

## A statisztikai próbák gondolatvilága

---

### Vita László

CSc, a Budapesti Corvinus  
Egyetem egyetemi tanára

E-mail: laszlo.vita@uni-  
corvinus.hu

A szerző sorra veszi a hipotézisvizsgálat lépéseit, kitér azok szerepére, logikájára, vitatott pontjaira és buktatóira. Ezután a hipotézisvizsgálat során elkövethető hibákkal, majd a szignifikanciaszint-választás problémáival, lehetőségeivel foglalkozik. Ezen belül említést tesz az értelmezés leggyakoribb hibáiról, valamint a statisztikai és szakmai szignifikancia megkülönböztetésének fontosságáról. Ennek kapcsán kitér a hatásvizsgálatok végzésének fontosságára, amit egy egyszerű példával illusztrál is.

TÁRGYSZÓ:  
Statisztikai próba.  
Hipotézisvizsgálat.

A hipotézisvizsgálat alkalmazott statisztikán belüli használata igen elterjedt és széles körű. Olykor már talán túlzottan és indokolatlanul is az, mint arra később még utalunk.

A hipotézisvizsgálatra a gyakorlatban minden olyan esetben szükség van, amikor valamely sokaság(ok), illetve eloszlások jellemzőivel kapcsolatban bizonyos feltevéseink vagy elvárásaink vannak, s azok teljesülését nem teljes körű adatfelvételtől, hanem csak a sokaságból vett mintából nyert információkra támaszkodva tudjuk vizsgálni.

A hipotézisvizsgálat egyik legkézenfekvőbb gyakorlati alkalmazása talán az, amikor a vizsgálat eredménye alapján azt mérlegeljük, hogy a mintából nyert adatok igényelnek-e külön magyarázatot, elemzést. Erre ugyanis csak akkor van szükség, ha a mintából vagy mintákból nyert adatok eléggé különböznek egymástól vagy egy valamilyen alapon elvárt értéktől, és az eltérés nem csak a mintavétellel szükségszerűen együtt járó véletlenszerű ingadozásoknak tulajdonítható. A próbák pedig éppen ennek a mérlegelésére szolgáló eszközök.

A próbák ugyancsak fontos szerepet játszanak a különféle gazdasági vagy társadalmi jelenségek leírására törekvő statisztikai, ökonometriai stb. modellek építése és használata során. A modellépítésnek ugyanis igen lényeges eleme mind a modell egyes komponenseire vonatkozó bizonyos feltevések teljesülésének ellenőrzése, mind a modellel nyert becsléseknek a valóság tényeivel való egybevetése. Mindkét feladat megfelelő hipotézisvizsgálatok segítségével oldható meg a leghatékonyabban. Elegendő itt a regressziós modellek feltételrendszerének ellenőrzésére szolgáló tesztekre, valamint a regressziós együtthatók bizonyos kitüntetett értékekkel való összehasonlítására használatos tesztekre utalni.

E meglehetősen általános alkalmazásokon túl a hipotézisvizsgálat talán legismertebb felhasználási területe a minőségellenőrzés, valamint a piac- és közvélemény-kutatás. A minőségellenőrzés esetében már a gyártás közben vagy a gyártás végén ellenőrizni kívánjuk, hogy a termelés eleget tesz-e bizonyos előírásoknak, követelményeknek. Ezt a gyártósorról lekerülő termékekből vett véletlen minta vizsgálatára alapozzuk. A piackutatás esetében arra keressük a választ, ugyancsak mintavételes adatgyűjtések eredményeire támaszkodva, hogy mitől függenek, illetve miként befolyásolhatók a fogyasztók vásárlási szokásai. A közvélemény-kutatás során az állampolgárok vélekedését kívánjuk megismerni bizonyos kérdésekről ugyancsak mintavétellel nyert adatokra támaszkodva. Mindhárom alkalmazási területre a kis, illetve közepesen nagy minták használata a jellemző.

