

1. **Olvassuk be a sleep adatbázist a datasets csomagból. Teszteljük 5%-os szignifikancia szinten azt a nullhipotézist, hogy az extra alvásidő (extra) várható értéke független a csoporttól (group). Az alábbi állítások közül melyik igaz?**  
 $p > 0,05$ , ezért nem tudjuk elutasítani a nullhipotézist, nem tudjuk bizonyítani, hogy az extra alvásidő várható értéke függ a csoporttól
2. **Olvassuk be a faithful adatbázist a datasets csomagból, és írjuk fel a két gejzírkitörés között eltelt idő hosszát (waiting) a gejzír kitörések hosszának (eruptions) lineáris függvényeként. Várhatóan milyen hosszú a várakozási idő 2 perc hosszúságú vulkánkitörés után? Adjuk meg két tizedes jegyre kerekítve.**  
54,94
3. **Hogyan határozzuk meg két klaszter távolságát a legközelebbi szomszéd módszer (single-linkage) alapján?**  
Az egymáshoz legközelebb eső elemeik távolsága.
4. **Olvassuk be a datasets csomagból az USJudgeRatings adatbázist. Ha az adatbázist csak két korrelálatlan változóval jellemezzük, hány százalék az információvesztés mértéke? Két tizedes jegyre kerekítve adjuk meg százalékkal.**  
6,35
5. **Az alábbi állítások közül melyik hamis? (főkomponens)**  
A legutolsó főkomponens magyarázza a variancia legnagyobb részét.
6. **Az alábbiak közül melyik mennyiség adja meg, hogy a független változó milyen mértékben magyarázza a függő változó ingadozását?**  
A Pearson-féle korrelációs együttható négyzete.
7. **Olvassuk be az Anscombe adatbázist a car csomagból. Standardizáljuk a folytonos változókat, majd csoportosítsuk a megfigyeléseket a rendelkezésre álló folytonos változók alapján úgy, hogy jól szeparált klasztereket kapjunk. Melyik az az egyetlen megfigyelés, mely teljesen elkülönül a többitől? Idézőjelek nélkül add meg (pl. LA)**  
AK
8. **Az alábbiak közül melyik igaz főkomponens analízis esetén?**  
A főkomponensek összegzik az eredeti adathalmaz varianciáját.
9. **Olvassuk be az Angell adatbázist a car csomagból. Soroljuk be a megfigyeléseket a region változó által meghatározott csoportokba. Térjünk át arra az új koordinátarendszerre, melyben a lehető legjobban szétválnak a csoportok. Adjuk meg a hetero változó első dimenzióhoz tartozó együtthatójának abszolút értékét három tizedes jegyre kerekítve.**  
0,053

- 10. Két folytonos változó közötti lineáris kapcsolat vizsgálatakor milyen módszer segítségével tudjuk meghatározni, hogy a független változó milyen mértékben magyarázza a függő változó ingadozását?**  
Regressziószámítás
- 11. Olvassuk be a datasets csomagból a USArrests adatbázist. Jellemezzük az adatbázist kevesebb változóval. Hogyan áll elő a legfontosabb új változó az eredeti változók lineáris kombinációjaként? Három tizedes jegyre kerekítve adjuk meg az UrbanPop változó együttthatójának abszolút értékét.**  
0,278
- 12. A függetlenségvizsgálatra vonatkozó khi-négyzet próba alapján levont következtetés nem megbízható, ha:**  
a várt gyakoriságok között van 5-nél kisebb.
- 13. A khi-négyzet próbák:**  
a megfigyelt és a nullhipotézis fennállása esetén várt gyakoriságokat hasonlítják össze.
- 14. Olvassuk be a state.x77 adatbázist a datasets csomagból. Standardizáljuk a folytonos változókat, majd csoportosítsuk a megfigyeléseket a rendelkezésre álló folytonos változók alapján úgy, hogy jól szeparált klasztereket kapjunk. Mekkora távolságnál került sor a legelső összevonásra? Adjuk meg két tizedes jegyre kerekítve.**  
0,49
- 15. Milyen módszert alkalmaz, ha egy diszkrét változó értékét szeretné megbecsülni több folytonos változóhoz tartozó megfigyelés alapján?**  
Diszkriminancia analízis
- 16. Milyen módszert alkalmaz, ha egy folytonos változó értékét szeretné megbecsülni egy másik folytonos változó bizonyos értékéhez?**  
Regressziószámítás
- 17. Az alábbiak közül melyik igaz divizív hierarchikus klaszterezési eljárásokra?**  
Kezdetben minden megfigyelés egyetlen klaszterbe kerül, majd ezt a klasztert minden lehetséges módon felbontjuk két részre, és azt a felbontást tartjuk meg, ahol a klaszterek közötti távolság a legnagyobb.
- 18. Olvassuk be az attitude adatbázist a datasets csomagból. Írjuk fel a rating változót a complaints változó lineáris függvényeként. Várhatóan mekkora lesz a rating változó értéke a complaints változó 71-es értéke mellett? Adjuk meg két tizedes jegy értékre kerekítve.**  
67,95
- 19. Egyszerű lineáris regresszió esetén az egyenes meredekségére vonatkozó hipotézisvizsgálat megegyezik:**  
a Pearson-féle korrelációs együttthatóra vonatkozó hipotézisvizsgálattal.

