Metode numerice și statistică

Curs 13

lect. Ciprian Deliu
cdeliu@tuiasi.ro
moodle.deliu.ro

Universitatea Tehnică "Gh. Asachi" Iași
Facultatea de Hidrotehnică, Geodezie și Ingineria Mediului

2019

Pentru testarea unei ipoteze statistice se parcurg următoarele etape:

1. Se alege ipoteza nulă H_0 și ipoteza alternativă H_a
2. Se alege un criteriu de testare $C(X,n)$
3. Se alege un nivel de semnificație α
4. Se determină regiunea de respingere W cu proprietatea că
$$P(C(X,n) \in W | H_0 \text{ este adevărată}) = \alpha \ (\text{sau } \le \alpha)$$
5. Se face o selecție de volum n, obținându-se valorile $x_1, x_2, \ldots, x_n$ ale variabilelor aleatoare de selecție $X_1, X_2, \ldots, X_n$ corespunzătoare caracteristicii studiate
6. Se calculează $C(x_1, x_2, \ldots, x_n)$:
   - Dacă $C(x_1, x_2, \ldots, x_n) \in \bar{W}$ atunci H_0 este acceptată
   - Dacă $C(x_1, x_2, \ldots, x_n) \in W$ atunci H_0 este respinsă (H_a este acceptată)

Test bilateral pentru medie (σ cunoscut)

Fie X o variabilă aleatoare teoretică cu repartiție normală $N(m, \sigma^2)$ cu media necunoscută și m_0 o estimație a acestei medii. Se face o selecție de volum n, obținându-se variabilele aleatoare de selecție $X_1, X_2, \ldots, X_n$.

1. Dorim să verificăm ipoteza
$$H_0 : \ m = m_0$$
față de alternativa
$$H_a : \ m \neq m_0.$$
2. Alegem criteriul de testare
$$C(X,n) = \frac{\bar{X} - m_0}{\frac{\sigma}{\sqrt{n}}}$$
deoarece dacă H_0 este adevărată, atunci $C(X,n)$ are repartiția complet specificată $N(0,1)$.
3. Alegem un nivel de semnificație α (de obicei 0.1, 0.05 sau 0.01).

4. Se determină $z =$ cuantila de ordin $1 - \frac{\alpha}{2}$ a repartiției normale standard, corespunzătoare nivelului de semnificație ales α. Avem
$$P(|C(X,n)| > z | m = m_0) = \alpha$$
deci regiunea de respingere este
$$W = (-\infty, -z) \cup (z, \infty).$$
5. Se face o selecție de volum n, obținându-se valorile $x_1, x_2, \ldots, x_n$ și se calculează
$$\bar{X} = \frac{x_1 + x_2 + \cdots + x_n}{n}.$$
6. Dacă $\dfrac{\bar{X} - m_0}{\frac{\sigma}{\sqrt{n}}} \in [-z, z]$ atunci H_0 este acceptată.
Dacă $\dfrac{\bar{X} - m_0}{\frac{\sigma}{\sqrt{n}}} \in (-\infty, -z) \cup (z, \infty)$ atunci H_0 este respinsă și se acceptă H_a.

Exemplu

La o fabrică patronul asigură că durata medie de funcționare a unui produs este de 800 ore cu abaterea medie pătratică de 50 ore. Pe un eșantion de 20 bucăți s-a obținut că durata medie de funcționare este 790 ore. Presupunând că durata de funcționare este normal distribuită să se verifice cu nivelul de semnificație de 4% dacă afirmația patronului este adevărată.

