Laborator 6 - Statistică inferenţială

I. Inferenţă asupra mediei - Testul Z pentru media unei populaţii cu dispersia cunoscută

Se consideră o populaţie statistică căreia i se cunoaşte dispersia σ^2. Pentru un eşantion aleator simplu cu media de selecţie $\overline{x}_n$, dacă populaţia urmează o lege normală sau dimensiunea eşantionului este suficient de mare, scorul $z = \dfrac{\overline{x}_n - \mu}{\sigma/\sqrt{n}}$ este distribuit normal standard: $N(0,1)$.

Testul Z decurge astfel:

1. se formulează ipoteza nulă, care susţine că media populaţiei ia o valoare particulară:
$$\boxed{H_0 : \ \mu = \mu_0}$$
2. se formulează o ipoteză alternativă care poate fi de trei feluri:
$$\boxed{H_a : \ \mu < \mu_0} \quad \text{(ipoteză asimetrică la stânga) sau}$$
$$\boxed{H_a : \ \mu > \mu_0} \quad \text{(ipoteză asimetrică la dreapta) sau}$$
$$\boxed{H_a : \ \mu \neq \mu_0} \quad \text{(ipoteză simetrică)}$$
3. se fixează nivelul de semnificaţie: α (care uzual poate fi 1% sau 5%);
4. se calculează scorul testului:
$$\boxed{z = \frac{\overline{x}_n - \mu_0}{\sigma/\sqrt{n}}}$$
5. se determină valoarea critică z^*:
$$\boxed{z^* = qnorm(\alpha, 0, 1)} \quad \text{pentru ipoteză } H_a \text{ asimetrică la stânga } (z^* < 0),$$
$$\boxed{z^* = qnorm(1-\alpha, 0, 1)} \quad \text{pentru ipoteză } H_a \text{ asimetrică la dreapta } (z^* > 0),$$
$$\boxed{z^* = -qnorm(\alpha/2, 0, 1) = qnorm(1-\alpha/2, 0, 1)} \quad \text{pentru ipoteză } H_a \text{ simetrică } (z^* > 0).$$
6. **ipoteza nulă H_0 este respinsă dacă**
$$\boxed{z < z^*} \quad \text{pentru ipoteză } H_a \text{ asimetrică la stânga sau}$$
$$\boxed{z > z^*} \quad \text{pentru ipoteză } H_a \text{ asimetrică la dreapta sau}$$
$$\boxed{|z| > |z^*|} \quad \text{pentru ipoteză } H_a \text{ simetrică,}$$
dacă nu suntem într-una din aceste situaţii, atunci se spune că **nu există suficiente dovezi pentru a respinge ipoteza nulă H_0 şi a accepta ipoteza alternativă H_a.**

Exerciţiu rezolvat. Un producător de becuri doreşte sa testeze cu 5% nivel de semnificaţie afirmaţia că media de viaţă a acestora este de cel puţin 810 de ore (se ştie că deviaţia standard a populaţiei este $\sigma = 50$ de ore). Se alege un eşantion de 200 de becuri a căror medie de viata este gasita 816 ore. Poate fi acceptată ipoteza producătorului?

```
> alfa = 0.05
> population_mean = 810
> sample_mean = 816
> n = 200
> sigma = 50
> critical_z = qnorm(1- alfa)
> z_score = (sample_mean - population_mean)/(sigma/sqrt(n))
> critical_z
> z_score
```

Scorul va fi $z = 1.69705 > z^* = 1.64485$ şi ipoteza nulă poate fi respinsă, se acceptă ipoteza că media populaţiei este mai mare decât 810.

Exerciţii propuse (I.1 şi încă două dintre I.2-I.6)

I.1 Scrieţi o funcţie (numită **z_test**) care să calculeze şi să returneze valoarea critică şi scorul testului (parametrii funcţiei vor fi: tipul ipotezei alternative, n, μ_0, $\overline{x}_n$, α, σ etc). Funcţia aceasta va fi utilizată apoi la rezolvarea exerciţiilor de mai jos.

