

Esettanulmányok és gyakorlati példák

a Békés-Kézdi könyvben

Adatelemzés workshop, 2024.10.10

Bárdits Anna

Esettanulmányok a tankönyvben

- Minden fejezetben 2-3 esettanulmány
- A tanult fogalmak bemutatása valódi példákon keresztül
- Kitér a gyakorlati problémákra, amik előjöhetnek, néhány példa:
 - Eltávolítunk-e szélsőséges értékeket?
 - Hogyan válasszunk rekesz-szélességet?
 - Milyen esetben transzformáljunk egy változót?
 - Milyen modellt válasszunk - lowess regresszió, szakaszonként lineáris szplájn, négyzetes függvény
 - Alkalmazzunk-e súlyokat az adatunkon
 - ...
- Világossá és átélhetővé válik, hogy a konkrét adat határozza meg a döntéseinket, és tipikusan nincs egy abszolút jó válasz
- Sorvezetőként használható a diákoknak saját elemzés elkészítésénél

ESETANULMÁNY – VÁLLALATVEZETÉSI MINŐSÉG ÉS A VÁLLALATOK MÉRETE: A KAPCSOLATI MINTÁZATOK LEÍRÁSA

Feltételes átlag és együttes eloszlás

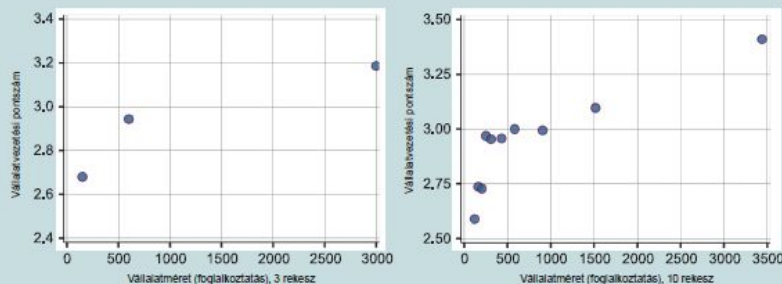
Térjünk vissza a vállalatvezetés minőségéről és a vállalat méretéről szóló esettanulmányunkhoz, amelyben mexikói cégek adatait használjuk. Az x a vállalatvezetési pontszám, az x pedig a foglalkoztatás. Emlékezzünk arra, hogy adataink olyan cégeket tartalmaznak, amelyek 100–5000 alkalmazottat foglalkoztatnak, és a foglalkoztatás eloszlása ferde és jobbra elnyújtott (4.2a. ábra).

A következő két ábrán (4.4a. és 4.4b. ábra) két rekesz-pontdiagramot láthatunk, amelyek az átlag y -t mutatják három x rekesz és tíz x rekesz feltétele mellett. A három rekeszt a korábbiakhoz hasonló módon hoztuk létre, 100–199, 200–999 és 1000+ alkalmazottal. A rekeszekben 72, 156, illetve 72 cég szerepel. A tíz rekesz kialakításakor más megközelítést alkalmaztunk: a szétválasztásukhoz nem kerestük a számokat, hanem az adatokat az alkalmazottak száma szerint tíz egyenlő méretű részmintára osztottuk. A legtöbb statisztikai szoftverrel egyszerűen tudunk ilyen rekeszeket alkotni. Mindkét rekesz-pontdiagramon az átlagos vállalatvezetési pontszámot egy pontként ábrázoljuk a foglalkoztatási rekesz felezőpontjánál (például a 100-199-es rekesz esetében a 150-nél).

A háromrekeszes rekesz-pontdiagram egyértelmű pozitív kapcsolatra utal: a nagyobb vállalatokat átlagosan jobban irányítják.

A kisvállalkozások esetében az átlagos vállalatvezetési pontszám 2,68, a közepes méretű cégeknél 2,94, a nagy méretűeknél pedig 3,19. A nagy és a kis cégek átlagos vállalatvezetési minősége között tehát mérsékelt különbség van (az 1–5-ig terjedő skálán körülbelül 0,5). A tízrekeszes rekesz-pontdiagram kevésbé egyenes lineáris kapcsolatot mutat és a 4–8. rekeszekben nagyon hasonló átlagokat láthatunk. Összességében azonban ez az ábra is magasabb átlagokat mutat a nagyobb vállalatokat tartalmazó rekeszekben: a két szélső rekeszben lévő vállalatoknál jelentős, 0,8 feletti különbséget látunk.

Végül figyeljük meg, a rekesz-pontdiagramok tükrözik azt is, hogy a foglalkoztatás nagyon ferde eloszlású: a nagyobb vállalatméreteknek nagyobb távolságok vannak a rekeszek felezőpontjai között. A rekesz-pontdiagramokat úgy is ábrázolhattuk volna, hogy ugyanezekhez a rekeszekhez az x tengelyen a foglalkoztatás helyett a foglalkoztatás logaritmusát ábrázoljuk. Ebben az esetben a rekeszek egyenletesebb eloszlását és lineárisabb mintázatot láthattunk volna. Pontosan ugyanazt a kapcsolatot



(a) A foglalkoztatás három rekesze

(b) A foglalkoztatás tíz rekesze

4.4. ábra Az átlagos vállalatvezetési pontszám és a vállalatméret

Megjegyzés: Rekesz-pontdiagramok.

Forrás: wms-management-survey adatbázis. 100–5000 fős mexikói vállalatok. $N = 300$.

