Verificarea ipotezelor statistice[1]
de I.Văduva

Notaţii si noţiuni preliminare

Variabila aleatoare: X,Y,U,V,etc., descrisă de funcţie de repartiţie. Variabila aleatoare este asaociată unei *populaţii statistice; valorile ei corespund indivizilor populaţiei.*

Funcţie de repartiţie: $F(x) = P(X < x)$. P =Probabilitate.
Repartiţie continuă, când există $F'(x)$.
Densitate de repartiţie: $f(x) = F'(x)$

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

$0 \leq F(x) \leq 1, F(-\infty) = 0, F(+\infty) = 1,$

Deci $f(x) \geq 0$, $\int_{-\infty}^{+\infty} f(u)du = 1$

$a, b \in R$, $a < b \Rightarrow F(a) \leq F(b)$

Variabila aleatoare discreta este dată de repartiţia sa

$$X : \begin{pmatrix} a_1, & a_2, & ..., & a_n \\ p_1, & p_2, & ..., & p_n \end{pmatrix}, p_i = P(X = a_i), 1 \leq i \leq n, \sum_{i=1}^{n} p_i = 1.$$

$$F(x) = \sum_{a_i < x} p_i, p_i = probabilitati.$$

[1]Conferinţă prezentată la deschiderea seminarului ştiinţific "Nicolas Georgescu Roegen" al Societăţii Române de Econometrie, 4 iulie 2012

Notă. n poate fi si ∞.

Definiţie. *Selecţie (Bernoulliană) de volum n asupra variabilei aleatoare X este mulţimea de variabile aleatoare $\{X_1, X_2, ..., X_n\}$ $n \in \mathcal{N}$,independente şi identic repartizate ca şi X.*

Notă. Selecţia este rezultatul unor observaţii sau măsurători independente (stochastic) efectuate asupra a n indivizi din populaţie. Daca variebilele aleatoare X, Y au respectiv funcţiile de repartiţie F, G, iar *funcţia lor comună de repartiţie* este

$$H(x, y) = P(X < x, Y < y)$$

atunci ele sunt independente dacă $H(x, y) = F(x)G(y)$.

Valori medii. Momente. Dacă considerăm funcţia reală

$\varphi(x)$ masurabilă (!) atunci numim *valoare medie a variabilei aleatoare* $\varphi(X)$ mărimea

$$E[\varphi(X)] = \int_{-\infty}^{\infty} \varphi(u)f(u)du$$

când integrala există, iar in cazul discret, dacă $n = \infty$,

$$E[\varphi(X)] = \sum_{i=1}^{\infty} \varphi(a_i)p_i$$

când seria este convergentă.

Cazuri particulare:

Momente de ordinul $r, r \in \mathcal{N}$:

$$m_r = E[X^r] = \int_{-\infty}^{\infty} x^r f(x)dx, \; in \, cazul \, continuu$$

$$m_r = E[X^r] = \sum_{i=1}^{\infty} a_i^r p_r,\ in\ cazul\ discret.$$

$m_1 = E[X] = (notat) = m$ se numeşte *medie* sau *valoare medie* a lui X.

Momente centrate de ordinul $r, r \in \mathcal{N}$:

$$\mu_r = E[(X - m)^r]$$

Momentul centrat de ordinul al doilea se numeşte *dispersie* sau *varianţă* şi se notează

$$\sigma^2 = \mu_2 = Var(X)$$

iar $\sigma = \sqrt{\sigma^2}$ se numeşte *abatere medie pătratică* sau *abatere standard* sau *deviaţie standard.*

Inegalitatea lui Cebysheff. Dacă esistă momentele de ordinul 1 şi 2, atunci are loc inegalitatea

$$P(|X - m| \leq t\sigma) \geq 1 - \frac{1}{t^2},\ \forall t \in \mathcal{R}^+.$$

Notă. Aceasta inegalitate permite determinarea unui interval de *concentraţie* al valorilor variabilei aleatoare X. De ex. dacă $t = 4$, atunci in intervalul $(m - 4\sigma, m + 4\sigma)$ se gasesc peste 94% din valorile variabilei aleatoare X.

Cazul multidimensional.

Vector aleator: $\mathbf{X} = (X_1, X_2, ..., X_k)'$=vector coloană de dimensiune k.

Funcţie de repartiţie:

$$F(\mathbf{x}) = F(x_1, x_2, ..., X_k) = P(X_1 < x_1, ..., X_k < x_k)$$

Densitate de repartiţie (cazul continuu) când ea există:

$$f(\mathbf{x}) = f(x_1, x_2, ..., x_k) = \frac{\partial^k F(x_1, ..., x_k)}{\partial x_1 ... \partial x_k}$$

Proprietăţi:

$$F(-\infty, ..., -\infty) = 0,\ F((\infty...., \infty) = 1,\ 0 \leq F(x_1, ..., x_k) \leq 1$$

$$\forall i, -\infty < a_i < b_i < \infty \Rightarrow F(x_1, ..., a_i, ..., x_k) \leq F(x_1, ..., b_i, ..., X_k)$$

(adica monotonia crescatoare pe componente).

Proprietăţi ale densităţii de repartiţie:

$$f(\mathbf{x}) \geq 0,\ \int\limits_{R^k} f(\mathbf{u}) d\mathbf{u} = 1$$

$$F(\mathbf{x}) = \int\limits_{-\infty}^{\mathbf{x}} f(\mathbf{u}) d\mathbf{u}$$

Fie $\mathbf{X} = (\mathbf{X_1}, \mathbf{X_2})', Dim\mathbf{X_1} = r, Dim\mathbf{X_2} = s,\ r + s = k$

$\mathbf{X_1}, \mathbf{X_2}$ subvectori ai lui $\mathbf{X}$

Funcţia de repartiţie marginală a lui $\mathbf{X_1}$ este

$$F_1(\mathbf{x_1}) = F(\mathbf{X_1}, \infty = \mathbf{x_2})$$

Densitatea marginală a lui $\mathbf{X_1}$ este

$$f_1(\mathbf{x_1}) = \frac{\partial^r F_1(\mathbf{x_1})}{\partial x_1, ..., \partial x_r}$$

Momente:

$$\forall i, E[X_i] = m_i = \int_{R^k} x_i f(\mathbf{x}) d\mathbf{x} = \int_{-\infty}^{\infty} x_i f_i(x_i) dx_i$$

unde $f_i(x_i)$ este densitatea marginala a componentei aleatoare X_i cand integrela există.

Momentul mixt

$$m_{ij} = E[X_i X_j] = \int_{-infty}^{+\infty} \int_{-\infty}^{+\infty} x_i x_j f_{ij}(x_i, x_j) dx_i dx_j$$

f_{ij}=densitate marginala a lui $(X_i, X_j)'$.

Covarianţa. cand există este:

$$cov(X_i, X_j) = E[(X_i - m_i)(X_j - m_j)] = m_{ij} - m_i m_j = \sigma_{ij}$$

Se observa că $Var(X_i) = cov(X_i, X_i) = \sigma_i^2 = \sigma_{ii}$.

Inegalitatea lui Schwarz

$$|\sigma_{ij}| \leq \sigma_i \sigma_j.$$

Coeficientul de corelaţie al variabilelor aleatoare X_i **şi** X_j este

$$\rho_{ij} = corr(X_i, X_j) = \frac{cov(X_i, X_j)}{\sqrt{Var(X_i) Var(X_j)}} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}$$

Notă. *Ineg. Schwarz* $\Rightarrow \rho_{ij} \in [-1, 1]$

Interpretarea lui ρ : măsoară gradul de dependenţă stochastică al variabilelor aleatoare X_i şi X_j.

Notaţii:

Vectorul valoare medie al lui $\mathbf{X}$ este $\mu = (m_1, m_2, ..., m_k)' = \mathcal{E}(\mathbf{X})$.

Matricea de covarianţă a vectorului X este

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & ,.... & ,\sigma_{1k} \\ . & . & ... & . \\ \sigma_{k1} & \sigma_{k2} & ,... & ,\sigma_{kk} \end{pmatrix} = \mathcal{C}ov(\mathbf{X}, \mathbf{X}')$$

Este pozitiv definită $\Sigma \succ 0$, adică $\mathbf{x}'\Sigma\mathbf{x} > 0$, şi deci inversabilă.

Ipoteză statistică.

$\mathcal{F}$ = multimea funcţiilor de repartiţie.
$\mathcal{F}_0 \subset \mathcal{F}$.
X= variabilă aleatoare $X \rightsquigarrow F$ = funcţie de repartiţie.

Definiţie. *Ipoteză statistică este o afirmaţie asupra lui F de forma $H_0 : F \in \mathcal{F}_0$ ce trebuie verificată cu ajutorul unei selecţii de volum n , $X_1, X_2, ..., X_n$, dată. (Se mai numeste ipoteza nula!)*

Ipoteză *simpla* când $Card\mathcal{F}_0 = 1$; ipoteză *compusă*, când $Card\mathcal{F}_0 > 1$.
Ipoteza alternativă: $H_1 : F \in \mathcal{F}_1, \mathcal{F}_1 \subset \mathcal{F}, \mathcal{F}_1 \neq \mathcal{F}_0$.
Cea mai generală alternativă $H_1 : F \in \mathcal{F} \setminus \mathcal{F}_1$.

