatoare este:

$$D\left(\overline{X}\right) = \frac{\sigma^2}{n}\frac{N-n}{N-1} = \frac{\sigma^2}{n}\left(1 - \frac{n-1}{N-1}\right)$$

Observatie:

Este de notat că dacă $N \to \infty$, atunci dispersia lui $\overline{X} \to \frac{\sigma^2}{n}$, forma ei obişnuită pentru o populație infinită, sau pentru experimentul de tip extracție din urnă cu întoarcerea bilelor extrase în urnă.

Testul Wilcoxon

Testul de rang Wilcoxon este un test cu ipoteza nulă că două populații sunt identice, fată de ipoteza alternativă că ele diferă printr-o translație lineară.

Testul înlocuieşte observațiile prin rangurile lor. Rangurile sunt repartizate la valorile din selecții în ordinea creşterii mărimii fără să țină cont de probele cărora le aparțin.

Să presupunem că o probă este de mărime n şi alta de mărime N-n. Testul presupune că orice combinație de ranguri în aceste două grupuri este egal probabilă. Numărul total de moduri de grupare a rangurilor este C_N^n.

Nu este uşor să calculăm toate posibilitățile, astfel încât vom folosi faptul că media rangurilor unei probe este distribuită aproximativ normal cu urmatorii parametri:

$$E(\overline{R}) = \frac{N+1}{2} \text{ si } D(\overline{R}) = \frac{(N+1)(N-n)}{12n}$$

Sunt disponibile tabelele care dau limitele de acceptare a ipotezei H_0 pentru suma obținută, ca o funcție de n, N şi riscul asumat.

Fie R suma rangurilor şi $\overline{R}$ media rangurilor probei de mărime n.

In cazul nostru in loc de x_i avem rangurile de N valori însemnând numerele 1,2,….,N, deci $\sum_1^N x_i = \sum_1^N i = \frac{N(N+1)}{2}$ si

$$\sum_1^N x_i^2 = \sum_1^N i^2 = \frac{N(N+1)(2N+1)}{6}$$

Valoarea medie a lui $\overline{R}$ va fi

$$E(\overline{R}) = \frac{1}{N}\sum_1^N x_i = \frac{1}{N} * \frac{N(N+1)}{2} \Rightarrow E(\overline{R}) = \frac{N+1}{2}$$

Dispersia lui $\overline{R}$ este $D(\overline{R}) = \frac{\sigma^2}{n}\left(1 - \frac{n-1}{N-1}\right)$ unde

$$\sigma^2 = D(X) = E(X^2) - (E(X))^2 = \frac{1}{N}\sum_1^N x_i^2 - \frac{1}{N^2}\left(\sum_1^N x_i\right)^2 =$$

$$= \frac{1}{N}\sum_1^N i^2 - \frac{1}{N^2}\left(\sum_1^N i\right)^2 = \frac{1}{N}\frac{N(N+1)(2N+1)}{6} - \frac{1}{N^2}\left(\frac{N(N+1)}{2}\right)^2 = \frac{N^2-1}{12}$$

Deci , $D(\overline{R}) = \dfrac{N^2-1}{12n}\dfrac{N-n}{N-1} = \dfrac{(N+1)(N-n)}{12n}$

În concluzie, variabila aleatoare $Z = \dfrac{\overline{R} - E(\overline{R})}{\sqrt{D(\overline{R})}} = \dfrac{\overline{R} - \frac{N+1}{2}}{\sqrt{\frac{(N+1)(N-n)}{12n}}}$ va fi repartizată aproximativ $N(0,1)$.

Notatii alternative:

a) Daca $N = n_1 + n_2$; $n = n_1$ si $N - n = n_2$, $(n_1 \le n_2)$ se obtine:

$z = \dfrac{\overline{R} - \frac{n_1+n_2+1}{2}}{\sqrt{\frac{(n_1+n_2+1)n_2}{12n_1}}}$ (testul Mann – Whitney)

b) Se amplifica cu n_1, se obtine $\overline{R}n_1 = R$ si $z = \dfrac{R - \frac{n_1(n_1+n_2+1)}{2}}{\sqrt{\frac{(n_1+n_2+1)n_1n_2}{12}}}$

c) Kruskal si Wallis au observat ca aproximația este îmbunătățită când valoarea α este mai mare de 0,02 prin aducerea lui $\overline{R}$ mai aproape de media lui cu $\dfrac{1}{2n}$.

$$z = \frac{\overline{R} - \frac{N+1}{2} + \frac{1}{2n}}{\sqrt{\frac{(N+1)(N-n)}{12n}}}$$

Ajustarea pentru valori egale în testul Wilcoxon

Dacă apar egalități, o alternativă pentru neglijarea lor este de a repartiza la aceste observații media rangurilor pe care le-ar fi primit dacă nu erau egale.

Să considerăm un grup de k egalități. Rangurile corespunzatoare m+1, m+2, …, m+k sunt înlocuite cu media lor.

$$\frac{(m+1)+(m+2)+...+(m+k)}{k}=\frac{km+\frac{k(k+1)}{2}}{k}=m+\frac{k+1}{2}$$

In acest caz $D(\overline{R})=\frac{N(N^2-1)-T}{12nN}*\frac{N-n}{n-1}$ unde $T=(k-1)k(k+1)$

În concluzie, variabila aleatoare $Z=\frac{\overline{R}-E(\overline{R})}{\sqrt{D(\overline{R})}}=\frac{\overline{R}-\frac{N+1}{2}+\frac{1}{2n}}{\sqrt{\frac{N(N^2-1)-T}{12nN}*\frac{N-n}{n-1}}}$

va fi repartizată aproximativ $N(0,1)$.

Testul Wilcoxon pereche

Wilcoxon a propus deasemenea un test pentru determinări pare în care rangurile sunt atribuite mărimii absolute a diferențelor şi apoi se dă rangurilor semnul diferențelor.

Ipoteza nulă este că distribuția diferențelor este simetrică față de zero, astfel orice rang este pozitiv sau negativ cu aceiaşi probabilitate. Valorile egale primesc ca rang media rangurilor grupului.

Numărul total de moduri de sume de ranguri ce se pot obține este 2^N.

Să ataşăm rangurilor i variabilele aleatoare d_i , unde

$$d_i=\begin{cases}1\ ,\ daca\ i\ este\ pozitiv\\ 0\ ,\ daca\ i\ este\ negativ\end{cases},$$

Se foloseste insa cea mai mica valoare dintre suma rangurilor pozitive si a celor negative.

Să considerăm suma rangurilor positive $s=\sum d_i i$.

În acest caz, variabila aleatoare $Z=\frac{s-E(s)}{\sqrt{D(s)}}=\frac{s-\frac{N(N+1)}{4}}{\sqrt{\frac{N(N+1)(2N+1)}{24}}}$ va fi repartizată aproximativ $N(0,1)$.

Testul H, Krusskal – Wallis, de analiza a variatiei "pe o cale" aplicata rangurilor

Testul H, sau testul Kruskal – Wallis este o generalizare a testului Wilcoxon în cazul a k probe, $k \rangle 2$. La fel ca şi în testul Wilcoxon, observațiile primesc ranguri, şi media rangurilor R_i se calculează pentru fiecare grup.

$$E\left(\overline{R_i}\right) = \frac{N+1}{2} \text{ şi } D^2\left(\overline{R_i}\right) = \frac{(N+1)(N-n_i)}{12n_i}$$

Raportul $\dfrac{\overline{R_i} - E\left(\overline{R_i}\right)}{\sqrt{D^2\left(\overline{R_i}\right)}}$ va fi repartizat $N(0,1)$, conform teoremei limita centrala.

Kruskal şi Wallis au arătat că suma pătratelor lor, cu un factor de ponderare $\left(1 - \dfrac{n_i}{N}\right)$ are aproximativ distribuția $\chi^2(k-1)$

$$H = \sum_{i=1}^{R} \left[\frac{\overline{R_i} - \frac{N+1}{2}}{\sqrt{\frac{(N+1)(N-n_i)}{12n_i}}} \right]^2 \left(1 - \frac{n_i}{N}\right) \cong \chi^2(k-1)$$

Dacă apar valori egale, H trebuie să fie împărțit la factorul $1 - \dfrac{\sum T}{N^3 - N}$ unde $T = (k-1)k(k+1) = k^3 - k$ este calculat pentru fiecare grup de legături.

Aplicatii:

1. Sa se verifice urmatoarea ipoteza $H_0 : \mu_1 = \mu_2$, $H_A : \mu_2 \neq \mu_1$ folosind datele

x_1	1,1	2,2	3,1	4,3
x_2	10	2,4	3,3	

si avand riscul $\alpha = 0.05$

Solutie:

Ordonam crescator valorile acordandu-le rangul corespunzator:

Rangurile	*1*	*2*	*4*	*6*
x_1	1,1	2,2	3,1	4,3
Rangurile	*7*	*3*	*5*	
x_2	10	2,4	3,3	

$$\overline{R_1} = \frac{1+2+4+6}{4} = \frac{13}{4} = 3.5 \qquad \overline{R_2} = \frac{7+3+5}{3} = \frac{15}{3} = 5$$

Aplicand testul Wilcoxon obtinem $z = \dfrac{5 - \dfrac{7+1}{2} + \dfrac{1}{2*3}}{\sqrt{\dfrac{(7+1)(7-3)}{12*3}}} \cong 1.1$

Deoarece $z_{0.95} = 1.64$ si $z = 1.1 \in (-1.64\,, 1.64)$ se accepta ipoteza.

2. Se dau datele urmatoare date:

x_1: 1.1; 2.2; 3.1; 4.3; 2.5 si x_2: 10; 2.4; 3.3; 2.5; 2.5

Sa se verifice ipoteza $H_0 : \mu_1 = \mu_2$, $H_A : \mu_2 \rangle\, \mu_1$ asumandu-ne riscul $\alpha = 0.05$.

Solutie:

	x_1					x_2				
Valoare	1.1	2.2	3.1	4.3	2.5	10	2.4	3.3	2.5	2.5
Rang	1	2	7	9	5	10	3	8	5	5
Media rangurilor	4.8					6.2				

$$\overline{R_1} = \frac{1+2+7+9+5}{5} = 4.8 \qquad \overline{R_2} = \frac{10+3+8+5+5}{5} = 6.2$$

$$z=\frac{\overline{R}-\frac{N+1}{2}+\frac{1}{2n}}{\sqrt{\frac{N\left(N^{2}-1\right)-T}{12n-N}*\frac{N-n}{n-1}}}$$ unde $T=(k-1)k(k+1)$

Deci, $T=4*5*6=120$ (rangurile care au valori egale).

$$z=\frac{6.2-\frac{10+1}{2}+\frac{1}{2*5}}{\sqrt{\frac{10\left(10^{2}-1\right)-120}{12*5-10}*\frac{10-5}{5-1}}}\cong 0.17$$

Deoarece $z_{0.95}=1.64$ si $z=0.17\in(-1.64\,,1.64)$ se accepta ipoteza

3. Sa se verifice urmatoarea ipoteza $H_0:\mu_1=\mu_2$,
$H_A:\mu_2\rangle\mu_1$ folosind datele:
x_1: 1.1; 2.2; 3.1; 4.3; 2.5 si x_2: 10; 2.4; 2.8; 2.5; 2.6
si considerand riscul $\alpha=0.05$
Solutie:
Vom face diferenta dintre valorile x_2 si x_1 si vom aloca rangurile corespunzatoare valorilor obtinute:

x_1	1.1	2.2	3.1	4.3	2.5
x_2	10	2.4	2.8	2.5	2.6
x_2-x_1	8.9	0.2	-0.3	-1.8	0.1
Rangul	5	2	-3	-4	1
Suma rangurilor pozitive: 8					

$$z=\frac{S-\frac{N(N+1)}{4}}{\sqrt{\frac{N(N+1)(2N+1)}{24}}}\qquad z=\frac{8-\frac{5*(5+1)}{4}}{\sqrt{\frac{5*(5+1)*(2*5+1)}{24}}}\cong 0.01$$

Deoarece $z_{0.95}=1.64$ si $z=0.01\in(-1.64\,,1.64)$ se accepta ipoteza

4. Se dau datele urmatoare:
x_1: 1.1; 2.2; 3.1; 4.3; 2.5
x_2: 10; 2.4; 2.8; 2.5; 2.6
x_3: -; 2.2; 3; 3.5; 2.7

Si dandu-se riscul $\alpha = 0.05$ sa se verifice ipoteza

$$H_0 : \mu_1 = \mu_2 = \mu_3 \,,\; H_A : \mu_1 \neq \mu_2 \neq \mu_3$$

Solutie:

	x_1					x_2					x_3				
x	1.1	2.2	3.1	4.3	2.5	10	2.4	2.8	2.5	2.6	-	2.2	3	3.5	2.7
R	1	2	10	13	5	14	4	9	6	7		3	11	12	8
$\overline{R}$	6.4					8					8.25				

$$\chi_2^2 = \left(\frac{6.4 - 7.5}{\sqrt{\frac{15*9}{12*5}}} \right)^2 * \left(1 - \frac{5}{14}\right) + \left(\frac{8 - 7.5}{\sqrt{\frac{15*9}{11*5}}} \right)^2 * \left(1 - \frac{5}{14}\right) +$$

$$+ \left(\frac{8.25 - 7.5}{\sqrt{\frac{15*10}{12*4}}} \right)^2 * \left(1 - \frac{4}{14}\right) \cong 3.2$$

Deoarece $\chi_{0.95}^2 = 5.99$ si $\chi_2^2 = 3.2 \langle\, 5.99$ se accepta ipoteza.

REGRESIA LINIARA

Dacă reprezentarea grafică a două mărimi ce sunt observate simultan sugerează o dependență liniară, ajungem la problema determinării dreptei ce descrie “cel mai bine” această dependență.

O soluție a acestei probleme o constituie “dreapta prin cele mai mici pătrate”, dreapta pentru care suma pătratelor distantelor de la ea la punctele experimentale este minimă.

Examinăm cazul cel mai simplu când valorile variabilei x (care în cele mai multe cazuri corespunde timpului) nu sunt afectate de erori şi, pentru fiecare valoare a lui x corespund un număr de valori y, determinate într-un singur experiment printr-o metodă afectată de erori întâmplătoare:

$y_{11}, y_{12}, ..., y_{1n_1}$, pentru x_1

.

.

$y_{i1}, y_{i2}, ..., y_{in_i}$, pentru x_i, i=1,2,…,k

Să admitem că pentru un x fixat, valoarea măsurată y este o variabilă aleatoare cu urmatoarea structură:

$$y = \eta + \varepsilon = x + \beta x + \varepsilon$$

distribuită normal cu dispersia σ^2 şi media $\eta = \alpha + \beta x$

Problema care ne-o punem este aceea ca, din datele experimentale y_i, să obținem nişte estimări a, b şi s^2 pentru α, β si σ^2, şi să determinăm distribuțiile acestor estimații.

Estimarea ecuației de regresie o notam : $Y = a + bx$

Metoda celor mai mici pătrate dă valorile a şi b care minimizează suma pătratelor deviațiilor (erorilor) între valorile observate y_i şi cele prezise de ecuația de regresie:

$$SS_E = \sum (y_i - Y_i)^2 = \sum (y_i - a - bx_i)^2$$

Metoda este în principal datorată lui Gauss.