A hipotézisvizsgálat gyakorlati alkalmazása során meglehetősen sok a helytelen és/vagy nem elég körültekintő felhasználás. Emiatt a szakirodalomban – különösen a

pszichológusok, szociológusok és orvosi-egészségügyi kutatók körében – elég gyakoriak a hipotézisvizsgálat gyakorlati alkalmazásával kapcsolatos viták is. Ezek áttekintése vagy összegezése meglehetősen reménytelen vállalkozás egy rövid cikk keretében. Ezért ehelyett arra vállalkozunk, hogy a hipotézisvizsgálat menetének és logikájának rövid áttekintéséhez kötve utalunk a vitatott kérdésekre, téves értelmezésekre, illetve néhány megszívlelendő tanácsot adunk a hipotézisvizsgálat gyakorlati alkalmazói számára.

Anélkül, hogy ennek részleteibe belemennénk, megjegyezzük, hogy a hipotézisvizsgálatnak több – legalább három – egymástól markánsan elkülönülő irányzata, „iskolája” van.<sup>1</sup> Az első a *Ronald Fisher* által követett gyakorlat, ami kizárólag a kétoldali  $p$ -értékre<sup>2</sup> támaszkodva ítéli meg a nullhipotézis minta általi támogatottságát, a második a bayesi megközelítés, aminek részleteiről ebben a számban is olvashatunk *Hunyadi* [2011] tanulmányában, a harmadik pedig a ma talán legáltalánosabban követett Neyman–Pearson-féle megközelítés. Az ezután következő összes fejtegetés ezt veszi alapul.

A statisztikai hipotézisvizsgálat arra irányul, hogy egy vagy több sokaságra vonatkozó olyan feltevések – ún. hipotézisek – helyességét vizsgálja, ellenőrizze mintavételi eredményekre támaszkodva, melyek fennállásában nem vagyunk biztosak. A hipotézisek a vizsgált sokaság(ok) eloszlására vagy az adott eloszlás(ok) egy vagy több paraméterére vonatkozhatnak. A hipotézisek helyességének ellenőrzésére különféle tesztek, próbákat használunk.

Már most érdemes leszögezni: egy hipotézisvizsgálat eredménye sohasem annak kimondása, hogy a kitüntetett hipotézis – az ún. nullhipotézis – igaz vagy nem igaz, hanem mindig csak az, hogy mennyire hihető az a mintavétel eredményeinek tükrében. Arról ugyanis, hogy egy hipotézis igaz-e vagy sem, száz százalékos bizonyossággal csak egy teljes körű adatfelvétel alapján lehetne meggyőződni.

Minden hipotézisvizsgálat menete és logikája ugyanaz. Ezért minden hipotézisvizsgálat ugyanazokból a lépésekből áll, s az egyes próbák kizárólag csak bizonyos technikai elemeikben különböznek egymástól. A továbbiakban e lépéseket veszem sorra, azokhoz bizonyos kiegészítő megjegyzéseket fűzve.

## 1. Hipotézisek megfogalmazása

Minden hipotézisvizsgálat két egymásnak ellentmondó feltevés: egy  $H_0$ -lal jelölt nullhipotézis és egy  $H_1$ -gyel jelölt ellenhipotézis – más néven alternatív hipotézis – megfogalmazásával kezdődik. A két hipotézisnek olyannak kell lennie, hogy azok

<sup>1</sup> Ezekről jó áttekintés található *Berger* [2003] tanulmányában.

<sup>2</sup> A  $p$ -értékről és annak meghatározási módjáról a 7. pontban esik majd szó.

- a formális logika szabályai szerint kizárják egymást, azaz ne lehessenek egyszerre igazak;
- bármelyikét is tekintjük majd a másikkal hihetőbbnek, megválaszolható legyen a bennünket érdeklő kérdés.

A hipotézisvizsgálat közvetlenül mindig a nullhipotézis helyességének ellenőrzésére irányul. Ezért a két hipotézis nem játszik szimmetrikus szerepet, nem cserélhető fel tetszés szerint. A nullhipotézis szinte mindig azt mondja ki, hogy valami – egy adat vagy egy eloszlás – nem tér el valami mástól: egy másik adattól vagy eloszlástól. Erre utal a nullhipotézis elnevezés is. A hipotézisvizsgálat végzőjét igen gyakran nem az érdekli, hogy a nullhipotézis fennáll-e vagy sem, hanem sokszor az alternatív hipotézisben szereplő állítás helyessége. Azonban, ha betartják a  $H_0$  és  $H_1$  megfogalmazásával kapcsolatos első követelményt, a nullhipotézis helyességéről való döntés egyben döntést jelent az alternatív hipotézis helyességéről is. Ha mód van rá,  $H_0$ -t és  $H_1$ -t célszerű úgy megfogalmazni, hogy  $H_0$  elvetése legyen igazán fontos számunkra. Ennek indoklására később – a hipotézisvizsgálat során elkövethető hibák tárgyalása kapcsán – még visszatérünk.