I.2 Dintr-o populaţie normală cu dispersia $\sigma^2 = 144$ se selectează 49 de indivizi a căror medie este 88; să se testeze ipoteza că media populaţiei este mai mică decât 90.

I.3 Din experienţă se ştie că rezultatele studenţilor la un test de matematică urmează o lege normală cu media 75 şi dispersia 17. Catedra de matematică doreşte să afle dacă studenţii din anul curent au un comportament atipic. Media rezultatelor unui grup de 36 studenţi este 85 de puncte. Cu 1% nivel de semnificaţie se poate trage concluzia că studenţii din anul curent sunt atipici?

I.4 Pe cutiile de un anumit tip de detergent este indicată o greutate de 21oz. O agenţie de protecţie a consumatorilor doreşte să testeze această greutate cu 1% nivel de semnificaţie. Pentru 100 de cutii găseşte o greutate medie de 20.5oz. Dacă se ştie că deviaţia standard a greutăţii este 2.5oz, agenţia poate pretinde mărirea cantităţii de detergent dintr-o cutie?

I.5 O firma producatoare de tuburi fluorescente doreste sa stie daca poate pretinde ca media de viata a acestora este 1000 de ore. Pentru aceasta fabrica 100 de tuburi si masoara pentru ele o medie de viata de 970 de ore. Firma respectiva cunoaste ca deviatia standard a vietii tuburilor este 85 de ore. Cu 5% nivel de semnificatie se poate trage concluzia ca media de viata este mai mica de 1000 de ore? Dar cu 1%?

I.6 Se cere ca media de viaţă a unui tip de baterii sa fie 22 de ore. Se ştie (din procesul de fabricaţie) că durata de viaţă bateriilor urmează o lege normală cu deviaţia standard 3 ore. Un eşantion de 16 baterii are o medie de viaţă măsurată de 20 de ore. Se poate trage concluzia că media de viaţă a bateriilor este diferită de 22 de ore?

II. Inferenţă asupra mediei - Testul t pentru media unei populaţii cu dispersia necunoscută

Se consideră o populaţie statistică distribuită normal căreia nu i se cunoaşte dispersia. Pentru un eşantion aleator simplu cu media de selecţie $\overline{x}_n$ şi deviaţia standard s, scorul $t = \dfrac{\overline{x}_n - \mu}{s/\sqrt{n}}$ este distribuit Student cu $n-1$ grade de libertate: $t(n-1)$.

Testul t decurge astfel:

1. se formulează ipoteza nulă, care susţine că media populaţiei ia o valoare particulară:

$$\boxed{H_0 : \ \mu = \mu_0}$$

2. se formulează o ipoteză alternativă care poate fi de trei feluri:

$$\boxed{H_a : \ \mu < \mu_0} \quad \text{(ipoteză asimetrică la stânga) sau}$$

$$\boxed{H_a : \ \mu > \mu_0} \quad \text{(ipoteză asimetrică la dreapta) sau}$$

$$\boxed{H_a : \ \mu \neq \mu_0} \quad \text{(ipoteză simetrică)}$$

3. se fixează nivelul de semnificaţie: α (care uzual poate fi 1% sau 5%);

4. se calculează scorul testului:

$$\boxed{t = \frac{\overline{x}_n - \mu_0}{s/\sqrt{n}}}$$

5. se determină valoarea critică t^*:

$$\boxed{t^* = qt(\alpha, n-1)} \quad \text{pentru ipoteză } H_a \text{ asimetrică la stânga } (t^* < 0),$$

$$\boxed{t^* = qt(1-\alpha, n-1)} \quad \text{pentru ipoteză } H_a \text{ asimetrică la dreapta } (t^* > 0),$$

$$\boxed{t^* = -qt(\alpha/2, n-1) = qt(1-\alpha/2, n-1)} \quad \text{pentru ipoteză } H_a \text{ simetrică } (t^* > 0).$$

