

R gyakorló feladatok

1. Nyissuk meg a kolcson.sav-ot! Adjuk meg a fedezet gyakoriságait és relatív gyakoriságait! Ábrázoljuk a fedezet gyakoriságait oszlopdiagramon és kördiagramon is! Mentjük az ábrákat pdf fájlba!
2. Nyissuk meg az emberek.txt-t! Ábrázoljuk a magasság sűrűséghisztogramját! Mentjük el a diagramot egy file-ba!
 - * 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) Ábrázoljuk boxploton a testtömegindexet!
3. Nyissuk meg az emberek.txt-t! Adjunk becslést azon jellemzők (változók) mediánjára és kvartiliseire, amelyekre értelmes! Ábrázoljuk ezeket boxploton is!
4. Nyissuk meg az emberek.txt-t! Ábrázoljuk a férfiak és nők testmagasságának kvartiliseinek becslését boxploton!
5. Nyissuk meg az emberek.txt-t! Adjunk becslést a férfiak és nők testmagasságának várható értékére! * Ábrázoljuk ezeket az értékeket oszlopdiagramon!
6. Nyissuk meg a kolcson.sav-ot! Adjunk becslést azon jellemzők (változók) várható értékére és szórására, amelyekre lehet! A várható értékekre adjunk 99%-os megbízhatósági szintű konfidencia-intervallumot is!
 - * Nyissuk meg a fold.xls-t! Mennyi az egy főre jutó nemzeti összetermék a Földön?
7. Nyissuk meg a brain.sav-ot! Teszteljük azt a nullhipotézist, hogy az intelligencia (FSIQ) várható értéke 100.
 - * Nyissuk meg a kolcson.sav-ot! Feltételezve, hogy a jövedelem normális eloszlást követ teszteljük le azt a nullhipotézist, hogy a jövedelme várhatóan 35.
8. 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!
9. 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!
10. Nyissuk meg a brain.sav-ot! Igaz-e, hogy a férfiak és nők magasságának várható értéke megegyezik?
11. Nyissuk meg a kolcson.sav-ot! Eltér-e a régi és új ügyfelek munkaviszonyának várható értéke (feltételezve, hogy a munkaviszonyok normális eloszlást követnek)?
 - * Nyissuk meg a kolcson.sav-ot! Adjunk becslést a legalább 40 évesek és a 40 alattiak várható jövedelmére! Ellenőrizzük azt a nullhipotézist, amely szerint ezek nem térnek el egymástól (feltételezve, hogy a jövedelem változók normális eloszlást követnek)! Adjunk 95%-os megbízhatósági szintű konfidencia-intervallumot a két csoport várható jövedelmének különbségére!

12. Nyissuk meg a kolcson.sav-ot! 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!
 - * 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!
13. Nyissuk meg a kolcson.sav-ot! 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!
14. Nyissuk meg a kolcson.sav-ot! Adjunk becslést a korábbi ügyfelek várható életkorára! Adjunk becslést azok várható életkorára is, akik nem voltak korábban ügyfelek! Feltételezve, hogy az életkorok normális eloszlást követnek 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?
15. Nyissuk meg a fold.xls-t! Mely skála változók között van a legerősebb lineáris kapcsolat? (A country és regio változók figyelmen kívül hagyásával.)
16. Nyissuk meg az emberek.txt-t! Mely skála változók 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!
 - * Nyissuk meg az emberek.txt-t! A testtömegindex kvadratikus kapcsolatra utal a magasság és súly között, adjunk becslést erre a kapcsolatra!
 - * Nyissuk meg a chemical.txt-t! (Az x1 változó a tartályok száma, az x2 változó a hajórakomány súlya és az y változó a kirakodás ideje percekben.) Ábrázoljuk az y értékeit az x1 és x2 függvényében! Adjunk többváltozós regressziós modellt y függő és x1 és x2 független változókra (regressziós sík)!
17. Nyissuk meg a kolcson.sav-ot! Teszteljük azt a nullhipotézist, hogy az életkor normális eloszlású!
18. Nyissuk meg a kolcson.sav-t! Teszteljük azt a nullhipotézist, hogy a különböző statuszok 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!
 - * Nyissuk meg a kolcson.sav-ot! Teszteljük azt a nullhipotézist, hogy azok, akik korábban ügyfelek voltak és akik nem voltak (adostip) életkorának eloszlása megegyezik!
 - * Nyissuk meg a kolcson.sav-ot! Teszteljük azt a nullhipotézist, hogy azok, akik korábban ügyfelek voltak és akik nem voltak (adostip) családi állapotának eloszlása megegyezik! Mely tesztek használhatjuk a kérdés eldöntésére?
19. Készítsünk gyakorisági táblázatot a kolcson.sav 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?

20. Nyissuk meg a kolcson.sav-ot! Mondhatjuk-e, hogy a statusok eltérnek a régi és új ügyfeleknél?
- * 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 végezzünk el minden erre vonatkozó vizsgálatot, amit lehet!
 - * Nyissuk meg a kolcson.sav-ot! 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?
21. Nyissuk meg a kolcson.sav-ot! A változók közül melyek tekinthetők függetlennek?
22. Nyissuk meg a brain.sav-ot! Az intervallumváltozókon alkalmazott főkomponens-analízissel hozzunk létre három változót és mentsük el őket! Adjuk meg az így kapott főkomponensek korrelációit és kovarianciáit!
23. 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 változót kapunk?
- * Nyissuk meg a brain.sav-ot! Az intervallumváltozók alapján hozzunk létre 3 faktort! A faktorok értelmezésének könnyítése érdekében alkalmazzunk forgatást!
24. Nyissuk meg a brain.sav-ot! Határozzuk meg a nemeket elválasztó diszkriminanciafüggvényt (adjuk meg az együtthatókat)! Hány esetben hibás az előrejelzett csoportbatartozás?
25. Nyissuk meg az autok.txt-t! Szeparáljuk a kategóriákat az első két főkomponens által meghatározott térben! Hány esetben hibás az előrejelzett csoportbatartozás?
26. Nyissuk meg 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?
27. Nyissuk meg az orszagok.xls-t! A legközelebbi szomszéd módszerével klaszterezzük hierarchikusan az országokat! Olvassuk le a dendrogramról, mely ország alkot önmagában egyelemű klasztert a legtovább!
28. Nyissuk meg a fold.xls-t! Klaszterezzük hierarchikusan az országokat az átlagos távolsággal a skála változók segítségével! Két klasztert meghatározva hány elemű lesz a kisebbik?
29. 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?
30. Nyissuk meg a brain.sav-ot! Klaszterezzük hierarchikusan az embereket centroid módszerrel az euklideszi távolságot használva a skála változók értékei szerint! Legalább mekkora klaszterszám esetén lesz olyan ember, aki önmagában egy klasztert alkot, ki az?