A két hipotézis megfogalmazásával kapcsolatban érdemes megjegyezni, hogy a vizsgálat végzőjét valójában érdeklő szakmai hipotézis<sup>3</sup> gyakran még nem olyan formában van megfogalmazva, hogy annak helyessége a statisztika eszközeivel közvetlenül vizsgálható. Ezért a statisztikai úton kezelhető ún. statisztikai  $H_0$  és  $H_1$  megfogalmazása előtt gyakran még arra is szükség van, hogy operacionalizáljuk a kutatót érdeklő kérdést, feltevést, majd csak ezután és ezzel összhangban fogalmazzuk meg a két statisztikai hipotézist. Ez azt jelenti, hogy a hipotézisvizsgálat eredménye valójában még a szakmai hipotézist a statisztika eszközeivel kezelhetővé tevő operacionalizálás során tett különféle feltevésektől, megoldásoktól is függ. Az operacionalizálás sokszor többféleképpen is elvégezhető. Ez főképpen a hipotézisvizsgálat pszichológiai és szociológiai alkalmazásaira igaz, de nem egyszer még a közgazdaságiakra is. Ezért a hipotézisvizsgálat eredményeinek közlésekor feltétlenül meg kell adni a szakmai hipotézis operacionalizálása során tett feltevéseket is. Ez legtöbbször a szakmai hipotézisben szereplő jelenségek, változók mérési módjának és szintjének megadását, illetve tisztázását igényli.

Ha például azt akarjuk vizsgálni, hogy a gyermek társadalmi státusa függ-e a szüleitől, akkor a statisztikai hipotézisvizsgálat elvégzése előtt még természetesen statisztikailag kezelhetővé kell tenni, valahogyan mérni kell a társadalmi státust. Ugyanez a helyzet azonban a legtöbb orvosi kezelés, gyógyszer hatásvizsgálata esetén is, mert azt, hogy valamely kezelés, illetve gyógyszer hatásos-e vagy sem, mérni kell valahogyan.

<sup>3</sup> Egyes szerzők a szakmai hipotézist kutatási hipotézisnek nevezik.

Fontos végül még azt is megjegyezni, hogy az ellenőrizni kívánt hipotézisek – a  $H_0$  nullhipotézis és a  $H_1$  alternatív hipotézis – megfogalmazása mindig meg kell előzze a mintavételt. Ez fokozottan igaz azokra az esetekre, amikor  $H_0$ -al szembe nem egy, azt általánosan tagadó  $H_1$  alternatívát állítunk, hanem azt valamely irányban tagadó ún. egyoldali alternatívát. Ennek részleteiről a megtartási és visszautasítási tartomány megválasztása kapcsán lesz majd szó.

Nyilvánvaló ugyanis, hogy a hipotézisek ugyanazon minta alapján történő megfogalmazása és ellenőrzése indokolatlanul növeli a nekünk tetsző állítások alátámasztásának esélyét. Ha esetleg a  $H_0$  vagy  $H_1$  hipotézist mégis mintavételi eredményekre támaszkodva vagyunk kénytelenek megfogalmazni, akkor azok helyességét mindig más mintára alapozva törekedjünk ellenőrizni. Amennyiben ez valamilyen oknál fogva nem lehetséges, akkor a hipotézis megfogalmazása és ellenőrzése előtt célszerű a mintát véletlenszerűen két almintára bontani, s a hipotézisek megfogalmazására az egyik, azok ellenőrzésére pedig a másik almintát használni.

## 2. Próbafüggvény-választás

A következő lépés a nullhipotézis helyességének ellenőrzésére alkalmas  $T(y_1, y_2, \dots, y_n)$  – röviden csak –  $T$  próbafüggvény képzése vagy választása. A próbafüggvény a hipotézisvizsgálat elvégzéséhez szükséges mintabeli információ kinyerésére szolgál.

