

Töltsd le az elsohazivaz.R file-t. A scriptben levő kommenteket követve fejezd be a függvény megírását. Futtasd le a függvényt tartalmazó script-et a ctrl+alt+R kombinációkkal (Rstudio-ban működik ez a kombináció).

Ezután azt szeretném ha lefuttatnátok a függvényt darabszam=500, nev="haziSAJATNEV.csv" argumentumokkal úgy, hogy a visszatérési értékeket mentsetek el egy pertekek nevű változóba. A változó ha mindent jól csináltatok, akkor az egyes futások p-értékeit fogja tartalmazni.

Gondoljátok át újra, hogyan is számolódik ki a p-érték. Ez alapján sejtsétek meg, hogy a pertekek vektorba került bejegyzések milyen eloszlást követnek. Végül teszteljétek 99%-os szignifikanciával Kolmogorov-Szmirnov próbával, hogy igazatok volt-e (vagyis végezzetek KS tesztet a p-értékek vektorán).

A fenti kiegészítő kérdésekhez kötődő R parancsokat az elsohazivaz.R script file-ba írjátok be alulra (a függvényen kívülre). Míg a pertekek vektoron végzett teszt p-értékét és a teszt kiértékelését kézzel (értsd szövegszerkesztő segítségével) írjátok bele a függvény által kiírt csv file-ba. Megoldásképpen a kiegészített elsohazivaz.R file-t és a függvény által kiírt és általatok kiegészített csv file-t várom.

Segítség:

Az egymintás Kolmogorov-Szmirnov tesztel folytonos háttérváltozók illeszkedését lehet tesztelni (a feladat megoldásához elég a p-érték koncepció ismerete, de érdeklődöknek megjegyzem, hogy a tankönyv új kiadásának 179. oldalán találhatóak a matematikai részletek). Itt találjátok a tesztet futtató függvény dokumentációját:

<https://stat.ethz.ch/R-manual/R-patched/library/stats/html/ks.test.html>

Az alternative és exact nevű paramétereket ne adjátok meg, hagyjátok default-an őket. Az első argumentum a tesztelendő vektor. A második argumentumban amint látjátok a tesztelendő eloszlás nevét kell megadni. Normális eloszlás esetén ez „pnorm”. (a főoldalon belinkelt angol R bevezető 17-18. oldalain megtaláljátok az elnevezéseket). A második argumentum után kell felsorolni a tesztelendő eloszlás paramétereit.

Megjegyzés:

Természetesen ha valós adataink vannak, akkor mindössze egy Kolmogov-Szmirnov tesztet végzünk. A házi a többszöri szimulálással majd tesztel azt szemlélteti, hogy a statisztikai döntések a véletlentől függenek. Előfordulhat, hogy igaz a nullhipotézis mégis elutasítunk.

Több statisztikai programban a Kolmogorov-Szmirnov tesztnél van egy kis csalás. Csak azt kell megadni, hogy milyen eloszláshoz való illeszkedést teszteljen a program. Ezután például normális-hoz való illeszkedés esetén a program kiszámolja az empirikus várható értéket és szórását, és az ilyen paraméterű normális-hoz hasonlítja az adatokat. Mindezt úgy lehet megfogalmazni, hogy a program egy kicsit módosított Kolmogorov-Szmirnov próbát hajt végre, ami hajlik arra, hogy elfogadjuk a nullhipotézist. Ennek az előnye az, hogy általában lehet tesztelni azt, hogy a változó például normális-e, anélkül, hogy ismernénk a várható értéket és a szórását, hátránya a leírt torzítás. Léteznek az említett torzítást korrigáló statisztikai eljárások.

Az R-ben nincsenek beéptve a fenti torzítást korrigáló eljárások (legalábbis én így tudom). Így érdemes megadni a tesztelendő eloszlás paramétereit.

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