Valorile lui a şi b care minimizează suma pătratelor erorilor sunt soluțiile sistemului

$$\frac{\partial SS}{\partial a} = 0 \text{ şi } \frac{\partial SS}{\partial b} = 0$$

$$\frac{\partial SS}{\partial a} = -2\sum (y_i - a - bx_i) = 0 \quad \text{şi} \quad \frac{\partial SS}{\partial b} = -2\sum (y_i - a - bx_i)x_i = 0$$

ceea ce este echivalent cu

$na + b\sum x_i = \sum y_i$ şi $a\sum x_i + b\sum x_i^2 = \sum x_i y_i$

Avem un sistemul Cramer cu doua ecuatii si doua necunoscute prin rezolvarea caruia se obțin a si b ca estimatori pentru α şi β:

$$\Delta = \begin{vmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{vmatrix} = n\sum x_i^2 - \left(\sum x_i\right)^2$$

$$\Delta_a = \begin{vmatrix} \sum y_i & \sum x_i \\ \sum x_i y_i & \sum x_i^2 \end{vmatrix} = \sum y_i \sum x_i^2 - \sum x_i \sum x_i y_i$$

$$\Delta_b = \begin{vmatrix} n & \sum y_i \\ \sum x_i & \sum x_i y_i \end{vmatrix} = n\sum x_i y_i - \sum x_i \sum y_i$$

Deci,

$$a = \frac{\sum y_i \sum x_i^2 - \sum x_i \sum x_i y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2} \text{ şi } b = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2}$$

Numărătorul expresiei lui b poate fi scris şi în forma

$$n\sum x_i y_i - \sum x_i \sum y_i = n\left(\sum x_i y_i - \sum \frac{x_i}{n} \sum y_i\right) = n\sum \left(x_i - \overline{X}\right) y_i$$

Deoarece $\sum \left(x_i - \overline{X}\right) = 0$ şi $\overline{Y}\sum \left(x_i - \overline{X}\right) = 0$, mai putem scrie

$$\sum \left(x_i - \overline{X}\right) y_i = \sum \left(x_i - \overline{X}\right) y_i - \overline{Y}\sum \left(x_i - \overline{X}\right) = \sum \left(x_i - \overline{X}\right)\left(y_i - \overline{Y}\right)$$

Pe de alta parte,

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

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

Deci, $b = \dfrac{n\sum \left(x_i - \overline{X}\right) y_i}{n\sum \left(x_i - \overline{X}\right)^2} = \dfrac{\sum \left(x_i - \overline{X}\right) y_i}{\sum \left(x_i - \overline{X}\right)^2}$ sau, scris intr-o formă alternativă, avem: $b = \dfrac{\sum \left(x_i - \overline{X}\right)\left(y_i - \overline{Y}\right)}{\sum \left(x_i - \overline{X}\right)^2}$.

Vom calcula in continuare media si dispersia pentru b si vom demonstra ca b este un estimator nedeplasat al lui β adica: $E(b) = \beta$

$$E(b)=E\left(\frac{\sum(x_i-\overline{X})y_i}{\sum(x_i-\overline{X})^2}\right)=\frac{\sum(x_i-\overline{X})E(y_i)}{\sum(x_i-\overline{X})^2}=\frac{\sum(x_i-\overline{X})(\alpha+\beta x_i)}{\sum(x_i-\overline{X})^2}=$$

$$=\alpha\frac{\sum(x_i-\overline{X})}{\sum(x_i-\overline{X})^2}+\beta\frac{\sum(x_i-\overline{X})x_i}{\sum(x_i-\overline{X})^2}=0+\beta\frac{\sum x_i^2-\frac{1}{n}\left(\sum x_i\right)^2}{\sum(x_i-\overline{X})^2}=\beta$$

Dispersiile lui a şi b pot fi obținute direct, deoarece sunt funcții liniare de y_i, care valori sunt presupuse independente şi distribuite normal, cu dispersia σ^2:

$$D(b)=D\left[\frac{\sum(x_i-\overline{X})y_i}{\sum(x_i-\overline{X})^2}\right]=\frac{\sum(x_i-\overline{X})^2D(y_i)}{\left(\sum(x_i-\overline{X})^2\right)^2}=\frac{\sigma^2}{\sum(x_i-\overline{X})^2}$$

Pentru a calcula dispersia lui a se tine cont ca: $a=\overline{Y}-b\overline{X}$ si se obtine

$$S_a=\sqrt{\frac{\sum x_i^2}{n}S_b^2}$$

Estimații şi ipoteze asupra coeficientului b

Coeficientul b este repartizat normal cu media β şi dispersia $\dfrac{\sigma^2}{\sum(x_i-\overline{X})^2}$

Dacă y_i sunt punctele experimentale, iar Y_i estimările lor teoretice, $Y_i=a+bx_i$, atunci suma pătratelor erorilor va fi $SS_E=\sum(y_i-Y_i)^2$.

Deci,

$$SS_E=\sum\left[y_i-(a+bx_i)\right]^2=$$

$$=\sum\left[(y_i-\overline{Y})+(\overline{Y}-a-bx_i)\right]^2=\sum\left[(y_i-Y)+(a+b\overline{X}-a-bx_i)\right]^2=$$

$$=\sum\left[(y_i-\overline{Y})-b(x_i-\overline{X})\right]^2=$$

$$=\sum(y_i-\overline{Y})^2-2b\sum(x_i-\overline{X})(y_i-\overline{Y})+b^2\sum(x_i-\overline{X})^2$$

Dar, conform formulei alternative, avem

$$b=\frac{\sum(x_i-\overline{X})(y_i-\overline{Y})}{\sum(x_i-\overline{X})^2}\Rightarrow\sum(x_i-\overline{X})(y_i-\overline{Y})=b\sum(x_i-\overline{X})^2.$$

Inlocuind in formula suma patratelor erorilor se obtine:

$$SS_E = \sum (y_i - \overline{Y})^2 - b^2 \sum (x_i - \overline{X})^2 = A - B \Rightarrow E(SS_E) = E(A) - E(B)$$

$$E(A) = E\left[\sum (y_i - \overline{Y})^2\right] = E\left(\sum y_i^2 - n\overline{Y}^2\right) = E\left(\sum y_i^2\right) - nE\left(\overline{Y}^2\right)$$

În continuare, folosind:

- “teorema lui Pitagora” $D(Y) = E(Y^2) + (E(Y))^2$
- $E(\overline{Y}) = \alpha + \beta\overline{X}$ şi
- $D(\overline{Y}) = D\left(\frac{\sum Y_i}{n}\right) = \frac{1}{n^2}\sum D(Y_i) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$

se poate obține:

$$E(A) = \sum\left[(\alpha + \beta x_i)^2 + \sigma^2\right] - n\left[(\alpha + \beta\overline{X})^2 + \frac{\sigma^2}{n}\right] =$$

$$= \sum\left(\alpha^2 + 2\alpha\beta x_i + \beta^2 x_i^2\right) + \sum \sigma^2 - n\left[\left(\alpha^2 + 2\alpha\beta\overline{X} + \beta^2\overline{X}^2\right) + \frac{\sigma^2}{n}\right] =$$

$$= n\alpha^2 + 2\alpha\beta\sum x_i + \beta^2\sum x_i^2 + n\sigma^2 - n\alpha^2 - 2n\alpha\beta\overline{X} - n\beta^2\overline{X}^2 - \frac{n\sigma^2}{n} =$$

$$= (n-1)\sigma^2 + 2\alpha\beta\sum(x_i - \overline{X}) + \beta^2\sum(x_i - \overline{X})^2 = (n-1)\sigma^2 + \beta^2\sum(x_i - \overline{X})^2$$

Deci, $E(A) = (n-1)\sigma^2 + \beta^2\sum(x_i - \overline{X})^2$

$$E(B) = E\left(b^2\sum(x_i - \overline{X})^2\right) = \sum(x_i - \overline{X})^2 E(b^2) = \sum(x_i - \overline{X})^2\left[D(b) + (E(b))^2\right] =$$

$$= \sum(x_i - \overline{X})^2\left(\frac{\sigma^2}{\sum(x_i - \overline{X})^2} + \beta^2\right)$$

Inlocuind in $E(SS_E)$ vom obtine:

$$E(SS_E) = (n-1)\sigma^2 + \beta^2\sum(x_i - \overline{X})^2 - \beta^2\sum(x_i - \overline{X})^2 - \sigma^2 = (n-2)\sigma^2$$

Am demonstrat astfel ca $E\left(\frac{SS_E}{n-2}\right) = \sigma^2$, deci variabila aleatoare $\frac{SS_E}{\sigma^2}$ este repatizată $\chi^2(n-2)$.

Vom putea estima intervalele de încredere pentru β şi verifica ipoteze asupra valorilor sale.

Vom studia cele doua cazuri: cand dispersiile sunt cunoscute si cand nu cunoastem dispersiile.

a) Cazul dispersiilor cunoscute

În cazul în care se cunoaşte dispersia erorilor de măsurare $D(\varepsilon_i) = D(y_i) = \sigma^2$ se foloseşte faptul că variabila aleatoare $z = \frac{b-\beta}{\sqrt{D(b)}} = \frac{b-\beta}{\sqrt{\left[\frac{\sigma^2}{\sum\left(x_i - \overline{X}\right)^2}\right]}}$ este repartizată $N(0,1)$.

b) Cazul dispersiilor necunoscute

În acest caz se înlocuieşte dispersia lui b: $\sigma_b = \frac{\sigma^2}{\sum\left(x_i - \overline{X}\right)^2}$ " dispersia de selecție": $S_b = \frac{\frac{SS_E}{n-2}}{\sum\left(x_i - \overline{X}\right)^2}$.

Variabila aleatoare $T = \frac{b-\beta}{\sqrt{\left[\frac{SS_E}{(n-2)\sum\left(x_i - \overline{X}\right)^2}\right]}}$ este repartizată Student cu n-2 grade de libertate.

Intervalul de încredere pentru β este $b - t_{n-2,1-\frac{\alpha}{2}} S_b \langle \beta \langle b + t_{n-2,1-\frac{\alpha}{2}} S_b$

Aplicatie:

1. Sa se calculeze dreapta prin cele mai mici patrate care aproximeaza punctele

x_i	1	2	3	4	5
y_i	2	5	5	8	11

si care trece prin origine si sa se determine intervalul de incredere 90% pentru panta dreptei.

Solutie:

REGRESIA LINIARA

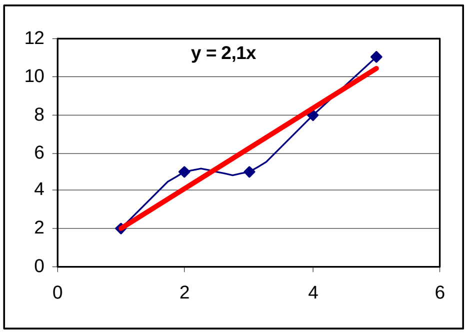
Dreapta care aproximeaza punctele este de forma $y = \beta * x$.

Din metoda celor mai mici patrate, panta dreptei care aproximeaza dependenta liniara $y(x)$ este $b = \frac{\sum (x_i - \overline{x}) y_i}{\sum (x_i - \overline{x})^2}$

In cazul nostru avem:

x_i	y_i	$\overline{x}$	$x_i - \overline{x}$	$(x_i - \overline{x}) y_i$	$(x_i - \overline{x})^2$	$\sum (x_i - \overline{x})^2$	$\sum (x_i - \overline{x}) y_i$
1	2	3	-2	-4	4	10	21
2	5		-1	-5	1		
3	5		0	0	0		
4	8		1	8	1		
5	11		2	22	4		

$b = \frac{21}{10} = 2.1$ deci, dreapta de regresie este $Y = 2{,}1 * x$



Intervalul de incredere 90% pentru β se calculeaza pornind de la distributia t a variabilei aleatoare.

$$T_{n-2} = \frac{b - \beta}{\sqrt{\frac{SS_E}{(n-2)\sum (x_i - \overline{x})^2}}}$$

$$SS_E = \sum [y_i - (a + b * x_i)]^2 = \sum (y_i - b * x_i)^2 =$$
$$= (2 - 2.1*1)^2 + (5 - 2.1*2)^2 + (5 - 2.1*3)^2 + (8 - 2.1*4)^2 + (11 - 2.1*5)^2 =$$
$$= 0.01 + 0.36 + 1.69 + 0.16 + 0.25 = 2.77$$

$$T_3 = \frac{2.1 - \beta}{\sqrt{\frac{2.77}{3*10}}} = \frac{2.1 - \beta}{\sqrt{0.09}} = \frac{2.1 - \beta}{0.3}$$

Deoarece $t_{3;0.05} = -2.35$, avem $P\left(-2.35 \langle \frac{2.1 - \beta}{0.3} \langle 2.35 \right) = 0.9$

$$P(-0.7 \langle 2.1 - \beta \langle 0.7) = 0.9$$
$$P(-2.8 \langle -\beta \langle -1.4) = 0.9$$
$$P(1.4 \langle \beta \langle 2.8) = 0.9$$

METODE STATISTICE DE ANALIZA FACTORILOR DE VARIABILITATE IN EXPERIMENTUL BIOLOGIC (ANOVA)

Se pune problema comparării mai multor selecții provenite din aceiasi populatie sau din populații pe care le ştim ca fiind normal repartizate, de exemplu concentrațiile plasmatice realizate de tablete care conțin diferiți excipienți, dar care au aceeaşi substanță activă, în aceeaşi doză.

Vrem să verificăm ipoteza compusă că acestea provin de fapt din aceiaşi populație, având media μ şi dispersia σ, deci că excipienții folosiți nu influențează semnificativ cedarea şi absorbția substanței active:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$

față de ipoteza alternativă că cel puțin două medii nu sunt egale.

O variantă de rezolvare a problemei ar fi compararea mediilor de selecție două câte două prin teste asupra mediei (de ex. testul t).

Fie, de exemplu, $\overline{X_1} = 5$, $\overline{X_2} = 6$, $\overline{X_3} = 7$ si $\overline{X_4} = 7$. La un nivel de incredere α putem conclude ca $\mu_1 = \mu_2$ şi $\mu_2 = \mu_3$ şi $\mu_3 = \mu_4$, dar este evident greşit a aplica o relație de tranzitivitate şi a spune că:

$$\mu_1 = \mu_2 \text{ şi } \mu_2 = \mu_3 \text{ şi } \mu_3 = \mu_4 \Rightarrow \mu_1 = \mu_2 = \mu_3 = \mu_4 .$$

Problema este ca in compararea acestor medii rezultatul depinde de dispersie, testul t analizand cazul dispersiilor egale.

Problema comparării mai multor selecții (loturi) poate fi rezolvata prin alte metode care permit compararea tuturor selecțiilor în acelasi timp. Cea mai populara metoda se bazează pe compararea dispersiilor de selecție şi se numeşte <u>analiză dispersională,</u> dar aproape toti cercetatorii o stiu sub forma prescurtarii denumirii sale in engleza ANOVA (de la Analysis of Variance).

Analiza dispersională este o metodă fundamentală a statisticii care, în plus față de mijloacele de calcul a "tendinței centrale" a rezultatelor experimentelor repetate, caracterizează mai ales variabilitatea acestora. Privita din unghiul de vedere al factorilor de variabilitate ANOVA se transforma in « analiza factoriala » care porneste de la aceeasi metode dar incearca sa rezolve alte probleme.