Ipoteza parametrica: H_0 se referă la un parametru al funcţiei de repartiţie. De ex. $\mathcal{F}_0$ este familia repartiţiilor normale $N(m, \sigma)$ şi ipoteza este de forma $H : m = m_0$ (ipoteză simplă); aici alternativa poate fi simplă de forma $H_1 : m =$

$m_1, m_1 \neq m_0$, sau altternativa compusă de forma $H_1 : m \neq m_0$. In acest caz ipoteza simplă poate fi de forma $H_0 : |m - m_0| < \lambda$, iar alternativa va fi de forma $H_1 : |m - m_0| \geq \lambda$. Aici $m = E[X]$ este *adevărata medie* a variabilei aleatoare X, m_0 este o valoare dată (de *referinţă*), iar $\lambda > 0$ este *eroarea* cu care apreciem că m poate fi egal cu m_0.

Ipoteză de concordanţă: $H_0 : F \in \mathcal{F}_0$, (adica se specifică *tipul* funcţiei de repartiţie (de ex normală exponenţială Cauchy, Poisson, binomoală etc.) Majoritatea funcţiilor de repartiţie depind de *parametri* θ, adică $F(x) = F(x, \theta)$ unde θ este un parametru *uni* sau *multidimensional.* Dacă θ este cunoscut, atunci ipoteza de concordanţă se numeşte **complet specificată,** iar in caz contrar, se numeşte **nespecificată.**

Notă. Fiind data o selecţie $\mathbf{X}' = X_1, X_2, ..., X_n$ de volum n asupra variabilei aleatoare X, vectorul $\mathbf{X}'$ are o repartiţie de probabilitate pe R^n, a cărui densitate f, (când F are densitate) este

$$L(x_1, x_2, ..., X_n) = \prod_{i=1}^{n} f(x_i)$$

Funcţia $L(x_1, ..., x_n)$ se numeşte **funcţie de verosimilitate.**

Să mai observăm că $L(X_1, ..., X_n)$, cu argumente X_i= valori de selecţie este o variabilă aleatoare!.

Definiţie. Un **test** de verificare a unei ipoteze statistice, este o regulă prin care spaţiul R^n al selecţiilor se descompune in două părţi $W = R_1^n$, şi $\overline{W} = R_2^n = R^n \setminus R_1^n$ (complementarul lui R_1^n) astfel incât, dacă vectoerul de selecţie $\mathbf{X}' \in W$ atunci **se respinge ipoteza** H_0, (adică se acceptă alternativa H_1), iar in caz contrar (adica dacă $\mathbf{X}' \in \overline{W}$,) atunci se acceptă ipoteza H_0. Mulţimea $W = R_1^n$ se numdeşte **domeniu critic al ipotezei** H_0, iar $\overline{W} = R_2^n$ se numeşte **domeniu de acceptare al ipotezei** H_0.

Observaţie importantă. Deoarece o selecţie de volum finit n nu asigură o informaţie completă, **decizia** care se ia pe baza acestei selecţii asupra validităţii sau nu a ipotezei H_0 ne poate conduce la următoarele rezultate: să acceptăm H_0 cand ea este adevărată (notată $(H_0|H_0)$), să accepotăm H_0 când ea nu este adevărată (notată $(H_0|H_1)$), să respingem H_0 când ea este adevărată (notată $(H_1|H_0)$ sau să respingem H_0 când ea nu este adevărată (notată $(H_1|H_1)$). Evident, deciziile bune sunt in primul şi ultimul caz, pe cand celelalte două cazuri constituie **erori ce se comit fiecare cu o probabilitate.** Aceste probabilităţi sunt

$$\alpha = P(H_1|H_0) = P(\mathbf{X}' \in W|H_0),\ \beta = P(H_0|H_1) = P(\mathbf{X}' \in \overline{W})$$

α este **probabilitatea erorii de genul intâi** sau **riscul ,** de genul intâi, in timp ce β este **probabilitatea erorii de genul doi** sau **riscul de genul doi.** α se mai numeşte şi **prag de semnificaţie.**

Probabilitatea

$$\pi = P(H_1|H_1) = 1 - \beta$$

se numeşte **puterea testului.**

Un test bun este acela pentru care α şi β sunt mici (de ex. 0.05 sau mai mici, sau α este mic şi puterea testului π este mare). Din păcate, pentru o selecţie de volum n dată, dacă se impune un rtisc α dat, atunci nu există un test pentru care β sa fie de asemenea oricât mic. **Testul pentru care la un**

prag de semnificaţie dat α există o limitare inferioară a riscului de genul doi β (sau corespunzător există o limitare superioară a lui π), se numeşte test uniform cel mai puternic. Existenţa acestui lucru a este stipulată de următoarea

Lema lui Neyman-Pearson. *Fie $X \rightsquigarrow f(x,\theta)$ şi fie ipoteza parametrică simplă $H_0 : \theta = \theta_0$ şi alternativa $H_1 : \theta = \theta_1$. Atunci pentru un prag α dat, există un test uniform cel mai puternic a cărui regiune critică este de forma*

$$W = \{(X_1, ..., X_n)' | \frac{L_1}{L_0} \geq c > o, \}$$

c este o constantă şi unde

$$L_1 = L(X_1, ..., X_n, \theta_1) = \prod_{i=1}^{n} f(X_i, \theta_1),$$

$$L_0 = L(X_1, ..., X_n, \theta_0) = \prod_{i=1}^{n} f(X_i, \theta_0)$$

, adică L_1, L_0 sunt respectiv **funcţiile de verosimilitate** *ale lui X in ipotezele H_1, H_0.*

Definiţie. Numim **statistică** o funcţie $t(X_1, ..., X_n)$ (care depinde de datele de selecţie). Depinzând de repartiţia de probabilitate a lui X, statistica t are o **repartiţie de probabilitate.** Dacă riscul α este dat atunci, pentru o statistică t convenabil aleasă se poate construi un test pentru ipoteza H_0 a cărui regiune critica este de forma

$$W_\alpha = \{(X_1, X_2, ..., X_n)' : P(t(X_1, ..., X_n) > c_\alpha | H_0) = \alpha\},$$

unde repartiţia statisticii t este considerată in ipoteza H_0. Regiumea critică a testului, W_α, se numeşte **regiune critică de nivel α.**

O statistică t cu ajutorul căreia se construieşte un test pentru o ipoteză nulă H_0 se numeşte **statistică test.**

Din lema lui Neyman-Pearson rezultă că pentru verificarea ipotezei H_0 cu alternativa H_1 **statistica test** este raportul de verosimililităţi

$$t(X_1, X_2, ..., X_n, \theta_0, \theta_1) = \frac{L(X_1, ..., X_n; \theta_1)}{L(X_1, X_2, ..., X_n; \theta_0)}$$

Testul, se numeşte **testul raportului de verosimilităţi.**

Exemplu. Fie $X \rightsquigarrow N(m.\sigma)$ variabila normală, cu abaterea medie pătratică σ, cunoscută. Fie de verificat ipoteza parametrică $H_0 : m + m_0$ cu alternativa $H_1 : m = m_1 > m_0$. (Ambele ipoteze sunt simple). Testul raportului de verosimilităţi conduce,după calcule, la statistica

$$t = \frac{L_1}{L_0} = e^{\overline{X}.n(\frac{m_1-m_0}{\sigma^2} - \frac{m_1^2-m_0^2}{2\sigma^2})},$$

unde $\overline{X}$ este *media aritmetică* a datelor de selecţie, sau *media de selecţie.*

Regiunea critică de nivel α se obţine din relaţia

$$P(\frac{L_1}{L_0} \geq c) = \alpha = P(\overline{X}(\frac{m_1 - m_0}{\frac{\sigma^2}{n}} - \frac{m_1^2 - m_2^2}{\frac{2\sigma^2}{n}}) \geq \log c) = \alpha,$$

adică regiunea critică a testului este in final de forma

$$W_\alpha = \{(X_1, ..., X_n)' : P(\overline{X} \geq \frac{\frac{2\sigma^2}{n} \log c + (m_1^2 - m_0^2)}{2(m_1 - m_0)}) = \alpha\}. \quad (1)$$

Regiunea critică W_α se poate deduce sub o formă echivalentă astfel. In ipoteza H_0, statistica

$$U = \frac{\overline{X} - m_0}{\frac{\sigma}{\sqrt{n}}} \rightsquigarrow N(0, 1).$$

Deci, pentru un α dat , alegem z_α astfel incat

$$P(Z \geq z_\alpha) = \int\limits_{z_\alpha}^{\infty} e^{-\frac{t^2}{2}} dt = \alpha,$$

de unde domeniul critic este

$$W_\alpha = \{(X_1, X_2, ..., X_n)' | \overline{X} \geq m_0 + \frac{z_\alpha}{\sqrt{n}}\}. \tag{2}$$

Mărimea z_α se numeşte α-cuantila superioară a repartiţiei normale $N(0, 1)$.

Observăm că cele două forme ale domeniului critic W_α date de (1) şi (2) coincid, deoarece au acelaş nivel α.