**20. Az alábbi állítások közül melyik igaz a függetlenségvizsgálatra vonatkozó khi-négyzet próba esetén?**

Ha  $p < \alpha$ , akkor elvetjük a nullhipotézist, a két diszkrét változó nem független.

**21. Milyen módszert alkalmaz, ha egy adatbázis folytonos változóinak számát szeretné csökkenteni?**

Főkomponens analízis

**22. Olvassuk be a state.x77 adatbázist a datasets csomagból. Standardizáljuk a folytonos változókat, majd csoportosítsuk a megfigyeléseket a rendelkezésre álló folytonos változók alapján úgy, hogy jól szeparált klasztereket kapjunk. Melyik az egyetlen megfigyelés, mely teljesen elkülönül a többitől? Idézőjelek nélkül adjuk meg (pl. Colorado)**

Alaska

**23. Milyen módszert használ annak a nullhipotézisnek a tesztelésére, hogy az átlagos fizetés független az iskolai végzettségtől?**

Egyszempontos ANOVA

**24. Olvassuk be a datasets csomagból az USJudgeRatings adatbázist. Ha az adatbázist csak három korrelálatlan változóval jellemezzük, hány százalék az információvesztés mértéke? Két tizedes jegy értékre kerekítve adjuk meg százalékjel nélkül.**

3,58

**25. Milyen módszert használ annak a nullhipotézisnek a tesztelésére, hogy a fizetés és az alkalmazásban eltöltött idő között nincs lineáris kapcsolat?**

t-próba a korrelációs együtthatókra

**26. Olvassuk be az InsectSprays adatbázist a datasets csomagból. Teszteljük 5%-os szignifikancia szinten azt a nullhipotézist, hogy a rovarok számának (count) várható értéke független a rovarirtó szer típusától (spray). Az alábbi állítások közül melyik igaz?  $p < 0,05$  ezért elutasítjuk a nullhipotézist, a rovarok számának várható értéke függ a rovarirtó szer típusától**

**27. Milyen módszert alkalmaz, ha a megfigyeléseket csoportosítani szeretné előre adott osztályokba?**

Diszkriminancia analízis

**28. Hogyan határozzuk meg két klaszter távolságát a legtávolabbi szomszéd módszer (complete-linkage) alapján?**

Az egymástól legtávolabb eső elemek távolsága.

**29. Olvassuk be a PlantGrowth adatbázist a datasets csomagból. Teszteljük 5%-os szignifikancia szinten azt a nullhipotézist, hogy az átlagos tömeg (weight) független a csoporttól (group). Az alábbi állítások közül melyik igaz?**

$p < 0,05$  ezért elutasítjuk a nullhipotézist, az átlagos tömeg függ a csoporttól