6. **ipoteza nulă H_0 este respinsă dacă**

$$\boxed{t < t^*} \quad \text{pentru ipoteză } H_a \text{ asimetrică la stânga sau}$$

$$\boxed{t > t^*} \quad \text{pentru ipoteză } H_a \text{ asimetrică la dreapta sau}$$

$$\boxed{|t| > |t^*|} \quad \text{pentru ipoteză } H_a \text{ simetrică,}$$

altfel se spune că **nu există suficiente dovezi pentru a respinge ipoteza nulă H_0 şi a accepta ipoteza alternativă H_a.**

Exerciţiu rezolvat. Pentru un experiment asupra metabolismului 5 insecte sunt hrănite cu zahăr. Valorile nivelului de glucoză (care urmează o lege normală) obţinute din măsurători sunt:

55.95 68.24 52.73 21.5 23.78

Sa se testeze cu 5% nivel de semnificaţie ipoteza că media nivelului de glucoză este mai mare de 40.

```
> alfa = 0.05
> x = c(55.95, 68.24, 52.73, 21.5, 23.78)
> population_mean = 40
> sample_mean = mean(x)
> n = 5
> s = sd(x)
> se = s/sqrt(n)
> critical_t = qt(1 - alfa, n - 1)
> t_score = (sample_mean - population_mean)/se
> critical_t
> t_score
```

Rezultatul va fi $t^* = 2.13184 > t = 0.47867$, ipoteza nulă nu poate fi respinsă.

Exerciţii propuse (II.1 şi încă două dintre II.2-II.6)

II.1 Scrieţi două funcţii (de tipul **t_test**) care să calculeze şi să returneze valoarea critică şi scorul testului t: una dintre funcţii va trebui să citească eşantionul dintr-un fişier, iar cealaltă va primi ca argumente tipul ipotezei alternative, media de selecţie, deviaţia standard a eşantionului etc. Funcţiile acestea vor fi utilizate, după caz pentru rezolvarea exerciţiilor de mai jos.

II.2 Se măsoară pentru un esantion provenit dintr-o populaţie normală următoarele valori

$$36\ 32\ 28\ 33\ 41\ 28\ 31\ 26\ 29\ 34$$

Cu 1% nivel de semnificaţie să se testeze ipoteza că media are o valoare diferită de 34.

II.3 Pe pachetele unui tip de ţigări este trecută o concentraţie de nicotină (care urmează o lege normală) de 11.4 mg. Datorită unor reclamaţii, o agenţie neguvernamentală se hotărăşte să testeze această concentraţie. Pentru 100 de pachete de ţigări este găsită o medie a concentraţiei de 11.9 mg cu o deviaţie standard $s = 0.25$mg. Să se testeze cu 1% şi 5% nivel de semnificaţie dacă reclamaţiile primite sunt îndreptăţite.

II.4 Media rezultatelor unui test la istorie este de 80 de puncte. Catedra de istorie doreste sa afle daca studentii actuali au un comportament tipic la acest test. Pentru un esantion aleator simplu rezultatele se găsesc în fişierul history.txt . Să se formuleze şi să se testeze ipoteza alternativă corespunzătoare (cu 1% si 5% nivel de semnificaţie).

II.5 Se consideră un eşantion de dimensiune 64 cu media 52 şi dispersia $s^2 = 89.5$, care provine dintr-o populaţie distribuită normal. Să se testeze ipoteza că media populaţiei este 49 versus ipoteza că media este diferită de 49.

II.6 Să se aplice unul dintre testele cunoscute pentru următoarele date obţinute dintr-un eşantion aleator simplu

Media esantionului $\overline{x}_n = 29$

Dispersia esantionului $s^2 = 5$

Dimensiunea esantionului $n = 40$

Ipoteza nula $\mu_0 = 30$

Ipoteza alternativă $\mu_0 < 30$

Nivelul de semnificaţie $\alpha = 0.05$

III. Inferenţă asupra mediei - Testul Z pentru diferenţa mediilor unor populaţii cu dispersii cunoscute

Se consideră o două populaţii statistice cărora li se cunoasc dispersiile σ_1^2 şi σ_2^2. Se aleg două eşantioane aleatoare simple şi independente între ele cu mediile de selecţie $\overline{x}_{n_1}$ şi $\overline{x}_{n_2}$. Dacă populaţiile urmează o lege normală sau dimensiunea eşantioanelor este suficient de mare, scorul

$$z = \frac{(\overline{x}_{n_1} - \overline{x}_{n_2}) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

este distribuit (eventual cu aproximaţie) normal standard: $N(0, 1)$.

