

Az employeedatahazihoz.sav állományban fogjuk lineáris regresszióval becsülni a jelenlegi fizetést (salary) a kezdő fizetés (salbegin), munkában töltött idő (jobtime), korábbi tapasztalat (prevexp), és a beosztás változókkal (jobcat). Az employeedatahazihoz.sav annyiban különbözik a megszokott employeedata-tól, hogy véletlenszerűen megkevertem az esetek sorrendjét.

Megszokott dolog becslőmodell építésnél, hogy az adathalmaz egyik részén kerül felépítésre a modell (tanuló rész), míg a másik részén vizsgáljuk meg a teljesítményét (tesztelő rész). Ezzel elkerülhető a tútanulás jelensége. Ennek megfelelően az első részfeladat az, hogy hozzatok létre egy „tanulo” nevű változót, ami az első 300 esetenél 1, egyébként pedig 0. Mindezt a Transform/Compute variable parancsával tudjátok megtenni, úgy, hogy használjátok az alul található if lehetőséget a &Casenum (esetszám) függvénnyel (&Casenum <= 300, illetve >300 feltételeket kell megadni). Összesen kétszer kell futtatni a Compute variable parancsot, a második futtatással felül kell írni a régit. Így az első 300 eseten fogjuk felépíteni a lineáris modellt, tesztelni pedig a maradékon fogunk.

A jobcat változót is szerepeltetni akarjuk a regresszióban. Ez egy kategorikus változó. Simán szerepeltetve is növelheti a modell becslési képességét, de érdekesebb a következő módon szerepeltetni. Hozzatok létre két új bináris változót, egy „biztonsagie” es egy „managere” változót, amelyek akkor 1-ek, ha a személy a névben található beosztású, egyébként pedig 0-k. Vagyis eggyel kevesebb bináris változót hozunk létre, mint ahány kategória van. A két új változót a Compute Variable parancsral létre tudjátok hozni (1,2,3 a három beosztás kódja). A regresszióban a jobcat helyett a két új változót fogjuk szerepeltetni. Ezzel voltaképpen megengedtük, hogy a lineáris regresszió konstansa a három munkakategóriában különbözzön.

A harmadik lépés magának a lineáris regresszióknak az elvégzése. Az első bekezdésben leírt szereposztással kell kitölteni a lineáris regressziós menüt azzal a különbséggel, hogy a jobcat helyett a „biztonsagie” es a „managere” változókat szerepeltetjük. Két kiegészítést kell tenni. A Selection Variable menüpont alá be kell vinni a „tanuloe” változót „equal to 1” szabállyal, továbbá a save almenüben kérni kell az unstandardized becslés elmentését. Így futtatva a regressziót a program azt csinálja, hogy azokon az eseteken ahol a „tanuloe” változó 1 építi fel a regressziós modellt, de a kapott modellel a többi esetet is becsli. A kapott statisztikák az építő részre vonatkoznak, egy kivétellel, szerepelni fog a teszt részen számolt R statisztika, amely mutatja, hogy mennyire működik jól a modell a tesztalacson.

Negyedik részfeladatként elemezzétek ki az eredményt. Az együtthatókból illetve a beta oszlopból mit tudsz mondani a változók jelentőségével kapcsolatban?

Ötödik részfeladatként azt kérem, hogy a Data/Select cases parancsral válasszátok ki azokat az eseteket, amelynél a „tanuloe” változó 0, vagyis az adathalmaz tesztelésre szánt részét. Majd a Graph/Chart builder-ben ábrázoljátok a Scatter/Dot első ábratípusán a salary és a becslése változókat (egyik az x, másik az y tengelyre kerül).

Megoldásképpen az adatfájlt, egy táblázatokkal, magyarázó szöveggel és ábrával ellátott doc vagy pdf fájlt (felesleges dolgok kiszűrésével) és a command syntacsot várom.

Megjegyzés: A fenti regresszióban megengedtük, hogy jobcat kategóriánként más-más legyen a regressziós modell konstansa. Ha a használt bináris változókat összeszorozzuk egyéb változókkal, akkor arra is lehetőségünk van, hogy megengedjük, hogy egyes változók előjele különbözzön kategóriánként. Utóbbi úgy szokták megfogalmazni, hogy interakciós lehetőséggel bővíthetjük a modellt. De ez most nem feladat.

Jó munkát kívánok! Bármilyen kérdésetek van, forduljatok hozzám bizalommal!