

Próba ZH feladatok

1. Egy munkaügyi állomány adatrendszerében minden dolgozóról az alábbi adatokat tárolják:

sorszám	Változónév	Definíció
1	NEV	A dolgozó neve
2	ANYA	Anyja neve
3	SZEV	A születés éve
4	SZHO	A születési hónap
5	SZNAP	A születési nap
6	ADO	Adóazonosító jel
7	TAJ	TAJ-szám
8	OTP	OTP bankszámla szám
9	FIZ1	Januári fizetés
:	:	:
20	FIZ12	Decemberi fizetés
21	JUT1	Januári jutalom
:	:	:
32	JUT12	Decemberi jutalom
33	PREM1	Januári prémium
:	:	:
44	PREM12	Decemberi jutalom
45	TAPNAP	A táppénzes napok száma
46	BEO	Munkaköri beosztás kódja
47	ISKOLA	Legmagasabb iskola kódja
48	NEM	A dolgozó neme
49	CSAL	Családi állapot
50	GYEREK	Eltartottak száma
51	CSALPOT	Családi pótlék

Összesen 384 dolgozó adatait tartalmazza a nyilvántartás.

- a.) Melyek a numerikus változók?
- b.) Melyek a nominális változók?
- c.) Melyek az ordinális változók?
- d.) Mennyi sora és oszlopa lesz az adatmátrixnak?
- e.) Hogyan és milyen módszert alkalmazva válaszolna arra a kérdésre, hogy a családok nők vajon több táppénzes napot vesznek-e ki, mint a nőtlen férfiak, vagy sem?

- 2.)
 - a.) Milyen kérdésre keressük a választ homogenitásvizsgálatkor?
 - b.) Mit jelent, hogy a minták függetlenek, illetve, hogy összetartozóak?
 - c.) Mondjon példát olyan statisztikai problémára, amikor kettőnél több összetartozó minta homogenitásáról kell dönten!
 - d.) Mondjon példát olyan statisztikai problémára, amikor kettőnél több független minta homogenitásáról kell dönten!
 - e.) Mi a próbastatisztika a kétmintás Kolmogorov-Szmirnov próbánál?

3.) Adott két gépsor napi leállásainak jegyzőkönyve. A LE1 változó azokat az x_1, x_2, \dots, x_n időpontokat jelöli, amikor az első gépsor valamilyen műszaki probléma miatt leállt. A UJRA1 pedig azokat az y_1, y_2, \dots, y_n időpontokat tartalmazza, amikor az első gépet újra lehetett indítani. Nyilván $8 \leq x_1 < y_1 \leq x_2 < y_2 \leq \dots \leq x_n < y_n \leq 16$. Ugyanezek az adatok a második gépsor esetében $8 \leq u_1 < v_1 \leq u_2 < v_2 \leq \dots \leq u_m < v_m \leq 16$. Ellenőrizni akarjuk, hogy a gépsorok leállásai exponenciális eloszlást követnek-e, és azt is, hogy a két gépsor leállásai azonos eloszlást követnek-e.

- a.) Milyen adatokkal számoljunk?
 - b.) Milyen módszerrel lehet az illeszkedést ellenőrizni?
 - c.) Milyen módszerrel lehet ellenőrizni az azonos eloszlást?
 - d.) Mivel lehet az exponenciális eloszlás paraméterét becsülni?
 - e.) Hogyan kell a jegyzőkönyv forrásadatait adatmátrixba rendezni?
- 4.
- a.) Sorolja fel, milyen megkötéseknek kell teljesülnie a függő változóra szórásanalízisnél!
 - b.) Mit jelent az, hogy a faktorok között interakciót tételezünk fel? Hogyan ellenőrizzük az interakció fennállását?
 - c.) Mit értünk kezelésen (treatment, cell)?
 - d.) Mi a különbség az egyszeres és a többszörös osztályozás között? Mondjon példát olyan statisztikai problémára, ahol az egyszeres, és olyanra, ahol a többszörös osztályozást kell alkalmazni!
 - e.) Mit értünk beállított és véletlen faktoron? Mondjon példát!
5. Az első feladatban leírt adatállomány esetén, az alábbi grafikonokat és statisztikákat melyik változóra van értelme kérni?
- a.) Oszlopdiagramot vagy kördiagramot.
 - b.) Steam-and-leaf diagrammot vagy hisztogramot.
 - c.) Kontingencia-táblázatot.
 - d.) Gauss papirosos normalitás grafikont (P-P plot).
 - e.) Ferdeséget, lapultságot.