Puterea testului, este

$$\pi(m_1) = P(\overline{X} \geq m_0 + z_\alpha \frac{\sigma}{\sqrt{n}} | H_1) = P(\frac{\overline{X} - m_1}{\frac{\sigma}{\sqrt{n}}} \geq \frac{m_0 - m_1}{\frac{\sigma}{\sqrt{n}}} + z_\alpha) =$$

$$= P(Z \geq \frac{m_0 - m_1}{\frac{\sigma}{\sqrt{n}}} + z_\alpha)$$

Deoarece $\pi(m_1) = 1 - \beta$ rezultă că

$$1 - \phi(z_\alpha + \frac{\sigma}{\sqrt{n}}) = 1 - \beta$$

deci

$$z_\alpha + \frac{m_o - m_1}{\frac{\sigma}{\sqrt{n}}} = z_{1-\beta}$$

Ultima formulă conduce la faptul că dacă se dau riscurile α şi β atunci volumul **minim** de selectie necesar pentru realizarea acestor riscuri este

$$n = (z_{1-\beta} - z_\alpha)^2 \frac{\sigma^2}{(m_1 - m_0)^2}$$

ceea ce conduce si la o altă consecinţă a lemei Neyman-Pearson.

Notă. Din cele de mai sus, observăm că dacă considerăm parametrul $\lambda = |m_0 - m_1|$ ca o **distanţă** intre ipotezele H_0 şi H_1 şi considerăm că pentru o distanţă λ_0 dată $H_1 \sim H_0$ atunci puterea π se exprimă in funcţie de λ si anume

$$\pi(\lambda) = 1 - \phi(z_\alpha + \frac{\lambda}{\sigma\sqrt{n}}).$$

Forma generală a testului raportului de verosimilităţi.

Să considerăm ipoteza $H : F \in \omega \subset \Omega$, unde Ω este o clasă de funcţii de repartiţie si ω o submulţime a sa.Alternativa este $\mathcal{N}H : F \in \Omega \setminus \omega$. Să notăm $(L)_\Omega, (L)_\omega$ **valorile maxime** ale funcţiei de verosimilitate in ip[otezele Ω, ω şi să cosniderăm **raportul de verosimilitate**

$$\Lambda(\mathbf{X}) = \frac{(L)_\omega}{(L)_\Omega}, \mathbf{X} = \textit{vectorul de selecţie}.$$

Deoarece $\omega \subset \Omega$ rezultă că $\Lambda(\mathbf{X}) \leq 1$, iar cand ω este adevărată, $\Lambda(\mathbf{X}) = 1$. (Caz ideal!).

Deci domeniul critic pentru testarea ipotezei H este de forma

$$W(c) = \{\mathbf{X}|\Lambda(\mathbf{X}) \leq c < 1\},\ P(\Lambda(\mathbf{X} \leq c) = \alpha. \qquad (3)$$

Lema lui Neyman-Pearson este valabilă şi aici; regiunea critică $W(c)$ dată de (3) corespunde **testului uniform cel mai puternic.**

Pentru a construi testul raportului de verosimilităţi pentru o ipoteză H va trebui mai intai să calculăm valorile maxime $(L)_\Omega, (L)_\omega$ ale funcţiei de verosimilitate.

Exemplu. Fie $X \rightsquigarrow N(m, \sigma)$ cu σ-**cunoscut** si fie de verificat ipoteza $H : m = m_0$ cu alternativa $\mathcal{N}H : m \neq m_0$. Maximul funcţiei de verosimilitate in ipoteza Ω conduce la

$$(L)_\Omega = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2}$$

iar

$$(L)_\omega = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum\limits_{i=1}^{n}(X_i-m_0)^2}.$$

Raportul de verosimilităţi este

$$\Lambda(\mathbf{X}) = e^{-\frac{n}{2\sigma^2}(\overline{X}-m_0)^2}$$

iar domeniul critic este de forma (3) unde $c = c_\alpha$ satisface relaţia

$$\alpha = P[-\frac{n}{2\sigma^2}(\overline{X} - m_0)^2 \leq \log c_\alpha] = P[\left|\frac{\overline{X} - m_0}{\frac{\sigma}{\sqrt{n}}}\right| \geq \sqrt{-2\log c_\alpha}].$$

Deoarece

$$\frac{\overline{X} - m_0}{\frac{\sigma}{\sqrt{n}}} = Z \rightsquigarrow N(0, 1)$$

rezultă că folosind $z_{\frac{\alpha}{2}}$ dat de relaţia

$$\int\limits_{-z_{\frac{\alpha}{2}}}^{z_{\frac{\alpha}{2}}} e^{-\frac{u^2}{2}} du = 1 - \alpha,$$

domeniul c ritic este de forma

$$W_\alpha = \{\mathbf{X}' : \left|\frac{\overline{X} - m_0}{\frac{\sigma}{\sqrt{n}}}\right| \geq z_{\frac{\alpha}{2}}\}. \qquad (3')$$

Puterea testului $\pi(m)$ se calculează cu formula

$$P(\left|\frac{\overline{X}-m_0}{\frac{\sigma}{\sqrt{n}}}\right| \geq z_{\frac{\alpha}{2}}|\mathcal{N}H) = \pi(m). \tag{4}$$

Testul prezentat se numeste testul U.

Problema celor două selecţii.

Fie $X \rightsquigarrow N(m_1, \sigma_1)$, $Y \rightsquigarrow N(m_2, \sigma_2)$ cu σ_1, σ_2 cunoascute. Se dă o selectie de volum n_1 pentru X si o selecţie de volum n_2 pentru Y. Pentru verificarea ipotezei $H : m_1 = m_2$ cu alternativa $\mathcal{N}H : m_1 \neq m_2$ se foloseşte statistica

$$U = \frac{\overline{X} - \overline{Y} - m_1 + m_2}{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}. \tag{5}$$

care in ipoteza H are repartiţia normală $N(0, 1)$.

Domeniul critic se determină pe baza statisticii U dată de (5) şi el este de forma

$$W_\alpha = \{\mathbf{X}' : |U| \geq z_{\frac{\alpha}{2}}\},$$

iar Puterea testului se calculează cu formula

$$\pi(m_1 - m_2) = P(|U| \geq z_{\frac{\alpha}{2}}|\mathcal{N}H).$$

Cazul dispersiilor necunoscute. Repartiţii inrudite cu repartiţiile normale.

Fie de testat $H : m = m_0, \mathcal{N}H : m \neq m_0$, cu σ necunoscut. Determinarea raportului de verosimilităţi, conduce mai intâi la estimarea lui m cu $\overline{X}$ şi a lui σ^2 cu formula

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \tag{6}$$

după care se calculează $(L)_\Omega$ şi $(L)_\omega$

In final testul raportului de verosimilităţi conduce la **statistica t a lui Student**, adică

$$t = \frac{\overline{X} - m_0}{\frac{s}{\sqrt{n}}} \tag{6'}$$

care in ipoteza H are **repartiţia Student cu $f = n - 1$ grade de libertate,** ce are densitatea de repartiţie

$$g(x) = \frac{1}{\sqrt{\pi}} \frac{\Gamma(\frac{f+1}{2})}{\Gamma(\frac{f}{2})} \frac{1}{(1+x^2)^{\frac{f+1}{2}}}. \tag{7}$$

Variabila Student se defineşte cu formula

$$t_f = \frac{Z}{\frac{\chi_f}{\sqrt{f}}}, t_f \in R,\ Z \rightsquigarrow N(0,1)$$

unde $\chi_f^2 = \sum\limits_{i=1}^{f} Z_i^2$, iar Z_i sunt variabile $N(0,1)$ independete şi Z e independent de χ_f^2.

Densitatea de repartiţie a lui χ_f^2 este

$$h(x) = \frac{1}{2^{\frac{f}{2}}\Gamma(\frac{f}{2})} x^{\frac{f}{2}-1} e^{-\frac{x}{2}}, x > 0, h(x) = 0\, daca\, x \leq 0. \tag{8}$$

Dacă $E[Z_i] = m_i \neq 0$ măcar pentru un i atunci

$$\sum_{i=1}^{f} Z_i^2 = \chi_{f,\delta}^2$$

cu $\delta^2 = \sum\limits_{i=1}^{f} m_i^2$ se numeşte **variabilă χ^2 necentrată, cu f grade de libertate şi cu parametru de excentricitate δ.**

Nu precxizăm densitatea de repartiţie (complicată!) a acestei variabile.

Definiţie. Variabila aleatoare $F_{f_1,f_2} > 0$ este definită astfel

$$F_{f_1,f_2} = \frac{f_2\chi^2_{f_1}}{f_1\chi^2_{f_2}}, \tag{9}$$

Variabila F_{f_1,f_2} are o densitate de repartiţie pe care nu o prezentăm aici. Sunt utilizate si variabile F necentrate de forma
$F_{f_1,f_2;\delta_1,0}, F_{f_1,f_2;0,\delta_2}, F_{f_1,f_2;\delta_1,\delta_2}$.
Cea mai utilizată după cum vom vedea, este prima formă de F-necentrataă.

Intre variabila F si variabila t este valabilă relaţia

$$t_f^2 = F_{1,f}.$$

Forme ale testului t.