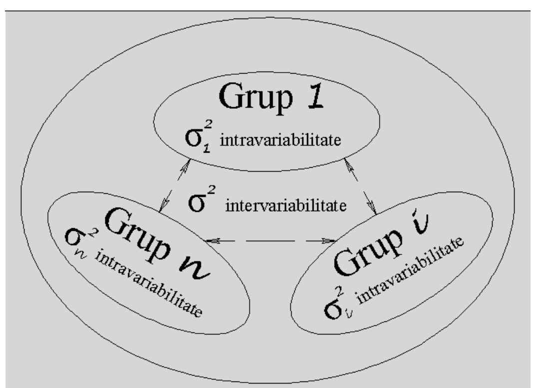
Variabilitatea se poate datora existenței unor factori cu influențe sistematice, a unor factori aleatori de fluctuație mai pronunțată sau a unor factori locali, inevitabili, determinând o fluctuație mai mică, definita "ca fluctuație experimentală". Analiza dispersională îşi propune separarea "variabilității totale" în: componentele datorate factorilor sistematici, in componentele datorate factorilor cu efecte aleatoare, ceea ce ramane reprezinta o variabilitate "reziduală" (diferența până la variabilitatea totală), care este de fapt variabilitatea experimentală. Din aceste variabilități se evaluează dispersiile parțiale corespunzătoare diferiților factori, calculându-se semnificația rapoartelor lor prin aplicarea testului F.

Datele experimentale se grupează în funcție de diferite criterii şi se urmăresc efectele asupra variabilitatii în funcție de aceste criterii, efecte care se compara in raport cu variablitatea reziduală.

Cea mai simplă analiză dispersională, numită analiză dispersională unidimensională sau unifactorială (numită în literatura engleză şi "one-way ANOVA") sau "experiment complet aleator", "experiment cu grupuri paralele", corespunde testului t de analiză a două eşantioane independente şi compară două sau mai multe grupuri.

In ipoteza că toate grupurile aparțin aceleiaşi populații, ideea testului este aceea că variabilitatea în interiorul grupurilor trebuie să fie de acelaşi ordin cu variabilitatea între mediile grupurilor.

În consecință, dispersia totală, evaluată ca suma a pătratelor diferențelor între valorile individuale şi media întregii populații selectate SS_T, este separată într-o parte datorită variației între grupuri (within), sau variabilității "interioare" şi o parte datorită variabilității "dintre" (between) grupuri: $SS_T = SS_W + SS_B$.



Daca comparam intra si intervariabilitatea cu testul Fischer – Snedecor si se obtine un rezultat pozitiv, atunci vom putea spune ca grupurile nu difera intre ele.

ANOVA

Fie modelul statistic : $x_{ij} = \mu + \alpha_i + \varepsilon_{ij}$ unde :

- μ este media populatiei
- $\mu + \alpha_i$ este media grupului respectiv
- ε_{ij} este eroarea de masurare

Dacă numărul de grupuri este k şi numărul de subiecți în grupul i este n_i atunci eroarea totala poate fi scrisa sub forma:

$$SS_T = \sum_{i}^{n} \sum_{j=1}^{n_i} \left(x_{ij} - \overline{X} \right)^2$$

Dupa cum s-a demonstrat la curs, aceasta relatie poate fi scrisa sub forma *identitatii analizei dispersionale:*

$$SS_T = \sum_{i}^{k} n_i \left(\overline{X_i} - \overline{X} \right)^2 + \sum_{i}^{k} \sum_{j}^{n_i} \left(x_{ij} - \overline{X_i} \right)^2 = SS_B + SS_W$$

si raportul $F = \dfrac{\dfrac{SS_B}{k-1}}{\dfrac{SS_W}{N-k}} = \dfrac{MS_B}{MS_W}$ este distribuit $F\left(k-1, N-k\right)$.

Dupa cum se observă, $\dfrac{SS_B}{k-1} = \dfrac{\sum_{i}^{k} n_i \left(\overline{X_i} - \overline{X} \right)^2}{k-1} = s_{\bar{x}}^2$ reprezintă dispersia de selecție ponderată a mediilor de grup față de marea medie.

Abaterile mediilor grupurilor față de media generală depind atât de hazardul măsuratorilor cât şi de factori ce țin de însăsi natura grupurilor. Abaterile în interiorul grupurilor sunt independente de aceşti factori, deoarece fiecare valoare măsurată este raportată la însăşi media grupului respectiv. Ele reprezintă fluctuații aleatoare.

Variabilitatea în interiorul grupurilor reprezintă diferența între variabilitatea totală şi variabilitatea între grupuri.

Aplicatie:

Consideram trei grupe de determinari conform tabelului de mai jos:

I	2	3	4
II	3	4	5
III	4	5	6

a caror medii au fost respectiv 3, 4 si 5:

Considerand un risc $\alpha = 0{,}05$ se poate spune ca diferentele obtinute sunt aleatoare si in fapt cele trei selectii sunt din aceeasi populatie omogena?

Solutie:

Avem de verificat urmatoarele ipoteze:

$$H_0 : \mu_1 = \mu_2 = \mu_3 \text{ vs. } H_A : \textit{mediile nu sunt toate egale}$$

Calculam $SS_W = \sum_i^k \sum_j^{n_i} \left(x_{ij} - \overline{X}_i\right)^2$, adica:

$$SS_W = (2-3)^2 + (3-3)^2 + (4-3)^2 + (3-4)^2 + (4-4)^2 + (5-4)^2 + \\ + (4-5)^2 + (5-5)^2 + (6-5)^2 = 6$$

Conform definitiei $SS_T = \sum_{i,j} \left(x_{ij} - \overline{X}\right)^2$ unde

$$\overline{X} = \frac{\sum x_{ij}}{N} = \frac{2+3+4+3+4+5+4+5+6}{9} = \frac{36}{9} = 4 \text{ deci,}$$

$$SS_T = (2-4)^2 + (3-4)^2 + (4-4)^2 + (3-4)^2 + (4-4)^2 + (5-4)^2 + \\ + (4-4)^2 + (5-4)^2 + (6-4)^2 = 12$$

O alta metoda de calcul pentru SS_T este folosirea urmatoarei formule prescurtate:

$$SS_T = \sum X^2 - \frac{\left(\sum X\right)^2}{N} = \left(2^2 + 3^2 + 4^2 + 3^2 + 4^2 + 5^2 + 4^2 + 5^2 + 6^2\right) - \\ - \frac{(1+2+3+3+4+5+4+5+6)^2}{9} = 156 - \frac{36^2}{9} = 156 - 144 = 12$$

Pentru determinarea lui SS_B putem tine cont de relatia $SS_T = SS_B + SS_W$, adica $SS_B = SS_T - SS_W = 12 - 6 = 6$ sau putem pleca de la definitia lui

$$SS_B = \sum_i^k n_i \left(\overline{X}_i - \overline{X}\right)^2 = 3(3-4)^2 + 3(4-4)^2 + 3(5-4)^2 = 3+0+3 = 6$$

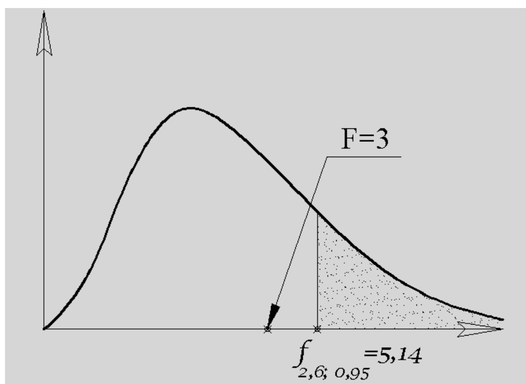
O alta metoda de calcul pentru SS_B foloseste formula echivalenta:

$$SS_B = \sum_i \frac{\left(\sum_j x_{ij}\right)^2}{n_i} - \frac{\left(\sum x_{ij}\right)^2}{N} =$$

$$= \frac{(2+3+4)^2}{3} + \frac{(3+4+5)^2}{3} + \frac{(4+5+6)^2}{3} - \frac{36^2}{9} = 150 - 144 = 6$$

Deci, $F = \dfrac{\dfrac{SS_B}{k-1}}{\dfrac{SS_W}{N-k}} = \dfrac{MS_B}{MS_W} = \dfrac{\dfrac{6}{3-1}}{\dfrac{6}{9-3}} = 3$ este distribuit Fischer cu 2, 6 grade de libertate

Conform tabelelor Fischer avem $f_{2,6;\,0,95} = 5{,}14$.



Cum $3 \langle 5{,}14$ se accepta ipoteza H_0 ceea ce inseamna ca cele trei selectii sunt din aceeasi populatie omogena.

TESTAREA EFECTELOR DE FORMULARE, SECVENTA SI PERIOADA IN EXPERIMENTUL INCRUCISAT CU 2 PERIOADE SI 2 SECVENTE

Pornind de la modelul statistic :

$$Y_{ijk} = \mu + S_{ik} + P_j + F_{(j,k)} + C_{(j-1,k)} + e_{ijk}$$

se formeaza combinatiile liniare ("contrastele"):

- $U_{ik} = Y_{i1k} + Y_{i2k}$, $i = \overline{1, n_k}$, $k = 1,2$ (R+T si respective T+R)
- $d_{ik} = \frac{1}{2}(Y_{i2k} - Y_{i1k})$, $i = \overline{1, n_k}$, $k = \overline{1,2}$.
- $O_{ik} = \begin{cases} d_{ik}, \textit{pentru.subiectii in secventa } 1\,(T-R) \\ -d_{ik}, \textit{pentru.subiectii in secventa } 2\,(R-T) \end{cases}$

prezenta sau absenta efectelor rezultand din aplicarea testului t acestor variabile aleatoare, conform tabelului de mai jos:

	Variabila test	$(1-\alpha)/100\%C.I.$	Test statistic
Carry - over	$C = \overline{U_{.2}} - \overline{U_{.1}} = \left(\overline{Y_{.11}} + \overline{Y_{.21}}\right) - \left(\overline{Y_{.12}} + \overline{Y_{.22}}\right)$	$C \pm t_{\left(\frac{\alpha}{2}, n_1+n_2-2\right)} \sigma_u \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$	$T_c = \frac{C}{\sigma_u \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$
Efect formulare	$F = \overline{d_{.1}} - \overline{d_{.2}} = \frac{1}{2}\left[\left(\overline{Y_{.21}} - \overline{Y_{.11}}\right) - \left(\overline{Y_{.22}} - \overline{Y_{.12}}\right)\right]$	$F \pm t_{\left(\frac{\alpha}{2}, n_1+n_2-2\right)} \sigma_d \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$	$T_d = \frac{F}{\sigma_d \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$
Perioada	$P = \overline{O_{.1}} - \overline{O_{.2}} = \frac{1}{2}\left[\left(\overline{Y_{.21}} - \overline{Y_{.11}}\right) + \left(\overline{Y_{.12}} - \overline{Y_{.22}}\right)\right]$	$P \pm t_{\left(\frac{\alpha}{2}, n_1+n_2-2\right)} \sigma_d \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$	$T_p = \frac{P}{\sigma_d \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

Aplicatie :

Sa se verifice ipoteze privind lipsa efectelor de perioada, secventa si formulare intr-un experiment incrucisat cu 2 perioade si 2 secvente, comparand doua medicamente testat (T) si referinta (R) si in care s-au obtinut rezultatele de mai jos :

	P_I	P_{II}
Secventa1 RT	2	3
	3	2
	2	3
Secventa 2 TR	4	2
	3	2
	3	3

Solutie :

	P_I	P_{II}	u_{ik}	$\overline{u_{\bullet 1}}$	d_{ik}	$\overline{d_{\bullet 1}}$	O_{ik}	$\overline{O_{\bullet 1}}$
RT	2	3	5	5	1/2	1/6	1/2	1/6
	3	2	5		-1/2		-1/2	
	2	3	5		1/2		1/2	
TR	4	2	6	5,6	-2/2	-1/2	2/2	1/2
	3	2	5		-1/2		1/2	
	3	3	6		0		0	

- Existența <u>efectelor carry – over</u> inegale poate fi determinată prin testarea următoarelor ipoteze:

$$H_0 : C = 0 \Leftrightarrow C_T = C_R$$

$$H_A : C \neq 0 \Leftrightarrow C_T \neq C_R$$

Respingerea ipotezei nule duce la concluzia prezenței efectelor carry – over inegale. Pentru testarea ipotezelor asupra lui C se folosesc următoarele medii de selecție corespunzând fiecărei secvențe:

$$\overline{U_{.k}} = \frac{1}{n_k}\sum_{i=1}^{n_k} U_{ik}\ ,\ k = 1,2 \text{ unde } U_{ik} = Y_{i_1 k} + Y_{i_2 k}$$

Deci, $U_{11} = 2 + 3 = 5$, $U_{21} = 3 + 2 = 5$, $U_{31} = 2 + 3 = 5$

$U_{42} = 4 + 2 = 6$, $U_{52} = 3 + 2 = 5$ si $U_{62} = 3 + 3 = 6$

ceea ce implica $\overline{U_{\bullet 1}} = \frac{5+5+5}{3} = 5$ si $\overline{U_{\bullet 2}} = \frac{6+5+6}{3} = 5,6$

In ipoteza H_0, variabila $T_c = \frac{\hat{C}}{\hat{\sigma}_u \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ are o repartiție Student cu $n_1 + n_2 - 2$ grade de libertate.

$$\hat{C} = \overline{U_{\bullet 2}} - \overline{U_{\bullet 1}} = 5,6 - 5 = 0,6$$

$$\hat{\sigma}_u = \sqrt{\frac{\sum (U_{i1} - U_{\bullet 1})^2 + \sum (U_{i2} - U_{\bullet 2})^2}{n_1 + n_2 - 2}} =$$

$$= \sqrt{\frac{\left[(5-5)^2 + (5-5)^2 + (5-5)^2\right] + \left[(6-5,6)^2 + (5-5,6)^2 + (6-5,6)^2\right]}{3+3-2}} =$$

$$= \sqrt{\frac{0,16 + 0,36 + 0,16}{4}} = \sqrt{\frac{0,68}{4}} \cong \sqrt{0,17} \cong 0,4$$

Obtinem astfel variabila $T_c = \frac{0,6}{0,4\sqrt{\frac{1}{3} + \frac{1}{3}}} \cong \frac{6}{4*0,7} \cong 2,14$

Vom respinge ipoteza nulă $H_0 : C_T = C_R$ în favoarea ipotezei alternative $H_A : C_T \neq C_R$ la un nivel $\alpha = 0,05$ de semnificație, dacă $|T_c| \rangle t_{\left(\frac{\alpha}{2}, n_1+n_2-2\right)}$.

Din tabele distributiei Student avem $t_{\left(\frac{\alpha}{2}, n_1+n_2-2\right)} = t_{0,975;\,4} = 2,78$

Cum $2,14 \langle 2,78$ acceptam ipoteza H_0, deci nu avem efecte reziduale.

