

Gyakorló feladatok

Őzek születési mérete

Töltsük le az oz.csv mátrixot innen:

<http://phon.nytud.hu/mady/courses/statistics/materials>

Nyissuk meg a fájlt a Jegyzettömb vagy más szoftverrel. Mi választja el az oszlopokat, és hogyan vannak jelölve a tizedesek? Milyen függvénnnyel tudjuk beolvasni az adatsort?

Három különböző területen mérték újszülött hím és nőstény őzek tömegét és testhosszát: Árpádhalmon, Babaton és Pitvaroson.

TERULET: a három adatmérési helyszín

SEX: az őz neme

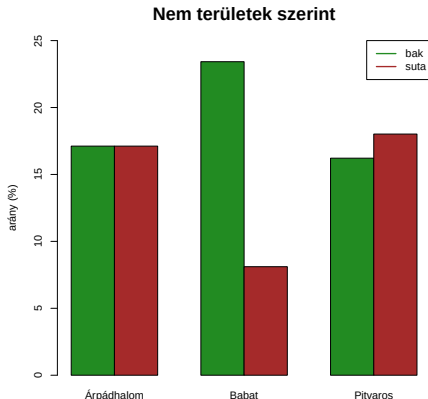
TOMEG: testtömeg kilogramban

TESTH: testhossz centiméterben

Jelmagyarázat készítése

Az ábra színeihez tartozó kategóriák megjelenítését jelmagyarázatban tudjuk megadni.

```
legend("topright", legend=c("bak", "suta"),  
col=c("forestgreen", "brown"), lwd=2)
```



Függvény: `legend()`, paraméterek:

- ▶ A jelmagyarázat helye, pl. `"top"`: felül középen, `"topleft"`: felül balra, `"topright"`: felül jobbra stb.
- ▶ A jelmagyarázatban szereplő kategóriák, pl. bak és suta:
`legend=c("bak","suta")`.
- ▶ A kategóriákhoz tartozó vonalak színe, pl. zöld és barna:
`col=c("green","brown")`.
- ▶ A kategóriákhoz tartozó vonalak vastagsága: `lwd = 2`.

Adatok szűrése

Ha az adatok valamelyik alcsoportjával akarunk számolni, logikai vektorokat alkalmazunk. Ezekről részletesen a 7. óra anyagának végén lesz szó. Itt csak egy egyszerű, de tipikus esetet nézünk meg.

Az egyes területeken élő őzek átlagos testtömegét akarjuk kiszámolni külön-külön. Az Árpádhalmon mért őzek csoportját így tudjuk elkülöníteni:

```
arp = oz$TERULET == "Arpadhalom"
```

Az arp egy logikai vektor. Amelyik sorra igaz, hogy a TERULET oszlopban Arpadhalom áll, ott TRUE jelenik meg, a többinél FALSE. Ez a vektor annyi elemből áll, ahány sort tartalmaz az oz dataframe (111).

A TOMEK értékek kiszámítása csak azokra az értékekre, amelyekre igaz, hogy a TERULET oszlopban "Arpadhalom" szerepel:

```
mean(oz$TOMEK[arp])
```

Fontos: a szűrés KÉT egyenlőségjellel történik.

1.

Ábrázoljuk a hím és nőstény őzek testtömegét a három mérési hely szerint dobozdiagrammal, nemek szerint eltérő színnel. Lássuk el az ábrát magyar nyelvű feliratokkal, és készítsünk jelmagyarázatot a színekhez. Mentsük el az ábrát pdf-ként vagy png-ként.

2.

Ábrázoljuk az őzek testtömegének eloszlását sűrűségfüggvénnyel.

3.

Ábrázoljuk a hím és nőstény csoportra jellemző testtömegek eloszlását egyazon sűrűségfüggvényen, különböző színekkel. Az x és y tengely skálázása azonos legyen.

4.

Igaz-e, hogy a három területen azonos arányú a bakok és a suták aránya? Készítsünk oszlopdiagramot, amiről leolvasható a válasz. Az y-tengelyen ne a megfigyelések száma, hanem százalékos aránya szerepeljen. Ezt jelezzük az y tengely feliratán. A területek neveit lássuk el ékezetekkel. Az ábra tartalmazzon jelmagyarázatot is. Ez az ábra megegyezik a 3. dián látható ábrával.

5.

Számoljuk ki az őzek testtömegének és testhosszának korrelációs együtthatóját. Melyik együtthatóval jellemezhetjük az összefüggést? Milyen összefüggést jelez az együttható értéke? Ábrázoljuk az adatpontokat pontdiagramon.

6.

Számoljuk ki a bakok és suták testtömegének és testhosszának az átlagát. Hogyan térnek el a két nem adatai?