

Név:

Hallgatói azonosító:

A 3 feladatsor kitöltésére háromszor 15, összesen 45 perc áll rendelkezésre.
Összesen 82 pont érhető el. Az elégségeshez kb. 30% (24 pont) elérése szükséges.
A vizsga során segédeszköz nem használható.
Jó munkát!

Figyelem! A tesztfeladatok kitöltése során a rossz válasz -1 pontot ér.

1. feladatsor (34 pont)

1. feladat. (24 pont)

Jelölje be, hogy Igaz, vagy Hamis-e az állítás (helyes válasz: 2 pont, hiányzó válasz: 0 pont, rossz válasz, vagy egyszerre 2 válasz: -1 pont) A színes jelek a javítást segítik, kérem ne használják!

1) Egy nemnegatív valószínűségi változó négyzetgyökének feltételes várható értéke (feltéve a Z valószínűségi változót) kisebb, mint a feltételes várható érték négyzetgyöke.

☒ ☒ ☐ Igaz Hamis

2) A valószínűségi változók kopulája meghatározza együttes eloszlásukat.

☒ ☒ ☒ Igaz Hamis

3) Nemnegatív valószínűségi változók korrelációja meghatározza kopulájukat.

☒ ☒ ☐ Igaz Hamis

4) A regresszióban a magyarázó változók által megmagyarázott variancia arányát a determinációs együttható adja meg.

☒ ☒ ☐ Igaz Hamis

5) A faktoranalízisben az első néhány faktor szórásnégyzetének összege az ezen faktorok által megmagyarázott variancia arányát adja meg.

☒ ☒ ☐ Igaz Hamis

6) A főkomponens analízisben az első néhány főkomponens szórásnégyzetének összege ezen főkomponensek által megmagyarázott variancia arányát adja meg.

☒ ☒ ☐ Igaz Hamis

7) A lineáris regresszió a magyarázó változók exponensei (e^{x_i}) és a válasz közti additív zajjal terhelt lineáris kapcsolatot nem képes feltárni.

☒ ☒ ☐ Igaz Hamis

8) A regresszióban a legkisebb négyzetes becslés a predikciós hibát minimalizálja.

☒ ☒ ☒ Igaz Hamis

9) A regresszióban legkisebb négyzetes becslés a reziduálisok négyzetösszegét minimalizálja.

☒ ☒ ☒ Igaz Hamis

10) A tolerancia azt méri, hogy az egyes magyarázó változóink tartalmazzak-e a többiektől eltérő információt.

☒ ☒ ☒ Igaz Hamis

11) Cross validation-nal felmérhetjük, hogy a magyarázó változók információtartalma mennyire releváns a válaszra nézve.

☒ ☒ ☐ Igaz Hamis

12) A logisztikus regresszió jóságát a ROC görbe alatti terület nagyságával tudjuk jellemezni; minél nagyobb, annál lényegesebb a feltárt kapcsolat.

☒ ☒ ☐ **Igaz** Hamis

13) Az elsőrendű autoregressziós folyamat oksági stacionárius megoldása pontosan akkor létezik, ha paraméterének abszolút értéke 1-nél kisebb.

☒ ☒ ☐ **Igaz** Hamis

14) Diszkrét eloszlású zajból generálva diszkrét eloszlású az autoregressziós folyamat stacionárius megoldása is.

☒ ☒ ☐ **Igaz** Hamis

pont

2. feladat (10 pont)

Melyik teszttel vagy eljárással döntene a következő problémákról?

Adja meg a helyes választ! (Egyetlen teszt vagy eljárás megnevezését fogadom csak el, minden más 0 pont. Helyes válasz: 2 pont)

a)

100 tesztalany közül 50 férfi, 50 nő, a férfiak és a nők közül is 25-25 dohányzik. Megadják, közülük kinek van magas vérnyomása. Eltérő módon hat-e a nők, ill. a férfiak magas vérnyomás kockázatára a dohányzás?

Válasz: *Khi négyzet próba.*

☒ ☒ ☐

b)

Egy kísérletben 100-100-100-100 növénynek külön-külön bórt, magnéziumot, cinket, vasat adnak (egy növénycsoport csak egyféle kiegészítő tápanyagot kap), illetve egy ugyancsak 100-as kontrol csoport nem kap kiegészítőt, és egyébként azonos körülmények közt nevelik őket. Feljegyzik mennyit nőttek egy hét alatt. Hogyan állapítaná meg, hat-e egyáltalán valamelyik anyag?

Válasz: *Két mintás t-próba*

☒ ☒ ☐

c)

Autógumi gyártmányfejlesztése során a gumi fékútjának hosszát mérik adott sebességen, két komponens különböző arányának függvényében. Egy-egy rögzített arány mellett 100 gumi fékhosszát jegyzik fel. Az optimális komponensarányt szeretnék meghatározni, a fékút, mint a komponensarány egy polinomiális függvénye minimumhelyeként.

Hogyan illesszék az adott fokszámú polinomot az adatokra.