- Pentru testarea <u>efectelor de formulare </u>se vor verifica urmatoarele ipoteze :

$$H_0 : F_R - F_T = 0 \Leftrightarrow F_R = F_T$$

$$H_A : F_R - F_T \neq 0 \Leftrightarrow F_R \neq F_T$$

Pentru testarea acestor ipoteze vom folosi variabila $T_d = \dfrac{F}{\hat{\sigma}_d \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ distribuita Student cu $n_1 + n_2 - 2$ grade de libertate, unde $F = \overline{d_{.1}} - \overline{d_{.2}}$ si $\hat{\sigma}_d^2 = \dfrac{\sum\left(d_{i1} - \overline{d}_{\bullet 1}\right)^2 + \sum\left(d_{i2} - d_{\bullet 2}\right)^2}{n_1 + n_2 - 2}$.

Dar, mediile diferențelor între perioade în interiorul fiecărei secvențe sunt $\overline{d_{.k}} = \dfrac{1}{n_k}\sum_{i=1}^{n_k} d_{ik}$, unde $d_{ik} = \dfrac{1}{2}\left(Y_{i2k} - Y_{i1k}\right)$.

In cazul nostru avem:

$d_{11} = \dfrac{3-2}{2} = \dfrac{1}{2}$, $d_{21} = \dfrac{2-3}{2} = -\dfrac{1}{2}$, $d_{31} = \dfrac{3-2}{2} = \dfrac{1}{2}$

$d_{42} = \dfrac{2-4}{2} = -\dfrac{2}{2} = -1$, $d_{52} = \dfrac{2-3}{2} = -\dfrac{1}{2}$, $d_{62} = \dfrac{3-3}{2} = 0$

ceea ce implica: $\overline{d}_{\bullet 1} = \dfrac{\dfrac{1}{2} - \dfrac{1}{2} + \dfrac{1}{2}}{3} = \dfrac{1}{6}$ si $\overline{d}_{\bullet 2} = \dfrac{-1 - \dfrac{1}{2} + 0}{3} = -\dfrac{1}{2}$

Facand inlocuirile obtinem:

$$\hat{\sigma}_d^2 = \frac{\sum\left(d_{i1} - \overline{d}_{\bullet 1}\right)^2 + \sum\left(d_{i2} - d_{\bullet 2}\right)^2}{n_1 + n_2 - 2} =$$

$$= \frac{\left[\left(\frac{1}{2} - \frac{1}{6}\right)^2 + \left(-\frac{1}{2} - \frac{1}{6}\right)^2 + \left(\frac{1}{2} - \frac{1}{6}\right)^2\right]}{3+3-2} +$$

$$+ \frac{\left[\left(-1 - \frac{1}{2}\right)^2 + \left(-\frac{1}{2} - \frac{1}{2}\right)^2 + \left(0 - \frac{1}{2}\right)^2\right]}{3+3-2} =$$

$$= \frac{\left(\frac{1}{9} + \frac{4}{9} + \frac{1}{9}\right) + \left(\frac{9}{4} + \frac{4}{4} + \frac{1}{4}\right)}{4} = \frac{\frac{6}{9} + \frac{14}{4}}{4} = \frac{1}{4} * \frac{25}{6} \cong 1{,}04$$

Rezulta ca $\hat{\sigma}_d \cong 1{,}02$.

Vom obtine astfel variabila aleatoare:

$$T_d = \frac{F}{\hat{\sigma}_d \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{\frac{1}{6} + \frac{1}{2}}{1{,}04 * \sqrt{\frac{1}{3} + \frac{1}{3}}} \cong 0{,}79$$

Aceasta variabila aleatoare este repartizata Student cu $n_1 + n_2 - 2$ grade de libertate.

Cum $2{,}14 \langle 2{,}78 = t_{2;\,0{,}975}$ acceptam ipoteza H_0, deci nu avem efecte formulare.

- Existența <u>efectelor de perioada</u> poate fi determinată prin testarea următoarelor ipoteze:

$$H_0 : P_I - P_{II} = 0 \Leftrightarrow P_I = P_{II}$$

$$H_A : P_I - P_{II} \neq 0 \Leftrightarrow P_I \neq P_{II}$$

Dupa cum am definit $O_{i1} = d_{i1}$, respectiv $O_{i2} = -d_{i2}$ si $\overline{O_{\bullet 1}} = \overline{d_{\bullet 1}}$, respectiv $\overline{O_{\bullet 2}} = -\overline{d_{\bullet 2}}$

Pentru testarea ipotezei H_0 vom folosi variabila $T_p = \frac{P}{\hat{\sigma}_d \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ repartizata Student cu $n_1 + n_2 - 2$ grade de libertate, unde

$P = \overline{O_{.1}} - \overline{O_{.2}}$ si $\hat{\sigma}_d^2 = \frac{\sum\left(d_{i1} - \overline{d}_{\bullet 1}\right)^2 + \sum\left(d_{i2} - d_{\bullet 2}\right)^2}{n_1 + n_2 - 2}$

Vom obtine astfel variabila $T = \frac{\frac{1}{6} - \frac{1}{2}}{1{,}02 * \sqrt{\frac{1}{3} + \frac{1}{3}}} = \frac{-\frac{1}{3}}{1{,}02 * 0{,}7} \cong -0{,}45$ si deoarece $t_{2;\,0{,}025} = -2{,}78 \langle -0{,}45 \langle 2{,}78 = t_{2;\,0{,}975}$ acceptam ipoteza H_0, deci nu avem efecte de perioada.

ANALIZA DISPERSIONALA CU INTERACTIUNI INTRE FACTORI

Analiza dispersionala a datelor de bioechivalenta urmeaza un model statistic ANOVA pe doua cai cu interactiuni de tipul

$$Y_{ijk} = \mu + \alpha_k + \beta_j + \gamma_{jk} + \varepsilon_{ijk}$$

unde $i = \overline{1, I}$, $j = \overline{1, J}$, $k = \overline{1, K}$

Acest modelul biostatistic se poate aplica unui experiment incrucisat, cu 2 perioade si 2 secvente.

Se consideră două medicamente, unul de testat (T) şi unul de referință (R), administrate unui lot de voluntari sănătoşi în două secvențe (RT) şi (TR) şi două perioade (I şi II). Fiecare subiect este asignat aleator fie secvenței 1 (RT), fie secvenței 2 (TR). Subiecții în cadrul secvenței RT (TR) primesc medicamentul R (T) în prima perioadă şi medicamentul T (R) în cea de a doua perioadă. Perioadele de administrare sunt, de obicei, separate printr-o perioadă de „spalare ” de cel puțin de trei ori timpul de înjumătățire al substanței active din medicamentul administrat.

Rezultatele experimentului sunt variabile aleatoare Y_{ijk} pe care le considerăm având următoarea structură:

$$Y_{ijk} = \mu + C_{j-1,k} + P_j + F_{jk} + \varepsilon_{ijk} + S_{ik}$$

unde:

- μ este media totala,
- i este indicele pentru subiect, $i = \overline{1, n_k}$,
- j este indicele pentru perioadă
- k este indicele pentru secvența.
- F_{jk} este efectul direct, fix, al medicamentului (formulării) administrat în perioada j, în secvența k (Observație: efectul este de fapt cantitatea de medicament măsurată sau un parametrul farmacocinetic calculat pornind de la aceasta).
- $C_{j-1,k}$ este efectul carry – over (fix) al medicamentului administrat în perioada j-1, (de exemplu concentratia medicamentului ramas in organism in perioada II din administrarea in perioada I). Considerăm că, datorită existenței unui interval de timp “de spălare” suficient intre administrari, efectul carry – over nu depăşeşte perioada consecutiva celei in care a fost administrat medicamentul.

LEGATURA ANOVA – TESTUL T

- Observam ca pe post de α apare efectul $C_{j-1,k}$ (carry – over sau “de secventa”), pe post de β_j apare perioada, si cand efectul de formulare apare efectul medicamentului.
- Considerăm ca variabilele S_{ik} (“efectul de subiect”) sunt repartizate identic pentru toate formularile administrate, cu media 0 şi dispersia σ_s^2,
- e_{ijk} reprezintă eroarea aleatoare în măsurarea valorilor individuale pentru fiecare subiect, variabilele e_{ijk} sunt repartizate $N\left(0,\sigma_e^2\right)$.

Modelul se mai poate scrie si sub forma :

$$Y_{ijk} = \mu_{jk} + S_{ik} + \varepsilon_{ijk}$$

unde efectul fix μ_{jk} este de forma prezentata mai jos:

Secvența	**Perioada** (j=1,2)	
(k=1,2)	I	II
1 (RT)	$\mu_{11} = \mu + \mathrm{P}_1 + \mathrm{F}_\mathrm{R}$	$\mu_{12} = \mu + \mathrm{P}_2 + \mathrm{F}_\mathrm{T} + \mathrm{C}_\mathrm{R}$
2 (TR)	$\mu_{21} = \mu + \mathrm{P}_1 + \mathrm{F}_\mathrm{T}$	$\mu_{22} = \mu + \mathrm{P}_2 + \mathrm{F}_\mathrm{R} + \mathrm{C}_\mathrm{T}$

unde

F_R (F_T) reprezintă efectul direct al administrării medicamentului R (T);

P_1 (P_2) reprezintă efectul administrării în perioada I (II);

C_R (C_T) reprezintă efectul rezidual („carry-over”) al administrării medicamentului R (T).

şi $P_1+P_2 = F_R+F_T = C_R+C_T = 0$.

Scopul experimentului este de a stabili bioechivalența dintre cele două medicamente („formulations” în literatura engleză).

Vom folosi urmatoarele notatii:

$$Y_{\bullet\bullet\bullet} = \sum_i \sum_j \sum_k Y_{ijk} \text{ si cu } \overline{Y_{\bullet\bullet\bullet}} = \frac{1}{IJK} \sum_i \sum_j \sum_k Y_{ijk}$$

$$Y_{\bullet jk} = \sum_i Y_{ijk} \text{ şi cu } \overline{Y}_{\bullet jk} = \tfrac{1}{I} \sum_i Y_{ijk}$$

$$Y_{\bullet j\bullet} = \sum_i \sum_k Y_{ijk} \text{ şi cu } \overline{Y}_{\bullet j\bullet} = \tfrac{1}{IK} \sum_i \sum_k Y_{ijk}$$

$$Y_{\bullet\bullet k} = \sum_i \sum_j Y_{ijk} \text{ şi cu } \overline{Y}_{\bullet\bullet k} = \tfrac{1}{IJ} \sum_i \sum_j Y_{ijk}$$

LEGATURA ANOVA – TESTUL T

Ipoteza de verificat este aceea a lipsei efectelor factorilor si, ca urmare si a interactiunii intre acestia.

$$H_0: \begin{cases} H_A: \quad \alpha_k = 0 \quad unde \quad 1 \le k \le K \\ H_B: \quad \beta_j = 0 \quad unde \quad 1 \le j \le J \\ H_{AB}: \quad \gamma_{jk} = 0 \quad unde \quad 1 \le j \le J \quad si \quad 1 \le k \le K \end{cases}$$

$$H_{alternativa}: \quad cel\ putin\ un\ factor\ este\ diferit\ de\ 0$$

Statistica F pentru verificarea ipotezelor H_A, H_B, H_{AB} are la numărător respectiv mediile sumelor de pătrate MS_α, MS_β, MS_γ, iar la numitor întotdeauna media sumei pătratelor rezidualelor MS_R.

Gradele de libertate sunt respectiv (ν_A, ν_R), (ν_B, ν_R), (ν_{AB}, ν_R) unde:

- $\nu_A = K - 1$,
- $\nu_B = J - 1$,
- $\nu_{AB} = (J-1)(K-1)$ si
- $\nu_R = JK(I-1)$

Evaluarea efectelor se face prin descompunerea erorilor in componente care sa estimeze contributia efectelor.

$$SS_T = \sum \left(Y_{ijk} - \overline{Y_{\bullet\bullet\bullet}}\right)^2 =$$

$$= \sum \left(\overline{Y_{\bullet\bullet k}} - \overline{Y_{\bullet\bullet\bullet}}\right)^2 + \sum \left(\overline{Y_{\bullet j\bullet}} - \overline{Y_{\bullet\bullet\bullet}}\right)^2 + \sum \left(\overline{Y_{\bullet jk}} - \overline{Y_{\bullet j\bullet}} - \overline{Y_{\bullet\bullet k}} + \overline{Y_{\bullet\bullet\bullet}}\right)^2 +$$

$$\sum \left(Y_{ijk} - \overline{Y_{\bullet jk}}\right)^2 = SS_\alpha + SS_\beta + SS_\gamma + SS_R$$

Termenii acestei diferențe corespund respectiv efectelor principale, interacțiunilor şi unei fluctuații aleatoare.

Deoarece indicele i se refera la subiectii experimentului $\left(Y_{ijk} - \overline{Y_{\bullet jk}}\right)$,diferenta intre valoarea subiectului i si media subiectilor din perioada j si secventa k, reprezinta o „intervariabilitate”. In biologie aceasta este foarte mare si ipoteza unor subiecti „identici” este departe de realitate sau se poate lua in calcul in cazul unor studii pe populatii foarte mari.

In cazul studiilor cross – over lui SS_α ii va corespunde $SS_{carry-over}$, lui SS_β ii va corespunde $SS_{perioada}$, iar lui SS_γ ii va corespunde $SS_{formulare}$.

SS_R , dupa cum s – a aratat in curs, se descompune intr-o componenta de variabilitate intraindividuala si o componenta interindividuala.

LEGATURA ANOVA – TESTUL T

Aplicatie:

1. Sa se testeze ipotezele

$$H_0 : \begin{cases} H_A : \quad \alpha_k = 0 \quad unde \quad 1 \le k \le K \\ H_B : \quad \beta_j = 0 \quad unde \quad 1 \le j \le J \\ H_{AB} : \quad \gamma_{jk} = 0 \quad unde \quad 1 \le j \le J \quad si \quad 1 \le k \le K \end{cases}$$

vs. $H_{alternativa}$: *cel putin un factor este diferit de* 0 cu riscul $\alpha = 0,05$ pentru datele din tabelul de mai jos:

	P_I	P_{II}
S_1	1	2
	2	3
	3	4
S_2	3	5
	4	6
	5	4

Solutie:

Conform notatiilor precedente vom calcula $\overline{Y}_{\bullet jk} = \frac{1}{I}\sum_i Y_{ijk}$, deci:

$$\overline{Y_{\bullet 11}} = \frac{1+2+3}{3} = 2 \qquad \overline{Y_{\bullet 21}} = \frac{2+3+4}{3} = 3$$

$$\overline{Y_{\bullet 12}} = \frac{3+4+5}{3} = 4 \qquad \overline{Y_{\bullet 22}} = \frac{5+6+4}{3} = 5$$

Avem $\overline{Y}_{\bullet\bullet k} = \frac{1}{IJ}\sum_i \sum_j Y_{ijk}$, sau altfel scris:

$$\overline{Y_{\bullet\bullet 1}} = \frac{\overline{Y_{\bullet 11}} + \overline{Y_{\bullet 21}}}{2} = \frac{2+3}{2} = 2,5 \text{ si } \overline{Y_{\bullet\bullet 2}} = \frac{\overline{Y_{\bullet 12}} + \overline{Y_{\bullet 22}}}{2} = \frac{4+5}{2} = 4,5$$

Avem $\overline{Y}_{\bullet j \bullet} = \frac{1}{IK}\sum_i \sum_k Y_{ijk}$, sau altfel scris:

$$\overline{Y_{\bullet 1\bullet}} = \frac{\overline{Y_{\bullet 11}} + \overline{Y_{\bullet 12}}}{2} = \frac{2+4}{2} = 3 \text{ si } \overline{Y_{\bullet 2\bullet}} = \frac{\overline{Y_{\bullet 21}} + \overline{Y_{\bullet 22}}}{2} = \frac{3+5}{2} = 4$$

Si $\overline{Y_{\bullet\bullet\bullet}} = \dfrac{1}{IJK}\sum_i \sum_j \sum_k Y_{ijk}$, sau altfel scris:

$$\overline{Y_{\bullet\bullet\bullet}} = \frac{\overline{Y_{\bullet 11}} + \overline{Y_{\bullet 12}} + \overline{Y_{\bullet 21}} + \overline{Y_{\bullet 22}}}{4} = \frac{2+4+3+5}{4} = 3,5$$

LEGATURA ANOVA – TESTUL T

Calculand eroarea totala conform definitiei obtinem:

$$SS_T = \sum_{i,j,k}\left(Y_{ijk} - \overline{Y_{\bullet\bullet\bullet}}\right)^2 = (1-3{,}5)^2 + (2-3{,}5)^2 + (3-3{,}5)^2 +$$

$$+ (3-3{,}5)^2 + (4-3{,}5)^2 + (5-3{,}5)^2 + (2-3{,}5)^2 + (3-3{,}5)^2 + (4-3{,}5)^2 +$$

$$+ (5-3{,}5)^2 + (6-3{,}5)^2 + (4-3{,}5)^2 = ...$$

ceea ce este o metoda foarte lunga de calcul.