A  $T$  próbafüggvény a mintavétel előtt mintáról mintára ingadozó valószínűségi változó, a mintavétel után pedig az adott valószínűségi változónak egy konkrét értéke, realizációja. Valamely próbafüggvény akkor alkalmas a  $H_0$  helyességének ellenőrzésére, ha annak eloszlása  $H_0$  igazságát feltételezve és bizonyos további feltételek fennállását biztosnak véve teljesen konkrét. E további feltételek az adott próba alkalmazásának feltételei. Az alkalmazási feltételek részben a sokaság(ok) eloszlásának típusára és/vagy bizonyos paramétereire vonatkozhatnak, részben a mintavétel módjára, több minta együttes használata esetén pedig még a minták egymáshoz való viszonyára is. Az olyan  $H_0$  hipotézist – aminek fennállása valamely próba alkalmazási feltételeinek teljesülése, teljesen konkrétá és ismertté teszi a próbafüggvény eloszlását – egyszerű hipotézisnek szokás nevezni.

A próbafüggvények egymással való összehasonlítása és konstruálása egy-egy konkrét nullhipotézis és alkalmazási feltételrendszer mellett alapvetően elvi, matematikai feladat. A próbafüggvények összehasonlítása, minősítése és konstruálása

sokban hasonlít a becslőfüggvények esetében követett gyakorlathoz. Az összehasonlítás, minősítés azonban rendszerint nem olyan könnyen kivitelezhető, mint a becslőfüggvények esetében.

A próbafüggvények fontos kismintás tulajdonságai a torzítatlanság és az erő. Egy tesztet akkor nevezünk torzítatlannak, ha a hibás nullhipotézis visszautasításának valószínűsége nagyobb, mint a helyesé. A később említésre kerülő erőfüggvény felhasználásával ez úgy fogalmazható meg, hogy kétoldali próba esetén az erőfüggvény  $H_0$ -ban veszi fel a minimumát. Egy  $T_1$  próbát erősebbnek nevezünk valamely  $H_1$  pontban egy  $T_2$  próbánál, ha erőfüggvénye e pontban magasabb, mint a  $T_2$  próbáé. Ha ez egy egész tartományra kimondható, akkor egyenletesen erősebb próbáról beszélünk. Ha valamely tartományon egy próba minden más próbánál (minden pontban) erősebb, akkor azt egyenletesen legerősebb próbának (uniformly most powerful – UMP) nevezzük. A nagymintás tulajdonságok közül legfontosabb a konzisztencia. Egy próbát akkor nevezünk konzisztensnek, ha erőfüggvénye minden  $H_0$ -tól különböző pontban tetszőlegesen közel jut az 1-hez a mintanagyság minden határon túli növelése esetén.

Konzisztens próbák készítésére viszonylag jól kezelhető eszköztár áll rendelkezésre: a három  $\chi^2$ -alapú próbakészítési elv (likelihood arány, Lagrange-multiplikátor vagy score test és a Wald-féle elv) konzisztens próbákat eredményez. Ezek alkalmazása viszonylag kényelmes még összetettebb feladatok esetén is. Ezért, ha a mintanagyság növelése reálisnak tűnik, ezek a nagymintás eredmények jól hasznosíthatók. A próbafüggvények tulajdonságainak, minősítésének és konstruálásának további részletei megtalálhatók Hunyadi [2001] könyvének 11. fejezetében.

A hipotézisvizsgálat végzőjére a gyakorlatban rendszerint csak az általa ellenőrizni kívánt hipotézis(ek) vizsgálatára alkalmas próba kiválasztásának feladata hárul. A választás részben attól függ, hogy mi a nullhipotézis, részben pedig attól, hogy az adott esetben milyen alkalmazási feltételek teljesülésére lehet számítani.

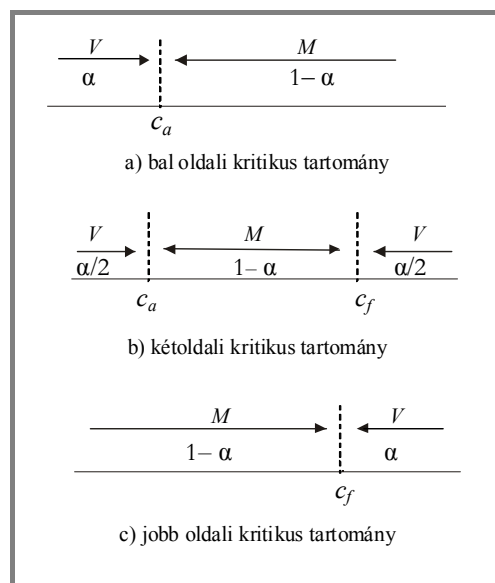
### 3. Megtartási és visszautasítási (kritikus) tartomány választása

