Laborator 5 - 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
```

Rezultatul 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 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 90; 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 mediilor a două populaţii - Testul t pentru media unei populaţii cu dispersia necunoscută

Se consideră o populaţie statistică 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 z^*:

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

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

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

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 la insecte 5 indivizi sunt hrăniţi 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 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 media de selecţie şi deviaţia standard a eşantionului. 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 de 11.4mg. 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 11.9mg 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 $s^2 = 89.5$. 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 $\alpha = 0.05$