Pentru a simplifica modul de determinare pentru SS_T vom folosi urmatoarea formula prescurtata:

$$SS_T = \sum_{i,j,k} Y_{ijk}^2 - \frac{\left(\sum_{i,j,k} Y_{ijk}\right)^2}{N}$$

unde $\sum_{i,j,k} Y_{ijk} = 1+2+3+2+3+4+3+4+5+5+6+4 = 42$;

$\sum_{i,j,k} Y_{ijk}^2 = 1^2 + 2^2 + 3^2 + 2^2 + 3^2 + 4^2 + 3^2 + 4^2 + 5^2 + 5^2 + 6^2 + 4^2 = 170$ si

$N = 12$

deci $SS_T = 170 - \frac{42^2}{12} = 23$

a. Pentru testarea <u>efectelor de secventa</u> vom calcula $SS_\alpha = \sum_{i,j,k}\left(\overline{Y_{\bullet\bullet k}} - \overline{Y_{\bullet\bullet\bullet}}\right)^2$.

Trebuie avut in vedere ca sumarea se face dupa toti indicii de sumare, respectiv dupa $j = 1{,}2$; $k = 1{,}2$ si $i = \overline{1, n_k}$, deci nu este corect sa se calculeze:

$SS_\alpha = \left(\overline{Y_{\bullet\bullet 1}} - \overline{Y_{\bullet\bullet\bullet}}\right)^2 + \left(\overline{Y_{\bullet\bullet 2}} - \overline{Y_{\bullet\bullet\bullet}}\right)^2 = (2{,}5-3{,}5)^2 + (4{,}5-3{,}5)^2 = 2 \Rightarrow$ fals

Calculul corect:

$$SS_\alpha = \sum_{i,j,k}\left(\overline{Y_{\bullet\bullet k}} - \overline{Y_{\bullet\bullet\bullet}}\right)^2 = \sum_{j=1}^{2}\sum_{i=1}^{n_1}\left(\overline{Y_{\bullet\bullet 1}} - \overline{Y_{\bullet\bullet\bullet}}\right)^2 + \sum_{j=1}^{2}\sum_{i=1}^{n_2}\left(\overline{Y_{\bullet\bullet 2}} - \overline{Y_{\bullet\bullet\bullet}}\right)^2 =$$

$$= \sum_{j=1}^{2}\sum_{i=1}^{n_1}(2{,}5-3{,}5)^2 + \sum_{j=1}^{2}\sum_{i=1}^{n_2}(4{,}5-3{,}5)^2 = 2n_1 * 1 + 2n_2 * 1 = 2(n_1 + n_2) =$$

$$= 2(3+3) = 12$$

Deci, $MS_\alpha = \frac{SS_\alpha}{K-1} = \frac{12}{2-1} = 12$

b. Pentru testarea <u>efectelor de perioada</u> vom calcula $SS_\beta = \sum_{i,j,k} \left(\overline{Y_{\bullet j \bullet}} - \overline{Y_{\bullet\bullet\bullet}}\right)^2$.

In mod analog,

$$SS_\beta = \sum_{i,j,k} \left(\overline{Y_{\bullet\bullet k}} - \overline{Y_{\bullet\bullet\bullet}}\right)^2 = \sum_{k=1}^{2}\sum_{i=1}^{n_k} \left(\overline{Y_{\bullet 1 \bullet}} - \overline{Y_{\bullet\bullet\bullet}}\right)^2 + \sum_{k=1}^{2}\sum_{i=1}^{n_k} \left(\overline{Y_{\bullet 2 \bullet}} - \overline{Y_{\bullet\bullet\bullet}}\right)^2 =$$

$$= \sum_{k=1}^{2}\sum_{i=1}^{n_k} (3-3{,}5)^2 + \sum_{k=1}^{2}\sum_{i=1}^{n_k} (3-3{,}5)^2 = 2(n_1+n_2)(-0{,}5)^2 = 2*6*0{,}25 = 3$$

Deci, $MS_\beta = \dfrac{SS_\beta}{J-1} = \dfrac{3}{2-1} = 3$

c. Pentru testarea <u>efectelor de interactiune</u> vom calcula

$$SS_\gamma = \sum_{i,j,k} \left(\overline{Y_{\bullet jk}} - \overline{Y_{\bullet\bullet k}} - \overline{Y_{\bullet j \bullet}} + \overline{Y_{\bullet\bullet\bullet}}\right)^2$$

j	k	$SS_\gamma = \sum_{i,j,k} \left(\overline{Y_{\bullet jk}} - \overline{Y_{\bullet\bullet k}} - \overline{Y_{\bullet j \bullet}} + \overline{Y_{\bullet\bullet\bullet}}\right)^2$	$SS_\gamma = 0$
1	1	$= \left(\overline{Y_{\bullet 11}} - \overline{Y_{\bullet\bullet 1}} - \overline{Y_{\bullet 1 \bullet}} + \overline{Y_{\bullet\bullet\bullet}}\right)^2 = (2-2{,}5-3+3{,}5)^2 = 0$	
2	1	$= \left(\overline{Y_{\bullet 21}} - \overline{Y_{\bullet\bullet 1}} - \overline{Y_{\bullet 2 \bullet}} + \overline{Y_{\bullet\bullet\bullet}}\right)^2 = (3-2{,}5-4+3{,}5)^2 = 0$	
1	2	$= \left(\overline{Y_{\bullet 12}} - \overline{Y_{\bullet\bullet 2}} - \overline{Y_{\bullet 1 \bullet}} + \overline{Y_{\bullet\bullet\bullet}}\right)^2 = (4-4{,}5-3+3{,}5)^2 = 0$	
2	2	$= \left(\overline{Y_{\bullet 22}} - \overline{Y_{\bullet\bullet 2}} - \overline{Y_{\bullet 2 \bullet}} + \overline{Y_{\bullet\bullet\bullet}}\right)^2 = (5-4{,}5-4+3{,}5)^2 = 0$	

Deci, $MS_\gamma = \dfrac{SS_\gamma}{(J-1)(K-1)} = 0$

d. Pentru determinarea lui SS_R nu vom pleca de la definitia sa $SS_R = \sum \left(Y_{ijk} - \overline{Y_{\bullet jk}}\right)^2$ deoarece am avea prea multe sume de calculat (9 sume) ci vom tine cont de uratoarea relatie:

$$SS_R = SS_T - SS_\alpha - SS_\beta - SS_\gamma$$

Inlocuind in formula precedenta obtinem: $SS_R = 23 - 12 - 3 - 0 = 8$ ceea ce implica $MS_R = \dfrac{SS_R}{JK(I-1)} = \dfrac{8}{2*2*(3-1)} = \dfrac{8}{8} = 1$

Vom testa ipoteza

LEGATURA ANOVA – TESTUL T

$$H_0: \begin{cases} H_A: \quad \alpha_k = 0 \quad unde \quad 1 \le k \le K \\ H_B: \quad \beta_j = 0 \quad unde \quad 1 \le j \le J \\ H_{AB}: \quad \gamma_{jk} = 0 \quad unde \quad 1 \le j \le J \quad si \quad 1 \le k \le K \end{cases}$$

vs $H_{alternativa}$: *cel putin un factor este diferit de* 0

aplicand testul Fischer astfel:

- Pentru efectul de secventa testul Fischer este:

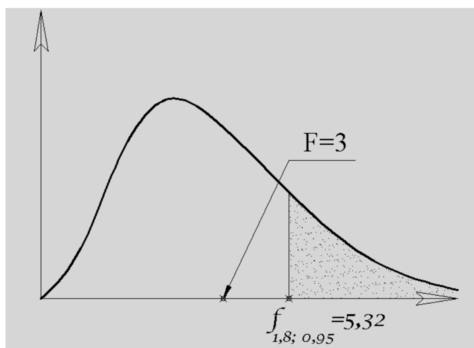
$$\frac{MS_\alpha}{MS_R} = F \Rightarrow F = \frac{12}{1} = 12\,; \quad f_{1,8;\,0,95} = 5{,}32$$

Deoarece $5{,}32 \langle 12$ se respinge ipoteza $H_0 : \alpha_k = 0$, $\forall k = 1{,}2 \Rightarrow$ avem efect de secventa

- Pentru efectul de perioada testul Fischer este:

$$\frac{MS_\beta}{MS_R} = F \Rightarrow F = \frac{3}{1} = 3\,; \quad f_{1,8;\,0,95} = 5{,}32$$

Deoarece $3 \langle 5{,}32$ nu se respinge ipoteza $H_0 : \beta_j = 0$, $\forall j = 1{,}2 \Rightarrow$ nu avem efect de perioada

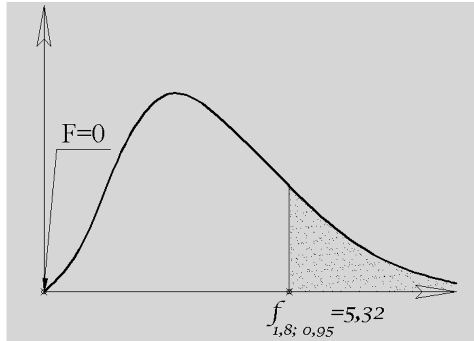


- Pentru interactiuni:

$$\frac{MS_\gamma}{MS_R} = F \Rightarrow F = \frac{0}{1} = 0\,; \quad f_{1,8;\,0,95} = 5{,}32$$

Deoarece $0 \langle 5{,}32$ nu se respinge ipoteza $H_0 : \gamma_{jk} = 0$, $\forall j = 1{,}2; \forall k = 1{,}2$

$\Rightarrow$ nu avem efect datorat interactiunilor



Modelul mult mai adevarat este cel care foloseste $MS_{\text{int}\,ra}$ in loc de MS_R, fiecare subiect fiind in acest fel propriul sau martor.

$$MS_{\text{int}\,ra} = \frac{SS_{\text{int}\,ra}}{\nu_{\text{int}\,ra}}$$

unde :

- $SS_{\text{int}\,ra} = \sum_{i,j,k} Y_{ijk}^2 - \sum \frac{Y_{\bullet jk}^2}{n_k} - \sum \frac{Y_{i\bullet k}^2}{2} + \sum \frac{Y_{\bullet\bullet k}^2}{2n_k}$
- $\nu_{\text{int}\,ra} = n_1 + n_2 - 2$

In cazul nostru avem :

$$\sum_{i,j,k} Y_{ijk}^2 = 1^2 + 2^2 + 3^2 + 2^2 + 3^2 + 4^2 + 3^2 + 4^2 + 5^2 + 5^2 + 6^2 + 4^2 = 170$$

$$n_1 = n_2 = 3$$

j	k	$\sum \frac{Y_{\bullet jk}^2}{n_k}$	$\sum \frac{Y_{\bullet jk}^2}{n_k} = 12 + 27 + 48 + 75 = 162$
1	1	$= \frac{1}{3}(1+2+3)^2 = \frac{36}{3} = 12$	
2	1	$= \frac{1}{3}(2+3+4)^2 = \frac{81}{3} = 27$	
1	2	$= \frac{1}{3}(3+4+5)^2 = \frac{144}{3} = 48$	
2	2	$= \frac{1}{3}(5+6+4)^2 = \frac{225}{3} = 75$	

LEGATURA ANOVA – TESTUL T

$$\sum \frac{Y_{i\bullet k}^2}{2} = \frac{(1+2)^2}{2} + \frac{(2+3)^2}{2} + \frac{(3+4)^2}{2} + \frac{(3+5)^2}{2} + \frac{(4+6)^2}{2} + \frac{(5+4)^2}{2} =$$

$$= \frac{9}{2} + \frac{25}{2} + \frac{49}{2} + \frac{64}{2} + \frac{100}{2} + \frac{81}{2} = \frac{328}{2} = 164$$

$$\sum \frac{Y_{\bullet\bullet k}^2}{2n_k} = \sum \frac{Y_{\bullet\bullet 1}^2}{2n_1} + \sum \frac{Y_{\bullet\bullet 2}^2}{2n_2} = \frac{(1+2+3+2+3+4)^2}{2*3} + \frac{(3+4+5+5+6+4)^2}{2*3} =$$

$$= \frac{225}{6} + \frac{729}{6} = \frac{954}{6} = 159$$

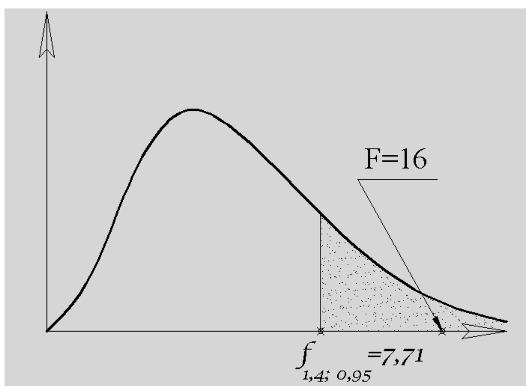
Deci, $SS_{\text{int}\,ra} = 170 - 162 - 164 + 159 = 3 \Rightarrow MS_{\text{int}\,ra} = \frac{3}{3+3-2} = \frac{3}{4} = 0{,}75$

In aceasta situatie vom obtine urmatoarele teste Fischer :

- Pentru efectul de secventa testul Fischer este:

$$\frac{MS_\alpha}{MS_{\text{int}\,ra}} = F \Rightarrow F = \frac{12}{0{,}75} = 16; \; f_{1,4;0,95} = 7{,}71$$