Ezt követően el kell végezni a  $T$  próbafüggvény teljes értékészletének egy megtartási ( $M$ ) és egy visszautasítási – más néven kritikus – ( $V$ ) tartományra bontását oly módon, hogy  $H_0$  fennállása esetén  $P(T \in M) = 1 - \alpha$ , illetve  $P(T \in V) = \alpha$  álljon fenn, ahol  $\alpha$  egy 0-hoz közeli érték. Az  $1 - \alpha$  értéket a próba (megbízhatósági) szintjének, az  $\alpha$  értéket pedig szignifikanciaszintnek hívjuk. Mind a próba szintjét, mind a szignifikanciaszintet százalékává alakítva szokás megadni. Az  $M$  és  $V$  tartományt egymástól elhatároló pontok a kritikus értékek. Magukat a kritikus értékeket

mindig a kritikus tartomány részének szokás tekinteni. A két tartomány kijelölése a hipotézisvizsgálat igen fontos lépése, mert a  $H_0$ -ról hozott döntés végső soron a  $V$  tartomány megválasztásától függ.

Az  $M$  és  $V$  tartomány egymáshoz képesti elhelyezkedési lehetőségeit az ábra mutatja.

A kritikus tartomány lehetséges helyzete



Az ábrán látható számegyenesek a  $T$  próbafüggvény lehetséges értéktartományát jelképezik, ami speciális esetben véges is lehet.

Bal vagy jobb oldali – összefoglaló néven egyoldali – kritikus tartomány kijelölésére olyankor van szükség, ha nem közömbös számunkra, hogy a valóságos helyzet milyen irányba – balra vagy jobbra – tér el a nullhipotézisben rögzített helyzettől. Ilyenkor  $T$  túlságosan kicsi (vagy túlságosan nagy) értékei jelzik  $H_0$  helytelenségét. Ez azt jelenti, hogy ilyenkor  $H_0$ -l szembe nem egyszerűen egy annak teljes tagadását jelentő „általános”  $H_1$  alternatívát állítunk, hanem vagy egy a  $H_0$ -ban megadott helyzettől balra esést kimondó  $H_1^b$  bal oldali alternatívát, vagy egy attól jobbra esést kimondó  $H_1^j$  jobb oldali alternatívát.

Mivel azt akarjuk elérni, hogy a próbafüggvény  $M$  tartományba esésének valószínűsége  $1 - \alpha$ , kritikus tartományba esésének valószínűsége pedig  $\alpha$  legyen, bal oldali kritikus tartomány esetén a próbafüggvény eloszlásának  $p = \alpha$  rendű

kvantilisét,<sup>4</sup> jobb oldali kritikus tartomány esetén pedig az adott eloszlás  $p = 1 - \alpha$  rendű kvantilisét kell határpontnak, kritikus értéknek ( $c_a$ , illetve  $c_f$ ) választani.

Ha ezzel szemben közömbös számunkra, hogy a valóság milyen irányba tér el a  $H_0$ -ban rögzített helyzettől, akkor a kritikus tartományt megosztjuk  $T$  lehetséges értéktartományának bal és jobb széle között. Ilyenkor a  $V$  tartományba esés teljes  $\alpha$  valószínűségét legtöbbször egyenlő arányban szokás megosztani a kritikus tartomány két része között. Ez úgy érhető el, hogy az alsó határpont ( $c_a$ ) a próbafüggvény eloszlásának  $p = \alpha/2$  rendű kvantilise, a felső határpont ( $c_f$ ) pedig a szóban forgó eloszlás  $p = 1 - \alpha/2$  rendű kvantilise.

