

<p>Intervale de încredere pentru media unei populații cu dispersia cunoscută (σ^2)</p> $(\bar{x}_n - z^* \cdot \frac{\sigma}{\sqrt{n}}, \bar{x}_n + z^* \cdot \frac{\sigma}{\sqrt{n}})$ $z^* = -qnorm\left(\frac{\alpha}{2}\right) = qnorm\left(1 - \frac{\alpha}{2}\right)$ <p>\bar{x}_n – media de selecție $\alpha = 1 - \text{confidence_level}$</p>	<p>Intervale de încredere pentru media unei populații cu dispersia necunoscută (s)</p> $(\bar{x}_n - t^* \cdot \frac{s}{\sqrt{n}}, \bar{x}_n + t^* \cdot \frac{s}{\sqrt{n}})$ $t^* = -qt\left(\frac{\alpha}{2}, n - 1\right) = qt\left(1 - \frac{\alpha}{2}, n - 1\right)$ <p>\bar{x}_n – media de selecție $\alpha = 1 - \text{confidence_level}$</p>
<p>Testul z asupra proporțiilor</p> <p>$H_0: p = p_0$, p ia o val particulară $H_a: p < p_0 \Rightarrow z^* = qnorm(\alpha)$ (asim st) $H_a: p > p_0 \Rightarrow z^* = qnorm(1 - \alpha)$ (asim dr) $H_a: p \neq p_0 \Rightarrow z^* = -qnorm\left(\frac{\alpha}{2}\right) = qnorm\left(1 - \frac{\alpha}{2}\right)$ (sim)</p> <p>Scorul testului $z = \frac{p' - p_0}{\sqrt{\frac{p_0 \cdot (1 - p_0)}{n}}}$, $p' = \frac{X}{n}$</p> <p>Ipoteza nulă H_0 este respinsă dacă: - $z < z^*$ pt ipoteză H_a asim la st - $z > z^*$ pt ipoteză H_a asim la dr - $z > z^*$ pt ipoteză H_a sim</p>	<p>Testul Z pentru media unei populații cu dispersia cunoscută (σ^2)</p> <p>$H_0: \mu = \mu_0$, μ ia o val particulară $H_a: \mu < \mu_0 \Rightarrow z^* = qnorm(\alpha)$ (asim st) $H_a: \mu > \mu_0 \Rightarrow z^* = qnorm(1 - \alpha)$ (asim dr) $H_a: \mu \neq \mu_0 \Rightarrow z^* = -qnorm\left(\frac{\alpha}{2}\right) = qnorm\left(1 - \frac{\alpha}{2}\right)$ (sim)</p> <p>Scorul testului $z = \frac{\bar{x}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}}$</p> <p>Ipoteza nulă H_0 este respinsă dacă: -Vezi test z asupra proporțiilor</p>
<p>Testul t pentru media unei populații cu dispersia necunoscută (s)</p> <p>$H_0: \mu = \mu_0$, μ ia o val particulară $H_a: \mu < \mu_0 \Rightarrow t^* = qt(\alpha, n - 1)$ (asim st) $H_a: \mu > \mu_0 \Rightarrow z^* = qt(1 - \alpha, n - 1)$ (asim dr) $H_a: \mu \neq \mu_0 \Rightarrow z^* = -qt\left(\frac{\alpha}{2}, n - 1\right) = qt\left(1 - \frac{\alpha}{2}, n - 1\right)$ (sim)</p> <p>Scorul testului $t = \frac{\bar{x}_n - \mu_0}{\frac{s}{\sqrt{n}}}$</p> <p>Ipoteza nulă H_0 