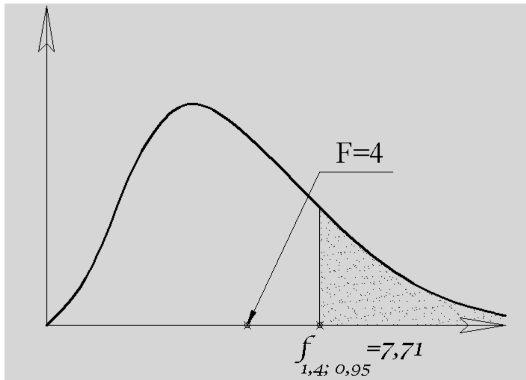
Deoarece $7{,}71 \langle 16$ se respinge ipoteza $H_0 : \alpha_k = 0$, $\forall k = 1{,}2 \Rightarrow$ avem efect de secventa



- Pentru efectul de perioada testul Fischer este:

$$\frac{MS_\beta}{MS_{\text{int}\,ra}} = F \Rightarrow F = \frac{3}{0{,}75} = 4; \; f_{1,4;0,95} = 7{,}71$$

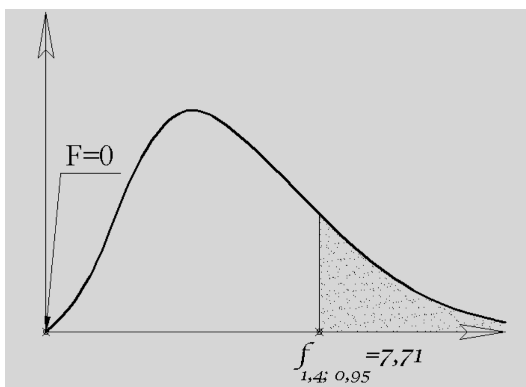
Deoarece $4 \langle 7{,}71$ nu se respinge ipoteza $H_0 : \beta_j = 0$, $\forall j = 1{,}2 \Rightarrow$ nu avem efect de perioada



- Pentru interactiuni:

$$\frac{MS_\gamma}{MS_{\text{int}\,ra}} = F \Rightarrow F = \frac{0}{0{,}75} = 0; \quad f_{1,4;0,95} = 7{,}71$$

Deoarece $0 \langle 7{,}71$ nu se respinge ipoteza $H_0 : \gamma_{jk} = 0$, $\forall j = 1{,}2; \forall k = 1{,}2$ $\Rightarrow$ nu avem efect datorat interactiunilor



ANALIZA DISPERSIONALA IN TESTAREA CORELATIEI SI REGRESIEI LINIARE

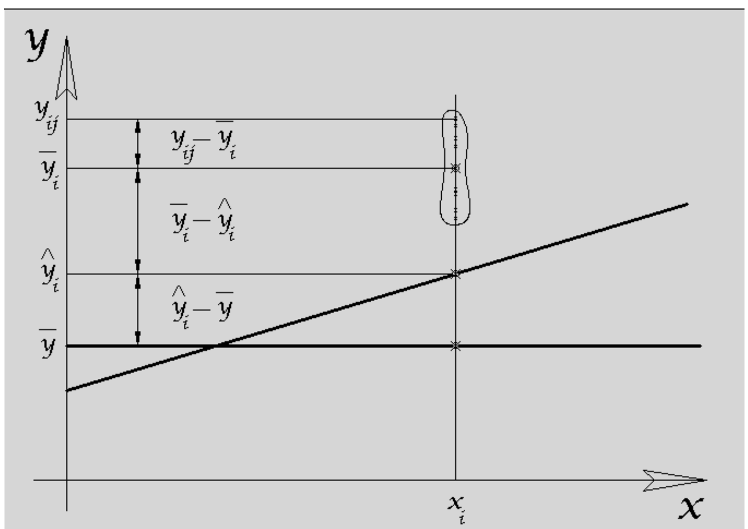
Consideram o variabila aleatoare y care depinde liniar de variabila aleatoare x:

$$y = \alpha + \beta x$$

Atunci cand facem determinarile experimentale noi nu stim nici daca cele doua variabile se coreleaza liniar si nici care este dreapta care descrie dependenta lor. Putem insa, prin analiza datelor experimentale sa determinam, prin metoda celor mai mici patrate, o estimare a dreptei

$$\hat{y} = a + bx$$

daca vom considera un set de determinari $(y_{ij})_{j=1,N_j}$ corespunzatoare pentru un x_i dat :



Distanta de la un punct dat y_{ij} la $\overline{y}$ se poate descompune in trei componente: distanta pana la $\overline{y_i}$ - media punctelor y_{ij}, distanta de la media grupului la valoarea estimata prin dreapta $\hat{y}_i$ si distanta de la punctele de pe dreapta la media totala $\overline{y}$:

$$y_{ij} - \overline{y} = \left(y_{ij} - \overline{y_i}\right) + \left(\overline{y_i} - \hat{y}_i\right) + \left(\hat{y}_i - \overline{y}\right)$$

Ridicand la patrat, sumand si tinand cont ca sumele de produse mixte sunt zero, se obtine :

$$\sum \left(y_{ij} - \overline{y}\right)^2 = \sum \left(y_{ij} - \overline{y_i}\right)^2 + \sum N_i \left(\overline{y_i} - \hat{y}_i\right)^2 + \sum N_i \left(\hat{y}_i - \overline{y}\right)^2$$

sau

$$SS_T = SS_{eroare} + SS_{deviatie\ de\ la\ linearitate} + SS_{linearitate}$$

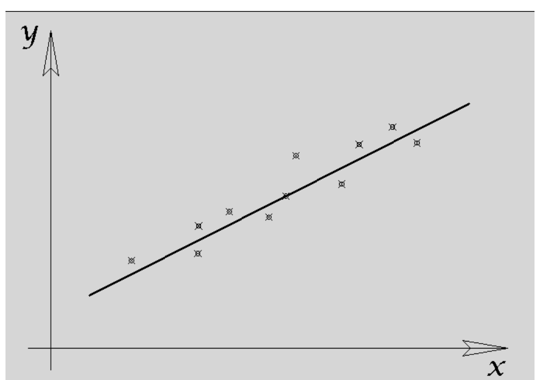
Observam ca, daca toate punctele ar fi pe o dreapta $SS_{deviatie\ de\ la\ linearitate}$ va fi zero, deci aceasta suma este o masura a corelarii liniare.
Intr-adevar :

$$\hat{y} - \overline{y} = a + bx - a - bx = b\left(x - \overline{x}\right) = r\frac{S_y}{S_x}\left(x - \overline{x}\right)$$

Facem observatia ca datele pot fi aproximate foarte bine dupa o alta lege (de exemplu $y = k\sqrt{x}$ cum este in cazul in care se aplica la dizolvare legea lui Higuchi).

Observam din definitiile coeficientului de corelatie si a raportului de corelare ca :

$$r^2 = \frac{SS_{linear}}{SS_{total}} \text{ si } \eta^2 = \frac{s_Y^2 - s_{Y/X}^2}{s_Y^2} = \frac{SS_{linear} + SS_{deviatie\ de\ la\ linearitate}}{SS_{total}}$$



Deci daca SS_{linear} are o valoare pozitiva avem un semn de necorelare liniara, dar nu obligatoriu de non – corelare.

ANALIZA DISPERSIONALA BIFACTORIALA CU INTERACTIUNI INTRE FACTORI

Legatura intre panta dreptei de regresie si coeficientul de corelatie

Avem dupa definitie

$$r = \frac{1}{N}\sum\left(\frac{x_i - \overline{x}}{S_x}\right)\left(\frac{y_i - \overline{y}}{S_y}\right)$$

In cazul in care punctele y_i sunt toate pe o dreapta $y_i = a + bx_i$

$$r = \frac{1}{N}\sum\left(\frac{x_i - \overline{x}}{S_x}\right)\left(\frac{a + bx_i - a - b\overline{x}}{S_y}\right) = \frac{1}{N}\frac{b\sum\left(x_i - \overline{x}\right)^2}{S_x S_y}$$

dar, $S_y^2 = \dfrac{\sum\left(a + bx_i - a - b\overline{x}\right)^2}{N} = \dfrac{b^2\sum\left(x_i - \overline{x}\right)^2}{N} = b^2 S_x^2$

Deci, inlocuind mai sus

$$r = \frac{1}{N}\sum\left(\frac{x_i - \overline{x}}{S_x}\right)\left(\frac{a + bx_i - a - b\overline{x}}{S_y}\right) = \frac{1}{N}\frac{b\sum\left(x_i - \overline{x}\right)^2}{S_x S_y}$$

$$r = \frac{1}{N}\frac{b\sum\left(x_i - \overline{x}\right)^2}{S_x b S_x} = \frac{S_x^2}{S_x^2} = 1$$

Cand punctele nu sunt pe dreapta, panta dreptei prin cele mai mici patrate b este:

$$b = \frac{\sum\left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)}{\sum\left(x_i - \overline{x}\right)^2} = \frac{\sum\left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)}{S_x^2} = \frac{\sum\left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)}{S_x S_y}\frac{S_y}{S_x} = r\frac{S_y}{S_x}$$

Deci, $b = r\dfrac{S_y}{S_x}$

Testarea linearitatii :

Observam ca SS_{eroare} are $N - I$ grade de libertate si deci $MS_{eroare} = \frac{SS_{eroare}}{N - I}$ avem ca $E(MS_{eroare}) = \sigma_e^2$

Media sumei abaterilor de la medie datorita linearitatii este : $E(MS_{linear}) = \sigma_e^2 + N\rho^2\sigma_y^2$

In cele ce urmeaza vom calcula media sumei MS_{linear} **;**

$$E(MS_{linear}) = E\left(\sum(\hat{y}_i - \overline{y})^2\right) = E\left(\sum(a + bx_i - a - b\overline{x})^2\right) = \sum(x_i - \overline{x})^2 E(b^2)$$

Dar, $E(b^2) = D(b) + [E(b)]^2 = \frac{\sigma_y^2}{\sum(x_i - \overline{x})^2} + [E(b)]^2$

Folosind relatia $b = r\frac{S_y}{S_x} \Rightarrow E(b) = \rho\frac{\sigma_x}{\sigma_y}$ si

$$E(MS_{linear}) = \sum(x_i - \overline{x})\left(\frac{\sigma_y^2}{\sum(x_i - \overline{x})^2} + \rho^2\frac{\sigma_y^2}{\sigma_x^2}\right) =$$

$$= \sigma_y^2 + \frac{\sum(x_i - \overline{x})^2\rho^2\sigma_y^2}{\sigma_x^2} = \sigma_y^2 + N\rho^2\sigma_y^2$$

In fapt aici am presupus ca pentru fiecare punct x_i valorile corespunzatoare y_{ij} au o dispersie $\sigma_{y/x}^2$ care este aceeasi pentru toate punctele x_i si deci putem sa o notam cu σ_y^2 sau σ_e^2 .

Lucrurile nu se intampla intotdeauna in acest fel. De exemplu in cazul dreptei de etalonare in bioanalitica dispersiile sunt practic semnificativ mai mari la limita de cuantificare (pana la 20%) – fata de restul concetratiilor la care limita admisa pentru « precizie » este de 15%.

Ipotezele de verificat sunt :

$H_0 : \rho = 0$ echivalenta cu $H_0 : \beta = 0$ folosind variabila aleatoare $F_{1,N-I} = \frac{MS_{linear}}{MS_{eroare}}$.

ANALIZA DISPERSIONALA BIFACTORIALA CU INTERACTIUNI INTRE FACTORI

Testarea ipotezei de nonlinearitate : $H_0 : \eta^2 - \rho^2 = 0$

Pentru aceasta se compara valorile testului $F_{I-2,\, N-I} = \dfrac{MS_{deviatie\ de\ la\ linearitate}}{MS_{eroare}}$ cu valorile din distributia Fischer.

Pentru a intelege mai clar principiul testului reamintim definitiile lui η si ρ :

- Raportul de corelare η^2 este proportia de variabilitate a lui Y atribuabila covariantei cu X ;
- Coeficientul de determinare (corelatie) este proportia de variabilitate a lui Y atribuabila covariantei liniare cu X .

ANALIZA DISPERSIONALA BIFACTORIALA CU INTERACTIUNI INTRE FACTORI

Aplicatii :

1. Consideram urmatoarele puncte de pe dreapta $y = 2x + 1$

x_i	y_i
0	1
1	3
2	5
3	7

Sa se calculeze r si b .

Indicatie:

x_i	y_i	$\overline{x}$	$\overline{y}$	$x_i - \overline{x}$	$y_i - \overline{y}$	$(x_i - \overline{x})(y_i - \overline{y})$	$(x_i - \overline{x})^2$
0	1	1,5	4	-1,5	-3	4,5	2,25
1	3			-0,5	-1	0,5	0,25
2	5			0,5	1	0,5	0,25
3	7			1,5	3	4,5	2,25

2. Consideram in continuare ca punctele y_i sunt afectate de erori :

x_i	y_i
0	1
1	2
2	5
3	6

Sa se calculeze r si b

3. Sa se testeze daca datele de la exercitiul anterior sunt corelate dar nu sunt corelate linear.

XII. TESTAREA BIOECHIVALENTEI A DOUA PRODUSE

Una din aplicatiile cele mai importante ale metodei de studiu clinic incrucisat priveste studiile de bioechivalenta care urmaresc testarea ipotezei privind bioechivalenta unui medicament generic cu medicamentul inovator.

Definitia cantitativa a bioechivalentei, aprobata prin lege federala in SUA si adoptata de toate autoritatile de reglemetare a medicamentului in lume, este dupa cum urmeaza:

Doua medicamente sunt bioechivalente daca intervaul de incredere 90% pentru raportul mediilor parametrilor farmacocinetici esentiali (aria de sub curba $AUC_{0-\infty}$ si concentratia maxima $C_{\max}$) se incadreaza in intervalul $(0{,}8\,,\,1{,}25)$.

Este de retinut faptul ca se recomanda de regula, analiza datelor dupa logaritmare considerandu-se ca acestea urmeaza mai curand o distributie lognormala decat o distributie normala.

Aplicatie :

1. Consideram ca intr-un experiment de bioechivalenta au terminat experimentul 7 voluntari (4 in secventa RT si 3 in secventa TR), iar concentratiile maxime (mg/l) au fost dupa cum urmeaza:

	P_1	P_2
S_1	3	4
S_2	4	6
S_3	5	5
S_4	4	5
S_5	4	6
S_6	5	5
S_7	6	5

Sa se verifice ipoteza privind bioechivalenta celor doua medicamente prin compararea intervalului de incredere 90% pentru raportul mediilor $\frac{\mu_T}{\mu_R}$ cu limitele de acceptare.

