

Feladatok, tippek

Faktorok kinyerése fájlnevekből

filenames.RData a szokásos linkről.

Fájlnemek szerkezete:

beszélő kódjakísérlet neve_ismétlés_mondatszám.wav

`strsplit` függvénnyel string változóra alakítjuk a faktorokat, majd elvágjuk a "_" mentén. Alapértelmezett output: lista, ezt vektorra alakítjuk.

```
file.list =  
unlist(strsplit(as.character(filenames), "_"))
```

Utána a vektor 1., 4., 7. stb. elemét a `subj` változóba írjuk:

```
subj = file.list[seq(1,length(file.list),3)]
```

Ehhez hasonlóan az ismétléseket és a mondatazonosítókat is vektorokba írjuk.

merge() függvény további felhasználása

A fenti adatokban szereplő mondatazonosítóhoz rendelhető adatokkal való társítás.

sentence_code.txt fájl létrehozása:

```
sent.id;sent.type;target.word;focus.type  
1;sent1;Németországba;broad  
2;sent1;Németországba;narrow  
3;sent1;Németországba;contr  
4;sent2;Máltán;broad  
5;sent2;Máltán;narrow  
6;sent2;Máltán;contr
```

Fájl beolvasása `read.table()` függvénnyel. Utána a fájlnevekből kinyert mondatazonosítókhoz lehet merge-elni. Fontos: a mondatazonosítók oszlopnevének egyeznie kell az itt megadottal, vagyis `sent.id`-del!

Adatmátrixok átrendezése

pol dataframe alapján az ismételt méréses ANOVA bemeneteként ilyen formátumú mátrixra van szükségünk:

subject	attitude	frequency
F1	inf	246.7857
F1	pol	217.28571
F2	inf	270.0143
F2	pol	246.35714
...		

Nyerjük ki a cellaátlagokat a `tapply()` függvénnyel.

Feladat: hozzuk az adatokat az `aov()` függvény bemeneteként elfogadott formába. Az ismételt méréses ANOVA eredményeit hasonlítsuk össze a random intercept és random slope kevert modellek eredményeivel.

Két adatmátrix összefűzése

Feltétel: mindkettőben azonos nevű oszlop.

oz.csv: őzek élőhelye, neme, testtömege és magassága.

oz_terulet.csv: a területre jellemző időjárási tényezők.

Feladat: vonjuk össze a két adatbázist úgy, hogy az egyes őzek adatainak sorában az élőhelyükre jellemző csapadék- és hőmérsékletadatok is megjelenjenek. Hasznos függvény: `merge()`. Az új mátrix neve legyen `oz_komplex`.

Gyakorló feladatok

1.

oz.csv: három különböző területen mérték újszülött hím és nőstény őzgidák tömegét és testhosszát. Ábrázoljuk a hím és nőstény őzek testtömegét a három mérési hely szerint dobozdiagrammal. A hím őzek tömegének doboza legyen világkoscék, a nőstényeké rózsaszín. Lássuk el az ábrát magyar nyelvű feliratokkal. Készítsünk jelmagyarázatot (legend) a színek jelentéséről.

2.

A két csoportra jellemző eloszlást ábrázoljuk egyazon sűrűségfüggvényben, különböző színekkel. Az átlagok értékét jelöljük függőleges egyenessel az x -tengelyen.

3.

Teszteljük a megfelelő próbával, hogy az őzgidák neme, születési helye és az ott előforduló csapadékmennyiség befolyásolja-e méretüket. Ellenőrizzük, hogy teljesülnek-e a különféle próbák előfeltételei, és eszerint válasszuk meg a modellt.

Gyakorló feladatok

4.

Számoljuk ki az özek testmagasságának és testtömegének korrelációs együtthatóját. Melyik együtthatóval jellemezhetjük az összefüggést? Ábrázoljuk az adatpontokat, és rajzoljuk be a regressziós egyenest. Milyen százalékban magyarázza a testmagasság a testtömeg értékét?

5.

Ábrázoljuk az oz_komplex mátrix alapján a hőmérsékleti tényezők hatását az özek testtömegére. Miért nem egyformák a kapott ábrák? Hogyan tudjuk elérni, hogy minden esetben dobozdiagramokat kapjunk?