este respinsă dacă: - $t < t^*$ pt ipoteză H_a asim la st - $t > t^*$ pt ipoteză H_a asim la dr - $t > t^*$ pt ipoteză H_a sim</p>	<p>Testul Z pentru diferența mediilor unor populații cu dispersii cunoscute</p> <p>$H_0: \mu_1 - \mu_2 = m_0$, m_0 ia o val particulară $H_a: \mu_1 - \mu_2 < m_0 \Rightarrow z^* = qnorm(\alpha)$ (asim st) $H_a: \mu_1 - \mu_2 > m_0 \Rightarrow z^* = qnorm(1 - \alpha)$ (asim dr) $H_a: \mu_1 - \mu_2 \neq m_0 \Rightarrow z^* = qnorm\left(1 - \frac{\alpha}{2}\right)$ (sim)</p> <p>Scorul testului $z = \frac{(\bar{x}_{n_1} - \bar{x}_{n_2}) - m_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$</p> <p>Ipoteza nulă H_0 este respinsă dacă: -Vezi test z asupra proporțiilor</p>
<p>Testul F – inferență asupra dispersiilor a două populații</p> <p>$H_0: \frac{\sigma_1}{\sigma_2} = 1$, dispersii egale $H_a: \frac{\sigma_1}{\sigma_2} > 1 \Rightarrow F^* = qf(1 - \alpha, n_1 - 1, n_2 - 1)$ (asim dr) $H_a: \frac{\sigma_1}{\sigma_2} \neq 1 \Rightarrow F_s^* = qf\left(\frac{\alpha}{2}, n_1 - 1, n_2 - 1\right)$, $F_d^* = qf\left(1 - \frac{\alpha}{2}, n_1 - 1, n_2 - 1\right)$ (sim)</p> <p>Scorul testului $F = \frac{s_1^2}{s_2^2}$</p> <p>Ipoteza nulă H_0 este respinsă dacă: - $F > F^*$ pt ipoteză H_a asim la dr - $F < F_s^*$ sau $F > F_d^*$ pt ipoteză H_a sim</p>	<p>Testul T pentru diferența mediilor unor populații cu dispersii necunoscute</p> <p>Se folosește mai întâi testul F pentru a decide dacă dispersiile celor două populații sunt:</p> <ol style="list-style-type: none"> Diferite Egale <p>$H_0: \mu_1 - \mu_2 = m_0$, m_0 ia o val particulară $H_a: \mu_1 - \mu_2 < m_0 \Rightarrow t^* = qt(\alpha, df)$ (asim st) $H_a: \mu_1 - \mu_2 > m_0 \Rightarrow t^* = qt(1 - \alpha, df)$ (asim dr) $H_a: \mu_1 - \mu_2 \neq m_0 \Rightarrow t^* = -qt\left(\frac{\alpha}{2}, df\right) = qt\left(1 - \frac{\alpha}{2}, df\right)$ (sim)</p> <p>Scorul testului:</p> <p>a. $z = \frac{(\bar{x}_{n_1} - \bar{x}_{n_2}) - m_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$, $df = \min(n_1 - 1, n_2 - 1)$</p> <p>b. $z = \frac{(\bar{x}_{n_1} - \bar{x}_{n_2}) - m_0}{\sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}}$, $sf = n_1 + n_2 - 2$, $s = \frac{(n_1 - 1) \cdot s_1^2 + (n_2 - 1) \cdot s_2^2}{sf}$</p> <p>Ipoteza nulă H_0 este respinsă dacă: -Vezi testul t</p>