Válasz: *F-próba*

☒ ☒ ☐

d)

Az előző példában milyen eljárással választaná ki a legjobb fokszámot.

e)

A városban 200 helyen mintát vesznek a csapvízből és elemzik azt nitritre, nitrátra, keménységre, vezetőképességre, vas- és kloridtartalomra. Milyen eljárással próbálná besorolni a mintavételi helyeket a vízjellemzők hasonlósága szerint (hogyan pl. ezzel szennyeződést lokalizáljon)?

Válasz: *Wilks-Lambda (MANOVA)*

☒ ☒ ☐

pont

Összpontszám: *pont*

☐ ☒ ☐

A MANOVA esetén melyik teszt eloszlása áll elő, mint nem azonos, de független béta eloszlású valváltozók szorzatának eloszlása.

- a) Roy
- b) Pillai
- c) Wilks-Lambda
- d) Lawley-Hotelling

☒ ☒ ☐

Az egydimenziós szóráselemzés ANOVA esetén milyen próbával döntünk a nullhipotézisről

- a) t-próba
- b) F-próba
- c) chi-négyzet próba
- d) U-próba

☒ ☒ ☐

Hogyan kapja meg a főkomponenseket?

- a) Az adatmátrix spektrálfelbontásából.
- b) Az adatok variancia-kovariancia mátrixának normált sajátvektoraiként.
- c) Az adatmátrix normált sajátvektoraiként.
- d) Az adatok variancia-kovariancia mátrixának QR felbontásában szereplő Q ortogonális mátrix oszlopvektoraiként (Gram-Schmidt ortogonalizáció).

☒ ☒ ☐

pont

Összpontszám: *pont*

Név:

Hallgatói azonosító:

3. feladatsor (28 pont)

6. feladat (14 pont)

Válassza ki a helyes választ! (helyes válasz: 2 pont, hiányzó válasz: 0 pont, rossz válasz: -1 pont)

1 Hol lép fel a Wishart eloszlás?

- a) Normális vektor hosszára vonatkozó próbában
- b) Normális vektor várható értékére vonatkozó próbában
- c) Normális vektor likelihood becslésének eloszlásaként
- d) A főkomponensek meghatározásában

☒ ☒ ☐

2 Honnan következik, hogy ha normális együttes eloszlású valváltozók korrelálatlanok, akkor függetlenek is?

- a) Abból, hogy normális vektor lineáris leképezése normális vektor marad
- b) A normális vektor sűrűségfüggvényének formulájából
- c) Abból, hogy minden normális vektor független koordinátájába forgatható
- d) A normális vektor feltételes várható értékének linearitásából

☒ ☒ ☐

3 Honnan következik, hogy szórásElemzésben (az alapfeltevések igaz volta mellett) jogos az F próba, mert a két szórásbecslés független „mintából” történik?

- a) A Fisher-Cochran tételből
- b) A Wishart eloszlás tulajdonságából
- c) Normális vektor sűrűségfüggvényének formulájából
- d) A variancia kovariancia mátrix spektrálfelbontásából

☒ ☒ ☐

4 Honnan következik, hogy együttesen normálisak feltételes várható értéke a feltétel lineáris függvénye

- a) Abból, hogy normális vektor lineáris leképezése normális vektor marad
- b) A normális vektor sűrűségfüggvényének formulájából
- c) Abból, hogy minden normális vektor független koordinátájába forgatható
- d) Abból, hogy ha normális együttes eloszlású valváltozók korrelálatlanok, akkor függetlenek is

☒ ☒ ☐

5 Mi a feltételes várható érték geometriai jelentése?

- a) forgatás
- b) projekció
- c) skalárszorzat
- d) eltolás

☒ ☒ ☐

6 Melyik mennyiség használható változó szelekcióra a regresszióban?

- a) korreláció
- b) kommunalitás
- c) Akaike információs kritérium
- d) Cook távolság

☒ ☒ ☐

7 Melyik mennyiség használható outlier detektálására a regresszióban?

- a) korreláció
- b) kommunalitás
- c) Akaike információs kritérium
- d) Cook távolság



pont

Röviden válaszolja meg az alábbi kérdéseket.

Csak rövid, néhány mondatos, vagy egyszerű formulás válaszokat kérek/fogadok el.

7. feladat (4 pont)

Definiálja a Wishart eloszlást.

Legyen $S = ZZ^T$ egy $p \times p$ dimenziós pozitív-definit szimmetrikus véletlen mátrix. Ekkor S Wishart eloszlású, n szabadsági fokkal. $S \sim W_p(\Sigma, n)$.

pont

8. feladat (4 pont)

Mi alapján választja meg a faktorok/ megtartott főkomponensek számát?

- a megtartottak magyarázzák a szórás egy adott százalékát
- nagyobb főkomponens szignifikancia tesztelésével
- kivesszük azokat melyek az átlagnál kisebb sajátértékekhez tartoznak
- köömlés-diagrammon az első vagy második törésvonal utániakat hagyjuk el (nagyság szerint a sajátértékek csökkenő sorrend)

pont

9. feladat (6 pont)

Többváltozós problémákban mi a koordinátánként végzett tesztelés 3 fő problémája

pont

Összpontszám: *pont*