Solutie:

			d_{ik}	$\overline{d_{\bullet k}}$	$\overline{d_{\bullet 1}} - \overline{d_{\bullet 2}}$	$d_{ik} - \overline{d_{\bullet k}}$	$\left(d_{ik} - \overline{d_{\bullet k}}\right)^2$	$\sum\left(d_{ik} - \overline{d_{\bullet k}}\right)^2$	S_d^2	S_d
secv. 1	P_1	P_2								
RT	3	4	0.5	0.5	0.33	0	0	0.5	0.33	0.6
	4	6	1			0.5	0.25			
	5	5	0			-0.5	0.25			
	4	5	0.5			0	0			
secv. 2	P_1	P_2								
	4	6	1			0.83	0.69			
TR	5	5	0	0.17		-0.17	0.03	1.17		
	6	5	-0.5			-0.67	0.45			

$n_1 = 4$; $n_2 = 3$; $t_{5;0.90} = 1.48$

$$d_{ik} = \frac{1}{2} * \left(y_{i2k} - y_{i1k}\right)$$

Deci

$d_{11} = \frac{1}{2} * (4-3) = \frac{1}{2} = 0.5$; $d_{21} = \frac{1}{2} * (6-4) = \frac{2}{2} = 1$;

$d_{31} = \frac{1}{2} * (5-5) = \frac{0}{2} = 0$; $d_{41} = \frac{1}{2} * (5-4) = \frac{1}{2} = 0.5$

$d_{12} = \frac{1}{2} * (6-4) = \frac{2}{2} = 1$; $d_{22} = \frac{1}{2} * (5-5) = \frac{0}{2} = 0$;

$d_{32} = \frac{1}{2} * (5-6) = -\frac{1}{2} = -0.5$

$$\overline{d_{\bullet k}} = \frac{\sum_i d_{ik}}{n_k}$$

$\overline{d_{\bullet 1}} = \frac{d_{11} + d_{21} + d_{31} + d_{41}}{n_1}$; $\overline{d_{\bullet 1}} = \frac{0.5 + 1 + 0 + 0.5}{4} = 0.5$

$\overline{d_{\bullet 2}} = \frac{d_{12} + d_{22} + d_{32} + d_{42}}{n_2}$; $\overline{d_{\bullet 2}} = \frac{1 + 0 - 0.5}{3} = 0.17$

$\overline{d_{\bullet 1}} - \overline{d_{\bullet 2}} = 0.5 - 0.17 = 0.33$

$S_d^2 = \frac{\sum_{i=1}^{n_1}\left(d_{i1} - \overline{d_{\bullet 1}}\right)^2 + \sum_{i=2}^{n_2}\left(d_{i2} - \overline{d_{\bullet 2}}\right)^2}{n_1 + n_2 - 2}$; $S_d^2 = \frac{0.5 + 1.17}{4 + 3 - 2} = 0.33$;

Putem verifica ca $\overline{d_{\bullet 1}} - \overline{d_{\bullet 2}} = \overline{Y_T} - \overline{Y_R}$.

$$\overline{Y_R} = \frac{\overline{Y_{\bullet 11}} + \overline{Y_{\bullet 22}}}{2};\ \overline{Y_R} = \frac{\frac{3+4+5+4}{4} + \frac{6+5+5}{3}}{2} = 4.66$$

$$\overline{Y_T} = \frac{\overline{Y_{\bullet 21}} + \overline{Y_{\bullet 12}}}{2};\ \overline{Y_T} = \frac{\frac{4+6+5+5}{4} + \frac{4+5+6}{3}}{2} = 5$$

Observam ca $\overline{Y_T} - \overline{Y_R} = 5 - 4.66 = 0.33 = \overline{d_{\bullet 1}} - \overline{d_{\bullet 2}}$

$$IC_S^{90} = \overline{Y_T} - \overline{Y_R} - t_{n_1+n_2-2;0.90} * S_d * \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \qquad ;$$

$$IC_S^{90} = 0.33 - 1.48 * 0.6 * 0.76 = 0.33 - 0.67 = -0.33$$

$$IC_D^{90} = \overline{Y_T} - \overline{Y_R} + t_{n_1+n_2-2;0.90} * S_d * \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \qquad ;$$

$$IC_D^{90} = 0.33 + 1.48 * 0.6 * 0.76 = 0.33 + 0.67 = 1$$

Prin definitie avem $\mu_R = \mu + F_R$; $\mu_T = \mu + F_T$ si deci $\mu_T - \mu_R = F_T - F_R$ (in ipoteza absentei efectelor reziduale).

Deci, $P\left(IC_S^{90} \langle\ \mu_T - \mu_R\ \langle\ IC_D^{90}\right) = 0.90$

$$\mu_T - \mu_R \subset (-0.33;1) \qquad ; \qquad \frac{\mu_T - \mu_R}{\mu_R} \subset \left(-\frac{0.33}{4.66}, \frac{1}{4.66}\right) \qquad ;$$

$$\frac{\mu_T - \mu_R}{\mu_R} \subset (-0.07;0.21);\ \frac{\mu_T}{\mu_R} \subset (0.93;1.21) \subset (0.8;1.25)$$

Deoarece intervalul de incredere 90% pentru raportul mediilor $\frac{\mu_T}{\mu_R}$ este inclus in intervalul de acceptare a ipotezei privind bioechivalenta $(0.8\,;1.25)$ rezulta ca cele doua medicamente sunt echivalente.

2. Testarea bioechivalentei pe datele de mai sus, logaritmate.

Solutie:

			d_{ik}	$\overline{d_{\bullet k}}$	$\overline{d_{\bullet 1}} - \overline{d_{\bullet 2}}$	$d_{ik} - \overline{d_{\bullet k}}$	$\left(d_{ik} - \overline{d_{\bullet k}}\right)^2$	$\sum\left(d_{ik} - \overline{d_{\bullet k}}\right)^2$	S_d^2	S_d
secv. 1	P_1	P_2								
RT	1.1	1.39	0.14	0.11	0.08	0.03	0.0011	0.02	0.01	0.12
	1.39	1.79	0.20			0.09	0.0086			
	1.61	1.61	0			-0.11	0.0121			
	1.39	1.61	0.11			0	0			
secv. 2	P_1	P_2								
	1.39	1.79	0.20			0.16	0.0265			
TR	1.61	1.61	0	0.04		-0.04	0.0016	0.05		
	1.79	1.61	-0.09			-0.13	0.0172			

$$\overline{Y_R} = \frac{\frac{1.1+1.39+1.61+1.39}{4} + \frac{1.79+1.61+1.61}{3}}{2} = \frac{1.37+1.67}{2} = 1.52$$

$$\overline{Y_T} - \overline{Y_R} - err = 0.08 - 1.48 * 0.12 * 0.76 = 0.08 - 0.14 = -0.06$$

$$\overline{Y_T} - \overline{Y_R} + err = 0.08 + 1.48 * 0.12 * 0.76 = 0.08 + 0.14 = 0.22$$

$$\ln \mu_T - \ln \mu_R \subset (-0.06\,;\,0.22)$$

$$\ln \frac{\mu_T}{\mu_R} \subset (-0.06\,;\,0.22),\ \frac{\mu_T}{\mu_R} \subset \left(e^{-0.06}\,;\,e^{0.22}\right),\ \frac{\mu_T}{\mu_R} \subset (0.94\,;\,1.24)$$

Concluzie:

Produsele sunt bioechivalente, intervalul de incredere 90% fiind inclus in intervalul (0.8, 1.25).

XIII. COMPARAREA A DOUA DREPTE DE REGRESIE

In farmacologie se poate pune problema popularii a doua medicamente nu numai in ceea ce priveste efectele la o doza administrata ci la mai multe doze aceasta mai ales in cazul unor farmacocinetici neliniare.

Relatia biodisponibilitatea – doza sau efect – doza se poate insa liniariza in urma unor transformari (de regula logaritmice).

Se pune in acest caz problema popularii celor doua drepte obtinute cu medicamentele respective. Acestea pot fi paralele, sau mai mult, pot fi estimari diferite ale uneia si aceleiasi drepte.

In alt context se pot compara doua metode bioanalitice prin compararea curbelor de etalonare liniarizate.

Deci matematic avem doua probleme :

- verificarea ipotezei privind paralelismul si
- verificarea ipotezei privind identitatea dreptelor

1. Verificarea ipotezei privind paralelismul

Presupunem ca am obtinut experimental doua grupe de date (x_{1i}, y_{1i}) si (x_{2i}, y_{2i}).

Presupunem ca aceste date sunt corelate de dreptele :

$$y_1 = \alpha_1 + \beta_1\left(x - \overline{x_1}\right) \text{ si } y_2 = \alpha_2 + \beta_2\left(x - \overline{x_2}\right)$$

si deci noi avem sa verificam ipoteza :

$$H_0 : \beta_1 = \beta_2 = \beta \text{ vs } H_A : \beta_1 \neq \beta_2$$

Similar cu demonstrarea facuta pentru o simpla dreapta se arata ca:

$$D(b) = \frac{\sigma^2}{\sum\left(x_{i1} - \overline{x_1}\right)^2 + \sum\left(x_{i2} - \overline{x_2}\right)^2}$$

Deci, pentru testarea ipotezei egalitatii pantelor folosim testul t. Pentru aceasta observam ca,

$$D(b_1 - b_2) = \frac{\sigma_1^2}{\sum\left(x_{i1} - \overline{x_1}\right)^2} + \frac{\sigma_2^2}{\sum\left(x_{i2} - \overline{x_2}\right)^2} \text{ si}$$

$$E(b_1 - b_2) = \beta_1 - \beta_2 = 0$$

Deci statistica de calculat este:

$$T_{n_1+n_2-4} = \frac{b_1 - b_2}{S_p\sqrt{\frac{1}{\sum\left(x_{i1} - \overline{x_1}\right)^2} + \frac{1}{\sum\left(x_{i2} - \overline{x_2}\right)^2}}}$$

unde S_p este dispersia ponderata a reuniunii celor doua populatii :

$$S_p^2 = \frac{(n_1 - 2)S_1^2 + (n_2 - 2)S_2^2}{n_1 + n_2 - 4}$$

Exemplu :

Consideram urmatoarele selectii din doua populatii corelate respectiv prin dreptele :

$y = 2x + 1$ si $y = 2x + 2$

x_{i1}	y_{i1}
0	2
1	3
2	5
3	6

x_{i2}	y_{i2}
0	3
1	4
2	5
3	7

Sa se verifice ipoteza ca dreptele sunt paralele.

2. Verificarea ipotezei privind identitatea dreptelor.

Dupa verificarea ipotezei ca cele doua drepte sunt paralele, se pune problema daca ele sunt efectiv identice si deic dreptele obtinute prin metoda celor mai mici patrate sunt estimari ale uneia si aceiasi drepte.

Pentru aceasta vom considera dreptele

$$y_1 = \alpha_1 + \beta\left(x - \overline{x_1}\right) \text{ si } y_2 = \alpha_2 + \beta\left(x - \overline{x_2}\right)$$

si estimarile a_1, a_2 si b pentru α_1, α_2, respectiv β obtinute prin metoda celor mai mici patrate.

Deci vom calcula valorile a_1, a_2 si b care minimizeaza suma patratelor abaterilor :

$$SP(a_1, a_2, b) = \sum\left[y_{i1} - a_1 - b\left(x_{i1} - \overline{x_1}\right)\right]^2 + \sum\left[y_{i2} - a_2 - b\left(x_{i2} - \overline{x_2}\right)\right]^2$$

Conditia necesara ca $SP(a_1, a_2, b)$ sa fie minim este ca :

$$\frac{\partial SP}{\partial a_1} = \frac{\partial SP}{\partial a_2} = \frac{\partial SP}{\partial b} = 0$$

Calculand se obtine :

$$\frac{\partial SP}{\partial a_1} = -2\sum\left[y_{i1} - a_1 - b\left(x_{i1} - \overline{x_1}\right)\right] = 0$$

$$\frac{\partial SP}{\partial a_2} = -2\sum\left[y_{i2} - a_2 - b\left(x_{i2} - \overline{x_2}\right)\right] = 0$$

$$\frac{\partial SP}{\partial b} = -2\sum\left[y_{i1} - a_1 - b\left(x_{i1} - \overline{x_1}\right)\right]\left(x_{i1} - \overline{x_1}\right) -$$
$$-2\sum\left[y_{i2} - a_2 - b\left(x_{i2} - \overline{x_2}\right)\right]\left(x_{i2} - \overline{x_2}\right)$$

Deoarece $\sum\left(x_{i1} - \overline{x_1}\right) = \sum\left(x_{i2} - \overline{x_2}\right) = 0$ (suma diferentelor fata de medie), rezulta imediat din primele doua ecuatii:

$$a_1 = \frac{\sum y_{i1}}{n_1} = \overline{y_1} \text{ si } a_2 = \frac{\sum y_{i2}}{n_2} = \overline{y_2}$$

Inlocuind in a treia ecuatie se obtine:

$$-\sum\left(y_{i1} - \overline{y_1}\right)\left(x_{i1} - \overline{x_1}\right) + b\sum\left(x_{i1} - \overline{x_1}\right)^2 -$$
$$-\sum\left(y_{i2} - \overline{y_2}\right)\left(x_{i2} - \overline{x_2}\right) + b\sum\left(x_{i2} - \overline{x_2}\right)^2 = 0$$

de unde $b = \dfrac{\sum\left(y_{i1} - \overline{y_1}\right)\left(x_{i1} - \overline{x_1}\right) + \sum\left(y_{i2} - \overline{y_2}\right)\left(x_{i2} - \overline{x_2}\right)}{\sum\left(x_{i1} - \overline{x_1}\right)^2 + \sum\left(x_{i2} - \overline{x_2}\right)^2}$

Deoarece $\alpha_1 + \beta\left(x - \overline{x_1}\right) \equiv \alpha_2 + \beta\left(x - \overline{x_2}\right)$ si $\alpha_1 - \alpha_2 \equiv \beta\left(\overline{x_1} - \overline{x_2}\right)$ consideram estimarea $Z = \left(a_1 - a_2\right) - b\left(\overline{x_1} - \overline{x_2}\right)$ avand

- $E(Z) = 0$ deoarece $a_1 = \overline{y_1}$ si $a_2 = \overline{y_2}$
-

$$D(Z) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} + \left(\overline{x_1} - \overline{x_2}\right)^2 D(b) =$$
$$= \sigma^2\left[\frac{1}{n_1} + \frac{1}{n_2} + \frac{\left(\overline{x_1} - \overline{x_2}\right)^2}{\sum\left(x_{i1} - \overline{x_1}\right)^2 + \sum\left(x_{i2} - \overline{x_2}\right)^2}\right]$$

Astfel se compara valoarea raportului

$$T_{n_1+n_2-3} = \frac{\left(a_1 - a_2\right) - b\left(\overline{x_1} - \overline{x_2}\right)}{S\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2} + \dfrac{\left(\overline{x_1} - \overline{x_2}\right)^2}{\sum\left(x_{i1} - \overline{x_1}\right)^2 + \sum\left(x_{i2} - \overline{x_2}\right)^2}}}$$

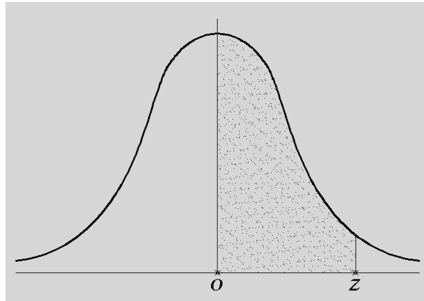
cu valorile pragului de acceptare (respingere) a ipotezei pentru un interval de incredere $1-\alpha$ fixat.

In ceea ce priveste estimarea S^2 a dispersiei, se poate lua suma erorilor $SP\left(a_1, a_2, b\right)$.

Exercitiu :

Sa se verifice ipoteza ca dreptele de la exercitiul anterior sunt identice.

