

## Gyakorló feladatsor a II. anyagrészhez <sup>1</sup>

1. Az Debreceni Egyetemen az egyik statisztika szemináriumvezető minden hétfőn, szerdán és pénteken autóval jár ki a Tócskertetből a város másik végén fekvő Kassai úti campusra. Otthonról mindig azonos időben indul el és ugyanazon az útvonalon autózik. Úgy érzi azonban, hogy a menetideje függ attól, hogy a hét melyik napján van órája. Ezért aztán márciusban, áprilisban és májusban véletlenszerűen kiválasztott 5-5 hétfőt, szerdát és pénteket és lejegyezte a menetidőket. Adatainak összegzését az alábbi táblázat tartalmazza:

Nap	Menetidő ( $x$ )				
Hétfő	28	34	29	34	30
Szerda	24	27	25	25	22
Péntek	25	28	27	26	21

Nem feltételezzük a menetidők normalitását, azaz az alábbi kérdésekben megfelelő nemparaméteres próbával döntsön az SPSS output alapján.

- (a) Igaz-e a szemináriumvezető sejtése, hogy a menetidő függ az utazás napjától?

SPSS: Analyze → Nonparametric Tests → Legacy Dialogs → K Independent Samples

- Adja meg a megfelelő próba próbastatisztikájának értékét!
- Döntsön 99%-os megbízhatósággal, elfogadja-e a  $H_0$  nullhipotézist, miszerint a menetidők mediánja a három napon megegyezik!
- Adja meg a szerdai menetidők rangösszegét!

- (b) Igaz-e, hogy a hétfői és pénteki menetidők azonos eloszlásúak? Vizsgálja meg a kérdést a Mann-Whitney, a Kolmogorov-Smirnov és a Wald-Wolfowitz próbák segítségével!

SPSS: Analyze → Nonparametric Tests → Legacy Dialogs → 2 Independent Samples: Mann-Whitney U, Kolmogorov-Smirnov-Z, Wald-Wolfowitz runs

- Adja meg a Mann-Whitney U-próba próbastatisztikájának értékét!
- Adja meg a Mann-Whitney próba aszimptotikus  $p$ -értékét!
- A Mann-Whitney próba alapján döntsön 95%-os megbízhatósággal, elfogadja-e a nullhipotézist, miszerint azonos a hétfői és pénteki menetidők eloszlása!
- Adja meg a kétmintás Kolmogorov-Smirnov próba próbastatisztikájának értékét!
- Adja meg a Kolmogorov-Smirnov próba aszimptotikus  $p$ -értékét!
- A Kolmogorov-Smirnov próba alapján döntsön 95%-os megbízhatósággal, elfogadja-e a nullhipotézist, miszerint azonos a hétfői és pénteki menetidők eloszlása!
- A Wald-Wolfowitz próba alapján is döntsön 95%-os megbízhatósággal, elfogadja-e a nullhipotézist, miszerint azonos a hétfői és pénteki menetidők eloszlása!

- (c) A szemináriumvezető a menetidők mediánját a feljegyzett értékek minimumának és maximumának átlagával becsülte. Eszerint a menetidők mediánja 27.5 perc, azaz az esetek legfeljebb felében ennél tovább tart az út a campusra, az esetek legfeljebb felében pedig kevesebb ideig.

SPSS: Analyze → Nonparametric Tests → Legacy Dialogs → Binomial

- Adja meg az előjel próba próbastatisztikájának  $p$ -értékét!
- Elfogadható-e 95%-os megbízhatósággal a  $H_0$  hipotézis, miszerint a menetidő mediánja 27.5 perc?

---

<sup>1</sup>A feladatok korábbi ZH feladatokból származnak.

2. Egy botanikus kertben azt vizsgálták, hogy a napi átlaghőmérséklet mennyire befolyásolja a látogatók számát. 8 tetszőlegesen választott napot néztek és a megfigyelt adatokat regresszió segítségével elemezték:

hőmérséklet (°C)	látogatók száma (fő)	hőmérséklet (°C)	látogatók száma (fő)
20	220	24	268
18	205	26	290
14	162	21	231
15	168	19	214

Néhány számolási segédérték:

$$\sum x = 157 \quad \sum y = 1758 \quad \sum xy = 35767 \quad \sum x^2 = 3199 \quad \sum \lg x = 10,2747$$

$$\sum \lg y = 18,6735 \quad \sum x \lg y = 368,9842 \quad \sum \lg x \lg y = 24,0398 \quad \sum (\lg x)^2 = 13,2558$$

