

Töltsétek le a munkakönyvtárba a `kuiper.Rdata` file-t, majd adjátok ki a `load(„kuiper.Rdata“)` parancsot. Ezzel a `kuiper` változóba betöltődik a negyedik óra végén megmutatott használt autók adatbázis.

Az első feladat az, hogy kérjétek le a `table` függvénnyel a `sound` (van-e hifi a kocsiban) és a `leather` (az ülések borítása bőr-e) változók keresztábráját, majd `chi-négyzet` próbával vizsgáljátok meg, hogy függetlenek-e ezek a változók.

Ezután a `Price` (ár) és `Mileage` (megtett mérföld) változók függetlenségét vizsgáljátok meg. Mivel folytonos változóról van szó, ezért közvetlenül nem használható a `chi-négyzet` teszt. Ezért azt kérem, hogy előbb hozzátok létre a változók diszkrétizált változatait (`dPrice`, `dMileage`). A `dPrice` változó értéke legyen 1 ha a `Price` 15000-nél szigorúan kisebb, 2 ha a `Price` 15000 és 30000 között van, 3 ha a `Price` szigorúan nagyobb 30000-nél. A `dMileage` változó értéke legyen 1 ha a `Mileage` 10000-nél szigorúan kisebb, 2 ha a `Mileage` 10000 és 20000 között van, 3 ha a `Mileage` szigorúan nagyobb 20000-nél. Ezután vizsgáljátok meg a diszkrétizált változók függetlenségét. Az eredmény ismeretében mit mondhatunk az eredeti `Price` és `Mileage` változók függetlenségével kapcsolatban?

A tesztek 95%-os szignifikanciával végezték. Megoldásképpen egy a megoldás minden lépését tartalmazó R scriptet kérek, amelybe megjegyzésként írjátok bele azt, hogy az egyes kérdéseknél milyen p-értékeket kaptatok és azt, hogy ez alapján, hogyan választok a feltett kérdésekre.

Megjegyzés: A `chi-négyzet` teszt határeloszlás tételén alapul. Fontos megfigyelés, hogy ha túl sok a közel üres cella a keresztábrában, akkor a határeloszlás közelítése még nem pontos. Emiatt hoztunk létre csak 3 értéket felvevő diszkrét változókat. Így elkerültük a sok kevés megfigyelést tartalmazó cellát.

Segítség: A `chisq.test()` függvény használata nagyon hasonló az órán lefuttatott `mcnemar.test()` változó működéséhez. A diszkrétizálásnál emlékeztetek titeket arra, hogy ha létrehozuk a `seged<-c(2,5,2,1,3,5)` listát, akkor a `seged[seged>4]<-7` parancs a lista 4-nél szigorúan nagyobb elemeit 7-esre cseréli. Megjegyzem, hogy ha alsó és felső korlát is van, akkor azokat külön kell beírni közöttük egy `&` (és) jellel (lásd bővebben jó angol nyelvű bevezető 25. oldala). Azt javaslom, hogy a `kuiper` `data.frame` `Price` és `Mileage` oszlopát másoljátok be az új `dPrice` és `dMileage` változóba (pl `kuiper$dPrice<-kuiper$Price`), majd az új változók értékeit diszkrétizáljátok.

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