- 30. Olvassuk be az Angell adatbázist a car csomagból. Soroljuk be a megfigyeléseket a region változó által meghatározott csoportokba. Az első megfigyelés mekkora valószínűséggel került a modell által besorolt csoportba? Adjuk meg két tizedes jegyre kerekítve.**  
0,98
- 31. Olvassuk be a Loblolly adatbázist a datasets csomagból, és írjuk fel a fák magasságát (height) az életkor (age) lineáris függvényeként. Két tizedes jegyre kerekítve adjuk meg az egyenes meredekségének a becslését.**  
2,59
- 32. Olvassuk be a datasets csomagból az USJudgeRatings adatbázist. Jellemezzük az adatbázist kevesebb változóval. Hogyan áll elő a legfontosabb új változó az eredeti változók lineáris kombinációjaként? Három tizedes jegyre kerekítve adjuk meg az DILG változó együttthatójának abszolút értékét.**  
0,304
- 33. Olvassuk be az Anscombe adatbázist a car csomagból. Standardizáljuk a folytonos változókat, majd csoportosítsuk a megfigyeléseket a rendelkezésre álló folytonos változók alapján úgy, hogy jól szeparált klasztereket kapjunk. Melyik az az egyetlen megfigyelés, mely teljesen elkülönül a többitől? Idézőjelek nélkül adjuk meg (pl. LA) AK**
- 34. Olvassuk be a survey adatbázist a MASS csomagból. Teszteljük 5%-os szignifikancia szinten azt a nullhipotézist, hogy a dohányzási szokás (Smoke) független a nemtől (Sex). Az alábbi állítások közül melyik igaz?**  
 $p > 0,05$  ezért nem tudjuk elutasítani a nullhipotézist, nem tudjuk bizonyítani, hogy a dohányzási szokás függ a nemtől.
- 35. Az alábbiak közül melyik mennyiség adja meg, hogy a független változó milyen mértékben magyarázza a függő változó ingadozását?**  
A Pearson-féle korrelációs együtttható négyzete.
- 36. Az alábbiak közül melyik hamis a diszkriminancia analízis esetén?**  
Eredménye alapján megjósolható egy folytonos változónak egy másik folytonos változóhoz tartozó értéke.
- 37. Olvassuk be az Angell adatbázist a car csomagból. Soroljuk be a megfigyeléseket a region változó által meghatározott csoportokba. Térjünk át arra az új koordináta-rendszerre, melyben a lehető legjobban szétválnak a csoportok. Adjuk meg az első megfigyelés y-koordinátáját ebben az új koordináta-rendszerben.**  
1,02523923
- 38. Milyen módszert használ ha függvényyszerű kapcsolatokat keres két folytonos változó között?**  
Regressziószámítás

39. Olvassuk be a faithful adatbázist a datasets csomagból, és írjuk fel a két gejzírkitörés között eltelt idő hosszát (waiting) a gejzír kitörések hosszának (eruptions) lineáris függvényeként. Várhatóan milyen hosszú a várakozási idő 2 perc hosszúságú vulkánkitörés után? Adjuk meg két tizedes jegyre kerekítve.

54,94

40. Olvassuk be a datasets csomagból a USArrests adatbázist. Jellemezzük az adatbázist kevesebb változóval. Hogyan áll elő a legfontosabb új változó az eredeti változók lineáris kombinációjaként? Három tizedes jegyre kerekítve adjuk meg a Murder változó együttthatójának abszolút értékét.

0,536

41. Két folytonos változó közötti lineáris kapcsolat vizsgálatakor milyen módszer segítségével tudjuk meghatározni, hogy a független változó milyen mértékben magyarázza a függő változó ingadozását?

Regressziószámítás

42. Olvassuk be a faithful adatbázist a datasets csomagból, és írjuk fel a két gejzírkitörés között eltelt idő hosszát (waiting) a gejzír kitörések hosszának (eruptions) lineáris függvényeként. Az eruptions változó hány százalékban magyarázza meg a waiting változót? Két tizedes jegyre kerekítve adjuk meg százalékkal nélkül.

81,12

43. Olvassuk be a datasets csomagból a stackloss adatbázist. Hány korrelálatlan változóval jellemezhető az adatbázis, ha legfeljebb 10 százalékos információvesztést szeretnénk elérni?

2

44. Olvassuk be az Angell adatbázist a car csomagból. Soroljuk be a megfigyeléseket a region változó által meghatározott csoportokba. A tizedik megfigyelés mekkora valószínűséggel került a modell által besorolt csoportba? Adjuk meg két tizedes jegyre kerekítve.

0,82

45. Olvassuk be az airquality adatbázist a datasets csomagból. Teszteljük 5%-os szignifikancia szinten azt a nullhipotézist, hogy az ózon mennyiség (Ozone) várható értéke független a hónaptól (Month). Az alábbi állítások közül melyik igaz?

$p < 0,05$  ezért elutasítjuk a nullhipotézist, az ózon mennyiség várható értéke függ a hónaptól

46. Milyen módszert alkalmaz, ha egy folytonos változó értékét szeretné megbecsülni egy másik folytonos változó bizonyos értékéhez?