Testul Z decurge astfel:

1. se formulează ipoteza nulă, care susţine că diferenţa mediilor celor două populaţii ia o valoare particulară:

$$\boxed{H_0 :\ \mu_1 - \mu_2 = m_0}$$

2. se formulează o ipoteză alternativă care poate fi de trei feluri:

$$\boxed{H_a :\ \mu_1 - \mu_2 < m_0} \quad \text{(ipoteză asimetrică la stânga) sau}$$

$$\boxed{H_a :\ \mu_1 - \mu_2 > m_0} \quad \text{(ipoteză asimetrică la dreapta) sau}$$

$$\boxed{H_a :\ \mu_1 - \mu_2 \neq m_0} \quad \text{(ipoteză simetrică)}$$

3. se fixează nivelul de semnificaţie: α (care uzual poate fi 1% sau 5%);

4. se calculează scorul testului:

$$\boxed{\frac{(\overline{x}_{n_1} - \overline{x}_{n_2}) - m_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}}$$

5. se determină valoarea critică z^*:

$$\boxed{z^* = qnorm(\alpha, 0, 1)} \quad \text{pentru ipoteză } H_a \text{ asimetrică la stânga,}$$

$$\boxed{z^* = qnorm(1 - \alpha, 0, 1)} \quad \text{pentru ipoteză } H_a \text{ asimetrică la dreapta,}$$

$$\boxed{z^* = qnorm(1 - \alpha/2, 0, 1)} \quad \text{pentru ipoteză } H_a \text{ simetrică.}$$

6. **ipoteza nulă H_0 este respinsă dacă**

$$\boxed{z < z^*} \quad \text{pentru ipoteză } H_a \text{ asimetrică la stânga sau}$$

$$\boxed{z > z^*} \quad \text{pentru ipoteză } H_a \text{ asimetrică la dreapta sau}$$

$$\boxed{|z| > |z^*|} \quad \text{pentru ipoteză } H_a \text{ simetrică,}$$

altfel se spune că **nu există suficiente dovezi pentru a respinge ipoteza nulă H_0 şi a accepta ipoteza alternativă H_a.**

Exerciţiu rezolvat. Se compara durata de viata a doua tipuri de baterii. Primul tip are o deviatie standard de 4 ore, al doilea tip are o deviatie standard de 3 ore. Se aleg doua esantioane fiecare de dimensiune de 100 de baterii. Pentru primul esantion media de viata este de 48 de ore, iar pentru cel de-al doilea de 47 de ore.

Sa se testeze diferenta mediilor de viata cu 5% nivel de semnificaţie.

Observaţie. Testarea diferenţei mediilor echivalează cu o ipoteză alternativă de tipul $\mu_1 - \mu_2 \neq 0 = m_0$.

```
> alfa = 0.05
> m0 = 0
> sample1_mean = 48
> sample2_mean = 47
> n1 = 100
> n2 = 100
> sigma1 = 4
> sigma2 = 3
> combined_sigma = sqrt(sigma1^2/n1 + sigma2^2/n2)
> critical_z = qnorm(1 - alfa/2
> z_score = (sample1_mean - sample2_mean - m0)/combined_sigma
> critical_z
> z_score
```

Rezultatul va fi $|z^*| = 1.95996 < |z| = 2.00$, ipoteza nulă va fi respinsă şi se acceptă că mediile celor două populaţii sunt diferite.

Exerciţii propuse

III.1 Scrieţi o funcţie (numită, de exemplu, **z_test _means**) care să calculeze şi să returneze valoarea critică şi scorul testului Z pentru diferenţa mediilor (parametrii vor fi: tipul ipotezei alternative, α, n_1, n_2, σ_1, σ_2 etc.). Funcţia aceasta va fi utilizată, pentru rezolvarea exerciţiilor de mai jos.