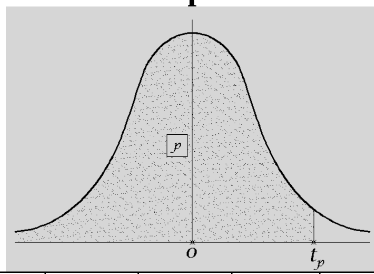
Tabele pentru z



z	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0754
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2258	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2518	0,2549
0,7	0,2580	0,2612	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2996	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4639
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990
3,1	0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993	0,4993

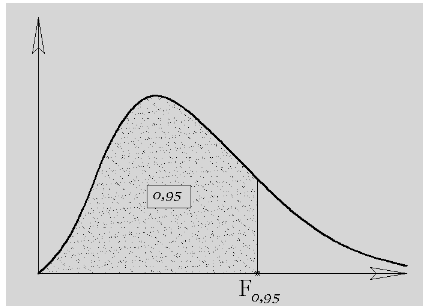
3,2	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995	0,4995
3,3	0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997
3,4	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998
3,5	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998
3,6	0,4998	0,4998	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,7	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,8	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,9	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000

Tabele pentru t

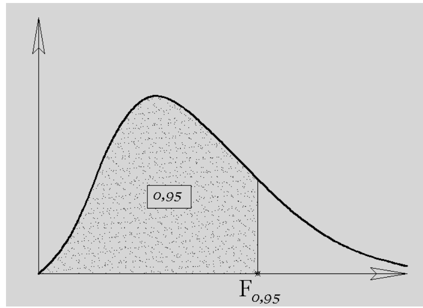


ν	$t_{0,55}$	$t_{0,60}$	$t_{0,70}$	$t_{0,75}$	$t_{0,80}$	$t_{0,90}$	$t_{0,95}$	$t_{0,975}$	$t_{0,99}$	$t_{0,995}$
1	0,158	0,325	0,727	1,000	1,376	3,08	6,31	12,71	31,82	63,66
2	0,142	0,289	0,617	0,816	1,061	1,89	2,92	4,30	6,96	9,92
3	0,137	0,277	0,584	0,765	0,978	1,64	2,35	3,18	4,54	5,84
4	0,134	0,271	0,569	0,741	0,941	1,53	2,13	2,78	3,75	4,60
5	0,132	0,267	0,559	0,727	0,920	1,48	2,02	2,57	3,36	4,03
6	0,131	0,265	0,553	0,718	0,906	1,44	1,94	2,45	3,14	3,71
7	0,130	0,263	0,549	0,711	0,896	1,42	1,90	2,36	3,00	3,50
8	0,130	0,262	0,546	0,706	0,889	1,40	1,86	2,31	2,90	3,36
9	0,129	0,261	0,543	0,703	0,883	1,38	1,83	2,26	2,82	3,25
10	0,129	0,260	0,542	0,700	0,879	1,37	1,81	2,23	2,76	3,17
11	0,129	0,260	0,540	0,697	0,876	1,36	1,80	2,20	2,72	3,11
12	0,128	0,259	0,539	0,695	0,873	1,36	1,78	2,18	2,68	3,06
13	0,128	0,259	0,538	0,694	0,870	1,35	1,77	2,16	2,65	3,01
14	0,128	0,258	0,537	0,692	0,868	1,34	1,76	2,14	2,62	2,98
15	0,128	0,258	0,536	0,691	0,866	1,34	1,75	2,13	2,60	2,95
16	0,128	0,258	0,535	0,690	0,865	1,34	1,75	2,12	2,58	2,92
17	0,128	0,257	0,534	0,689	0,863	1,33	1,74	2,11	2,57	2,90
18	0,127	0,257	0,534	0,688	0,862	1,33	1,73	2,10	2,55	2,88
19	0,127	0,257	0,533	0,688	0,861	1,33	1,73	2,09	2,54	2,86
20	0,127	0,257	0,533	0,687	0,860	1,32	1,72	2,09	2,53	2,84
21	0,127	0,257	0,532	0,686	0,859	1,32	1,72	2,08	2,52	2,83
22	0,127	0,256	0,532	0,686	0,858	1,32	1,72	2,07	2,51	2,82
23	0,127	0,256	0,532	0,685	0,858	1,32	1,71	2,07	2,50	2,81
24	0,127	0,256	0,531	0,685	0,857	1,32	1,71	2,06	2,49	2,80
25	0,127	0,256	0,531	0,684	0,856	1,32	1,71	2,06	2,48	2,79
26	0,127	0,256	0,531	0,684	0,856	1,32	1,71	2,06	2,48	2,78
27	0,127	0,256	0,531	0,684	0,855	1,31	1,70	2,05	2,47	2,77
28	0,127	0,256	0,530	0,683	0,855	1,31	1,70	2,05	2,47	2,76
29	0,127	0,256	0,530	0,683	0,854	1,31	1,70	2,04	2,46	2,76
30	0,127	0,256	0,530	0,683	0,854	1,31	1,70	2,04	2,46	2,75
40	0,126	0,255	0,529	0,681	0,851	1,30	1,68	2,02	2,42	2,70
60	0,126	0,254	0,527	0,679	0,848	1,30	1,67	2,00	2,39	2,66
120	0,126	0,254	0,526	0,677	0,845	1,29	1,66	1,98	2,36	2,62
∞	0,126	0,253	0,524	0,674	0,842	1,28	1,645	1,96	2,33	2,58

Tabele pentru $F_{0,95}$

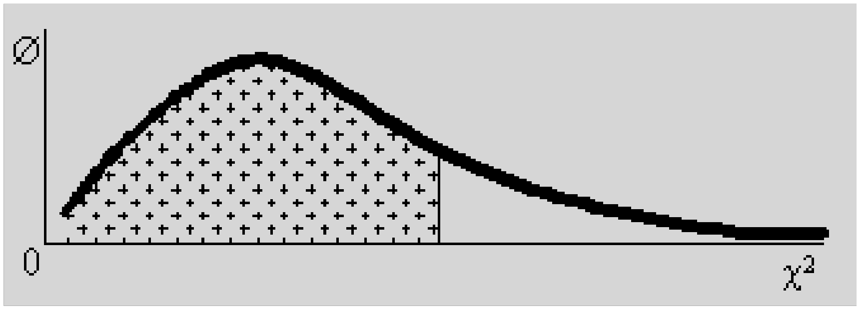


ν_1 / ν_2	1	2	3	4	5	6	7	8	9	10
1	161	200	216	225	230	234	237	239	241	242
2	18,5	19,0	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,4
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20
28	4,20	3,43	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99
120	3,92	3,07	2,68	2,45	2,29	2,18	2,09	2,02	1,96	1,91
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83



ν_1 / ν_2	12	15	20	24	30	40	60	120	∞
1	244	246	248	249	250	251	252	253	254
2	19,4	19,4	19,5	19,5	19,5	19,5	19,5	19,5	19,5
3	8,74	8,70	8,66	8,64	8,62	8,59	8,57	8,55	8,53
4	5,91	5,86	5,80	5,77	5,75	5,72	5,69	5,66	5,63
5	4,68	4,62	4,56	4,53	4,50	4,46	4,43	4,40	4,37
6	4,00	3,94	3,87	3,84	3,81	3,77	3,74	3,70	3,67
7	3,57	3,51	3,44	3,41	3,38	3,34	3,30	3,27	3,23
8	3,28	3,22	3,15	3,12	3,08	3,04	3,01	2,97	2,93
9	3,07	3,01	2,94	2,90	2,86	2,83	2,79	2,75	2,71
10	2,91	2,85	2,77	2,74	2,70	2,66	2,62	2,58	2,54
11	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,45	2,40
12	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,34	2,30
13	2,60	2,53	2,46	2,42	2,38	2,34	2,30	2,25	2,21
14	2,53	2,46	2,39	2,35	2,31	2,27	2,22	2,18	2,13
15	2,48	2,40	2,33	2,29	2,25	2,20	2,16	2,11	2,07
16	2,42	2,35	2,28	2,24	2,19	2,15	2,11	2,06	2,01
17	2,38	2,31	2,23	2,19	2,15	2,10	2,06	2,01	1,96
18	2,34	2,27	2,19	2,15	2,11	2,06	2,02	1,97	1,92
19	2,31	2,23	2,16	2,11	2,07	2,03	1,98	1,93	1,88
20	2,28	2,20	2,12	2,08	2,04	1,99	1,95	1,90	1,84
21	2,25	2,18	2,10	2,05	2,01	1,96	1,92	1,87	1,81
22	2,23	2,15	2,07	2,03	1,98	1,94	1,89	1,84	1,78
23	2,20	2,13	2,05	2,01	1,96	1,91	1,86	1,81	1,76
24	2,18	2,11	2,03	1,98	1,94	1,89	1,84	1,79	1,73
25	2,16	2,09	2,01	1,96	1,92	1,87	1,82	1,77	1,71
26	2,15	2,07	1,99	1,95	1,90	1,85	1,80	1,75	1,69
27	2,13	2,06	1,97	1,93	1,88	1,84	1,79	1,73	1,67
28	2,12	2,04	1,96	1,91	1,87	1,82	1,77	1,71	1,65
29	2,10	2,03	1,94	1,90	1,85	1,81	1,75	1,70	1,64
30	2,09	2,01	1,93	1,89	1,84	1,79	1,74	1,68	1,62
40	2,00	1,92	1,84	1,79	1,74	1,69	1,64	1,58	1,51
60	1,92	1,84	1,75	1,70	1,65	1,59	1,53	1,47	1,39
120	1,83	1,75	1,66	1,61	1,55	1,50	1,43	1,35	1,25
∞	1,75	1,67	1,57	1,52	1,46	1,39	1,32	1,22	1,00

Tabele χ^2



Numar grade de libertate	Aria			
	0,025	**0,950**	**0,975**	**0,990**
1	0	3,842	5,024	6,635
2	0,0501	5,992	7,378	9,210
3	0,216	7,815	9,348	11,345
4	0,484	9,488	11,143	13,277
5	0,831	11,071	12,833	15,086
6	1,237	12,592	14,449	16,812
7	1,690	14,067	16,013	18,475
8	2,180	15,507	17,535	20,090
9	2,700	16,919	19,023	21,666
10	3.247	18,307	20,483	23,209
11	3,816	19,675	21,920	24,725
12	4,404	21,026	23,337	26,217
13	5,009	22,362	24,736	27,688
14	5,629	23,685	26,119	29,141
15	6,262	24,996	27,488	30,578
16	6,908	26,296	28,845	32,000
17	7,564	27,587	30,191	33,409
18	8,231	28,869	31,526	34,805
19	8,907	30,144	32,852	36,191
20	9,591	31,410	34,170	37,566
21	10,283	32,671	35,479	38,932
22	10,982	33,924	36,781	40,289
23	11,689	35,173	38,076	41,638
24	12,401	36,415	39,364	42,980
25	13,120	37,653	40,647	44,314
26	13,844	38,885	41,923	45,642
27	14,573	40,113	43,195	46,963
28	15,308	41,337	44,461	48,278
29	16,047	42,557	45,722	49,588
30	16,791	43,773	46,979	50,892

1. *W.J.Westlake: Use of confidence intervals in analysis of comparative biovalability trials, J. Pharm. Sci. , 61 (8), 1340 – 1, 1972.*
2. *F.Wilcoxon: Individual comparisons by ranking methods, Biometrics Bul.,180-83,1947*
3. *W.H.Kruskal, W.Allen Wallis: Use of ranks in one-criterion analysis of variance, J. Am. Stat. Assoc.,47,583-621,1952*
4. *Hollander, Wolfe DA; Non parametric statistical methods, J.Wiley, New York, 1973*
5. *Hollander, Wolfe DA; Non parametric statistical methods, J.Wiley, New York, 1973*
6. Chow, S.C. & Liu, J.P. (1992) *Design and analysis of bioavailability and bioequivalence studies*. New York, Marcel Dekker (cap. 3) [1].
7. Saporta, C. (1990) *Probabilité, Analyse des données et statistique*. Paris, Ed. Technip (cap. 15) [2].
8. Vaduva, I. (1970) *Analiză dispersională*. Bucureşti, Ed. Tehnică (cap. 4) [3].
9. K.A.Brownlee, *Statistical Theory and metodology in Science and Engineering*, J. Wiley, New – York, 1960
10. D. Ceausescu, *Tratarea statstica a datelor chimico – analitice*, Ed. Tehnica, Bucuresti, 1973
11. M. Tiron, *teoria erorilor de masurare si metoda celor mai mici patrate*, Ed. Tehnica, Bucuresti, 1972
12. F. Gremy, D. Salmon, *Bases statistiques pur la recherchemedicale et biologique*, Dunod, Paris, 1969
13. M. R. Spiegel, *Probability and statistique*, McGraw – Hill, New – York, 1980
14. D. Ceausescu, *Utilizarea statisticii matematice in chimia analitica*, Ed. Tehnica, Bucuresti, 1980
15. M. Iosifescu, T. Postelnicu, *Curs de biomatematica*, Univ. Ecologica, Bucuresti, 1990
16. M. Iosifescu, Gh. Mihoc, R. Teodorescu, *Teoria probabilitatilor si statistica matematica*, Ed. Tehnica, Bucuresti, 1966
17. S. Bolton, Statistics, *in Remington: The Science and Practice of Pharmacy*, 9 – th ed., Mark publ., Easton, Pennsylvania, 1995
18. *United States Pharmacopoeia*, ed. XXIII, cap. *Statistical Procedures for Bioequivalence Studies Using a Standard Two – treatment Crossover design*, 1995
19. P. G. Welling, F.L.S. tse, S. Dighe, *Pharmaceutical Bioequivalence*, cap. 3, C.M. Metzler: Statistical criteria, M. Dekker, New – York, 1991

20. V.W.Steinijans, D. Hauschke, *Update on the statistical analysis of bioequivalence studies*, Int. J.Clin.Pharmacol. Ther. Toxicol,. 28(3), 105 – 110, 1990
21. M. Rowland (ed), *Variability and Drug Therapy*: Description, Estimation and Control, Raven Press, New York, 1985
22. S.C. Chow, J.P.Liu, *Design and Analysis of Biovailability and Bioequivalence Studies*, M. Dekker, London, New York, 1992
23. A. Rescigno. A. Marzo, U. Thyroff – Friesinger, *A new measure of bioequivalence*, 1 –st European Congress of Pharmacology, Milano, june 1995, poster nr. 19
24. A Marzo, *Open questions in bioequivalence*, 1 –st European Congress of Pharmacology, Milano, june 1995, poster nr. 18
25. E. Beyssac, C. Lauro. Marty, H-l Chabard, J-M Aiache, *Study of bioequivalence metrics,* 6-th European Biopharmaceutics and Pharmacokinetics, Atena, aprilie 1997
26. C. Mircioiu, V. Voicu: *Degenerated, solutions of pharmacokinetics models for some lipophilic drugs,* Canad. J. Physiol, Pharmacol. 72 (suppl.1), 305, 1994
27. C. Mircioiu, V. Voicu, M. Jiquidi: *Mathematical algoritms and computer programs as source of variability in population drugs,* 1-st Congress of the European Association for Clinical Pharmacology and Therapeutics, September, 27-30, 1995, Paris
28. C. Mircioiu: *„Mathematical variability" in pharmacokinetics,* 6-th Europ. Congress of Biopharmaceutics and Pharmacokinetics, Atena, 22-24 April 1996, Europ. J. Drug Metab. Pharmacokin. (special issue), abstract 371