Pentru un risc α dat, să cosiderăm cuantila superioară $t_{f,\frac{\alpha}{2}} > 0$ care satisface relaţia

$$P(|t_f| \leq t_{f,\frac{\alpha}{2}}) = \int_{t_{f,\frac{\alpha}{2}}}^{t_{f,\frac{\alpha}{2}}} g(u)du = 1 - \alpha \tag{10}$$

Ca şi testul U, testul t, dedus din testul general al raportului de probabilităţi, capătă forme asemănătoare, după cum urmează:

t1.**Verificarea ipotezei** $H : m = m_0, \sigma - \mathit{necunoscut}$**, cu alternativa** $\mathcal{N}H : m \neq m_0$. Domeniul critic este

$$\left|\frac{\overline{X} - m_0}{\frac{s}{\sqrt{n}}}\right| \geq t_{f,\frac{\alpha}{2}}, f = n - 1, \tag{11}$$

Puterea testului se calculează cu formula

$$\pi(m) = P(\left|\frac{\overline{X} - m_0}{\frac{s}{\sqrt{n}}} \geq t_{f,\frac{\alpha}{2}}\right| : \mathcal{N}H) \qquad (11')$$

unde statistica din formulă are repartiţia t-necentrată adica

$$t^2_{f,\delta} = F_{1,f:\delta,0}, \delta^2 = \left(\frac{m_1 - m_0}{\frac{s}{\sqrt{n}}}\right)^2. \qquad (11")$$

t2.**Verificarea ipotezei** $H : m_1 = m_2$ **pentru două populaţii** $N(m_1, \sigma), N(m_2), \sigma), \sigma$ – *cunoscut* **cu** $\mathcal{N}H : m_1 \neq m_2$. Fie $X \rightsquigarrow N(m_1, \sigma_1)$, $Y \rightsquigarrow N)(m_2, \sigma_2)$ $\sigma_1 = \sigma_2 = \sigma$. si volumele de selecţie n_1, n_2. Dispersia σ^2 se estimează astfel

$$s^2 = \frac{1}{n_1 + n_2 - 2}\{\sum_{i=1}^{n_1}(X_i - \overline{X})^2 + \sum_{j=1}^{n_2}(Y_j - \overline{Y})^2\},\ f = n_1 + n_2 - 2.$$

Statistica t este in acest caz

$$t = \frac{\overline{X} - \overline{Y}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

domeniul critic de nivel α este de forma (11), iar puterea testului $\pi(m_1 - m_2)$ este de forma (11') cu

$$\delta^2 = \left(\frac{m_1 - m_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}\right)^2.$$

t3.**Verificarea ipotezei** H **din cazul precedent, cu** σ_1, σ_2 **necunoscute şi ne egale.** In acest caz testul t are o construcţie specială şi anume;

- se estimează dsispersiile cu formulele obişnuite

$$s_1^2 = \frac{1}{f_1}\sum_{i=1}^{n_1}(X_i - \overline{X})^2, f_1 = n_1 - 1; s_2^2 = \frac{1}{f_2}\sum_{j=1}^{n_2}(Y_j - \overline{Y})^2, f_2 = n_2 - 1; \tag{12}$$

- se calcullează gradele de libertate f cu formulele

$$c = \frac{\frac{s_1^2}{f_1}}{\frac{s_1^2}{f_1} + \frac{s_2^2}{f_2}}$$

$$f = \frac{1}{\frac{c^2}{f_1} + \frac{(1-c)^2}{f_2}}$$

(f este rotunjit la intreg)

-statistica t este

$$t = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

In continuare testul t se dezvoltă ca la t2.

Teste privind egalitatea dispersiilor populaţiilor normale.

Se dau $X \rightsquigarrow N(m_1, \sigma_1), Y \rightsquigarrow M(m_2, \sigma_2)$ şi selecţiile independente corespunzătoare de volume n_1, n_2. Ipoteza $H : \sigma_1 = \sigma_2$ cu alternativa $\mathcal{N}H : \sigma_1 \neq \sigma_2$ se verifică folosind testul F (al lui Snrdrcor) după cum urmează:

- se estimează σ_1^2, σ_2^2 cu formulele (12);

se calculează statistica

$$F = \frac{s_1^2}{s_2^2}$$

Statistica F are repartiţia F-centrată cu (f_1, f_2) grade de libertate. Deci **domeniul critic de nivel** α este

$$F \geq F_{f_1,f_2;\alpha}, \, unde \, P(F_{f_1,f_2} \geq F_{f_1,f_2;\alpha}) = \alpha,$$

adică $F_{f_1,f_2;\alpha}$ este α-cuantila superioară a repartiţiei F.

Testul lui Bartlett pentru egalitatea a mai multe dispersii. Se dau k populaţii normale $N(m_1, \sigma_i), 1 \leq i \leq k$ si selectţii corespunzătoare $X_{i,j}, .1 \leq i \leq k, 1 \leq j \leq n_i$ de volume $n_1, n_2, ..., n_k, n_i > 3$ respectiv. Se cere să se verifice ipoteza

$$K : \sigma_1^2 = ... = \sigma_k^2.$$

Testul lui Bartlett se realizează in următorii paşi:

- se estimează dispersiile cu formulele

$$S_i^2 = \frac{1}{n_i - 1}(\sum_{j=1}^{n_i} X_{ij}^2 - n_i \overline{X_i}^2), 1 \leq i \leq k$$

- se calculează s^2 cu formula

$$s^2 = \frac{1}{f}(\sum_{i=1}^{k} f_i s_i^2), \; f_i = n_i - 1, f = \sum_{i=1}^{k} f_i$$

se calculează statistica lui Bartlett

$$\chi^2 = -\frac{1}{B}\sum_{i=1}^{k} f_i \log \frac{s_i^2}{s^2}, \; B = \frac{\sum_{i=1}^{k} \frac{1}{n_i - 1} - \frac{1}{n-k}}{3(k-1)} + 1, \; n = \sum_i n_i. \quad (13)$$

Statistica χ^2 are $k-1$ grade de libertate, deci domeniul critic al testului lui Bartlett este

$$\chi^2 \geq \chi^2_{k-1,\alpha}, \, unde \, P(\chi^2_{k-1} \geq \chi^2_{k-1,\alpha}) = \alpha.$$

(aici α este riscul de genul intai).

Puterea testului se calculează pe baza repatriţiei necentrate

$$\chi^2_{k-1,\delta},\ \delta^2 = \frac{\sum\limits_{i=1}^{k} f_i\sigma_i^2}{f},$$

(σ_i diferite intre ele).

Teste de concordanţă. Presupunem că se dă o selecţie de volum n asupra lui X si se cere să verificăm ipoteza de concordanţă $H : X \rightsquigarrow F$. Prezentăm două teste **asimptotice** (când $n \to \infty$).

Testul de concordanţă χ^2.

Dacă ipoteza H este **complet specificată**, atunci testul χ^2 constă din următoarele etape:

- se consideră 0 diviziune a mulţimii Δ pe care variabila aleatoare X ia valori de probabilităţi pozitive, adică

$$\Delta = \bigcup_{i=1}^{k} \Delta_i,\ \Delta_i \cap \Delta_j = \emptyset,\ P(\Delta_i) > 0.$$

- se calculează probabilităţile $p_i = P(\Delta_i) > 0$, $1 \le i \le k$;

- pentru selecţia dată $X_1, X_1, ..., X_n, n = f.mare(n > 1000)$ se determină f_i = **numărul valorilor de selecţie ce aparţin lui** Δ_i, adică **frecvenţele absolute pe** Δ_i;

- se calculează statistica

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - np_i)^2}{np_i} \qquad (13-1)$$

Deoarece statistica χ^2 are repartiţia χ^2_{k-1}, domeniul critic al testului este

$$\chi^2 \ge \chi^2_{k-1,\alpha},\ P(\chi^2_{k-1} \ge \chi^2_{k-1,\alpha} = \alpha.$$

Puterea testului se determină ca deobicei cu χ^2 necentrat (repartiţia statisticii (13-1) in ipoteza $\mathcal{N}H$.)

Dacă H **este nespecificată,** atunci etapele testului χ^2 suferă o modificare şi anume dacă funcţia de repartiţie depinde de un parametru $\theta = (\theta_1, ..., \theta_c)'$, $c < k-1$, atunci $p_i = p_i(\theta)$ şi statistica (13-1) devine

$$\chi^2(\theta) = \sum_{i=1}^{k} \frac{(f_i - np_i(\theta)^2}{np_i(\theta)} \qquad (13-2)$$

iar parametrul θ trebuie estimat. Estimaţia $\overline{\theta}$ se obţine minimizând (13-2) in raport cu **theta**, dar cu condiţia ca numitorii din suma (13-2) sa fie asimptotic constanţi.(Această metodă de estimare se numeşte **metoda minimului lui** χ^2 **modificat.**) După estimarea celor c parametri, probabilităţile din (13-1) devin $p_i = \hat{p}_i = p_i(\overline{\theta})$, iar statistica devine

$$\chi^2(\overline{\theta}) = \sum_{i=1}^{k} \frac{(f_i - n\hat{p}_i)^2}{n\hat{p}_i}. \qquad (13-3)$$

Se ştie că statistica (13-3) are o repartiţie χ^2_{k-c-1} şi de aici se continuă paşii din cazul când H este complet specificată. Puterea testului se calculează tot cu χ^2-necentrat unde paqrametrul de excentricitate este

$$\delta^2 = \sum_{i=1}^{k} \frac{(p_i^H - p_i^{\mathcal{N}H})^2}{np_i^H},$$

unde $p_i^H, p_i^{\mathcal{N}H}$ sunt calculate in ipotezele respective.

Teste de concordanţă de tip Kolmogorov-Smirnov.