III.2 80 dintre angajaţii aleşi aleator ai unei firme foarte mari au un salariu mediu săptămânal de 160$ (deviaţia standard a întregii populaţii fiind 3.24$). 70 dintre angajaţii unei alte firme au în medie 155$ salariu pe săptămână (deviaţia standard a întregii populaţii fiind 2.25$). Să se testeze dacă salariul mediu săptămânal la cele doua firme diferă semnificativ (1% nivel de semnificaţie).

III.3 Un raport recent arată ca absolvenţii de universitate fără diplomă se căsătoresc mai repede decât cei cu diplomă. Sunt aleşi câte 100 de indivizi din cele două populaţii; pentru aceste două eşantioane absolvenţii fără diplomă se căsătoresc în medie la 22.8 ani (deviaţia standerd cunoscută populaţiei fiind $\sigma_1 = 1.3$ ani) iar cei cu diplomă la 23.3 ani (deviaţia cunoscută a populaţiei este $\sigma_2 = 1.9$ ani).

Cu 1% nivel de semnificaţie se poate trage concluzia că raportul este corect?

IV. Inferenţă asupra dispersiilor a două populaţii - Testul F

Se consideră două populaţii normale ; din cele două populaţii se extrag două eşantioane aleatoare simple (şi independente între ele) cărora li se calculează dispersiile s_1^2 şi s_2^2. Scorul $F = \frac{s_1^2}{s_2^2}$ este distribuit $F(n_1 - 1, n_2 - 1)$.

Testul F decurge astfel:

1. se formulează ipoteza nulă, care susţine că dispersiile celor două populaţii sunt egale:

$$\boxed{H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1}$$

2. se formulează ipoteza alternativă; putem avea două tipuri de ipoteză alternativă:

$$\boxed{H_a : \frac{\sigma_1}{\sigma_2} > 1} \quad \text{(asimetrică la dreapta) pentru un un test one-tailed}$$

$$\boxed{H_a : \frac{\sigma_1}{\sigma_2} \neq 1} \quad \text{(simetrică) pentru un un test two-tailed.}$$

3. se fixează nivelul de semnificaţie: α (care uzual poate fi 1% sau 5%);

4. se calculează scorul testului:

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

5. se determină valoarea critică (sau, după caz, valoarile critice)

$$\boxed{F^* = qf(1 - \alpha, n_1 - 1, n_2 - 1)} \text{ pentru } H_a \text{ asimetrică la dreapta ,}$$

$$\boxed{F_s^* = qf(\alpha/2, n_1 - 1, n_2 - 1), F_d^* = qf(1 - \alpha/2, n_1 - 1, n_2 - 1)}$$

pentru H_a simetrică.

6. **ipoteza nulă** H_0 **este respinsă** şi se acceptă H_a dacă

$$F > F^* \text{ pentru } H_a \text{ asimetrică la stânga,}$$

$$F < F_s^* \text{ sau } F > F_d^* \text{ pentru } H_a \text{ simetrică.}$$

altfel **nu există suficiente dovezi pentru a respinge ipoteza nulă**.

Exerciţiu rezolvat. Rezultatele unui test psihologic efectuat pe două eşantioane, unul de femei şi unul de bărbaţi sunt următoarele:

$$\text{bărbaţi: } n_1 = 120, s_1 = 5.05$$

$$\text{femei: } n_2 = 135, s_2 = 5.44$$

Se poate trage concluzia că dispersiile celor două populaţii diferă semnificativ (1%)?

```
> alfa = 0.01
> n1= 120
> n2 = 135
> s1 = 5.05
> s2 = 5.44
> critical_F_s= qf(alfa/2, n1 - 1, n2 - 1)
> critical_F_d= qf(1 - alfa/2, n1 - 1, n2 - 1)
> critical_F_s
> critical_F_d
> F_score
```

Scorul este $F = 0.86175$, valorile critice sunt $F_s^* = 0.62843$ şi $F_d^* = 1.58257$; deoarece $F \in [F_s^*, F_d^*]$ ipoteza nulă nu poate fi respinsă. În acest caz putem considera că nu există dovezi semnificative pentru a afirma că dispersiile sunt diferite.

