

# R Zárthelyi dolgozat

## B csoport

### Speciális szoftverek (GEMAK236-B2)

2032/24/I. félév

#### Beadás módja

A megoldásokat feladatonként .R script fájlokban várom e-mailben, *R ZH* tárggyal.

#### Pontozás

1.: 25 pont, 2.: 25 pont, 3.: 30 pont

Összesen: 80 pont

Értékelés: 40+ elégséges, 48+ közepes, 56+ jó, 64+ jeles

#### $\Sigma$ :25 pont 1. feladat

a) Az  $f(x) = 2 \cos(2x) - 1$  függvény ábrázolása 4 lépésben függvénytranszformációval

(1 pont) i. Dolgozzunk a  $(0, 4\pi]$  intervallumon.

(4 pont) ii. Ábrázoljuk a transzformáció 4 lépését!

(3 pont) iii. Minden lépésben használjunk különböző színt!

(2 pont) iv. Legalább az egyik lépésben definiáljunk (nem alapértelmezett) vonaltípust!

(3 pont) v. Legalább az egyik lépésben használjunk (nem alapértelmezett kör) markereket (avagy, ábrázoljuk a pontokat is, ne csak a töröttvonalat)!

(2 pont) vi. Ha szükséges, módosítsuk az ábrázolási tartományt, hogy minden függvény beleférjen az ábrába!

b) Jelmagyarázat

(2 pont) i. Keressünk üres területet az ábrán (vagy módosítsuk az ábrázolási tartományt, hogy legyen üres terület) jelmagyarázatnak.

(2 pont) ii. A feliratok legyenek a függvények képletei a transzformáció során.

(6 pont) iii. A jelmagyarázat tartalmazza ne csak a színeket, hanem a vonalstílust és markereket is!

#### $\Sigma$ :25 pont 2. feladat

a) Adatok generálása

(6 pont) i. Készítsünk egy `data.frame`-et. Az első két oszlopa legyen  $k1$  és  $k2$ , ahol generáljunk 200-200 db független kockadobást.

ii. Tipp: kockadobás generálásához `runif()` függvénnyel tudunk folytonos egyenletes eloszlású valószínűségi változót generálni, ezt skálázzuk és kerekítjük!

- (2 pont) iii. Új  $s$  oszlopként számoljuk ki  $k_1$  és  $k_2$  összegét.
- (3 pont) iv. Új  $m$  oszlopként számoljuk ki  $k_1$  és  $k_2$  abszolút különbségét (a különbség abszolút értékét)!
  - b) Adatok grafikusan
    - (3 pont) i. Készítsünk hisztogramot  $s$ -ről!
    - (2 pont) ii. Készítsünk hisztogramot  $s$  *relatív* gyakoriságairól!
    - (3 pont) iii. Ez utóbbi ábrában ábrázoljuk az empirikus sűrűségfüggvényt is!
  - c) Gyakoriság, függetlenség
    - (2 pont) i. Ellenőrizzük, hogy  $s$  és  $m$  faktor típusú változók-e! Ha nem, akkor konvertáljuk őket!
    - (2 pont) ii. Kérjük le  $s$  és  $m$  együttes gyakoriság-táblázatát/mátrixát!
    - (2 pont) iii. Vizsgáljuk meg  $\chi^2$  (khí-négyzet) próbával, hogy  $s$  és  $m$  függetlenek-e!

$\Sigma$  :30 pont **3. feladat**

- (5 pont) a) Olvassuk be az adatokat a mellékelt `adat3B.txt` fájlból egy `data.frame`-be. (Vagy – előbb mentjük az előző munkamenetet! – nyissuk meg a mellékelt `.RData` fájlt és használjuk a memóriában lévő `pontok` nevű `data.frame`-et.)
- (2 pont) b) Ábrázoljuk a kapott  $(x, y)$  pontokat, és/vagy az általuk kirajzolt függvényt!
  - c) 1. számú regressziós modell
    - (2 pont) i. Illesszünk egy regressziós modellt  $y \approx a \cdot x + b$  formában!
    - (2 pont) ii. Ábrázoljuk a regressziós egyenest az előző ábrában (új színnel)!
      - iii. Úgy tűnik, hogy az adatpontok által kirajzolt függvénynek van görbülete, próbáljunk meg egy magasabb fokú polinomot illeszteni rá!
  - d) 2. számú regressziós modell
    - (3 pont) i. Illesszünk egy másik regressziós modellt  $y \approx a \cdot x^2 + b \cdot x + c$  formában!
    - (2 pont) ii. Ábrázoljuk ezt a regressziós függvényt is az előző ábrában (új színnel)!
      - e) 2. számú modell „validálása”
        - (2 pont) i. Új ábrá(k)ban ábrázoljuk a 2. modell hibáit (residuals) az  $x$  és/vagy a becsült értékek (fitted) függvényében.
        - (2 pont) ii. Kérjük le a modell `summary()`-jét és elemezzük az együtthatókat. Van-e köztük olyan, ami 90%-os biztonsággal 0-nak tekinthető (azaz  $\Pr(>|t|) > 0.1$ )?
    - f) 3. számú regressziós modell
      - (3 pont) i. Az előző `summary` elemzés alapján válasszuk ki, hogy melyik változót hagyjuk el a 2. modellből, és így generáljuk a harmadikat.
        - ii. (Vagy válasszunk véletlenszerűen, hogy ne adjunk el.)
      - (3 pont) iii. Ábrázoljuk ismét az eredeti  $(x, y)$  pontpárokat és a 3. modell regressziós függvényét egy ábrában.
    - g) ANOVA
      - (4 pont) i. ANOVA táblázattal hasonlítsuk össze az 1. és 2., illetve a 2. és 3. számú modellt.