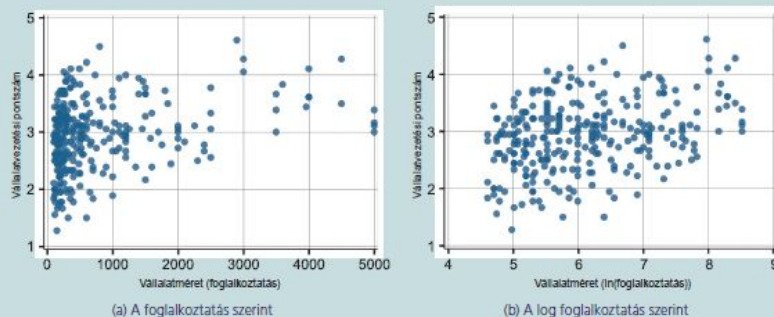
mutatnak tehát azok az ábrák is: a nagyobb cégeknél általában jobb minőségű a vállalatvezetés, és az abszolút foglalkoztatás tekintetében a vállalatvezetés minőségbeli különbsége egyre kisebb a több főt foglalkoztató cégeknél. (Az Adatfeladatok között ez a feladat is szerepel.)

A rekesz-pontdiagramok átlagosan egy pozitív kapcsolati mintázatot mutatnak. Vajon ez azt jelenti, hogy az adatainkban szereplő összes rekeszre jobban irányítják? Egyáltalán nem. Nézzük meg a pontdiagramot, hogy jobban megértsük, milyen a vállalatvezetési pontszám eloszlása a feltételes átlagértékei körül!

A bal oldali ábrán (4.5a. ábra) megfigyelhetjük a nagyon ferde eloszlás következményeit: a legtöbb megfigyelés az ábra bal szélső részén csoportosul. Ezen a grafikonon a nagyobb vállalatok között egy pozitív meredekséget láthatunk, de a kisebb cégek körében nehéz bármilyen mintázatot találni.

A jobb oldali ábrán (4.5b. ábra) a mintázatokat jobban láthatóvá tesszük: ugyanezt a pontdiagramot mutatjuk, de az x tengelyen a foglalkoztatás helyett a foglalkoztatás természetes alapú logaritmusát ábrázoljuk. Ez azt jelenti, hogy a cégek közötti foglalkoztatási különbségeket a kevesebb főt foglalkoztatóknál megnyújtjuk, a többet foglalkoztatóknál pedig összenyomjuk. (Később, a 8. fejezetben részletesen foglalkozunk azzal, mit tesz egy ilyen átalakítás a változókkal és a változók összehasonlításával.) Ezen a pontdiagramon már egy szétterjedtebb mintázat látható, ami az x változó szimmetrikusabb eloszlását tükrözi. Az átlagos vállalatvezetési pontszám és a foglalkoztatás logaritmusa közötti pozitív kapcsolat is jobban látható.

Mindenesetre az is látszik mindkét ábrán, hogy a vállalatvezetési pontszám sokféle értéket vesz fel a foglalkoztatás minden szintjén. A vállalatvezetési pontszám tehát az azonos méretű cégek között is nagy szóródást mutat. Az eloszlás többi jellemzőjét nézve a pontdiagramon nem látunk egyértelmű mintázatot a foglalkoztatás (vagy a log foglalkoztatás) és az eloszlások szóródása vagy ferdesége között. De ez azért is elképzelhető, mivel ezek a jellemzők nem mindig olvashatók le könnyen egy pontdiagramról.



(a) A foglalkoztatás szerint

(b) A log foglalkoztatás szerint

4.5. ábra A vállalatvezetési pontszám és a vállalatméret együttes eloszlása

Megjegyzés: Pontdiagramok.

Forrás: wms-management-survey adatbázis. 100–5000 fős mexikói vállalatok. $N = 300$.

A vállalatvezetési pontszámot három foglalkoztatási rekesz szerint doboz- és hegedűdiagramokkal ábrázoljuk (4.6. ábra), mivel még alaposabban meg akarjuk vizsgálni, eltérő-e – és ha igen, mennyire eltérő – a vállalatvezetési pontszám eloszlásának szóródása vagy ferdesége a foglalkoztatás különböző szintjein.

Esettanulmányok változatos témákban

- Könnyű az adott csoport igényeire szabni
- Egy adat több témán is végigvonulhat, és minden témára több megfelelő adat
- <https://gabors-data-analysis.com/casestudies/>
- Szállás árak Bécsben, online- és offline árak összehasonlítása, menedzsment minőség és vállalatméret, átooltottsági arányok országonként, sikeres fociedzők, nemek közti bérkülönbségek, veszteségek egy részvényportfóliónál....
- Én most a (kiscsoportos) kurzusomon: óra előtt elolvassák a fejezetet, az órán egy esettanulmányt nézünk végig, élőben futtatom a kódokat. Adatos házit kapnak (fejezetek végén) az esettanulmány továbbgondolása

Egy konkrét példa - Szállásárak Bécsben

<https://gabors-data-analysis.com/data-and-code/>

1. Adat letöltése
2. Kódok letöltése
3. Milyen szoftver? (**R**, stata, python?)
4. A kódot elejétől végéig futtatva visszacapjuk a tankönyv ábráit (set_working_directory scriptbe kell csak belenyúlni). Könnyű a saját igényeinkre formálni