Exerciţii propuse

IV.1 Scrieţi o funcţie (numită **F_test**) care să calculeze şi să returneze valoarile critice şi scorul testului F (parametrii funcţiei vor fi: tipul ipotezei alternative, α, n_1, n_2, s_1, s_2 etc., eşantioanele se pot extrage din fişier ca mai jos). Funcţia aceasta va fi utilizată apoi la rezolvarea exerciţiilor care urmează.

```
> x1 = read.table("program.txt", header = TRUE)[['A']]
> x2 = read.table("program.txt", header = TRUE)[['B']]
> n1 = length(x1)
> s1 = sd(x1)
> ...
```

IV.2 Un profesor crede că un anumit program de lectură imbunătăţeşte abilităţile şi dorinţa copiilor de a citi. Pentru aceasta el alege două grupuri de elevi: unul de 22 de elevi care urmează programul prescris (A) şi unul de 22 de elevi care nu urmează acest program (B). Rezultatele sunt date în fişierul *program.txt*.

Să se decidă cu 1% şi 5% nivel de semnificaţie dacă dispersiile celor două populaţii sunt diferite.

IV.3 Cercetătorii studiază amplitudinea mişcării obţinută prin stimularea nervoasă a şoarecilor. Pentru şoarecii drogaţi se obţin următoarele date:

12.512 12.869 19.098 15.350 13.297 15.589

Pentru şoarecii normali se obţin următoarele date:

11.074 9.686 12.164 8.351 12.182 11.489

Influenţa drogurilor este semnificativă în ceea ce priveşte cele doua dispersii (5% nivel de semnificaţie)?

V. Inferenţă asupra mediilor a două populaţii - Testul T pentru diferenţa mediilor unor populaţii cu dispersii necunoscute

Se consideră o două populaţii statistice cărora nu li se cunosc dispersiile. Se aleg două eşantioane aleatoare simple şi independente între ele cu mediile de selecţie $\overline{x}_{n_1}$ şi $\overline{x}_{n_2}$ şi dispersiile σ_1^2 şi σ_2^2. Înaintea efectuării testului T se foloseşte testul F pentru a decide dacă dispersiile celor două populaţii sunt diferite. Dacă dispersiile nu sunt diferite $\boxed{t = \dfrac{(\overline{x}_{n_1} - \overline{x}_{n_2}) - m_0}{\sqrt{\dfrac{s^2}{n_1} + \dfrac{s^2}{n_2}}}}$ urmează o distribuţie Student[1]: $t(n_1+n_2-2)$. Dacă dispersiile sunt diferite, atunci $\boxed{t = \dfrac{(\overline{x}_{n_1} - \overline{x}_{n_2}) - m_0}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}}$ urmează o distribuţie Student: $t(\min(n_1-1, n_2-1))$.

Testul T decurge astfel:

1. se formulează ipoteza nulă, care susţine că diferenţa mediilor celor două populaţii ia o valoare particulară (de cele mai multe ori zero):

$$\boxed{H_0 : \ \mu_1 - \mu_2 = m_0}$$

2. se formulează o ipoteză alternativă care poate fi de trei feluri:

$\boxed{H_a : \ \mu_1 - \mu_2 < m_0}$ (ipoteză asimetrică la stânga) sau

$\boxed{H_a : \ \mu_1 - \mu_2 > m_0}$ (ipoteză asimetrică la dreapta) sau

$\boxed{H_a : \ \mu_1 - \mu_2 \neq m_0}$ (ipoteză simetrică)

3. se fixează nivelul de semnificaţie: α (care uzual poate fi 1% sau 5%);

4. se calculează scorul testului:

   a) dacă dispersiile sunt diferite:

$$\boxed{t = \frac{(\overline{x}_{n_1} - \overline{x}_{n_2}) - m_0}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}}$$

   iar numărul de grade de libertate este $df = \min(n_1-1, n_2-1)$.

