

Töltsd le az elsohazivaz2016.R file-t. A scriptben levő kommenteket követve fejezd be a függvény megírását. Futtasd le a függvényt!

Ezután azt szeretném ha lefuttatnátok a függvényt darabszam=500, nev="haziSAJATNEV.csv" argumentumokkal úgy, hogy a visszatérési értékeket mentsétek 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 sejtsetek 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 átalatott 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 érdeklődőknek megjegyzem, hogy a tankönyv új kiadásának 179. oldalán találhatóak a matematikai részletek). A nullhipotézis az, hogy a megadott eloszlást követi a minta. Itt találjátok a tesztet futtató függvény dokumentációját:

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

A futtatás és a p-érték kiolvasás részleteit az elsohazivaz2016.R file elején találjátok. Ha nem normális eloszlást teszteltek, akkor értelemszerűen le kell cserélni a "pnorm"-ot és a paramétereket.

Megjegyzés:

Természetesen ha valós adataink vannak, akkor mindössze egy Kolmogorov-Szmirnov tesztet végzünk. A házi a többszöri szimulálással majd tesztel azt szemlélteti, hogy a statisztikai döntések a véletlentől függnek. 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 normálishez való illeszkedés esetén a program kiszámolja az empirikus várható értéket és szórást, és az ilyen paraméterű normálishez 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.

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