Regressziószámítás

47. Milyen módszert alkalmaz, ha a megfigyeléseket előre nem adott osztályokba szeretné csoportosítani?

Klaszteranalízis

- 48. Milyen módszert alkalmaz két diszkrét változó függetlenségének tesztelésére?**  
Khi-négyzet-próba
- 49. A Pearson-féle korrelációs együttható -1-hez közeli értéke:**  
erős csökkenő lineáris kapcsolatot jelez
- 50. Olvassuk be az Angell adatbázist a car csomagból. Soroljuk be a megfigyeléseket a region változó által meghatározott csoportokba. Térjünk át arra az új koordinátarendszerre, melyben a lehető legjobban szétválnak a csoportok. Adjuk meg az első megfigyelés x-koordinátájának abszolút értékét ebben az új koordinátarendszerben.**  
3,71291656
- 51. Az alábbi állítások közül melyik igaz az egyszempontos ANOVA esetén?**  
ha  $p < \alpha$  akkor elutasítjuk a nullhipotézist, nem minden várható érték egyezik meg
- 52. Olvassuk be a survey adatbázist a MASS csomagból. Teszteljük 5%-os szignifikancia szinten azt a nullhipotézist, hogy a testmozgás gyakorisága (Exer) független a nemtől (Sex). Az alábbi állítások közül melyik igaz?**  
 $p > 0,05$  ezért nem tudjuk elutasítani a nullhipotézist, nem tudjuk bizonyítani, hogy a testmozgás gyakorisága függ a nemtől
- 53. Olvassuk be a datasets csomagból az USArrests adatbázist. Ha az adatbázist csak két korrelálatlan változóval jellemezzük, hány százalék az információvesztés mértéke? Két tizedes jegyre kerekítve adjuk meg százalékjel nélkül**  
13,25
- 54. Olvassuk be az attitude adatbázist a datasets csomagból. Írjuk fel a rating változót a complaints változó lineáris függvényeként. A complaints változó hány százalékban magyarázza meg a rating változót?**  
68,13
- 55. A khi-négyzet próba nullhipotézise:**  
Két diszkrét változó független.
- 56. Milyen próbát használ annak vizsgálatára, hogy az önkéntes munka vállalása függ-e a nemtől?**  
Khi-négyzet próba függetlenségvizsgálatra
- 57. Olvassuk be az Orange adatbázist a datasets csomagból, és írjuk fel a fák kerületét (circumference) az életkor (age) lineáris függvényeként. Két tizedes jegy értékre kerekítve adjuk meg az egyenes meredekségének a becslését.**  
0,11

**58. Olvassuk be a datasets csomagból az USArrests adatbázist. Jellemezzük az adatbázist kevesebb változóval. Hogyan áll elő a legfontosabb új változó az eredeti változók lineáris kombinációjaként? Három tizedes jegyre kerekítve adjuk meg az Assault változó együttthatójának abszolút értékét.**

0,583

**59. Milyen módszer segítségével tud a megosztott figyelemre, a figyelem pontosságára, a kockázatvállalási hajlandóságra és az észlelés gyorsaságára vonatkozó folytonos változók alapján előrejelzést adni arra, hogy egy adott személynek lesz-e balesete vagy sem?**

Diszkriminancia analízis

**60. Olvassuk be a Loblolly adatbázist a datasets csomagból, és írjuk fel a fák magasságát (height) az életkor (age) lineáris függvényeként. Várhatóan mekkorra lesz egy 12 éves fa magassága? Adjuk meg két tizedes jegyre kerekítve.**

29,77

**61. Olvassuk be a datasets csomagból az USArrests adatbázist. Ha az adatbázist csak három korrelálatlan változóval jellemezzük, hány százalék az információvesztés mértéke? Két tizedes jegyre kerekítve adjuk meg százalékkal nélkül (pl. 5,32).**

0,04

**62. Olvassuk be az Angell adatbázist a car csomagból. Soroljuk be a megfigyeléseket a region változó által meghatározott csoportokba. Térjünk át arra az új koordinátarendszerre, melyben a lehető legjobban szétválnak a csoportok. Adjuk meg a mobility változó első dimenzióhoz tartozó együttthatójának abszolút értékét három tizedes jegyre kerekítve.**

0.12846644