Aceste teste se aplică **numai când funcţia de repartiţie** F **este continuă.**

Definim mai intâi **estimaţia nedeplasaată a funcţiei de repartiţie** $F(x)$. Aceasta este

$$F_n(x) = \frac{\nu(x)}{n}, \qquad (13-4)$$

unde $\nu(x)$ =**numărul valorilor de selecţie mai mici decât** X. Ea se mai numeşte şi **funcţia de repartiţie empirică.**

Să notăm

$$D_n = \sup_x |F(x) - F_n(x)| = \max_{1 \le i \le n} |F(X_i) - F_n(X_i)|$$

$$D_n^+ = \max_{1 \le i \le n} [F_n(X_i) - F(X_i)],\ D_n^- = \max_{1 \le i \le n} [F(X_i) - F_n(X_i)].$$

Testele de tip Kolmogorov-Smirnov se bazează pe următoarele teoreme limită:

Teorema lui Kolmogorov. Dacă F este continuă, atunci

$$\lim_{n \to \infty} P(D_n < \frac{\lambda}{\sqrt{n}}) = \sum_{k=-\infty}^{+\infty} (-1)^k e^{-\lambda^2 k^2} = K(\lambda). \qquad (13-5)$$

Teorema lui Smirnov. Dacă F este continuă atunci

$$\lim_{n \to \infty} P(D_n^+ < \frac{\lambda}{\sqrt{n}}) = 1 - e^{2\lambda^2}. \qquad (13-6)$$

Testul lui Kolmogorov are domeniul critic de nivel $\alpha \le 0.05$ de forma

$$D_n > \frac{\lambda_\alpha}{\sqrt{n}},\ unde\, K(\lambda_\alpha) = 1 - \alpha. \qquad (13-7)$$

In mod asemănător, domeniul critic pentru **testul lui Smirnov** este

$$D_n^+ > \frac{\theta_\alpha}{\sqrt{n}},\ \ unde\ \ e^{-2\theta_\alpha^2} = \alpha. \qquad (13-8)$$

Puterea testului Kolmogorov se calculează pe baza repartiţiei asimptotice a statisticii

$$D_n^* = \sup_x |F_n(x) - G(x)|, \ unde \mathcal{N}H : X \rightsquigarrow G(x).$$

Nu există evaluări exacte privind puterea testului lui Kolmogorov.

Dacă pentru două variabile X având funcţia de repartiţie F şi Y având funcţia de repartiţie G (F, G necunoscute!), se dau două selecţii asupra lor, de volume n şi m respectiv, atunci se poate pune problema **testării ipotezei** $H : F = G$. Testarea acestei ipoteze se face pe baza următoarei teoreme

Teorema lui Smirnov. Dacă F şi G sunt continue şi notăm

$$D_{n,m} = \sup_x |F_n(x) - G_m(x)|,$$

atunci

$$lim_{n,m\to\infty, \frac{n}{m}=\rho=const.} P(D_{n,m} < \lambda(\sqrt{\frac{1}{n} + \frac{1}{m}})) = K(\lambda). \quad (13-9)$$

Domeniul critic al testului este

$$D_{n,m} > \lambda_\alpha \sqrt{\frac{1}{n} + \frac{1}{m}}, \ K(\lambda_\alpha) = 1 - \alpha. \qquad (13-10)$$

Puterea testului se determină ca şi in cazul testului Kolmogorov.

Teste pentru repartiţii multidimensionale.

Vom prezenta teste referitoare la mediile **repartiţiilor normale multidimensionale**. Vectorul $\underset{\sim}{X} = (X_1, X_2, ..., X_k)'$ are repartiţia normală k-dimensională $N(\mu, \Sigma)$ dacă densitatea sa de repartiţie este

$$f(\mathbf{x}, \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{k}{2}} det(\Sigma)^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)}. \qquad (14)$$

μ este vectorul medie al lui $\mathbf{X}$, iar Σ este mateicea de covarianţă a lui $\mathbf{X}$ notate respectiv

$$\mu = \mathcal{E}(\mathbf{X}),\ \Sigma = \mathcal{C}ov(\mathbf{X}, \mathbf{X}'),$$

vectorii, fiind vectori coloană, iar produsele matriceale sunt calculate conform regulii obişnuite "linii prin coloane". Matricea Σ este pozitiv definită, (notată $\Sigma \succ 0$), de unde rezultă că **forma pătratică** de la exponent in formula (14) este **pozitiv definită.** O selecţie de volum N asupra vectorului aleator $\mathbf{X}$ este de forma $\mathbf{X_1}, \mathbf{X_2}, ..., \mathbf{X_N}$ care de fapt este o matrice $N \times k$, $\mathbf{X_i}$ fiind coloanele acestei matrici: $\mathbf{X_i}$ sunt deci **valori de selectie efectuate asupra lui X**.

Estimaţiile **nedeplasate** ale parametrilor ν, Σ sunt respectiv

$$\overline{\mathbf{X}} = \sum_{i=1}^{N} \mathbf{X_i},\ S = \frac{1}{N-1} \sum_{i=1}^{N} (\mathbf{X_i} - \overline{\mathbf{X}})(\mathbf{X_i} - \overline{\mathbf{X}})', \qquad (16)$$

adică

$$\mathcal{E}[\overline{\mathbf{X}}] = \mu,\ \mathcal{E}[S] = \Sigma.$$

In cazul unidimensional testele asupra mediilor se bazau pe staistica U repatizată normal şi pe statistica t a lui Student. Asemănător, testele privind mediile repartiţiilor normale multi dimensionale se vor baza pe o **statistică** χ^2 si pe o statistică T^2 a lui Hoteling, cu n grade de libertate. Aceste statistici arată de forma

$$\chi_k^2 = \mathbf{Y}'\Sigma^{-1}\mathbf{Y},\ \mathbf{Y} \rightsquigarrow N(\mathbf{0}, \Sigma), \qquad (17)$$

$$T_n^2 = \mathbf{Y}'S^{-1}\mathbf{Y},\ \ \mathbf{Y} \rightsquigarrow N(\mathbf{0}, \Sigma), nS = \sum_{i=1}^{n} \mathbf{Z}_\alpha \mathbf{Z}_\alpha{}', \qquad (18)$$

unde $\mathbf{Z_i} \rightsquigarrow N(\mathbf{0}, \sigma), \mathbf{Z_i}\, ind\, \mathbf{Y}$. S este o matrice $Wishart$. Variabila T_n^2 are **repartiţia Hoteling cu n grade de libertate.**

Se arată că variabila T_n^2 este legată de variabila F prin relaţia

$$\frac{n-k+1}{k}\frac{T_n^2}{n} = F_{k,n-k+1}. \qquad (18)'$$

iar dacă in (18) $\mathbf{Y} \rightsquigarrow N(\mu, \Sigma)$, atunci T_n^2 din (18) are repartiţia **Hoteling necentrată** cu parametrul de excentricitate

$$\delta^2 = \mu'\Sigma^{-1}\mu,$$

relatia (18') ramânând valabilă si pentru variabile necentrate. Relaţia (18') se păstrează şi intre cuantilele variabilelor F şi T^2 şi anume

$$T_{n,\alpha}^2 = \frac{nk}{n-k+1}F_{k,n-k,\alpha}. \qquad (18'')$$

Verificarea ipotezelor asupra mediilor cand matricile de covarianţă sunt cunoscute.

$\mathbf{H_1}$. Ipoteza $H : \mu = \mu_\mathbf{0}$, cu alternativa $\mathcal{N}H; \mu \neq \mu_\mathbf{0}$. Se foloseşte selecţia de volum N. Deoarece in ipoteza, H $\overline{X} \rightsquigarrow N(\mu_\mathbf{0}, \frac{\Sigma}{N})$, rezultă că statistica

$$\chi^2 = N(\overline{\mathbf{X}} - \mu_\mathbf{0})'\Sigma^{-1}(\overline{\mathbf{X}} - \mu_\mathbf{0}) \qquad (19)$$

are repartiţia χ_k^2, deci domeniul critic de nivel α este conform (19)

$$\chi^2 \geq \chi_{k,\alpha}^2, \ \textit{unde}\ P(\chi_k^2 \geq \chi_{k,\alpha}^2) = \alpha. \qquad (19')$$

Puterea testului este dată de repartiţia χ^2-necentrată adică

$$\pi(m) = P(\chi_{k;\delta} \geq \chi_{k,\alpha}^2),\ \textit{unde}\ \delta^2 = N(\mu - \mu_\mathbf{0})'\Sigma^{-1}(\mu - \mu_\mathbf{0}). \quad (20)$$

Amintim faptul că **distanţa lui Mahalanobis** dintre repartiţiile normale $N(\mu_\mathbf{1}, \Sigma), N(\mu_\mathbf{2}, \Sigma)$ este

$$D^2 = (\mu_\mathbf{1} - \mu_\mathbf{2})'\Sigma^{-1}(\mu_\mathbf{1} - \mu_\mathbf{2})$$

deci δ^2 este o distanţă Mahalanobis.

