

Laborator 6 - Statistică inferențială

I. Inferență asupra mediei - Testul Z pentru media unei populații cu dispersia cunoscută

Se consideră o populație statistică căreia i se cunoaște dispersia σ^2 . Pentru un eșantion aleator simplu cu media de selecție \bar{x}_n , dacă populația urmează o lege normală sau dimensiunea eșantionului este suficient de mare, scorul $z = \frac{\bar{x}_n - \mu}{\sigma/\sqrt{n}}$ este distribuit normal standard: $N(0, 1)$.

Testul Z decurge astfel:

1. se formulează ipoteza nulă, care susține că media populației ia o valoare particulară:

$$H_0 : \mu = \mu_0$$

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

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

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

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

3. se fixează nivelul de semnificație: α (care uzual poate fi 1% sau 5%);
4. se calculează scorul testului:

$$z = \frac{\bar{x}_n - \mu_0}{\sigma/\sqrt{n}}$$

5. se determină valoarea critică z^* :

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

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

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

6. ipoteza nulă H_0 este respinsă dacă

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

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

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

dacă nu suntem într-una din aceste situații, atunci se spune că **nu există suficiente dovezi pentru a respinge ipoteza nulă H_0 și a accepta ipoteza alternativă H_a .**

Exercițiu rezolvat. Un producător de becuri dorește să testeze cu 5% nivel de semnificație afirmația că media de viață a acestora este de cel puțin 810 de ore (se știe că deviația standard a populației este $\sigma = 50$ de ore). Se alege un eșantion de 200 de becuri a căror medie de viață este găsită 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 $z = 1.69705 > z^* = 1.64485$ și ipoteza nulă poate fi respinsă, se acceptă ipoteza că media populației este mai mare decât 810.

Exerciții propuse (I.1 și încă două dintre I.2-I.6)

- I.1 Scrieți o funcție (numită **z_test**) care să calculeze și să returneze valoarea critică și scorul testului (parametrii funcției vor fi: tipul ipotezei alternative, n , μ_0 , \bar{x}_n , α , σ etc). Funcția aceasta va fi utilizată apoi la rezolvarea exercițiilor de mai jos.
- I.2 Dintr-o populație normală cu dispersia $\sigma^2 = 144$ se selectează 49 de indivizi a căror medie este 88; să se testeze ipoteza că media populației este mai mică decât 90.
- I.3 Din experiență se știe că rezultatele studenților la un test de matematică urmează o lege normală cu media 75 și dispersia 17. Catedra de matematică dorește să afle dacă studenții din anul curent au un comportament atipic. Media rezultatelor unui grup de 36 studenți este 85 de puncte. Cu 1% nivel de semnificație se poate trage concluzia că studenții din anul curent sunt atipici?
- I.4 Pe cutiile de un anumit tip de detergent este indicată o greutate de 21oz. O agenție de protecție a consumatorilor dorește să testeze această greutate cu 1% nivel de semnificație. Pentru 100 de cutii găsește o greutate medie de 20.5oz. Dacă se știe că deviația standard a greutății este 2.5oz, agenția poate pretinde mărirea cantității de detergent dintr-o cutie?
- I.5 O firma producătoare de tuburi fluorescente dorește să afle dacă poate pretinde ca media de viață a acestora este 1000 de ore. Pentru aceasta fabrică 100 de tuburi și măsoară pentru ele o medie de viață de 970 de ore. Firma respectivă cunoaște că deviația standard a vieții tuburilor este 85 de ore. Cu 5% nivel de semnificație se poate trage concluzia că media de viață este mai mică de 1000 de ore? Dar cu 1%?
- I.6 Se cere ca media de viață a unui tip de baterii să fie 22 de ore. Se știe (din procesul de fabricație) că durata de viață a bateriilor urmează o lege normală cu deviația standard 3 ore. Un eșantion de 16 baterii are o medie de viață măsurată de 20 de ore. Se poate trage concluzia că media de viață a bateriilor este diferită de 22 de ore?