	H_0 nu este respinsă	H_0 este respinsă
H_0 este adevărată	Corect	Eroare de tip I
H_0 este falsă	Eroare de tip II	Corect

Eroarea marginală: $z * \frac{\sigma}{\sqrt{n}}$	Lungimea unui interval de încredere pentru medie este $2 * \text{eroarea marginală}$.
Dacă dorim o lungime anume pentru acest interval, ω , atunci ne trebuie un eșantion de dimensiune $n = \frac{(2 * z * \sigma)^2}{\omega^2}$	Pe măsură ce n crește, lungimea intervalului și eroarea marginală scad.

Media populației = media de selecție = $\bar{x}_n = \frac{x_1 + x_2 + \dots + x_n}{n} = \mu$ (estimator nedepășat)	Media geometrică G = $\sqrt[n]{x_1 * x_2 * \dots * x_n}$ Media armonică H = $\frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}$
Mediana (Me) este valoarea din mijloc când datele din eșantion sunt sortate.	Modul este observația/valoarea cea mai frecventă din eșantion. Antimodul este opusul modului.
În cazul datelor grupate, se alege mai întâi clasa cu cea mai mare frecvență, clasa modală. Atunci $\text{Mod} = a_i + \frac{L * (f_i - f_{i-1})}{(f_i - f_{i-1}) + (f_i - f_{i+1})}$, L – lungimea comună a intervalelor, i – indexul acestei clase, a_i – marginea stângă.	Cvartilele sunt valori care împart domeniul (ordonat al) observațiilor în patru segmente egale: Q1, mediana, Q3. Cvartila medie este valoarea de mijloc dintre prima și a treia cvartilă: $\text{midq} = \frac{Q_1 + Q_3}{2}$ Domeniul intercvartilic este diferența dintre prima și a treia cvartilă: $\text{IQR} = Q_3 - Q_1$
Domeniul este diferența dintre cea mai mică și cea mai mare valoare din eșantion: $\text{range} = \text{max} - \text{min}$	Sumarul celor 5 numere: min, Q1, Me, Q3, max.
Dispersia eșantionului , s^2 , este: $s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x}_n)^2}{n-1}$ sau $s^2 = \frac{n * \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}{n * (n-1)}$, n – dimensiunea eșantionului.	Dispersia eșantionului este statistica asociată dispersiei populației , notată σ^2 . Deviația standard a eșantionului, s , este rădăcina pătrată a dispersiei eșantionului.
Valorile aberante nu aparțin intervalului: a. $(\bar{x}_n - 2 * s, \bar{x}_n + 2 * s)$ b. $(Q1 - 1.5 * \text{IQR}, Q3 + 1.5 * \text{IQR})$	Corelația este o metodă destinată studiului relației dintre două variabile, fiind parte a statisticii bivariate.
Coeficientul de corelație Pearson pentru două variabile aleatoare X și Y este $p[X, Y] = \frac{\text{cov}[X, Y]}{\sigma_X * \sigma_Y}$, σ_X^2, σ_Y^2 - dispersiile celor două variabile. Pentru două eșantioane aleatoare $x = \{x_1, x_2, \dots, x_n\}$ și $y = \{y_1, y_2, \dots, y_n\}$ este: $r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x}_n) * (y_i - \bar{y}_n)}{\sqrt{[\sum_{i=1}^n (x_i - \bar{x}_n)^2] * [\sum_{i=1}^n (y_i - \bar{y}_n)^2]}}$	Corelația este: a. Pozitivă dacă valorile sunt apropiate de prima bisectoare, în cadranele 1 și 3; b. Negativă dacă valorile sunt majoritare în cadranele 2 și 4; c. Aproape de zero dacă valorile sunt distribuite uniform în toate cadranele.
Linia deviației standard trece prin punctul mediilor și prin toate punctele care se află la un număr egal de deviații standard față de medie. Ecuația dreptei SD: $y - \bar{y}_n = m * (x - \bar{x}_n)$, $m = \pm \frac{s_Y}{s_X}$, în funcție de semnul corelației.	Împrăștierea de-a lungul liniei SD este aprox. $\sqrt{2 * (1 - r)} * s_Y$ pe verticală și $\sqrt{2 * (1 - r)} * s_X$ pe orizontală. Asocierea pozitivă/negativă perfectă, $r = \pm 1$, corespunde situației când între cele două variabile există o dependență liniară cu panta pozitivă/negativă: $Y = m * X + n$
Regresia încearcă să descrie cum una dintre variabile depinde de cealaltă. Unei creșteri cu o st dev a lui x îi corespunde o creștere de r st dev a lui y , în medie.	Linia de regresie: $y = p * x + q$, unde $p = r * \frac{s_Y}{s_X}$, $q = \bar{y}_n - \bar{x}_n * r$
Funcția de repartiție: $F: R \rightarrow [0, 1], F(a) = P(X \leq a)$	Funcția de densitate (de masă) , $f: R \rightarrow [0, \infty)$, astfel încât funcția de repartiție F poate fi descrisă astfel: $F(a) = P(X \leq a) = \int_{-\infty}^a f(t) dt$. Orice funcție $f: R \rightarrow [0, \infty)$ cu $\int_{-\infty}^{\infty} f(t) dt = 1$ este funcția de densitate pentru o anumită variabilă aleatoare continuă.
Media folosind funcția de densitate: $M(X) = \int_{-\infty}^{\infty} t f(t) dt$	Dispersia folosind funcția de densitate: $D^2(X) = \int_{-\infty}^{\infty} [t - M(X)]^2 f(t) dt$
Probabilitățile asociate unei variabile aleatoare continue: $P(a < X \leq b) = F(b) - F(a) = \int_a^b f(t) dt$. Dacă F este continuă: $P(X = a) = F(a) = 0$ și $P(a \leq X < b) = P(a < X \leq b) = P(a < X < b)$.	Standardizarea unei variabile aleatoare $X: \Omega \rightarrow R$ este $Y = \frac{X - M[X]}{D[X]}$. Standard pentru că $M[Y] = 0$ și $D^2[Y] = 1$.
Inegalitatea lui Markov: $X \geq 0$ variabilă aleatoare. Dacă $a > 0$, atunci $P(X \geq a) \leq \frac{M[X]}{a}$.	Inegalitatea lui Cebășev: X variabilă aleatoare cu media μ și dispersia σ^2 . Atunci $P(X - \mu \geq k) \leq \frac{\sigma^2}{k^2}$
Legea numerelor mari: spune că pe măsură ce crește numărul de variabile independente, identic distribuite, media lor de selecție se apropie de media lor comună.	Teorema limită centrală: pentru eșantioane suficient de mari ($n \geq 30$), variabila $\frac{\sum_{i=1}^n X_i - n\mu}{\sigma / \sqrt{n}}$ urmează o lege normală standard, $N(0, 1)$.

--	--