H_2. **Problema celor două saelecţii, pentru două populaţii normale $\mathbf{X} \rightsquigarrow N(\boldsymbol{\mu_1}, \Sigma), \mathbf{Y} \rightsquigarrow N(\boldsymbol{\mu_2}, \Sigma)$ cu Σ cunoscut.**

Presupunem că volumele celor două selecţii sunt N_1 respectiv N_2 si avem de testat ipoteza $H : \mu_1 = \mu_2$ cu alternativa $\mathcal{N}H : \mu_1 \neq \mu_2$. Deoarece in ipoteza H avem
$(\overline{X} - \overline{Y}) \rightsquigarrow N(\mathbf{0}, (\frac{1}{N_1} + \frac{1}{N_2})\Sigma)$, rezultă

$$\chi^2 = \frac{N_1 N_2}{N_1 + N_2}(\overline{\mathbf{X}} - \overline{\mathbf{Y}})'\Sigma^{-1}(\overline{\mathbf{X}} - \overline{\mathbf{Y}}) \tag{21}$$

Domeniul critic de nivel α este deci de forma (19') iar puterea testului se determină cu χ^2-necentrat cu parametrul de excentricitate

$$\delta^2 = \frac{N_1 N_2}{N_1 + N_2}(\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2). \tag{21'}$$

H_3. **Problema celor r selecţii.** Fie vectorii normali $\mathbf{X^{(i)}} \rightsquigarrow N(\mu^{(i)}, \Sigma)$, Σ – *cunoscut* şi selecţiile de volume N_i asupra lor, $1 \leq i \leq r$. Se dau constantele $\beta_i, 1 \leq i \leq r$ (ce pot fi numite **măsuri de ponderare**).

Se cere să se verifice ipoteza $H : \mu = \mu_0,\ \mu = \sum\limits_{i=1}^{r} \beta_i \mu_i$, numită problema celor r selectii. (In biologie μ este media caracteristicii unei specii ce provine din r ascendenţi; in economie, μ poate fi suma cheltuită de o familie pentru a-şi asigura r resurse necesare).

Deoarece in ipoteza H vectorul aleator $\sum\limits_{i=1}^{r} \beta_i \overline{\mathbf{X^{(i)}}} \rightsquigarrow N(\mu_0, (\sum\limits_{i=1}^{r} \frac{\beta_i^2}{N_i})\Sigma)$, rezultă ca testul se bazează pe statistica

$$\chi^2 = \left(\sum_{i=1}^{r} \frac{\beta_i^2}{N_i}\right)^{-1} (\sum_{i=1}^{r} \beta_i \overline{\mathbf{X^{(i)}}} - \mu_0)'\Sigma^{-1}(\sum_{i=1}^{r} \beta_i \overline{\mathbf{X^{(i)}}} - \mu_0). \tag{22}$$

Domeniul critic de nivel α este tot de forma (19') cu χ^2 dat de (22). Puterea testului se calculează tot cu χ^2 necentrat cu parametrul de excentricitate

$$\delta^2 = \left(\sum_{i=1}^{r} \frac{\beta_i^2}{N_i}\right)^{-1} (\sum_{i=1}^{r} \beta_i \mu^{(i)} - \mu_0)" \Sigma^{-1} (\sum_{i=1}^{r} \beta_i \mu_i - \mu_0). \qquad (22')$$

H_4. **Cazul matricilor de covarianţă neegale.** Dacă $\mathbf{X^{(i)}} \rightsquigarrow N(\mu_i, \Sigma_i)$, nu implică dificultăţi. In acest caz vectorul $\sum_{i=1}^{r} \beta_i \overline{\mathbf{X^{(i)}}} \rightsquigarrow N(\mu, \Sigma_*)$ unde

$$\mu = \sum_{i=1}^{r} \beta_i \mu_i,\ \Sigma_* = (\sum_{i=1}^{r} \frac{\beta_i^2}{N_i})\Sigma. \qquad (23)$$

Statistica testului este in acest caz

$$\chi^2 = (\sum_{i=1}^{r} \beta_i \overline{\mathbf{X^{(i)}}} - \mu_0)' \Sigma_*^{-1} (\sum_{i=1}^{r} \beta_i \overline{\mathbf{X^{(i)}}} - \mu_0), \qquad (23')$$

care are repartiţia χ_k^2, deci domeniul critic este de forma (19'), iar puterea testului se determină cu χ^2 necentrat cu parametrul de excentricitate

$$\delta^2 = (\mu - \mu_0)' \Sigma_*^{-1} (\mu - \mu_0). \qquad (23'')$$

H_5. **Problema simetriei.** Fie $X \rightsquigarrow N(\mu, \Sigma), \mu = (\mu_1, ..., \mu_k)'$. **Problema simetriei** constă in a verifica ipoteza $H : \mu_1 = ... = \mu_k$.

Fie $\epsilon = (1, 1, ..., 1)'$ vectorul k-dimensional cu toate componentele 1. Să considerăm o matrice $C^{k \times (k-1)}$, astfel incât $C\epsilon = \mathbf{0}$. O astfel de matrice există deoarece cele $k \times (k-1)$ elemente ale ei satisfac numai k ecuaţii. Cu aceste notaţii ipoteza H se poate scrie $H : C\mu = \mathbf{0}$. Deoarece $\overline{\mathbf{X}}$ este o estimaţie a lui μ,, rezultă că **statistica test**

$$\chi^2 = N(C\overline{\mathbf{X}})'(C\Sigma C')^{-1}(C\overline{\mathbf{X}}) \tag{24}$$

are repartiţia χ^2_{k-1} şi deci domeniul critic

$$\chi^2 \geq \chi^2_{k-1,\alpha},$$

iar puterea tesctului este

$$\pi = P(\chi^2_{k-1;\delta} \geq \chi^2_{k-1,\alpha}),\ \textit{unde}\ \delta^2 = N(C\mu)'(C\Sigma C')^{-1}(C\mu). \tag{24'}$$

Teste asupra mediilor repartiţiilor normale k-dimensionale, când matricile de covarianţă sunt necunosacute.

T_1. **Verificarea ipotezei $H : \mu = \mu_\mathbf{0}$ cu alternativa $\mathcal{N}H$: $\mu \neq \mu_\mathbf{0}$, cu Σ-necunoscut.**

Cu ajutorul selecţiei de volum N se estimează μ şi Σ astfel

$$\overline{\mathbf{X}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{X_i},\ S = \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{X_i} - \overline{\mathbf{X}})(\mathbf{X_i} - \overline{\mathbf{X}})'. \tag{25}$$

Matricea S fiind o matrice Wishart, rezultă că statistica

$$T^2 = N(\overline{\mathbf{X}} - \mu_0)'S^{-1}(\overline{\mathbf{X}} - \mu_0) \tag{26}$$

are, in ipoteza H, o repartiţie **Hoteling cu $N-1$ grade de libertate.** Deci domeniul critic de nivel α pentru verificarea ipotezei H este

$$T^2 \geq T^2_{N-1,\alpha},\ \textit{unde}\ P(T^2_{N-1} \geq T^2_{N-1,\alpha}) = \alpha.$$

Puterea testului se calculează cu ajutorul repartiţiei T^2 necentrate cu parametrul de excentricitate

$$\delta^2 = N(\mu - \mu_\mathbf{0})'\Sigma^{-1}(\mu - \mu_\mathbf{0}) \tag{26'}$$

adică

$$\pi(\mu) = P(T^2_{N-1;\delta} \geq T^2_{N-1,\alpha}). \tag{26''}$$

T_2. **Problema celor două selecţii când matricile de covarianţă sunt necunoscute si egale.** Fie $\mathbf{X^{(1)}} \rightsquigarrow N(\mu_{\mathbf{1}},\Sigma), \mathbf{X^{(2)}} \rightsquigarrow N(\mu_{\mathbf{2}}, \Sigma$ si două selecţii de volume N_1, N_2 respectiv. Se cere testarea ipotezei $H : \mu_{\mathbf{1}} = \mu_{\mathbf{2}}$ cu alternativa $\mathcal{N}H : \mu_{\mathbf{1}} \neq \mu_{\mathbf{2}}$.

Matricea de covarianţă comună se estimează cu

$$S =$$

$$\frac{1}{N_1+N_2-2}\left\{\sum_{i=1}^{N_1}(\mathbf{X_i^{(1)}} - \overline{\mathbf{X^{(1)}}})(\mathbf{X_i^{(1)}} - \overline{\mathbf{X^{(1)}}})' + \sum_{j=1}^{N_2}(\mathbf{X_j^{(2)}} - \overline{\mathbf{X^{(2)}}})(\mathbf{X_j^{(2)}} - \overline{\mathbf{X^{(2)}}})'\right\}. \tag{27)}$$

Deoarece $\overline{\mathbf{X^{(1)}}} - \overline{\mathbf{X^{(2)}}} \rightsquigarrow N(\mathbf{0}, \frac{N_1+N_2}{N_1N_2}\Sigma)$, rezultă că statistica

$$T^2 = \frac{N_1N_2}{N_1+N_2}(\overline{\mathbf{X^{(1)}}} - \overline{\mathbf{X^{(2)}}})'S^{-1}(\overline{\mathbf{X^{(1)}}} - \overline{\mathbf{X^{(2)}}}), \tag{28}$$

are repartiţia $T^2_{N_1+N_2-2}$. Atunci, domeniul critic al testului este

$$T^2 \geq T_{N_1+n_2-2,\alpha},\ unde\ T^2_{N_1+N_2-2,\alpha} = \frac{N_1+N_2-2)k}{N_1+n_2-k-1}F_{k,N_1+N_2-k-1<\alpha}, \tag{28'}$$

iar puterea testului se calculează cu T^2 necentrat cu parametrul de excentricitate

$$\delta^2 = \frac{N_1N_2}{N_1+N_2}(\mu_{\mathbf{1}} - \mu_{\mathbf{2}})'\Sigma^{-1}(\mu_{\mathbf{1}} - \mu_{\mathbf{2}}). \tag{28''}$$

T_3. **Problema celor două selecţii când matricile de covarianţă sunt necunoscute şi diferite.** Presupunem deci că se dau vectorii normali $\mathbf{X^{(1)}} \rightsquigarrow N(\mu_{\mathbf{1}}, \Sigma_1)$, $\mathbf{X^{(2)}} \rightsquigarrow N(\mu_{\mathbf{2}}, \Sigma_2)$, selecţiile corespunzătoare de volume N_1, N_2 şi se cere să se testeze ipoteza $H : \mu_{\mathbf{1}} = \mu_{\mathbf{2}} <$ cu alternativa $\mathcal{N}H : \mu_{\mathbf{1}} \neq \mu_{\mathbf{2}}$.