1. ipotezele $H_0 : \ m = 800, \qquad H_a : m \neq 800$
2. criteriul de testare $\frac{\bar{X} - 800}{\frac{50}{\sqrt{20}}}$
3. nivelul de semnificație $\alpha = 0.04$
4. cuantila de ordin $1 - \frac{\alpha}{2} = 0.98$ este $z = 2.05$
5. pentru $\bar{X} = 790$ avem $\frac{\bar{X} - 800}{\frac{50}{\sqrt{20}}} = -0.894$
6. ipoteza H_0 se acceptă deoarece $-0.894 \in [-2.05, 2.05]$

Teste unilaterale pentru medie (σ cunoscut)

Test unilateral dreapta:

- $H_0 : \ m = m_0, \qquad H_a : \ m > m_0$
- criteriul de testare $C(X,n) = \frac{\bar{X} - m_0}{\frac{\sigma}{\sqrt{n}}}$
- $z =$ cuantila de ordin $1 - \alpha$ a repartiției normale standard, corespunzătoare nivelului de semnificație ales α.
- regiunea de respingere este $W = (z, \infty)$.

Test unilateral stânga:

- $H_0 : \ m = m_0, \qquad H_a : \ m < m_0$
- criteriul de testare $C(X,n) = \frac{\bar{X} - m_0}{\frac{\sigma}{\sqrt{n}}}$
- $z =$ cuantila de ordin α a repartiției normale standard, corespunzătoare nivelului de semnificație ales α.
- regiunea de respingere este $W = (-\infty, z)$.

Exerciții

1. Procentul de grăsime al laptelui este o variabilă aleatoare normal distribuită cu $\sigma^2 = 0.09$. Se face un studiu statistic pe un eșantion de volum $n = 10$ și se obțin rezultatele următoare (în procente): 2.05, 2.1, 2.5, 3, 3.5, 2.75, 2.25, 2.85, 2.9, 3.4. Să se verifice cu un nivel de semnificație $\alpha = 0.01$ dacă procentul mediu de grăsime al laptelui este mai mic decât 2.5.
2. Într-o întreprindere consumul de timp pe unitatea de produs (în minute) este normal distribuit $N(20,3)$. Determinările făcute pe un număr de 25 muncitori dintr-o secție au condus la un consum mediu de timp pe unitatea de produs de 21 minute. Cu un nivel de semnificație $\alpha = 0.05$, acest rezultat indică faptul că muncitorii din secția respectivă au un consum mediu de timp pe unitatea de produs mai mare decât muncitorii din celelalte secții?

Teste pentru proporție

Considerăm problema testării ipotezei că proporția de indivizi dintr-o populație care au o anumită caracteristică are o valoare specificată. Considerăm variabila aleatoare teoretică Bernoulli $X : \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$, unde 1 înseamnă că individul are caracteristica respectivă. Media de selecție corespunzătoare $\hat{p} = \frac{X_1 + X_2 + \cdots + X_n}{n}$ este un estimator nedeplasat pentru proporția p și pentru valori mari ale lui n are repartiția $N\left(p, \frac{pq}{n}\right)$ unde $q = 1 - p$.

1. Se formulează ipotezele $H_0 : \ p = p_0, \qquad H_a : \ p \neq p_0$
2. Se alege criteriul de testare $C(X,n) = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$
3. Se alege un nivel de semnificație α
4. Se determină $z =$ cuantila de ordin $1 - \frac{\alpha}{2}$ a repartiției normale standard, corespunzătoare nivelului de semnificație ales α
5. Se află estimația punctuală a proporției $\hat{p}$ pe un eșantion de volum n
6. Dacă $\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \in [-z, z]$ atunci H_0 este acceptată.
Dacă $\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \in (-\infty, -z) \cup (z, \infty)$ atunci H_0 este respinsă.

Exemplu

Se studiază proporția produselor defecte dintr-un proces tehnologic. Pe un eșantion de 300 produse testate s-au obținut 13 defecte. Aceste date ne permit să afirmăm cu un nivel de semnificație $\alpha = 0.05$ că 5% din produse sunt defecte?