[1] În acest caz $s = \sqrt{\dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$

b) dacă dispersiile sunt "egale"

$$\boxed{t = \frac{(\overline{x}_{n_1} - \overline{x}_{n_2}) - m_0}{\sqrt{\dfrac{s^2}{n_1} + \dfrac{s^2}{n_2}}}}$$

unde $s^2 = \dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$, iar numărul de grade de libertate este $df = n_1 + n_2 - 2$.

5. se determină valoarea critică t^*:

$$\boxed{t^* = qt(\alpha, df)} \quad \text{pentru ipoteză } H_a \text{ asimetrică la stânga } (t^* < 0),$$

$$\boxed{t^* = qt(1-\alpha, df)} \quad \text{pentru ipoteză } H_a \text{ asimetrică la dreapta } (t^* > 0),$$

$$\boxed{t^* = -qt(\alpha/2, df) = qt(1-\alpha/2, df)} \quad \text{pentru ipoteză } H_a \text{ simetrică } (t^* > 0).$$

6. **ipoteza nulă H_0 este respinsă dacă**

$$\boxed{t < t^*} \quad \text{pentru ipoteză } H_a \text{ asimetrică la stânga sau}$$

$$\boxed{t > t^*} \quad \text{pentru ipoteză } H_a \text{ asimetrică la dreapta sau}$$

$$\boxed{|t| > |t^*|} \quad \text{pentru ipoteză } H_a \text{ imetrică,}$$

altfel se spune că **nu există suficiente dovezi pentru a respinge ipoteza nulă H_0 şi a accepta ipoteza alternativă H_a.**

Exerciţiu rezolvat. Rezultatele unui test psihologic efectuat pe două eşantioane, unul de femei şi unul de bărbaţi sunt următoarele:

$$\text{bărbaţi: } n_1 = 110, \overline{x}_1 = 25.84, s_1 = 4.25$$

$$\text{femei: } n_2 = 105, \overline{x}_1 = 21.53, s_2 = 3.85$$

Se poate trage concluzia că mediile celor două populaţii diferă semnificativ (1%)?

Observaţie. Testarea diferenţei mediilor echivalează cu o ipoteză alternativă de tipul $\mu_1 - \mu_2 \neq 0 = m_0$.

```
> alfa = 0.01
> m0 = 0
> sample1_mean = 25.84
> sample2_mean = 21.53
> n1 = 110
> n2 = 105
> s1 = 4.25
> s2 = 3.85
> critical_F_s= qf(alfa/2, n1 - 1, n2 - 1) # testul F
> critical_F_d= qf(1 - alfa/2, n1 - 1, n2 - 1)
> F_score = s1^2/ s2^2
> if(F_score < critical_F_s | F_score > critical_F_d) {
+ df = min(n1 - 1, n2 - 1)
+ combined_s = sqrt(s1^2/n1 + s2^2/n2)
+ } else {
+ df = n1 + n2 - 2
+ combined_s = sqrt(((n1 - 1)*s1^2 + (n2 - 1)*s2^2)/df)*sqrt(1/n1+1/n2)
+ }
> critical_t = qt(1 - alfa/2, df)
> t_score = (sample1_mean - sample2_mean - m0)/combined_sigma
> critical_t
> t_score
```

Rezultatul va fi $|t^*| = 2.6239 > |t| = 7.7994$, ipoteza nulă poate fi respinsă.

Exerciţii propuse

V.1 Scrieţi o funcţie (numită, de exemplu, **T_test _means**) care să calculeze şi să returneze valoarea critică şi scorul testului T pentru diferenţa mediilor; această funcţie va include şi testul F simetric (parametrii vor fi: tipul ipotezei alternative, α, n_1, n_2, s_1 etc.). Funcţia aceasta va fi utilizată, pentru rezolvarea exerciţiilor care urmează.

