

# Az R adatelemzési nyelv alapjai I.

## Gyakorlat

A feladat megoldása során egy olyan adattáblával fogunk dolgozni, amely hipotetikus személyek testmagasság és testsúly adatait tartalmazza.

Gyakorlandó ismeretek:

- adattábla létrehozása
- statisztikai függvények használata
- adattáblák összefűzése
- hiányzó adatok kezelése
- adatok származtatása
- leíró statisztikák készítése adattáblákról
- részhalmazok szűrése
- adattábla exportálása / importálása

1. Hozzon létre egy adattáblát, amely 50 hipotetikus férfi testmagasságát és testsúlyát tartalmazza. Az adatok előállításához sorsoljon normális eloszlású véletlen számokat (testmagasság átlaga = 175 [cm], szórása = 10 [cm]; testsúly átlaga = 80 [kg], szórása = 10 [kg]).

Parancsok: `data.frame`, `rnorm`

2. Hozzon létre egy másik adattáblát, amely 50 hipotetikus nő testmagasságát és testsúlyát tartalmazza. Az adatok előállításához sorsoljon normális eloszlású véletlen számokat (testmagasság átlaga = 165 [cm], szórása = 8 [cm]; testsúly átlaga = 65 [kg], szórása = 10 [kg]).

Parancsok: `data.frame`, `rnorm`

3. Mindkét táblához fűzzön hozzá egy új oszlopot, amely a személy nemét tartalmazza, majd fűzze össze a két táblát egy új táblává. A további feladatok megoldása során ezt az új táblát használja.

Parancsok: `rbind`

4. Generáljon véletlenszerű hiányzást az adattáblában a testmagasság és a testsúly oszlopba (~10%-nyi hiányzó adatot generáljon). Tipp: sorsoljon véletlen indexeket a megadott oszlopokban és töltsse ki "NA"-val.

Parancsok: `sample`

5. Készítsen leíró statisztikákat az adattábláról, összesítve és nemenkénti bontásban is. Figyeljen a hiányzó értékekre.

Parancsok: `summary`, `split`, `lapply`

6. Pótolja a hiányzó adatokat a nemeknek megfelelő átlagértékek alapján.

Parancsok: `mean`, `is.na`

7. Számítsa ki a BMI értékét minden személyre, és ezt fűzze az adattáblához egy új oszlopba ("BMI").

$BMI = \text{testsúly [kg]} / (\text{testmagasság [m]})^2$

8. A BMI értéke alapján sorolja be a személyeket az alábbi kategóriákba:

- 18.5 alatt: Alultáplált
- 18.5 és 25 között: Normális
- 25 és 30 között: Túlsúlyos
- 30 felett: Elhízott

Ezt fűzze hozzá az adattáblához egy új oszlopba ("BMI.Status").

Parancsok: `cut`

9. Szűrje le

- a túlsúlyos személyeket,
- az elhízott férfiakat,
- azokat a normális nőket, akiknek a testmagassága átlagon aluli
- azokat a túlsúlyos vagy elhízott nőket, akiknek a testsúlya meghaladja a normális férfiak átlagos testsúlyát

Parancsok: `pl.subset`, `mean.stb`.

10. Exportálja az adattáblát egy szöveges fájlba, majd töltsse be onnan.

Parancsok: `write.table`, `read.table`