- Illesszen az adatokra lineáris regressziót, írja fel a regressziós függvényt és értelmezze a kapott paramétereket!
  - Számolja ki a maradék (hiba) négyzetösszeget!
  - Becsülje meg, hogy várhatóan mennyi lesz a látogatók száma egy hűvös, 15 °C hőmérsékletű napon, majd szerkesszen konfidenciaintervallumot ugyanerre 98%-os megbízhatósági szinten!
  - Adjon 98%-os konfidenciaintervallumot a 15 °C-os napok átlagos látogatószámára!
  - Számítsa ki a két változó lineáris korrelációs együtthatóját!
  - Számítsa ki és értelmezze a lineáris regresszió determinációs együtthatóját!
  - Számolja ki és értelmezze az elaszticitást a 20 °C-os hőmérsékletre!
3. Egy fürdőruhabolt 2010-es és 2011-es negyedéves forgalmai (db) az alábbiak voltak:

év	I.	II.	III.	IV.
2010	300	365	430	370
2011	420	485	550	510

Néhány számolási eredmény:

$$\sum y_t = 3430, \quad \sum ty_t = 16740, \quad \sum t = 36, \quad \sum t^2 = 204.$$

- Illesszen lineáris trendet a kapott adatsorra és értelmezze a kapott paramétereket!
- Határozza meg a III. negyedév tisztított szezonális eltérését és értelmezze a kapott eredményt!
- Becsülje meg a forgalmat 2012. III. negyedévére!
- Írja fel a 2010. III. és IV. negyedéveihez tartozó trendértékeket 4-tagú szimmetrikus mozgóátlag segítségével!

4. Kézzel festett porcelánvázák árát vizsgálták. A modellalkotásnál figyelembe vették a váza magasságát (**Magasság** cm), festési idejét (**Idő** min), a felhasznált színek számát (**Szín** db), a festés típusát (**Típus**=1, ha a festés a máz feletti és **Típus**=2, ha festés a máz alatti), továbbá az égetések számát (**Égetés** db).

- (a) A magyarázó változók közül melyik kapcsolata legerősebb a célváltozóval?
- (b) Adja meg a függő változó és a vele legkevésbé korreláló magyarázó változó korrelációs együtthatóját!
- (c) Adja meg a végső modell illeszkedését leíró módosított mutató értékét!
- (d) Adja meg a legjobb modell többszörös korrelációs együtthatójának az értékét!
- (e) Mely magyarázó változó(k)tól nem függ az árát legjobban leíró lineáris modell?
- (f) A végső modellből kihagyott magyarázó változók közül adja meg a legkevésbé szignifikáns parciális korrelációját!
- (g) Írja fel az adatokra legjobban illeszkedő modellt és értelmezze annak paramétereit!
- (h) Adjon 95%-os konfidenciaintervallumot a legjobb modell konstans tagjára!
- (i) Adja meg a legjobban illeszkedő modellben az Idő együtthatójának szignifikanciájára vonatkozó teszt próbastatisztikájának értékét!
- (j) Mi történik a második lépésben? Miért?
- (k) A végső modell esetén milyen hipotézist tesztelünk az ANOVA segítségével?
- (l) Adja meg a legjobb modell reziduális és teljes négyzetösszegét!
- (m) Adja meg a végső modellhez tartozó globális F-próba próbastatisztikájának értékét!
- (n) Mikor áll le az algoritmus? Miért?
- (o) Mennyibe kerül várhatóan egy 15 cm magas váza, melyet 3 féle festéssel 25 percig festettek és 3-szor égették?
- (p) Az illesztett modell alapján mennyivel nő egy váza ára, ha a festésére 1 órával többet szánunk, a többi érték viszont változatlan marad?
- (q) Mit tud mondani a hiba eloszlásáról? Minden rendelkezésre álló információ alapján döntsön! Mit jelent ez a modellre vonatkozóan?
- (r) Adja meg a legjobban illeszkedő modell maradécai normalitásának ellenőrzésére szolgáló teszt aszimptotikus p-értékét! Amennyiben az SPSS alsó határt ad meg, úgy azt írja be!
- (s) 10%-os döntési szintet használva normálisnak tekinthetőek-e a végső modell maradécai?