1. ipotezele $H_0 : \ p = 0.05, \qquad H_a : p \neq 0.05$
2. criteriul de testare $\frac{\hat{p} - 0.05}{\sqrt{\frac{0.05 \cdot 0.95}{300}}}$
3. nivelul de semnificație $\alpha = 0.05$
4. cuantila de ordin $1 - \frac{\alpha}{2} = 0.975$ este $z = 1.96$
5. pentru $\hat{p} = \frac{13}{300}$ avem $\frac{\hat{p} - 0.05}{\sqrt{\frac{0.05 \cdot 0.95}{13}}} = -0.5298$
6. ipoteza H_0 se acceptă deoarece $-0.5298 \in [-1.96, 1.96]$

- Pentru test unilateral dreapta, se folosește $z =$ cuantila de ordin $1 - \alpha$ a repartiției normale standard, iar regiunea de respingere este $W = (z, \infty)$
- Pentru test unilateral stânga, se folosește $z =$ cuantila de ordin α a repartiției normale standard, iar regiunea de respingere este $W = (-\infty, z)$

Exerciții

1. Consiliul de administrație al unei fabrici afirmă că 90% dintre produsele sale sunt bune. Se face o cercetare selectivă pe 100 de produse obținute printr-un nou proces de fabricație și se găsesc 5 produse defecte. Cu un nivel de semnificație $\alpha = 0.05$ putem afirma că noul proces de fabricație este mai bun decât precedentul?
2. Directorul unei firme de asigurări a fost informat că 85% dintre clienți sunt mulțumiți de serviciile oferite. Șeful Departamentului de Relații cu Publicul crede că informația este exagerată și cere efectuarea unui studiu statistic. Pe un eșantion de 180 clienți s-a stabilit că 160 dintre ei sunt satisfăcuți de serviciile oferite. Ce concluzie rezultă dacă testul s-a făcut cu un nivel de semnificație $\alpha = 0.05$?

Test bilateral pentru medie (σ necunoscut)

Fie X o variabilă aleatoare teoretică cu repartiție normală $N(m, \sigma^2)$ cu media și dispersia necunoscute și m_0 o estimație a mediei. Se face o selecție de volum n, obținându-se variabilele aleatoare de selecție $X_1, X_2, \ldots, X_n$.

1. Se formulează ipotezele $H_0: \ m = m_0, \quad H_a: \ m \neq m_0$.
2. Se alege criteriul de testare $C(X, n) = \dfrac{\bar{X} - m_0}{\frac{s}{\sqrt{n}}}$
3. Se alege un nivel de semnificație α
4. Se determină $t =$ cuantila de ordin $1 - \frac{\alpha}{2}$ a repartiției Student cu $n - 1$ grade de libertate, corespunzătoare nivelului de semnificație α
5. Se află media de selecție $\bar{X}$ și dispersia de selecție modificată s pentru un eșantion de volum n
6. Dacă $\frac{\bar{X} - m_0}{\frac{s}{\sqrt{n}}} \in [-t, t]$ atunci H_0 este acceptată.
   Dacă $\frac{\bar{X} - m_0}{\frac{s}{\sqrt{n}}} \in (-\infty, -t) \cup (t, \infty)$ atunci H_0 este respinsă.

Exemplu

S-au luat aleator 12 bucăți dintr-un aliaj. În urma analizei chimice s-au obținut următoarele procente ale unei substanțe componente: 2.94, 2.75, 2.75, 2.81, 2.90, 2.90, 2.82, 2.95, 3.00, 2.95, 3.00, 3.05 (cantitatea de substanță este normal distribuită). Să se testeze cu un nivel de semnificație $\alpha = 0.05$ dacă procentul mediu este egal cu 2.95.

1. ipotezele $H_0: \ m = 2.95, \qquad H_a: m \neq 2.95$
2. criteriul de testare $\frac{\bar{X} - 2.95}{\frac{s}{\sqrt{12}}}$
3. nivelul de semnificație $\alpha = 0.05$
4. cuantila de ordin $1 - \frac{\alpha}{2} = 0.975$ cu 11 grade de libertate este $t = 2.201$
5. pentru $\bar{X} = 2.9017$ și $s = 0.0993$ avem $\frac{\bar{X} - 2.95}{\frac{s}{\sqrt{12}}} = -1.6853$
6. ipoteza H_0 se acceptă deoarece $-1.6853 \in [-2.201, 2.201]$

