

Statisztikai programcsomagok

gyakorló feladatok

A tesztelési feladatoknál a teljes pontszámért a következőket kell megadni: melyik próbát alkalmazzuk, mennyi a p -érték, illetve elfogadjuk vagy elvetjük a nullhipotézist. Ha a próbának van valamilyen előfeltétele, akkor azt is tessék ellenőrizni.

A `cars.txt` a '80-as években az amerikai piacon árúsított néhány autótípus fontosabb műszaki adatait tartalmazza. A változók:

MANUFACTURER: az autó gyártója

MODEL: az autó modellje

COUNTRY: származási ország

VOL: csomagtér térfogata (köbláb)

HP: teljesítmény (lóerő)

SP: végsebesség (mérőöld/óra)

MPG: fogyasztás (mérőöld/gallon)

WT: jármű tömege (100 font)

1. Adjon becslést a végsebesség várható értékére és szórására.
2. Tesztelje 5%-os szignifikancia szinten azt, hogy a végsebesség várható értéke 112 mérőöld/óra. Adjon 95%-os megbízhatósági szintű konfidencia intervallumot a végsebesség várható értékére.
3. Adjon 99% megbízhatósági szintű konfidencia intervallumot a végsebesség és a teljesítmény várható értéke közötti különbségre. (... mely értéknek a gyakorlatban semmi értelme, de ez kit érdekel.)
4. Tesztelje 10%-os szignifikancia szinten azt, hogy az amerikai és a japán autóknál azonos a fogyasztás elméleti várható értéke. Adjon 90% megbízhatóságú konfidencia intervallumot a várható értékek különbségére.
5. Függ az autók teljesítménye a származási országtól? ($\alpha = 0,05$)
6. Független a végsebesség a teljesítménytől? Ha nem, akkor végezzünk lineáris regressziót a két változóra. Adjuk meg a regressziós egyenes egyenletét is. Mennyire jó a regressziós egyenes illeszkedése? Várhatóan mekkora a végsebessége egy 120 lóerős autónak? ($\alpha = 0,01$)
7. Független egymástól az autó gyártója és a származási ország? (Elég hülye kérdés, de ezt a próbát is gyakorolni kell.) A szignifikancia szint legyen 10%.

A további feladatok azon magasabb dimenziós módszerekhez tartoznak, melyeket csak a Statisztikai programcsomagok kurzuson vettünk.

A `TOP.txt` néhány globális nagyvállalat fontosabb pénzügyi adatait tartalmazza. (Az adatsor régi, több évtizedes.) A változók:

8. Végezzünk főkomponens-analízist a következő változókon:

Vagyon, Eladas, Piaci_ertek, Profit, Cash_flow.

Hogyan írható fel a második főkomponens a változók segítségével? Hogyan írható fel a Profit változó a főkomponensek segítségével? Hány főkomponensre van szükség, ha a változókat legfeljebb 5 százalékos hibával akarjuk felírni?

9. Végezzünk diszkriminancia-analízist a Szektor változóra az előző feladatban megadott öt változó alkalmazásával, és határozzuk meg a diszkriminancia függvények egyenletét! Határozzuk meg, hogy az egyes megfigyeléseknek mik a koordinátáik a diszkriminanciafüggvények által meghatározott koordináta-rendszerben. Milyen hatékonysággal működik az algoritmus?
10. A `TOP2.txt` állomány egy új cég adatait tartalmazza. Az előző feladatban kapott modell melyik szektorba sorolja a vállalatot? Mennyire megbízható ez a besorolás?
11. Végezzünk hierarchikus klaszterezést a fenti öt változó segítségével. Törekedjünk arra, hogy a klaszterek ne legyenek nagy átmérőjűek. Melyik klaszterezési módszer alkalmazzuk? Hozzunk létre 5 klasztert. Hány eleme van a klasztereknek?