

Töltsétek le a munkakönyvtárba a kuiper.RData file-t, majd nyissátok meg és futtassátok le a negyedikhazi.R scriptet. A legvégén a summary() függvény kerül meghívásra, ami tartalmazza a kruiserepit data.frame-en lefuttatott lineáris regresszió R^2 statisztikáját. Azt kérem, hogy a következő feladatokat megoldó parancsokkal egészítsétek ki a scriptet:

- A kruiserepit data.frame-hez adjatok hozzá egy Priceb oszlopot, ami a Price becsült értékeit tartalmazza
- Kérjétek korrelációt a Price és a Priceb oszlopok között a cor() függvény segítségével. A cor() függvény outputjából nyerjétek ki a korrelációt majd emeljétek négyzetre. Írjátok bele a scriptbe megjegyzésként, hogy mit tapasztaltatok!
- Ugyan a summary() kiszámolta, de hasznos ha a mi magunk írunk for ciklussal egy a Price és a Priceb változókból R^2 statisztikát számoló scriptet. Arra kérlek titeket, hogy ezt tegyétek meg. Emlékeztetőül ideraktam az R^2 statisztika képletét. Kalap jelöli a becsült értékeket, míg a felülvonás átlagot jelöl. Addig próbálkozzatok amíg meg nem kapjátok a summary()-nél található értéket.
$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$$
- Ezután a kruiserepit-en felépített modell segítségével becsüljétek meg az órán is használt módon a kruiserteszt Price változóját (itt is hozzatok létre egy Priceb változót).
- Először számoljátok ki a Price és a Priceb változók korrelációjának a négyzetét a kruiserteszt data.frame-ben. Ezután a kézzel R^2 -t számoló parancsokkal számoljátok ki az R^2 statisztikát is. Mit vesztek észre?

Megoldásképpen egy a megoldás minden lépését tartalmazó R scriptet kérek.

Megjegyzés: Előnyösebb lett volna úgy két részre bontani az adathalmazt modellépítő és tesztelő részre, hogy kisorsoljuk azt, hogy melyik megfigyelés melyik részbe kerüljön.

Ha bármi kérdésetek van, keressetek nyugodtan! Jó munkát kívánok!