Zárthelyi dolgozat
A csoport

Specális 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

Σ :25 pont **1. feladat**

a) Az $f(x) = \frac{3}{2} \cdot \log(x/2) - 2$ függvény ábrázolása 4 lépésben függvénytranszformációval

(1 pont) i. Dolgozzunk a $(0, 4]$ 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!

Σ :25 pont **2. feladat**

a) Adatok generálása

(5 pont) i. Készítsünk egy `data.frame`-et. Az első három oszlopa $u1$, $u2$ és $u3$, tartalmazzon 100-100 db független, $[0, 2]$-n egyenletes valószínűségi változót.
(2 pont) ii. Új $v1$ és $v2$ oszlopokként számoljuk ki $u1$ és $u2$, ill. $u2$ és $u3$ összegét.

(5 pont) iii. Új $t1$ és $t2$ oszlopként diszkretizáljuk $v1$-et ill. $v2$-t: vágjuk őket intervallumokra az egész számok mentén (azaz $(0, 1], (1, 2]$, stb.).

(2 pont) iv. Ellenőrizzük, hogy $t1$ és $t2$ faktor típusú változók-e! Ha nem, akkor konvertáljuk őket!

b) Adatok grafikusan

(5 pont) i. Készítsünk hisztogramokat $v1$-ről, két különböző intervallum-bontásban!

(2 pont) ii. Készítsünk hisztogramot $t1$-ről is!

c) Gyakoriság, függetlenség

(2 pont) i. Kérjük le $t1$ és $t2$ együttes gyakoriság-táblázatát/mátrixát!

(2 pont) ii. Vizsgáljuk meg χ^2 (khí-négyzet) próbával, hogy $t1$ és $t2$ függetlenek-e!

Σ :30 pont **3. feladat**

(5 pont) a) Olvassuk be az adatokat a mellékelt `adat3.txt` fájlból egy `data.frame`-be.

(2 pont) b) Ábrázoljuk a kapott (x, y) pontokat, és/vagy az általuk kirajzolt függvényt!

i. Figyeljük meg, hogy ez – a véletlen zajtól eltekintve – egy periodikus függvény, ezért a regresszióban a trigonometrikus függvényeket, mint speciális periodikus függvényeket fogjuk használni.

c) 1. számú regressziós modell

(2 pont) i. Illesszünk egy regressziós modellt $y \approx a \cdot \sin(x) + b$ formában!

(2 pont) ii. Ábrázoljuk a regressziós függvényt az előző ábrában (új színnel)!

d) 2. számú regressziós modell

(3 pont) i. Illesszünk egy másik regressziós modellt $y \approx a \cdot \sin(x) + b \cdot \cos(x) + c$ formában!

(2 pont) ii. Ábrázoljuk ezt 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 akadjunk 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.