Laborator 9 - solutii

Exercitiul 1

IQ-ul populatiei este in medie de 100 cu o deviatie standard de 15. O echipa de cercetatori vrea sa testeze un nou medicament si sa verifice daca acesta are un efect pozitiv/negativ/niciun efect asupra inteligentei. S-au selectat 30 de participanti care au luat medicamentul si s-a observat ca media IQ-urilor lor este de 140. Le-a afectat acest medicament coeficientul de inteligenta?

Solutie:

1. Fixam ipotezele nula si alternativa:
   - medicamentul afecteaza inteligenta
   - medicamentul nu afecteaza inteligenta
2. Decidem tipul de test statistic pe care o sa-l utilizam. Deoarece marimea setului de date peste $n = 30$, vom folosi Z-test.
3. Calculam statistica corespunzatoare testului z: $z = \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}}$, unde $\bar{X}_n = \frac{X_1 + \cdots + X_n}{n}$ este media setului efectiv de date, μ_0 este media teoretica a datelor, iar σ este deviatia standard teoretica a datelor. In cazul nostru, $z = \frac{140-100}{15} \cdot \sqrt{30} = 14.606$.
4. Alegem nivelul de semnificatie α, determinand regiunea de respingere. Comparam valoarea z cu valoarea x care indeplineste $\mathbb{P}(Z < x) > 1 - \alpha \Leftrightarrow \mathbb{P}(Z \geq x) < \alpha$, unde $Z \sim N(0,1)$ este normala de medie 0 si devianta standard 1. Alegem $\alpha = 0.05$. In cazul de fata, din tabelul cu valorile cdf ale $N(0,1)$, obtinem ca $x = 1.96$.
5. Pentru ca $14.606 > 1.96$, respingem ipoteza nula.

Exercitiul 2

Un profesor vrea sa stie daca cursul sau de introducere in statistica a fost bine inteles de catre studenti. De aceea, au fost alesi 6 studenti la intamplare si li s-a dat un test. Profesorul considera un succes un scor de peste 70 de puncte. Cei 6 studenti au luat urmatoarele note: 62, 92, 75, 68, 83, 95. Poate avea profesorul 90% incredere ca media clasei este de peste 70?

Solutie:

1. Fixam ipotezele nula si alternativa:
   - $\mu_0 > 70$
   - $\mu_0 \leq 70$

2. Decidem tipul de test statistic pe care o sa-l utilizam. Intrucat marimea setului de date este $n = 6 < 30$ si varianta teoretica este necunoscuta, alegem testul T pentru o selectie de date.

3. Calculam valorile necesare statisticii corespunzatoare testului T:

   - $\bar{X}_n = \frac{X_1+\cdots+X_n}{n} = 79.167$ media setului de date.
   - $S_d^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n}(x_i - \bar{x}_n)^2 = 173.38$ este varianta setului de date. Astfel, devianta standard a setului de date este 13.167.

4. Calculam statistica $t = \frac{\bar{X}_n - \mu_0}{S_d} \cdot \sqrt{n} = \frac{79.167-70}{13.167} \cdot \sqrt{6} = 1.705$

5. Alegem nivelul de semnificatie α, determinand regiunea de respingere. Comparam valoarea t cu valoarea x care indeplineste $\mathbb{P}(T < x) > 1 - \alpha \Leftrightarrow \mathbb{P}(T \geq x) < \alpha$, unde T este repartitia Student cu $n - 1 = 6$ grade de libertate. Alegem $\alpha = 0.1$. In cazul de fata, din tabelul cu valorile cdf ale $T(5)$, obtinem ca $x = 2.1328$.

6. Pentru ca $t = 1.705 < 2.1328 = x$, acceptam ipoteza nula.

Exercitiul 3

Un studiu a fost efectuat pe 10 masini pentru a testa daca acestea au un kilometraj mai bun pe benzina premium sau normala. Fiecare din cele 10 masini a fost alimentata prima data cu benzina normala, ulterior cu benzina premium si kilometrajul(per litru) a fost masurat. Stiind ca s-a obtinut:

- Kilometraj pentru benzina normala: $16, 20, 21, 22, 23, 22, 27, 25, 27, 28$.
- Kilometraj pentru benzina premium: $19, 22, 24, 24, 25, 25, 26, 26, 28, 32$

determinati daca masinile obtin un kilometraj mai bun cu benzina premium. ***Solutie:***

1. Fixam ipotezele nula si alternativa:

   - $\mu_1 = \mu_2$
   - $\mu_1 \neq \mu_2$

2. Decidem tipul de test statistic pe care o sa-l utilizam. Intrucat marimea setului de date este $n = 10 < 30$, vom efectua testul T pentru 2 selectii de date.

3. Calculam valorile necesare statisticii corespunzatoare testului T cu 2 selectii:

   - $\bar{X} = \frac{X_1+\cdots+X_n}{n}$ media de selectie a primului set de date.

- $\bar{Y} = \frac{Y_1 + \cdots + Y_n}{n}$ media de selectie a celui de-al doilea set de date.
- $s_x^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x}_n)^2$ dispersia de selectie a primului set de date.
- $s_y^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (y_i - \bar{y}_n)^2$ dispersia de selectie a celui de-al doilea set de date.
- $s_p^2 = \left(\frac{1}{n} + \frac{1}{n}\right) \cdot \frac{(n-1)s_x^2 + (n-1)s_y^2}{n+n-1}$ este dispersia combinata de selectie a celor 2 seturi de date.

4. Calculam statistica $t = \frac{\bar{X} - \bar{Y}}{s_p} = 4.4721$ (inlocuim valorile).