II. Inferență asupra mediei - Testul t pentru media unei populații cu dispersia necunoscută

Se consideră o populație statistică distribuită normal căreia nu i se cunoaște dispersia. Pentru un eșantion aleator simplu cu media de selecție \bar{x}_n și deviația standard s , scorul $t = \frac{\bar{x}_n - \mu}{s/\sqrt{n}}$ este distribuit Student cu $n - 1$ grade de libertate: $t(n - 1)$.

Testul t decurge astfel:

1. se formulează ipoteza nulă, care susține că media populației ia o valoare particulară:

$$H_0 : \mu = \mu_0$$

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

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

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

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

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

4. se calculează scorul testului:

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

5. se determină valoarea critică t^* :

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

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

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

6. ipoteza nulă H_0 este respinsă dacă

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

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

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

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

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

55.95 68.24 52.73 21.5 23.78

Sa se testeze cu 5% nivel de semnificație 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.

Exerciții propuse (II.1 și încă două dintre II.2-II.6)

II.1 Scrieți două funcții (de tipul **t_test**) care să calculeze și să returneze valoarea critică și scorul testului t: una dintre funcții va trebui să citească eșantionul dintr-un fișier, iar cealaltă va primi ca argumente tipul ipotezei alternative, media de selecție, deviația standard a eșantionului etc. Funcțiile acestea vor fi utilizate, după caz pentru rezolvarea exercițiilor de mai jos.

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

36 32 28 33 41 28 31 26 29 34

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

II.3 Pe pachetele unui tip de țigări este trecută o concentrație de nicotină (care urmează o lege normală) de 11.4 mg. Datorită unor reclamații, o agenție neguvernamentală se hotărăște să testeze această concentrație. Pentru 100 de pachete de țigări este găsită o medie a concentrației de 11.9 mg cu o deviație standard $s = 0.25$ mg. Să se testeze cu 1% și 5% nivel de semnificație dacă reclamațiile primite sunt îndreptățite.

II.4 Media rezultatelor unui test la istorie este de 80 de puncte. Catedra de istorie dorește să afle dacă studenții actuali au un comportament tipic la acest test. Pentru un esantion aleator simplu rezultatele se găsesc în fișierul history.txt . Să se formuleze și să se testeze ipoteza alternativă corespunzătoare (cu 1% și 5% nivel de semnificație).

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

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

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

Dispersia esantionului $s^2 = 5$

Dimensiunea esantionului $n = 40$

Ipoteza nulă $\mu_0 = 30$

Ipoteza alternativă $\mu_0 < 30$

Nivelul de semnificație $\alpha = 0.05$

III. Inferență asupra mediei - Testul Z pentru diferența mediilor unor populații cu dispersii cunoscute

Se consideră o două populații statistice cărora li se cunoasc dispersiile σ_1^2 și σ_2^2 . Se aleg două eșantioane aleatoare simple și independente între ele cu mediile de selecție \bar{x}_{n_1} și \bar{x}_{n_2} . Dacă populațiile urmează o lege normală sau dimensiunea eșantioanelor este suficient de mare, scorul

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

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

Testul Z decurge astfel:

1. se formulează ipoteza nulă, care susține că diferența mediilor celor două populații ia o valoare particulară:

$$H_0 : \mu_1 - \mu_2 = m_0$$

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

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

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

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

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

4. se calculează scorul testului:

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

5. se determină valoarea critică z^* :

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

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

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

6. ipoteza nulă H_0 este respinsă dacă

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

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

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

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

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

Sa se testeze diferența mediilor de viața cu 5% nivel de semnificație.

Observație. Testarea diferenței 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 și se acceptă că mediile celor două populații sunt diferite.