A legtöbb próba mind egyoldali, mind kétoldali kritikus tartomány kijelölése mellett végrehajtható. A kritikus tartomány elhelyezkedését ugyanis mindig a  $H_0$  hipotézissel szembeállított ellenhipotézisben szereplő feltevés, pontosabban e feltevés  $H_0$ -ban feltételezett helyzettől való eltérésének iránya határozza meg. A kritikus tartomány helyzetére utaló bal oldali, kétoldali és jobb oldali jelzőket igen gyakran a megfelelő alternatív hipotézisekre, sőt a hipotézisvizsgálat módjára is vonatkoztatják, használják. Egyes fontos próbák – a próbafüggvény bizonyos sajátosságainál fogva – csakis jobb oldali kritikus tartományt felvéve hajthatók végre. Vannak akik – tudományetikai alapon – vitatják az egyoldali alternatívák jogosságát.<sup>5</sup> Ettől függetlenül számos gyakorlati esetben kifejezetten szükség van az egyoldali alternatívák használatára. Gondoljunk csak például a minőségellenőrzés tipikus eseteire!

Amennyiben a próbafüggvény valamely minta alapján nyert értéke a kritikus tartományba esik, azt szokás mondani, hogy a minta adatai  $100 \cdot \alpha$  százalékos szignifikanciaszinten ellentmondanak a nullhipotézisnek, vagy csak egyszerűen azt, hogy a próba eredménye  $100 \cdot \alpha$  százalékos szinten szignifikáns. A szignifikancia annál nagyobb, erősebb, minél kisebb  $\alpha$  értéke. Egy valamely szinten statisztikailag szignifikáns eredmény önmagában még nem feltétlenül szignifikáns szakmai értelemben is. Erre a kérdésre rövidesen visszatérünk még.

#### 4. Döntés a nullhipotézisről

A hipotézisvizsgálat utolsó lépése az, hogy egy vagy több mintát veszünk a vizsgált sokaság(ok)ból, meghatározzuk a  $T$  próbafüggvény értékét, majd azt tekintetbe

<sup>4</sup> Egy folytonos valószínűség-eloszlás  $p$ -edrendű kvantilise az az  $x_p$  érték, melyben az  $F(x)$  eloszlásfüggvény helyettesítési értéke éppen  $p$ , azaz  $F(x_p) = p$ .

<sup>5</sup> Lásd például a Vargha [2007] 153. oldalán olvasható, R. J. Harristól vett idézetet.



véve döntést hozunk a nullhipotézisről ( $H_0$ ). Maga a döntés igen egyszerű: ha a próbafüggvénynek a minta (minták) adataiból számított értéke a  $V$  visszautasítási tartományba esik, akkor elvetjük  $H_0$ -t, ellenkező esetben nem vetjük el, hanem megtartjuk azt az adott szignifikanciaszinten. E döntésnek az a logikája, hogy ha a próba alkalmazási feltételei mellett még  $H_0$  is igaz, akkor a próbafüggvény csak kis valószínűséggel eshet a visszautasítási tartományba, s ha ez mégis bekövetkezett, akkor kételkedni kezdünk  $H_0$  fennállásában. Ha viszont  $T$  értéke az  $M$  tartományba esik, akkor ez egy olyan esemény, aminek a bekövetkezésére nagy valószínűséggel számítottunk, s így nincs semmi okunk a kételkedésre. Az ilyen alapon hozott döntés hibalehetőségeire rövidesen visszatérünk. Ez az eljárás és logika igen hasonlít az indirekt bizonyításhoz. Mindössze annyiban tér el attól, hogy itt a kutatót valójában érdeklő  $H_1$  ellenkezőjét kimondó  $H_0$  fennállásából nem teljes bizonyosságú, hanem csak nagy valószínűséggel teljesülő következtetést végzünk.

A hipotézisvizsgálat eredménye többféle módon is megfogalmazható. A „ $H_0$ -t elvetjük”, „a minta adatai ellentmondanak  $H_0$ -nak”, „a minta adatai (vagy a „valóság”) és a nullhipotézis között szignifikáns eltérés van” megfogalmazások mindegyike nagyjából ugyanazt:  $H_0$  elvetését fejezi ki. Döntésünk megfogalmazási módját bizonyos fokig célszerű a mintanagyságtól is függővé tenni. Kisebb minták esetében  $H_0$  megtartásakor (el nem vetésekor) célszerűbb úgy fogalmazni, hogy a minta adatai nem mondanak ellent a nullhipotézisnek, vagy hogy  $H_0$ -t nem sikerült elvetni a minta alapján. Nagyobb minták lehetőséget adnak az ennél bátrabb –  $H_0$ -t elfogadjuk – megfogalmazásra is. Ennek hátterében az áll, hogy minél nagyobb egy minta, annál pontosabb, megbízhatóbb információk nyerhetők abból, s kisebb a lehetősége a rövidesen tárgyalásra kerülő másodfajú hiba elkövetésének. Ehhez még két megjegyzést érdemes hozzáfűzni:

