

## 6. fejezet

# KORRELÁCIÓ- ÉS REGRESSZIÓSZÁMÍTÁS

### A fejezet céljai

A fejezet célja, hogy az olvasó a fejezet elolvasása után

- ➔ felismerje a korreláció- és regressziószámítás alkalmazhatóságát és jelentőségét,
- ➔ ismerje a korreláció és regressziószámítás során szükséges változókat és a számítások elvégzésének feltételeit,
- ➔ tudja előállítani és elemezni a statisztikákat és a releváns mutatókat, képes legyen helyes következtetéseket levonni az elvégzett elemzések alapján.

A fejezet során a korreláció- és regresszióelemzéssel foglalkozunk<sup>1</sup>. A változók közötti összefüggés szorosságát, a kapcsolat erősségét, intenzitását korreláció-, míg az összefüggés jellegét regressziószámítással határozzuk meg. Utóbbi esetben az összefüggésekben lévő tendenciát, a kapcsolat természetét valamilyen függvényvel írjuk le<sup>2</sup>.

Először két gyakorlati példát nézünk át röviden, amelyekben a korreláció- és regressziószámítás jól alkalmazható. Ezt követően a mélyebb összefüggésekre térünk át: a kapcsolat meglétét és erősségét ellenőrizzük.

1. Tétélezzük fel, hogy Yamaha motort szeretnénk vásárolni. Az egyik internetes oldalon tizenkét különböző évjáratú (1993-tól 2004-ig) és négy különböző teljesítményű motort ( $75 \text{ cm}^3$ ,  $125 \text{ cm}^3$ ,  $250 \text{ cm}^3$ ,  $500 \text{ cm}^3$ ) találunk ebből a márkából. Ezenkívül a weboldal információt ad arról is, hogy hány

<sup>1</sup> A logisztikus regressziót a 9. fejezetben mutatjuk be.

<sup>2</sup> Ha a vizsgálatok során az idő a független változó, a számításokat trendszámításnak nevezzük, de a fejezetben idősorok elemzésével nem foglalkozunk.

korábbi tulajdonosa volt a motornak, történt-e vele baleset, garázsban tartották-e és milyen színű.

Szeretnénk megtudni: Van-e összefüggés (1) a motor teljesítménye és az ára; (2) az évjárata és az ára, (3) a teljesítménye és évjárata vagy akár (4) a törésszám és az ár között? Ha igen, akkor a kapcsolat milyen erős?

2. Rendelkezésünkre áll egy olyan adatbázis, amelyben Bagaméri, a fagyfalt-árus feljegyezte az előző nyár minden egyes napján a következőket: hány gombóc fagyfaltot sikerült eladnia, hány fok volt a napi legmagasabb hőmérséklet, milyen mennyiségű csapadék hullott aznap, mekkora volt a felhőtlen (napsütéses) órák száma.

Bagaméri a következő kérdésre keresi a választ: Milyen összefüggés lehet az eladott fagyimennyiség, a napsütéses órák, a csapadékmennyiség és a hőmérséklet között?

## 6.1. Korrelációs számítás

A korrelációs számítás a változók közötti lineáris kapcsolat szorosságának és irányának leírására szolgál. A bemutatandó korrelációelemzésnél metrikus változókkal dolgozunk<sup>3</sup> (6.1. táblázat).

		Független változó	
		Nem metrikus	Metrikus
Függő változó	Nem metrikus	Keresztábra elemzés	Diszkriminancia elemzés
	Metrikus	Variancia elemzés	<b>Korreláció, Regresszió elemzés</b>

6.1. táblázat. A struktúravizsgáló módszerek egy részének összefoglalása

### 6.1.1. Lineáris korreláció

Ennél a korrelációs számításnál két változó közötti kapcsolatot vizsgálunk. A lineáris korrelációs (vagy Pearson-féle) együttható (jele:  $r$ ) értékét a következő módon számíthatjuk ki:

<sup>3</sup> Nem metrikus változóknál használhatjuk a Spearman-féle rangkorrelációs együtthatót, amely rangsorskálák vagy olyan intervallumskálák esetén alkalmazható, amikor nem teljesül a normalitás feltétele. Hasonló elven működik a Kendall-féle tau-b együttható is, amely szintén ordinális skálák esetén használható. Az utóbbi két mutatószám értelmezése hasonló a Pearson-féle korrelációs együttható értelmezéséhez.