Exerciții propuse

- III.1 Scrieți o funcție (numită, de exemplu, **z_test_means**) care să calculeze și să returneze valoarea critică și scorul testului Z pentru diferența mediilor (parametrii vor fi: tipul ipotezei alternative, α , n_1 , n_2 , σ_1 , σ_2 etc.). Funcția aceasta va fi utilizată, pentru rezolvarea exercițiilor de mai jos.
- III.2 80 dintre angajații aleși aleator ai unei firme foarte mari au un salariu mediu săptămânal de 160\$ (deviația standard a întregii populații fiind 3.24\$). 70 dintre angajații unei alte firme au în medie 155\$ salariu pe săptămână (deviația standard a întregii populații fiind 2.25\$). Să se testeze dacă salariul mediu săptămânal la cele două firme diferă semnificativ (1% nivel de semnificație).
- III.3 Un raport recent arată ca absolvenții de universitate fără diplomă se căsătoresc mai repede decât cei cu diplomă. Sunt aleși câte 100 de indivizi din cele două populații; pentru aceste două eșantioane absolvenții fără diplomă se căsătoresc în medie la 22.8 ani (deviația standard cunoscută populației fiind $\sigma_1 = 1.3$ ani) iar cei cu diplomă la 23.3 ani (deviația cunoscută a populației este $\sigma_2 = 1.9$ ani).
- Cu 1% nivel de semnificație se poate trage concluzia că raportul este corect?

IV. Inferență asupra dispersiilor a două populații - Testul F

Se consideră două populații normale ; din cele două populații se extrag două eșantioane aleatoare simple (și independente între ele) cărora li se calculează dispersiile s_1^2 și 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 susține că dispersiile celor două populații sunt egale:

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

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

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

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

3. se fixează nivelul de semnificație: α (care uzual poate fi 1% sau 5%);
4. se calculează scorul testului:

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

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

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

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

pentru H_a simetrică.

6. ipoteza nulă H_0 este respinsă și se acceptă H_a dacă

$F > F^*$ pentru H_a asimetrică la stânga,

$F < F_s^*$ sau $F > F_d^*$ pentru H_a simetrică.

altfel nu există suficiente dovezi pentru a respinge ipoteza nulă.

Exercițiu rezolvat. Rezultatele unui test psihologic efectuat pe două eșantioane, unul de femei și unul de bărbați sunt următoarele:

bărbați: $n_1 = 120, s_1 = 5.05$

femei: $n_2 = 135, s_2 = 5.44$

Se poate trage concluzia că dispersiile celor două populații 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 $F_s^* = 0.62843$ și $F_d^* = 1.58257$; 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.

Exerciții propuse

IV.1 Scrieți o funcție (numită **F.test**) care să calculeze și să returneze valorile critice și scorul testului F (parametrii funcției vor fi: tipul ipotezei alternative, α , n_1 , n_2 , s_1 , s_2 etc., eșantioanele se pot extrage din fișier ca mai jos). Funcția aceasta va fi utilizată apoi la rezolvarea exercițiilor 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 îmbunătățește abilitățile și dorința copiilor de a citi. Pentru aceasta el alege două grupuri de elevi: unul de 22 de elevi care urmează programul prescris (A) și unul de 22 de elevi care nu urmează acest program (B). Rezultatele sunt date în fișierul *program.txt*.

Să se decidă cu 1% și 5% nivel de semnificație dacă dispersiile celor două populații sunt diferite.

IV.3 Cercetătorii studiază amplitudinea mișcării obținută prin stimularea nervoasă a șoarecilor. Pentru șoarecii drogați se obțin următoarele date:

12.512 12.869 19.098 15.350 13.297 15.589

Pentru șoarecii normali se obțin următoarele date:

11.074 9.686 12.164 8.351 12.182 11.489

Influența drogurilor este semnificativă în ceea ce privește cele două dispersii (5% nivel de semnificație)?