a) A hipotézisvizsgálat esetében kerülendő a túl nagy minták használata, mert azok alapján  $H_0$  még a tőle való igen kis – gyakorlati-szakmai szempontból jelentéktelen – eltérések alapján is elvethető. Ilyenkor – ha a minta tényleg a bennünket érdeklő sokaságból származó véletlen minta – valójában nincs szükség hipotézisvizsgálatra, mert a nagy minta alapján még igen magas megbízhatósági szint választása esetén is nagy pontosságú információk nyerhetők a sokaságról. Ekkor a hipotézisvizsgálatnál célszerűbb a sokasági jellemzők intervallumbecsléséhez folyamodni.

b) Gyakorlati szempontból rendszerint a  $H_0$ -tól való olyan eltérések az igazán lényegesek, melyek még kis minta alapján is szignifi-

kánsnak mutatkoznak. Ehhez ugyanis a kis mintákra jellemző viszonylag nagy mintavételi ingadozások következtében a  $H_0$ -tól való jelentős eltérésekre van szükség.

A hipotézisvizsgálat lépéseinek áttekintése és kommentálása után néhány további, ugyancsak fontos kérdést veszünk sorra.

## 5. Összetett nullhipotézisek vizsgálata

A hipotézisvizsgálat előbb vázolt technikája könnyen kiterjeszhető olyan esetekre is, amikor  $H_0$  nem egyszerű, hanem összetett hipotézis. Ez azért hasznos és fontos, mert a gyakorlatban minden olyan esetben összetett nullhipotézisek használatára van szükség, amikor  $H_1$  egyoldali, és azt olyan  $H_0$ -lal szembeállítva kívánjuk vizsgálni, amikor  $H_0$  és  $H_1$  együtt az összes lehetőséget kimeríti. Ekkor mind a  $H_0$ , mind a  $H_1$  hipotézis egyszerű hipotézisek kisebb-nagyobb halmazából álló összetett hipotézis, hiszen mindegyikük fennállása a  $T$  próbafüggvény sokféle, gyakran végtelen sok eloszlását engedi meg.

Az összetett nullhipotézisek helyességének vizsgálatára csak akkor van mód, ha az azt alkotó egyszerű hipotézisek halmazának eleme a vele szembe állított egyoldali alternatív hipotézisnek legkevésbé ellentmondó egyszerű hipotézis is. Ezt az egyszerű hipotézist gyakran technikai nullhipotézisnek nevezik, és  $H_0^T$ -vel jelölik.

Ekkor egy összetett nullhipotézis valamely egyoldali alternatív hipotézissel szembeni helyessége igen egyszerűen vizsgálható a  $H_0^T$  technikai nullhipotézis helyességének ellenőrzésére támaszkodva. Ha ugyanis  $H_0^T$  elvethető valamely egyoldali alternatív hipotézissel szemben, akkor vele együtt elvethető az adott egyoldali alternatív hipotézisnek  $H_0^T$ -nél jobban ellentmondó minden egyszerű hipotézis is, azaz maga a teljes összetett nullhipotézis. Ha viszont  $H_0^T$  nem vethető el valamely egyoldali alternatív hipotézissel szemben, akkor csak annyi állítható, hogy a vizsgált alternatív hipotézissel szemben legalább egy egyszerű hipotézis nem utasítható vissza. Emiatt  $H_0^T$  elvetése „kemény”, megtartása pedig „puha” döntésnek minősíthető. A technikai nullhipotézis megtartásakor (el nem vetésekor) sok esetben viszonylag egyszerűen meg lehet találni a vele együtt el nem vethető további egyszerű nullhipotéziseket is, ami valamelyest „keményíti” a  $H_0^T$  megtartását kimondó döntést.