V.2 Un profesor crede că un anumit program de lectură imbunătăţeşte abilităţile şi dorinţa copiilor de a citi. Pentru aceasta el alege două grupuri de elevi: unul de 22 de elevi care urmează programul prescris (A) şi unul de 22 de elevi care nu urmează acest program (B). Rezultatele sunt date în fişierul *program.txt*.

Să se decidă cu 1% şi 5% nivel de semnificaţie dacă mediile celor două populaţii diferă semnificativ.

V.3 Cercetătorii studiază amplitudinea mişcării obţinută prin stimularea nervoasă a şoarecilor. Pentru şoarecii drogaţi se obţin următoarele date:

12.512 12.869 19.098 15.350 13.297 15.589

Pentru şoarecii obişnuiţi se obţin următoarele date:

11.074 9.686 12.164 8.351 12.182 11.489

Influenţa drogurilor este semnificativă în ceea ce priveşte cele două medii (1% nivel de semnificaţie)?

Temă pentru acasă.

5 puncte [1p: E1 sau E2] + [1.5p: E3 sau E4] + [2.5p: E5 sau E6]

E1. (1 punct) Se măsoară nivelul de magneziu (care urmează o lege normală) din sângele unui pacient:

$$1.64 \;\; 1.54 \;\; 1.56 \;\; 1.57 \;\; 1.44 \;\; 1.48 \;\; 1.56 \;\; (mg/dl)$$

Putem trage concluzia că acest nivel este mai mic decât maximum posibil, 1.6 mg/dl? (1%)

E2. (1 punct) Folosind un eşantion aleator simplu format din 36 de indivizi se proiectează un test de semnificaţie în ceea ce priveşte media unei populaţii normale:

$$H_0 : \mu = 4$$

$$H_a : \mu \neq 4,$$

şi se calculează un scor $z = -2.35$. Este acest scor statistic semnificativ pentru 1% nivel de semnificaţie? Aceeaşi întrebare pentru 5% nivel de semnificaţie.

E3. (1.5 puncte) O firma producătoare de pastă de dinţi doreşte să ştie dacă introducerea unui nou compus chimic măreşte performanţele produsului său. Sunt alese doua eşantioane fiecare a câte 100 de indivizi. Pentru primul grup (care foloseste noul produs) se numără în medie 3 carii (deviaţia standard cunoscută este 0.6), pentru cel de-al doilea grup se numără în medie 3.5 carii (deviaţia standard este de 0.4).

Cu 1% nivel de semnificaţie se poate trage concluzia că prin introducerea noului compus pasta de dinţi îşi îmbunătăţeşte performanţele? Dar cu 5%?

E4. (1.5 puncte) Pentru două eşantioane provenind din două populaţii normale distincte A şi B se determină următoarele valori

$$(A) \quad \sigma_1 = 0.75, n_1 = 155, \overline{x}_{n_1} = 15$$

$$(B) \quad \sigma_2 = 0.78, n_2 = 150, \overline{x}_{n_2} = 14.5$$

Să se testeze diferenţa mediilor celor două populaţii cu 1% si cu 5% nivel de semnificaţie.

E5. (2 puncte) Pentru două eşantioane aleatoare simple provenind din două populaţii normale se obţin următoarele date

$$n_1 = 66 \quad \overline{x}_{n_1} = 21 \quad s_1 = 1.2$$

$$n_2 = 68 \quad \overline{x}_{n_2} = 20 \quad s_2 = 1.1$$

Să se testeze diferenţa mediilor ceor două populaţii (5%).

E6. (2 puncte) Pentru două eşantioane aleatoare simple provenind din două populaţii normale se obţin următoarele date

$$n_1 = 210 \quad \overline{x}_{n_1} = 3.5 \quad \sigma_1 = 0.6$$

$$n_2 = 190 \quad \overline{x}_{n_2} = 4 \quad \sigma_2 = 0.7$$

Putem trage concluzia că media celei de-a doua populaţii este mai mare? (1%).

Rezolvările acestor exerciţii (funcţiile R şi apelurile lor) vor fi redactate într-un script R.