

Alkalmazott statisztika gyakorló feladatok

Az elvégzett tesztek eredményeit és azok magyarázatait mentsük el egy valasz.txt, ha ábra is van, a valasz.xls nevű fájlba!

1. Nyissuk meg a kolcson.txt-t! Adjuk meg a fedezet gyakoriságait! Ábrázoljuk a fedezet gyakoriságait oszlopdiagramon és kördiagramon is!
2. Nyissuk meg az emberek.txt-t! Ábrázoljuk a magasság sűrűséghisztogramját! Exportáljuk ki csak a diagramot egy html file-ba!
3. Nyissuk meg az emberek.txt-t! Hozzunk létre egy ttidx változót, amely az emberek testtömegindexét tartalmazza! (testtömegindex: testtömeg osztva a testmagasság négyzetével kg/m^2) Hozzunk létre súlykat változót, amely értéke 1, ha sovány ($ttidx < 20$), 2, ha normál testtömegű ($ttidx$ 20 és 25 között) és 3, ha túlsúlyos ($ttidx > 25$)! Ábrázoljuk oszlopdiagramon az átlag életkorokat a súlykat változó szerint csoportosítva!
4. Nyissuk meg az emberek.txt-t! Hozzunk létre egy új változót, amely az emberek életkorát tartalmazza! Adjunk becslést azon jellemzők (változók) mediánjára, amelyekre lehet!
5. Nyissuk meg az emberek.txt-t! Ábrázoljuk a férfiak és nők testmagasságának mediánjának becslését oszlopdiagramon! Ábrázoljuk boxploton is a két csoport testmagasságát!
6. Nyissuk meg az emberek.txt-t! Adjunk becslést a férfiak és nők testmagasságának várható értékekre! Ábrázoljuk ezeket az értékeket oszlopdiagramon!
7. Nyissuk meg a kolcson.txt-t! Adjunk becslést azon jellemzők (változók) várható értékére és szórására, amelyekre lehet! A várható értékekre adjunk 99%-os megbízhatósági szintű konfidencia-intervallumot is!
8. Nyissuk meg a brain.sav-ot! Teszteljük azt a nullhipotézist, hogy az intelligencia (FSIQ) várható értéke 100.
9. Nyissuk meg a kolcson.txt-t! Teszteljük le azt a nullhipotézist, hogy a 40 év felettiek jövedelme várhatóan 35. Teszteljük le ugyanezt a hipotézist az egész mintán!
10. Nyissuk meg a brain.sav-ot! Teszteljük azt a nullhipotézist, hogy a verbális intelligencia (VIQ) és performációs intelligencia (PIQ) várhatóan nem tér el egymástól! Teszteljük a két változó különbségének várható értékének 0 voltát egymintás t-próbával is és ellenőrizzük, hogy a két eljárás ugyanazt az eredményt adja!
11. Nyissuk meg a brain.sav-ot! Teszteljük azt a nullhipotézist, hogy a férfiak és nők intelligenciájának (FSIQ) várható értéke megegyezik. Teszteljük azt a nullhipotézist is, hogy a férfiak esetén ez a várható érték 5-tel magasabb! Magyarázzuk meg a kapott eredményeket!
12. Nyissuk meg a kolcson.txt-t! Teszteljük, hogy a házasok és nem házasok munkaviszonyának várható értéke nem tér el egymástól!

13. Nyissuk meg a kolcson.txt-t! Adjunk becslést a legalább 40 évesek és a 40 alattiak várható jövedelmére! Ellenőrizzük azt a nullhipotézist, amely szerint ezek nem térnek el egymástól! Adjunk 95%-os megbízhatósági szintű konfidencia-intervallumot a két csoport várható jövedelmének különbségére!
14. Nyissuk meg a kolcson.txt-t! Tegyük fel, hogy a régi és új ügyfelek csoportjaiban a jövedelmek normális eloszlásúak. Teszteljük, hogy a régi és új ügyfelek jövedelmének szórása nem tér el egymástól!
15. Nyissuk meg a calc.sav-ot! Itt kalcium hatását vizsgálják a vérnyomásra. A kutató arra kíváncsi, van-e hatása a kezelésnek. Milyen próbákat alkalmazhatunk? Mik az adott próbák feltételei? Írjuk fel a nullhipotézist és az ellenhipotézist is minden esetben!
16. Nyissuk meg a kolcson.txt-t! Tegyük fel, hogy a különböző családi állapotúak csoportjaiban az életkorok normális eloszlásúak. Teszteljük azt a nullhipotézist, hogy a várható életkor megegyezik minden csoportban. Ha elvetjük a nullhipotézist, végezzünk kiegészítő vizsgálatot is!
17. Nyissuk meg a kolcson.txt-t! Adjunk becslést a korábbi ügyfelek várható életkorára! Adjunk becslést azok várható életkorára is, akik nem voltak korábban ügyfelek! Ellenőrizzük kétmintás t-próbával azt a nullhipotézist, hogy a két csoport várható életkora nem tér el egymástól! Más eredményt kapunk-e, ha ANOVA-val teszteljük ugyanezt?
18. Nyissuk meg az emberek.txt-t! Mely intervallumváltozók között lineáris van kapcsolat? Adjuk meg a legerősebb kapcsolathoz tartozó regressziós egyenes paramétereit! Ábrázoljuk az egyik változót a másik függvényében, az ábrán tüntessük fel a regressziós egyenest is! Adjunk becslést ezen modell segítségével a hiányzó adatokra!
19. Nyissuk meg az emberek.txt-t! A testtömegindex kvadratikusan kapcsolatra utal a magasság és súly között, adjunk becslést erre a kapcsolatra!
20. Nyissuk meg a chemical.txt-t! Címkezzük fel az x1 változót a "tartályok száma", az x2 változót a "hajórakomány súlya" és az y változót a "kirakodás ideje percekben" címkékkel! Ábrázoljuk az y értékeit az x1 és x2 függvényében! Adjunk többváltozós regressziós modellt y függő és x1 és x2 független változókra (regressziós sík)!
21. Nyissuk meg a kolcson.txt-t! Teszteljük azt a nullhipotézist, hogy az életkor normális eloszlású!
22. Nyissuk meg a kolcson.txt-t! Teszteljük azt a nullhipotézist, hogy a különböző státuszok azonos valószínűségűek! Teszteljük azt a nullhipotézist is, hogy a status 1-es értékének valószínűsége $1/2$, a 2-es és 3-as érték pedig $1/4$ - $1/4$ valószínűségű!
23. Nyissuk meg a kolcson.txt-t! Teszteljük azt a nullhipotézist, hogy azok, akik korábban ügyfelek voltak és akik nem voltak (adostip) életkorának eloszlása megegyezik!
24. Nyissuk meg a kolcson.txt-t! Teszteljük azt a nullhipotézist, hogy azok, akik korábban ügyfelek voltak és akik nem voltak (adostip) családi állapotának eloszlása megegyezik! Mely tesztek használhatók a kérdés eldöntésére?

