

## Laborator 14 – Probabilități și Statistică Matematică

### STATISTICA INFERENTIALA (2)

#### ***I. Inferenta asupra mediei - Testul Z pentru media unei populatii cu dispersia cunoscuta***

Exercitiu rezolvat. Un producator de becuri doreste sa testeze cu 5% nivel de semnificatie afirmatia ca media de viata a acestora este de cel putin 810 de ore (se stie ca deviatia standard a populatiei este  $\sigma = 50$  de ore). Se alege un esantion de 200 de becuri a caror medie de viata este gasita 816 ore. Poate fi acceptata ipoteza producatorului?

```
> 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  $z = 1.69705 > z^* = 1.64485$  si ipoteza nula poate fi respinsa, se accepta ipoteza ca media populatiei este mai mare decat 810.

#### **APLICATII:**

1. Dintr-o populatie normala cu dispersia  $\sigma^2 = 144$  se selecteaza 49 de indivizi a caror medie este 88; sa se testeze ipoteza ca media populatiei este mai mica decat 90.
2. Din experienta se stie ca rezultatele studentilor la un test de matematica urmeaza o lege normala cu media 75 si dispersia 17. Catedra de matematica doreste sa afle daca studentii din anul curent au un comportament atipic. Media rezultatelor unui grup de 36 studenti este 85 de puncte. Cu 1% nivel de semnificatie se poate trage concluzia ca studentii din anul curent sunt atipici?

## ***II. Inferenta asupra mediei - Testul t pentru media unei populatii cu dispersia necunoscuta***

Exercitiu rezolvat. Pentru un experiment asupra metabolismului 5 insecte sunt hranite cu zahar. Valorile nivelului de glucoza (care urmeaza o lege normala) obtinute din masuratori sunt:

55.95 68.24 52.73 21.5 23.78

Sa se testeze cu 5% nivel de semnificatie ipoteza ca media nivelului de glucoza 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 nula nu poate fi respinsa.

### **APLICATIE:**

Se masoara pentru un esantion provenit dintr-o populatie normala urmatoarele valori

36 32 28 33 41 28 31 26 29 34

Cu 1% nivel de semnificatie sa se testeze ipoteza ca media are o valoare diferita de 34.

### ***III. Inferenta asupra mediei - Testul Z pentru diferenta mediilor unor populatii cu dispersii cunoscute***

Exercitiu 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 semnificatie.

*Observatie. Testarea diferentei mediilor echivaleaza cu o ipoteza alternativa 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 nula va fi respinsa si se accepta ca mediile celor doua populatii sunt diferite.

#### **IV. Inferenta asupra dispersiilor a doua populatii - Testul F**

Exercitiu rezolvat. Rezultatele unui test psihologic efectuat pe doua esantioane, unul de femei si unul de barbati sunt urmatoarele:

barbati:  $n_1 = 120$ ,  $s_1 = 5.05$

femei:  $n_2 = 135$ ,  $s_2 = 5.44$

Se poate trage concluzia ca dispersiile celor doua populatii difera 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  $F_s^* = 0.62843$  ,si  $F_d^* = 1.58257$ ; deoarece

$F \in [F_s^*, F_d^*]$  ipoteza nula nu poate fi respinsa. In acest caz putem considera ca nu exista dovezi semnificative pentru a afirma ca dispersiile sunt diferite.

#### **APLICATIE:**

Cercetatorii studiaza amplitudinea miscarii obtinuta prin stimularea nervoasa a soarecilor.

Pentru soarecii drogati se obtin urmatoarele date: 12.512 12.869 19.098 15.350 13.297 15.589

Pentru soarecii normali se obtin urmatoarele date: 11.074 9.686 12.164 8.351 12.182 11.489

Influenta drogurilor este semnificativa in ceea ce priveste cele doua dispersii (5% nivel de semnificatie)?

***V. Inferenta asupra mediilor a doua populatii - Testul T pentru diferenta mediilor unor populatii cu dispersii necunoscute***

Exercitiu rezolvat. Rezultatele unui test psihologic efectuat pe doua esantioane, unul de femei si unul de barbati sunt urmatoarele:

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

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

Se poate trage concluzia ca mediile celor doua populatii difera semnificativ (1%)?

Observatie: Testarea diferentei mediilor echivaleaza cu o ipoteza alternativa 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 nula poate fi respinsa.