Dacă până acum construcţia testelor T^2 decurgea asemănător testelor t din statistica unidimensională aici construcţia presupune **un atificiu** ce va fi prezentat in continuare. Astfel să presupunem că $N_1 < N_2$. (In caz contrar schimbam notarea vectorilor normali!). Din selectiile $\mathbf{X_i^{(1)}}, 1 \leq i \leq N_1$ şi $\mathbf{X_j^{(2)}}, 1 \leq j \leq N_2$, construim o **nouă selecţie** $\mathbf{Y_i}, 1 \leq i \leq N_1$ astfel

$$\mathbf{Y_i} = \mathbf{X_i^{(1)}} - \sqrt{\frac{N_1}{N_2}} + \frac{1}{N_1 N_2} \sum_{s=1}^{N_1} \mathbf{X_s^{(1)}} - \frac{1}{N_2} \sum_{j=1}^{N_1} \mathbf{X_j^{(2)}}, 1 \leq i \leq N_1. \tag{29}$$

Se arată că valorile de selectie $\mathbf{Y_i}, 1 \leq i \leq N_1$ sunt independente stochastic şi repartizate normal

$$N(\mu_{\mathbf{1}} - \mu_{\mathbf{2}}, \Sigma),\ unde\ \Sigma = \Sigma_1 + \frac{N_1}{N_2}\Sigma_2. \tag{30}$$

Matricea Σ se estimează cu

$$S = \frac{1}{N_1 - 1} \sum_{j=1}^{N_1} (\mathbf{Y_j} - \overline{\mathbf{Y}})(\mathbf{Y_j} - \overline{\mathbf{Y}})' \tag{30'}$$

iar in ipoteza H statistica

$$T^2 = N_1 \overline{\mathbf{Y}}' S^{-1} \overline{\mathbf{Y}} \tag{31}$$

are $N_1 - 1$ grade de libertate. Domeniul critic al testului este in acest caz

$$T^2 \geq T^2_{N_1-1,\alpha}, \tag{31'}$$

iar puterea testului este

$$\pi(\mu_{\mathbf{1}} - \mu_{\mathbf{2}}) = P(T^2_{N_1-1;\delta} \geq T^2_{N_1-1,\alpha}); \delta^2 = N_1(\mu_{\mathbf{1}} - \mu_{\mathbf{2}})'\Sigma^{-1}(\mu_{\mathbf{1}} - \mu_{\mathbf{2}}). \tag{31''}$$

T_4. **Problema celor r selecţii când matricile de covarianţă sunt necunoscute şi egale.**Problema se tratează in paralel cu cazul $\mathbf{H_4}$. Fie $\mathbf{X}_\alpha^{(\mathbf{i})}, 1 \leq i \leq r, 1 \leq \alpha \leq N_i$ cele r selecţii, selecţia $\mathbf{X}_\alpha^{(\mathbf{i})}$ fiind efectuată asupra populeţiei normale $N(\mu_\mathbf{i}, \Sigma)$. Se cere testarea ipotezei $H : \mu = \sum_{i=1}^{r} \beta_i \mu_\mathbf{i} = \mu_\mathbf{0}$, cu alternativa $\mathcal{N}H : \mu \neq \mu_\mathbf{0}$. Matricea Σ se estimează in mod obişnuit adică

$$S = \frac{1}{\sum_{i=1}^{r} N_i - r} \sum_{i=1}^{r} \sum_{\alpha=1}^{n_i} (\mathbf{X}_\alpha^{(\mathbf{i})} - \overline{\mathbf{X}^{(\mathbf{i})}})(\mathbf{X}_\alpha^{(\mathbf{i})} - \overline{\mathbf{X}^{(\mathbf{i})}})', \qquad (32)$$

şi deoarece $\sum_{i=1}^{r} \beta_i \overline{\mathbf{X}^{(\mathbf{i})}} - \mu_\mathbf{0} \rightsquigarrow N(\mathbf{0}, \Sigma_*)$ unde

$$\Sigma_* = \left(\sum_{i=1}^{r} \frac{\beta_i^2}{N_i} \right) \Sigma,$$

rezultă statistica test

$$T^2 = \left(\sum_{i=1}^{r} \frac{\beta_i^2}{N_i} \right) (\sum_{i=1}^{r} \beta_i \overline{\mathbf{X}^{(\mathbf{i})}} - \mu_\mathbf{0})' S^{-1} (\sum_{i=1}^{r} \beta_i \overline{\mathbf{X}^{(\mathbf{i})}} - \mu_\mathbf{0}) \qquad (32')$$

care are $f = \sum_{i=1}^{r} N_i - k$ grade de libertate. Domeniul critic este deci

$$T^2 \geq T^2_{f,\alpha}, \ T^2 \ dat \ de \ (32'), \qquad (33)$$

iar puterea testului se calculează cu T^2 necentrat adică

$$\pi(\mu) = P(T_{f;\delta} \geq T^2_{f,\alpha}), \qquad (33')$$

cu

$$\delta^2 = \left(\sum_{i=1}^{r} \frac{\beta_i^2}{N_i} \right)^{-1} (\sum_{i=1}^{r} \beta_i \overline{\mu_\mathbf{i}} - \mu_\mathbf{0})' \Sigma^{-1} \sum_{i=1}^{r} \beta_i \mu_\mathbf{i} - \mu_\mathbf{0}). \qquad (33'')$$

T_5. **Problema celor r selecţii, cazul general.**

Presupunem că se dau r selecţii $\mathbf{X}_\alpha^{(\mathbf{i})}, 1 \le i \le r, 1 \le \alpha \le N_i$ din populaţiile normale independente $N(\mu_{\mathbf{i}}, \Sigma_i), 1 \le i \le r$, cu Σ_i necunoscute si ne egale. Se cere să se verifice ipoteza $H : \mu = \sum_{i=1}^{r} \beta_i \mu_{\mathbf{i}} = \mu_{\mathbf{0}}$ cu alternativa $\mathcal{N}H : \mu \ne \mu_{\mathbf{0}}$, unde $\mu_{\mathbf{0}}$ şi coeficienţii $\beta_i, 1 \le i \le r$ sunt daţi.

Şi aici se aplică un artificiu asemănător celui din cazul T_3. Presupunem că $N_1 = \min_{1 \le i \le r} N_i$. (In caz contrar schimbăm numerotarea astfel incât N_1 să fie cel mai mic). Construim selecţia

$$\mathbf{Y}_\alpha = \beta_1 \mathbf{X}_\alpha^{(\mathbf{1})} + \sum_{i=1}^{r} \sqrt{\frac{N_1}{N_i}} \left(\mathbf{X}_\alpha^{(\mathbf{i})} - \frac{1}{N_i} \sum_{\nu=1}^{N_1} \mathbf{X}_\nu^{(\mathbf{i})} + \frac{1}{\sqrt{N_1 N_i}} \sum_{\gamma=1}^{N_i} \mathbf{X}_\gamma^{(\mathbf{i})} \right). \tag{34}$$

Se arată că variabilele de selecţie $\mathbf{Y}_\alpha$ sunt independente stochastic şi repartizate normal

$$N(\mu, \Sigma_*),\ \Sigma_* = \sum_{i=1}^{r} \frac{\beta_i^2 N_1}{N_i} \Sigma_i. \tag{35}$$

Fie estimaţia lui Σ_*

$$S = \frac{1}{N_1 - 1} \sum_{i=1}^{N_1} (\mathbf{Y}_\alpha - \overline{\mathbf{Y}})(\mathbf{Y}_\alpha - \overline{\mathbf{Y}})'. \tag{35'}$$

Dacă notăm cu S_i estimatia lui $\Sigma_i, 1 \le i \le r$ se arată că

$$S = \sum_{i=1}^{r} \frac{\beta_i^2 N_1}{N_i} S_i. \tag{35''}$$

Statistica test pentru ipoteza H este

$$T^2 = N_1 (\overline{\mathbf{Y}} - \mu_{\mathbf{0}})' S^{-1} (\overline{\mathbf{Y}} - \mu_{\mathbf{0}}) \tag{36}$$

şi ea are (in ipoteza H) repartiţia Hoteling cu $N_1 - 1$ grade de libertate. De aici rezultă că domeniul critic al testului este

$$T^2 \geq T_{N_1-1,\alpha}, \tag{36}$$

iar puterea testului este

$$\pi(\mu) = P(T^2_{N_1-1;\delta} \geq T_{N_1-1,\alpha}),\ \delta^2 = N_1(\mu - \mu_{\mathbf{0}})'\Sigma_*^{-1}(\mu - \mu_{\mathbf{0}}). \tag{36$'$}$$