25. Készítsünk gyakorisági táblázatot a kolcson.txt adóstípus és a kölcsön típusa változók alapján. Végezzük el a χ^2 próbát. Milyen következtetést von le az eredmény alapján 0,05-ös szinten? Mi itt a nullhipotézis?
26. Nyissuk meg a kolcson.txt-t! Mondhatjuk-e, hogy a statusok eltérnek a régi és új ügyfeleknél?
27. Nyissuk meg a beteg.sav-ot! A következő kérdést vizsgáljuk: Egyformán reagálnak-e a kezelésre a férfiak és a nők? Írjuk fel a nullhipotézist, és végezzük el minden erre vonatkozó vizsgálatot, amit lehet!
28. Nyissuk meg a kolcson.txt-t! Hozzunk létre egy új változót "összjov" néven, amely tartalmazza a jelenlegi munkahelyen az eddigi összes jövedelmet (éves jövedelem szorozva azzal, hány éve dolgozik jelenlegi munkahelyén). Van-e összefüggés a kölcsön típusa és az összjövedelem közt?
29. Nyissuk meg a kolcson.txt-t! A változók közül melyek tekinthetők függetlennek?
30. Nyissuk meg a brain.sav-ot! Az intervallumváltozókon alkalmazott főfaktor analízissel hozzunk létre három változót és mentjük el őket! Adjuk meg az így kapott faktorok korrelációs mátrixát!
31. Nyissuk meg a brain.sav-ot! Az intervallumváltozók alapján (korrelációs mátrixszal számolva) a Kaiser kritériumot használva (legalább 1 sajátértékhez tartozó vektorok kiválasztása) csökkentsük a változók számát! Hány faktort kapunk?
32. Nyissuk meg a brain.sav-ot! Az intervallumváltozók alapján hozzunk létre 3 faktort! A faktorok értelmezésének könnyítése érdekében alkalmazzunk forgatást!
33. Nyissuk meg a brain.sav-ot! Határozzuk meg a nemeket elválasztó diszkriminanciafüggvényt (adjuk meg az együtthatókat) az előző feladatbeli faktorok által meghatározott térben! Hány esetben hibás az előrejelzett csoportbatartozás?
34. Nyissuk meg a birth.sav-ot! Egy anya nem adta meg, dohányzik-e! Adjunk előrejelzést arra, hogy a dohányzók csoportjába tartozik-e! A használt módszer hány esetet klasszifikált rosszul?
35. Nyissuk meg az országok.xls-t! A Manhattan-távolságot (Minkowski-távolság 1 hatvánnyal) használva a legközelebbi szomszéd módszerével klaszterezzük hierarchikusan az országokat! Olvassuk le a dendrogramról, mely ország alkot önmagában egyelemű klasztert a legtávolabb!
36. Nyissuk meg a brain.sav-ot! Csoportosítsuk k-közép módszerrel az embereket két csoportba FSIQ, VIQ, PIQ, Weight, Height változók alapján! Adjuk meg a klaszterközéppontokat! Hány eset tartozik egy-egy klaszterbe?
37. Nyissuk meg a brain.sav-ot! Klaszterezzük hierarchikusan az embereket centroid módszerrel az euklideszi távolságot használva az intervallumváltozók (scale változók) értékei szerint! Legalább mekkora klaszterszám esetén lesz olyan ember, aki önmagában egy klasztert alkot, ki az?