

A véletlen hatásokra vonatkozó eloszlásbeli feltételezések ellenőrzése

A rögzített faktorok szerinti varianciaanalízisnél az eloszlásbeli feltételek (függetlenség, konstans variancia, normális eloszlás) az ε véletlen hibára vonatkoznak. Az epszilon azonban nem ismert, hanem úgy járunk el, hogy az ε helyett az azzal nem pontosan azonos reziduumokra vizsgáljuk e feltételek teljesülését.

Ha több ingadozás-forrásunk (véletlen hatásunk) van, akkor mindegyiknek az eloszlásáról, függetlenségéről és varianciájuk konstans voltáról nyilatkoznunk kell, nemcsak a modell ε tagjáéről. A csoportokon belüli és közötti függetlenség az epszilon és a véletlen faktorok függetlenségére is értendő.

Vegyük például az egy véletlen faktor esetét! A modell

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

ahol α_i az A véletlen faktor hatásának i -edik szintjét jelöli.

A véletlen hatások (itt α_i) nagysága éppúgy nem ismert, mint ahogy ε -é. A véletlen hatásokkal analóg reziduumok közvetlenül nem adódnak az adatokból. Ha azonban az A véletlen faktor esetében úgy járunk el, hogy az ismétlések szerint átlagolunk, és az átlagokat tekintjük mérési adatnak, akkor elvégezve a varianciaanalízist, olyan reziduumokat kapunk, amelyek az A véletlen faktor egy-egy szintjéhez tartoznak. Igaz, hogy ezekben a reziduumokban az ε (átlaga) is benne van, de ha igaz az, hogy a véletlen faktor hatása (varianciája) nagyobb, mint az ε -é, pláne az ε -ok átlagáé, akkor ezen reziduumokra elvégezve a vizsgálatokat közelítőleg a véletlen faktorra végezzük el. Persze a közelítés annál jobb, minél inkább elhanyagolható az a része a reziduumnak, ami az ε -ok átlaga miatt kerül be, ez pedig akkor ilyen, ha n nagy és $\text{var}(\varepsilon)$ sokkal kisebb a véletlen hatás varianciájánál.

Ha a reziduumok eloszlására vagyunk kíváncsiak, meg kell adni, hogy mihez képesti reziduumokat akarunk, ez a modell.

Vegyünk egy kicsit bonyolultabb példát, legyen egy B faktorunk is (például a labor), és ennek minden szintjébe ágyazzuk az A faktort (pl. a mintát), az ismétlés az ismételt injekciót jelenti. Ha a modellt úgy adjuk meg, hogy $A(B)$, tehát minta(labor), akkor csak az ismétlés (injekció) nincs felsorolva, tehát az adja a reziduumot. Ha úgy adjuk meg, hogy B (labor), akkor az A (minta) és az ismétlés (injekció) miatti eltérés is a reziduumba kerül. Nem pontosan ezt kérdezzük, tehát jó lenne az ismétlés miatti eltérést kivenni. Ezt nem tudjuk megtenni, csak csökkenteni, ha kiszámítjuk minden mintára a több ismételt mérés átlagát, és ebből a táblázatból dolgozunk tovább, úgy, hogy az átlagokat tekintjük mérési adatoknak.