Soluţii la examenul de Statistică - semianul A, iunie 2017

1. (3 p) Care dintre următoarele statistici este estimator nedeplasat şi pentru ce parametru al populaţiei?

☐ mediana	☒ dispersia eşantionului	☐ eroarea standard a mediei de selecţie
☒ dispersia populaţiei	☐ media populaţiei	☐ deviaţia standard a populaţiei

2. (6 p) Zece studenţi obţin la *Probabilităţi şi statistică* următoarele note: $8, 7, 10, 6, 7, 6, 8, 5, 7, 5$.
(a) (3p) Calculaţi două măsuri ale tendinţei centrale. (b) (3p) Există valori aberante (folosind regula 1.5 IQR)?

(a) Media $(8 + 7 + 10 + 6 + 7 + 6 + 8 + 5 + 7 + 5)/10 = 6.9$, modul $= 7$.
(b) Valorile ordonate sunt: $5, 5, 6, 6, 7, 7, 7, 8, 8, 10$. $Q_1 = 6, Q_3 = 8, IQR = 8 - 6 = 2$. Valorile care nu sunt aberante se găsesc în intervalul: $[3, 11]$. Astfel, nu există valori aberante.

3. (6 p) Se consideră o variabilă aleatoare continuă cu funcţia de densitate de probabilitate

$$f(x) = \begin{cases} \dfrac{\alpha}{x^3}, & x \in [1, +\infty) \\ 0, & \text{altfel} \end{cases}$$. Determinaţi $\alpha \in \mathbb{R}$, funcţia de repartiţie a lui X şi $P(X \geqslant 2)$.

$$1 = \int_{-\infty}^{+\infty} f(x)\,dx = \int_{1}^{+\infty} f(x)\,dx = -\frac{\alpha}{2x^2}\Big|_1^{+\infty} = \frac{\alpha}{2} \Rightarrow \alpha = 2.\ F(a) = \begin{cases} 0, & a < 1 \\ \dfrac{1}{a^2}, & a \geqslant 1 \end{cases},\ P(X \geqslant 2) = \int_2^{+\infty} f(x)\,dx = \frac{1}{4}.$$

4. (5 p) Se consideră un algoritm aleator Las Vegas, A, pentru rezolvarea unei probleme P (pentru intrarea x, algoritmul returnează $A(x)$, în timp ce răspunsul corect este $P(x)$). Se ştie că probabilitatea ca algoritmul să returneze *"nu ştiu"* este $1/3$. De câte ori trebuie rulat algoritmul, în mod independent, astfel ca probabilitatea de a primi răspunsul *"nu ştiu"* să fie cel mult $1/30$?

$$P(A_k(x) = \text{"nu ştiu"}) = [P(A(x) = \text{"nu ştiu"})]^k = \frac{1}{3^k} \leqslant \frac{1}{30} \Leftrightarrow k \geqslant [\log_3 30] \Leftrightarrow k \geqslant 4.$$

5. (12 p) *Institutul de diabet, nutriţie şi boli metabolice* susţine că 1 din 5 copii este supraponderal. Pentru a verifica această afirmaţie se alege un eşantion (aleator simplu) reprezentativ la nivel naţional format din 400 de copii. Dintre aceştia 95 sunt găsiţi supraponderali.
(a) (8p) Cum ar putea fi atacată afirmaţia acestui institut? Formulaţi ipotezele statistice şi întreprindeţi un test corespunzător de semnificaţie ($\alpha = 5\%$). $\text{qt}(0.05, 399) = -1.965$, $-\text{qnorm}(0.025) = 1.959$, $-\text{qnorm}(0.05) = 1.645$

Se aplică testul Z al proporţiilor $H_0 : p = 0.2 \quad H_a : p > 0.2$, scorul este $z = \dfrac{95/400 - 0.2}{\sqrt{0.2 \cdot 0.8/400}} = \dfrac{15}{8} = 1.875$, valoarea critică este $z^* = \text{qnorm}(0.95) = -\text{qnorm}(0.05) = 1.645$. Deoarece $z > z^*$ ipoteza nulă se poate respinge şi se acceptă că proporţia este mai mare decât 0.2