V. Inferență asupra mediilor a două populații - Testul T pentru diferența mediilor unor populații cu dispersii necunoscute

Se consideră o două populații statistice cărora nu li se cunosc dispersiile. Se aleg două eșantioane aleatoare simple și independente între ele cu mediile de selecție \bar{x}_{n_1} și \bar{x}_{n_2} și dispersiile σ_1^2 și σ_2^2 . Înaintea efectuării testului T se folosește testul F pentru a decide dacă dispersiile celor două populații sunt diferite. Dacă dispersiile nu sunt diferite

$$t = \frac{(\bar{x}_{n_1} - \bar{x}_{n_2}) - m_0}{\sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}} \text{ urmează o}$$

distribuție Student¹: $t(n_1 + n_2 - 2)$. Dacă dispersiile sunt diferite, atunci

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

urmează o distribuție Student: $t(\min(n_1 - 1, n_2 - 1))$.

Testul T decurge astfel:

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

$$H_0 : \mu_1 - \mu_2 = m_0$$

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

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

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

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

3. se fixează nivelul de semnificație: α (care uzual poate fi 1% sau 5%);
4. se calculează scorul testului:

a) dacă dispersiile sunt diferite:

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

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

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

b) dacă dispersiile sunt "egale"

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

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

5. se determină valoarea critică t^* :

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

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

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

6. ipoteza nulă H_0 este respinsă dacă

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

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

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

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

Exercițiu rezolvat. Rezultatele unui test psihologic efectuat pe două eșantioane, unul de femei și unul de bărbați sunt următoarele:

$$\text{bărbați: } n_1 = 110, \bar{x}_1 = 25.84, s_1 = 4.25$$

$$\text{femei: } n_2 = 105, \bar{x}_2 = 21.53, s_2 = 3.85$$

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

Observație. Testarea diferenței 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.

Exerciții propuse

- V.1 Scrieți o funcție (numită, de exemplu, **T_test_means**) care să calculeze și să returneze valoarea critică și scorul testului T pentru diferența mediilor; această funcție va include și testul F simetric (parametrii vor fi: tipul ipotezei alternative, α , n_1 , n_2 , s_1 etc.). Funcția aceasta va fi utilizată, pentru rezolvarea exercițiilor care urmează.
- V.2 Un profesor crede că un anumit program de lectură îmbunătățește abilitățile și dorința copiilor de a citi. Pentru aceasta el alege două grupuri de elevi: unul de 22 de elevi care urmează programul prescris (A) și unul de 22 de elevi care nu urmează acest program (B). Rezultatele sunt date în fișierul *program.txt*.
- Să se decidă cu 1% și 5% nivel de semnificație dacă mediile celor două populații diferă semnificativ.
- V.3 Cercetătorii studiază amplitudinea mișcării obținută prin stimularea nervoasă a șoarecilor. Pentru șoarecii drogați se obțin următoarele date:

12.512 12.869 19.098 15.350 13.297 15.589

Pentru șoarecii obișnuiți se obțin următoarele date:

11.074 9.686 12.164 8.351 12.182 11.489

Influența drogurilor este semnificativă în ceea ce privește cele două medii (1% nivel de semnificație)?

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 semnificație în ceea ce privește media unei populații normale:

$$H_0 : \mu = 4$$

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

și se calculează un scor $z = -2.35$. Este acest scor statistic semnificativ pentru 1% nivel de semnificație? Aceeași întrebare pentru 5% nivel de semnificație.

E3. (1.5 puncte) O firma producătoare de pastă de dinți dorește să știe dacă introducerea unui nou compus chimic mărește performanțele produsului său. Sunt alese doua eşantioane fiecare a câte 100 de indivizi. Pentru primul grup (care folosește noul produs) se numără în medie 3 carii (deviația standard cunoscută este 0.6), pentru cel de-al doilea grup se numără în medie 3.5 carii (deviația standard este de 0.4).

Cu 1% nivel de semnificație se poate trage concluzia că prin introducerea noului compus pasta de dinți își îmbunătățește performanțele? Dar cu 5%?

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

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

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

Să se testeze diferența mediilor celor două populații cu 1% și cu 5% nivel de semnificație.

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

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

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

Să se testeze diferența mediilor celor două populații (5%).

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

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

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

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

Rezolvările acestor exerciții (funcțiile R și apelurile lor) vor fi redactate într-un script R.