

8. fejezet

KLASZTERELEMZÉS

A fejezet céljai

A fejezet célja, hogy az olvasó a fejezet elolvasása után:

- ➔ ismerje a klaszterelemzés folyamatát,
- ➔ képes legyen a megfigyelési egységek releváns szempontok alapján történő csoportosítására,
- ➔ elemezni tudja a kialakult csoportokat.

8.1. A klaszterelemzés gyakorlati alkalmazhatósága

A kutatók gyakran találkoznak olyan problémával, amely legjobban a valamilyen szempontból hasonló elemek csoportosításával oldható meg, legyen szó személyekről, vállalatokról, vagy akár magatartásformákról. A klaszterezés hasonló dolgok csoportosítását jelenti, s gyakorlatilag az osztályozás szinonimája. A klaszteranalízis alapvető célja, hogy a megfigyelési egységeket viszonylag homogén csoportokba rendezze, az elemzésbe bevont változók alapján. A folyamat akkor sikeres, ha az egységek hasonlítanak csoporttársaikhoz, azonban eltérnek a más csoportba tartozó elemektől.

A klaszterelemzés néhány gyakori alkalmazási területe:

- Piacszegmentálás. Gyakran piaci kudarcokhoz vezet az úgynevezett „a priori szegmentáció”, amikor a menedzsment adatok nélkül, csak saját elképzelései alapján próbálja csoportokra osztani a piacot. Alternatív módszerként jelent meg a klaszteranalízis alapján történő szegmentáció, amely valóban alkalmas a homogén csoportok azonosítására. Szegmentációs ismérv lehet például az érzékenység vagy a márkahűség.

- Piacszerkeze-elemzés: Struktúra megállapítása annak alapján, hogy a versenytárs márkákat mennyire tartják egymás helyettesítőjének.
- Új termék lehetőségeinek feltárása: A márkák és termékek csoportosításával meghatározhatók a piacon használható versenyeszközök (azonos klaszterbe tartozó márkák között élesebb lehet a verseny).
- Tesztpiacok kiválasztása: Homogén városokból álló klaszterek létrehozása különféle marketingstratégiák tesztelésére.
- Adatcsökkentés: A kutatók számos esetben óriási mintákkal rendelkeznek, ahol az elemzés egészen addig értelmetlen, míg nem hoznak létre kezelhető nagyságú csoportokat. Az információvesztés minimalizálása miatt a további többváltozós elemzéseket is a kialakított klasztereken végzik.

A sokrétű felhasználhatósága mellett a klaszterelemzésnek van néhány korlátja:

- Nem vonhatók le következtetések a mintából az alapsokaságra, vagyis elsősorban feltáró technikaként használható.
- Nincs egyetlen legjobb megoldás, mivel a klaszterbe való tartozás a választott eljárás függvénye.
- A klaszteranalízis minden esetben létrehoz klasztereket, függetlenül attól, hogy azok ténylegesen jelen vannak-e az adatokban.
- A klasztermegoldások teljes mértékben az elemzésbe bevont változóktól függenek. A releváns változók hozzáadása vagy törlése alapvetően befolyásolja az eredményeket. A kutatónak tehát körültekintően kell eljárnia minden ilyen döntésnél.

A klaszterelemzés a faktorelemzéstől eltérően nem a változók, hanem a megfigyelési egységek számát kívánja csökkenteni, mindazonáltal a klaszterelemzés is alkalmas a változók számának redukciójára. A klaszterelemzés a diszkriminációelemzéshez hasonlóan csoportosítással foglalkozik, azonban az utóbbi megköveteli a csoportokba tartozás előzetes ismeretét, s ez alapján alakít ki szabályokat minden megfigyelési egységre, míg a klaszteranalízisnél nem rendelkezünk előzetes (a priori) információkkal egyetlen elem klasztertagságáról sem.

A klaszterelemzésnél alapvető feladat azoknak a változóknak a megtalálása, amelyek a csoportok közötti különbséget okozzák, ezért gyakori megoldás az, hogy a klaszterelemzést a faktorelemzés során létrejött változók segítségével végzik. A változók kiválasztása történhet korábbi kutatások, elméleti megfontolások, hipotézisek, kutatói döntés, illetve (feltáró kutatásnál) intuíció alapján. A klaszteranalízis lépéseit és főbb kérdéseit az 8.1. táblázat foglalja össze.

	Lépések	Döntési helyzetek és tennivalók
1.	A kutatási probléma megfogalmazása	Mi a kutatási probléma? Mi a hipotézisünk? Mekkora legyen a vizsgálandó elemek nagysága? Milyen releváns változók alapján csoportosíthatunk?
2.	Klaszteranalízis feltételeinek vizsgálata	Vannak-e kiugró adatok? Ki kell-e zárni azokat az elemzésből? Azonos skálákat használtunk? Szükség van standardizálásra? Erősen korrelálnak az elemzésbe bevont változók? Reprezentatív a minta?
3.	Hasonlósági és távolságmérték meghatározása	Milyen hasonlóság- vagy távolságmértéket válasszunk? Bináris változó: <ul style="list-style-type: none"> ➤ Hasonlóságmértékek ➤ Távolságmértékek Gyakoriságok (counts): <ul style="list-style-type: none"> ➤ Távolságmértékek Metrikus változó: <ul style="list-style-type: none"> ➤ Távolságmértékek (ha a távolság a lényeges, gyakoribb) ➤ Hasonlóságmérték (ha a profilok hasonlósága lényeges, ritkább)
4.	Klasztermódszer kiválasztása	Hierarchikus vagy nem hierarchikus eljárást válasszunk, esetleg a kettő kombinációját?
5.	Döntés a klaszterek számáról	Mi alapján képezzünk klasztereket? Hány csoportot képezzünk? Hogyan változnak az eredmények különböző csoport-számoknál? Újra kell-e futtatni a klaszterelemzést?
6.	A klaszterek értelmezése és jellemzése	Miben különböznek egymástól a klaszterek? Értelmesen interpretálhatók az eredmények? Milyen egyéb, a klaszterek létrehozására nem használt változót vonjunk be a jellemzésbe? Hogyan nevezzük el a klasztereket?
7.	A klaszterelemzés érvényességének ellenőrzése	Milyen módszerrel ellenőrizzük az érvényességet?

8.1. táblázat. Lépések és döntési helyzetek a klaszterelemzés során