(b) (4p) Probabilitatea de a respinge afirmaţia institutului atunci când aceasta este de fapt adevărată este

☐ puterea testului ☐ probabilitatea erorii de tip II ☒ probabilitatea erorii de tip I

6. (5 p) Pentru o populaţie distribuită normal cu deviaţia standard $\sigma = 10$ se doreşte găsirea unui interval de încredere de 95% pentru media necunoscută a populaţiei. Cât de mare trebuie să fie dimensiunea eşantionului pentru ca intervalul rezultat să aibă o lungime de cel mult 2.8? ($\text{qnorm}(0.05) = -1.645$, $-\text{qnorm}(0.025) = 1.960$, $-\text{qnorm}(0.005) = 2.576$)

Pentru o lungime a intervalului de $w = 2.8$ avem $n = \left[(2z^*\sigma)^2/w^2\right] = \left[(2 \cdot 1.960 \cdot 10)^2/(2.8)^2\right] = 196$.

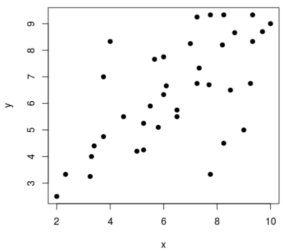
7. (10 p) O agenţie imobiliară susţine că un apartament scos la vânzare în Cityville este vândut în medie după 90 de zile. Pentru un eşantion aleator simplu de 16 apartamente se găseşte o medie de $\overline{x}_{16} = 75$ de zile cu deviaţia standard $s = 30$ de zile. Întreprindeţi un test de semnificaţie corespunzător acestor date (1%). (Numărul de zile până la vânzarea unui apartament scos pe piaţă urmează o lege normală.)

$\text{qt}(0.01, 15) = -2.602$, $-\text{qt}(0.05, 15) = 1.753$, $\text{qnorm}(0.01) = -2.326$, $-\text{qt}(0.005, 15) = 2.947$

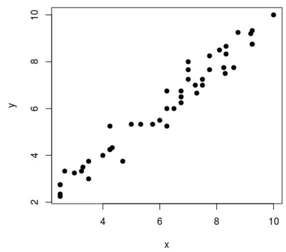
Se aplică testul T pentru media unei populaţii cu dispersia necunoscută: $H_0 : \mu = 90 \quad H_a : \mu < 90$, scorul este $t = \dfrac{75 - 90}{30/\sqrt{16}} = -2$, iar valoarea critică este $t^* = \text{qt}(0.01, 15) = -2.602$. Deoarece $t > t^*$ ipoteza nulă nu se poate respinge.

8. (8 p) Pentru diagramele de mai jos coeficienţii de corelaţie sunt $-1, -0.185, 0.637$ şi 0.977
(a) (4p) Găsiţi coeficientul corespunzător fiecărei diagrame în parte: (1): 0.637 (2): 0.977 (3): -0.185 (4): -1
(b) (4p) Care dintre diagrame corespunde asocierii negative perfecte şi care este dependenţa liniară a acestei asocieri?
Diagrama (4) corespunde unei asocieri liniare perfecte şi anume $X = -Y$.

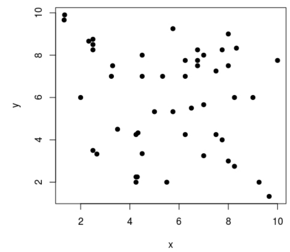
(1)


y
3
4
5
6
7
8
9
2
4
6
8
10
x

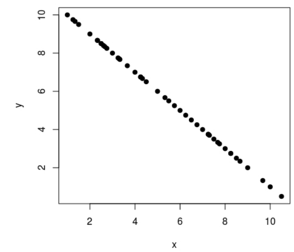
(2)


y
2
4
6
8
10
4
6
8
10
x

(3)


y
2
4
6
8
10
2
4
6
8
10
x

(4)


y
2
4
6
8
10
2
4
6
8
10
x