- Pentru test unilateral dreapta, se folosește $t =$ cuantila de ordin $1 - \alpha$ a repartiției Student cu $n - 1$ grade de libertate, iar regiunea de respingere este $W = (t, \infty)$
- Pentru test unilateral stânga, se folosește $t =$ cuantila de ordin α a repartiției Student cu $n - 1$ grade de libertate, iar regiunea de respingere este $W = (-\infty, t)$

Exerciții

1. La un magazin s-a primit un lot de sticle de Pepsi. Lotul este admis dacă volumul mediu este de cel puțin 2l. Se face o selecție de volum $n = 31$ sticle și se obține $\bar{X} = 1.89$ l, $s = 0.142$ l. Să se testeze cu un nivel de semnificație $\alpha = 0.05$ dacă lotul este admis (se presupune că volumul sticlelor este o variabilă aleatoare cu repartiție normală).
2. Directorul unei companii afirmă că numărul mediu al reclamațiilor săptămânale este egal cu 20, în timp ce membrii Consiliului de Administrație susțin că acesta este mai mare. Se face un studiu timp de 9 săptămâni asupra numărului reclamațiilor primite și se obțin următoarele rezultate: 20, 20, 22, 23, 21, 26, 23, 22, 21 (numărul reclamațiilor săptămânale este normal distribuit). Să se testeze cu un nivel de semnificație $\alpha = 0.01$ dacă afirmația directorului este adevărată.

Test bilateral pentru dispersie

Fie X o variabilă aleatoare teoretică cu repartiție normală $N(m, \sigma^2)$ cu media și dispersia necunoscute și σ_0 o estimație a dispersiei. Se face o selecție de volum n, obținându-se variabilele aleatoare de selecție $X_1, X_2, \ldots, X_n$.

1. Se formulează ipotezele $H_0: \ \sigma = \sigma_0, \qquad H_a: \ \sigma \neq \sigma_0$.
2. Se alege criteriul de testare $C(X, n) = \dfrac{(n-1) \cdot s^2}{\sigma_0^2}$
3. Se alege un nivel de semnificație α
4. Se determină $\chi_1, \chi_2 =$ cuantilele de ordin $\frac{\alpha}{2}$, respectiv $1 - \frac{\alpha}{2}$ ale repartiției χ^2 cu $n - 1$ grade de libertate;
5. Se află dispersia de selecție modificată s pentru un eșantion de volum n;
6. Dacă $\frac{(n-1) \cdot s^2}{\sigma_0^2} \in [\chi_1, \chi_2]$ atunci H_0 este acceptată.
   Dacă $\frac{(n-1) \cdot s^2}{\sigma_0^2} \in [0, \chi_1) \cup (\chi_2, \infty)$ atunci H_0 este respinsă.

Exemplu

Un număr de 10 containere au fost umplute cu următoarele cantități de substanță: 10.2, 9.7, 10.1, 10.3, 10.1, 9.8, 9.9, 10.4, 10.3, 9.8 litri (cantitatea de substanță este normal distribuită). Testați ipoteza că dispersia este $\sigma^2 = 0.03$ cu un nivel de semnificație $\alpha = 0.05$.

1. ipotezele $H_0: \ \sigma^2 = 0.03, \qquad H_a: \sigma^2 \neq 0.03$
2. criteriul de testare $C(X, n) = \dfrac{9 \cdot s^2}{0.03}$
3. nivelul de semnificație $\alpha = 0.05$
4. cuantila χ^2 cu 9 grade de libertate de ordin $\frac{\alpha}{2}$ este $\chi_1 = 2.7004$, iar cea de ordin $1 - \frac{\alpha}{2}$ este $\chi_2 = 19.0228$
5. pentru $s = 0.2459$ avem $\frac{9 \cdot s^2}{0.03} = 18.1333$
6. ipoteza H_0 se acceptă deoarece $18.1333 \in [2.7004, 19.0228]$