T_6. **Problema simetriei când Σ este necunoscut.** Se dă deci selecţia $\mathbf{X}_\alpha, 1 \leq \alpha \leq N$ asupra unei populaţii normale $N(\mu, \Sigma)$, $\mu = (\mu_1, ..., \mu_k)'$ cu Σ necunoscut. Se cere sa se testeze ipoteza $H : \mu_1 = ... = \mu_k$ cu alternativa $\mathcal{N}H$ care inseamnă că nu toate μ_i sunt egale. Ca şi in cazul H_5, se alege matricea $C^{k\times(k-1)}$ astfel incât

$$C\epsilon = 0,\ \epsilon = (1, 1, ..., 1)'$$

iar ipoteza H este echivalentă cu $C\mu = 0$. Dacă considerăm estimaţia obişnuită S a lui Σ,şi estimaţia $\overline{\mathbf{X}}$ a lui μ, atunci rezultă că $C\overline{\mathbf{X}} \rightsquigarrow N(C\mu, C\Sigma C')$ şi deci statistica

$$T^2 = N(C\overline{\mathbf{X}})'(CSC')^{-1}(C\overline{\mathbf{X}}) \tag{37}$$

are repartiţia Hoteling cu $N - 1$ grade de libertate (pe spaţiul $k - 1$ dimensional!).Domeniul critic al testului este

$$T^2 \geq T^2_{N-1,\alpha}, T^2_{N-1,\alpha} = \frac{(N-1)(k-1)}{(N-k)(k-1)} F_{k-1,N-k,\alpha}. \tag{37$'$}$$

Puterea testului se calculează cu variabila Hoteling necentrată (pe spaţiul $k - 1$ dimensional) şi anume

$$\pi(\mu) = P(T^2_{N-1;\delta} \geq T^2_{N-1,\alpha}),\ \delta^2 = N(C\mu)'(C\Sigma C')^{-1}(C\mu). \tag{37$''$}$$

Consideraţii finale.

1. Aici s-au prezentat numai consideraţii introductive privind verificarea ipotezelor statistice. Probleme ca: **verificarea ipotezelor folosind selecţiile cenzurate ce intervin in fiabilitate, etc; testele secvenţiale; analiza dispersională;teste bazate pe statistici de ordine; teste pentru serii dinamice cu multiple aplicaţii in activităţi bancare;etc** am coniderat că-şi au locul in prezentari speciale separate.

2. Pentru aplicarea testelor prezentate, se impun unele precizări legate de utilizarea tehnicilor moderne de calcul.

2.1. Toate funcţiile de repartiţie pot fi calculate cu pachetele de programe statistice existente. Astfel se pot determina atât cuantilele cat şi valorile acestor funcţii. Este o preoblema insă cu utilizarea funcţiilor de repartiţie ne centrate. Deoarece expresiile densităţilor de repartiţie ale lui t-necentrat, χ^2-necentrat şi F-necentrat sunt date de serii de puteri, folosirea acestor expresii la calculul numeric al functiilor de repartiţie sau al cuantilelor (când trebuie rezolvată o ecuaţie in x de forma $F(x) = p$), este complicată. O ieşire din impas o poate reprezenta **aproximarea lui Pathnaik** pentru repartiţia $\chi^2_{k;\delta}$ si anume: se aproximează repartiţia acestei variabile cu o variabilă repartiţie de forma $c\chi^2_{k^*}$, adică

$$\chi^2_{k;\delta} = c\chi^2_{k^*}. \tag{38}$$

Egalând mediile şi dispersiile celor două variabile din (38) rezultă

$$k + \delta^2 = ck^*, \;\; k + 2\delta^2 = c^2k^*, \tag{38'}$$

de unde

$$c = \frac{k + 2\delta^2}{k + \delta^2}, \; k^* = \frac{(k + \delta^2)^2}{k + 2\delta^2}. \tag{38''}$$

Soluţia k^* din (38") se **rotunjeşte la un intreg.** Pentru utilizarea repartiţiilor F şi T^2 necentrate se poate utiliza in prealabil aproximarea repartiţiei χ^2 necentrată ce intră in definiţia lui F necentrată. Trebuie subliniat faptul că aproximarea Pathnaik este ne recomandată, fiind prea laxă.

2.2. Simularea Monte Carlo oferă o alternativă facilă şi mai bună pentru determinarea puterii testului in cazul unei repartiţii necentrate (sau oricărei alte repartiţii) şi anume:

-in ipoteza $\mathcal{N}H$, se simulează o selecţie de volum mare n, a statisticii test g:

- cu această selecţie se determină estimaţia puterii testului

$$\pi_n = 1 - F_n(x_\alpha) \approx P(g > x_\alpha),$$

unde x_α este valoarea critică a statisticii test.

Dacă nu se poate utiliza uşor sau **nu se cunoaşte** expresia convenabilă a repartiţiei statisticii test g, atunci se procedează in mod asemănător,adică:

- se simulează o selecţie de volum mare n, a statisticii test g in ipoteza H;
- Se construieţe histograma lui g pe baza acestei selecţii;
- cu ajutorul histogramei se rezovă ecuaţia

$$P(g > x_\alpha) = \alpha,$$

unde α este riscul de genul intâi, x_α fiind valoarea critica a statisticii test.(Problema inversă celei precedente).

Când selecţiile de care dispunem au un volum mic, se poate folosi **metoda bootstrap** de re-selectie, care produce multe replici ale selecţiei iniţiale, ce pot permite o abordare **asimptotica** a analizei statistice a datelor originale ale selecţiei.

3. Verificarea ipotezei de normalitate unidimensională, nu ridică nicio problemă. Nu s-a menţionat ceva semnificativ privind **verificarea ipotezei de normalitate multidimensională.** In acest sens, recomandăm lucrările [3,4] de la bibliografie care prezintă **adaptarea** testului de concordanţă χ^2 in acest caz.

3.1. Cazul specificat. S-a văzut că testul χ^2 presupune ca spaţiul $R^p = \Delta$,care reprezintă mulţimea valorilor vectorului p-dimensional $\mathbf{X} \rightsquigarrow N(\mu, \Sigma)$, sa fie divizat in k părţi disjuncte, făra **a se impune** cum se alege diviziunea. In lucrările menţionate se pleacă de la ideea că forma pătratică

$$\Phi(\mathbf{x}) = (\mathbf{x} - \mu)'\Sigma^{-1}(\mathbf{x} - \mu)$$

permite divizarea spaţiului R^p in **coroane de elipsoizi,** determinate de $k-1$ constante $0 < \theta_0 < ... < \theta_{k-1}$. Astfel spaţiul R^p se divide in k mulţimi disjuncte de forma

$$\Delta_1 = \{\mathbf{x}|0 \leq \Phi(\mathbf{x}) \leq \theta_1\}, \Delta_i = \{\mathbf{x}|\theta_{i-1} < \Phi(\mathbf{x}) \leq \theta_i, 2 \leq i \leq k-1\},$$

$$\Delta_k = \{\mathbf{x}|\Phi(\mathbf{x}) > \theta_{k-1}. \tag{39}$$

Deoarece

$$(\mathbf{X} - \mu)'\Sigma^{-1}(\mathbf{X} - \mu) = \chi^2_p, \tag{40}$$

adica are o repartiţie χ^2_p, rezultă că

$$p_1 = P(\mathbf{X} \in \Delta_1) = P(\chi^2_p \leq \theta_1), p_i = P(\mathbf{X} \in \Delta_i) =$$

$$= P(\theta_{i-1} < \chi^2_p \leq \theta_i), 2 \leq i \leq k-1), p_k = P(\mathbf{X} \in \Delta_k) = P(\chi^2_p > \theta_{k-1}). \tag{40'}$$

Frecvenţele f_i care intervin in testul de concordanţă χ^2 se calculează simpu, numărând valorile de selecţie ce cad in $\Delta_i, 1 \leq i \leq k$.

3.2. Cazul nespecificat. In acest caz, construcţia statisticii testului de concordanţă χ^2 se realizeată in următorii paşi (pentru selecţia $\mathbf{X_1}, ..., \mathbf{X_n}$ de volum n mare):

-se separă o (sub)selecţie de volum $n_1 < n$;

- cu acestă selecţie se estimează parametri μ şi Σ cu formulele obişnuite (adica $\mu \approx \overline{\mathbf{X}}, \Sigma \approx S$); se observă că variabilele

$$(\mathbf{X_i} - \overline{\mathbf{X}})' S^{-1} (\mathbf{X_i} - \overline{X_i}), n_1 < i \leq n$$

sunt repartizate $T^2_{n_1 - 1}$ pe spaţiul R^p.

- construcţia continuă ca in cazul specificat, elipsoizii fiind de acelaş tip, dar probabilităţile teoretice p_i se calculează cu repartiţia $T^2_{n_1-1}$ pe R^p in loc de χ^2_p.

References

[1] Gheorghe MIHOC, Virgil CRAIU.(1977).**Tratat de statistică matematică, Vol.II. Verificarea ipotezelor statistice,** *Editura Academiei.*

[2] Ion VADUVA. (1970). **Analiză dispersională.** *Editura Tehnică.*

[3] Ion VĂDUVA and Nicolae POPOVICIU.(1979)."χ^2 test of goodness of fit for multivariate normal distribution. Specified case". *Econ.Comp.Econ.Cyb.St. and Res.,No. 2, 1979,p.93-109.*

[4] Ion VĂDUVA and Nicolae POPOVICIU.(1980)."χ^2 test of goodness of fit for multivariate normal distribution.Unspecified case". *Econ.Comp.Econ.Cyb.St. and res., No 1, 1980,p.33-42.*