## 6. A hipotézisvizsgálat során elkövethető hibák

A hipotézisvizsgálat menetét és logikáját átgondolva könnyű észrevenni, hogy a leírt módon eljárva a  $H_0$  hipotézis (vagy a  $H_0^T$  technikai nullhipotézis) helyességéről hozott döntésünk nem lesz mindig feltétlenül jó. Előfordulhat ugyanis az, hogy a  $H_0$  hipotézis helyes, de a  $T(y_1, y_2, \dots, y_n)$  próbafüggvény adott mintából számított értéke mégis a kritikus tartományba esik. Ilyenkor a  $H_0$  hipotézist annak ellenére el fogjuk vetni, hogy az a valóságban helyes. Ez nyilvánvalóan hibás döntés, amely az ún. elsőfajú hiba. Ez a fajta hibás döntés a megtartási és visszautasítási tartomány konstrukciójánál fogva  $\alpha$  valószínűséggel fordulhat elő, ami egyben azt is jelenti, hogy az elsőfajú hiba elkövetésének esélye tetszés szerint korlátozható. Ennek azonban gátat szab egy másik fajta hiba, az ún. másodfajú hiba elkövetésének az  $\alpha$  csökkentésével párhuzamosan emelkedő kockázata.

A másodfajú hiba nem más, mint a téves  $H_0$  megtartása. Elkövetésére az adhat alapot, hogy  $T \in M$  akkor is előfordulhat, ha  $H_0$  nem igaz. A másodfajú hiba elkövetésének valószínűségét  $\beta$ -val szokás jelölni. Ennek  $1-\beta$  kiegészítő valószínűsége, tehát annak valószínűsége, hogy nem követjük el a másodfajú hibát (nem tartjuk meg tévesen a nullhipotézist), a próba ereje, ami  $n \rightarrow \infty$  esetén elég általános feltételek mellett 1-hez tart. A próba annál jobb, minél gyorsabban közelíti meg az ereje az 1 értéket. Az is könnyen belátható, hogy a  $H_0$  hipotézisben feltételezett helyzettől távolodva a próbák ereje ugyancsak egyre közelebb kerül 1-hez. Ez azt jelenti, hogy a helytelen nullhipotézist annál könnyebb elvetni, minél távolabb esik az a valóságos helyzettől. Ha viszont a valóságban  $H_0$  tényleg fennáll, akkor nincs is mód a másodfajú hiba elkövetésére.

Könnyű észrevenni, hogy az első- és másodfajú hiba tartalmilag megegyezik a bírósági ítélezésben elkövethető kétféle hibával. Ha ugyanis a vádlott ártatlanságát nullhipotézisnek tekintjük, akkor a bírósági ítélezés elsőfajú hibája az ártatlan vádlott elítélése, míg a másodfajú hiba a bűnös vádlott felmentése. Érdeemes felfigyelni arra, hogy ha a vádlott bűnösségét tekintjük  $H_0$ -nak, akkor az első- és másodfajú hiba is szerepet cserél. Nyilvánvaló, hogy egy valamennyire is méltányos ítélezési gyakorlatban célszerű mindkét fajta hiba elkövetésének valószínűségét minél alacsonyabb szinten tartani. Természetesen erre érdemes törekedni a hipotézisvizsgálat során is.

A  $\beta$  valószínűség csak akkor határozható meg, ha pontosan tudjuk azt, hogy a valóságban a  $H_0$ -ban szereplő feltételezéssel szemben milyen egyszerű hipotézis áll fenn. Mivel rendszerint nem ez a helyzet, s ugyanakkor a  $\beta$  valószínűség a hipotézisvizsgálat minőségének fontos jellemzője, a valóság ismeretének hiányát úgy szo-

kás áthidalni, hogy a  $\beta$  valószínűségeket az egyszerű alternatív hipotézisek egész halmazára vonatkozóan vizsgáljuk. E vizsgálat jól bevált eszközei a jelleggörbe és erőfüggvény.<sup>6</sup> A kétféle hiba elkövetésének valószínűségéről általánosságban annyi azért elmondható, hogy adott  $n$  mellett az elkövetési valószínűségek egymással ellentétes irányba mozognak. Az elsőfajú hiba elkövetésének  $\alpha$  valószínűségét csökkentve ugyanis megnő az  $M$  tartomány, s ennek folytán annak valószínűsége is, hogy  $T \in M$  következzen be, akár igaz  $H_0$ , akár nem.

A kétféle hibáról és azok elkövetési valószínűségéről az 1. táblázat ad szemléletes áttekintést.

1. táblázat

*A hipotézisvizsgálat során elkövethető hibák (és azok elkövetési valószínűsége)*

$H_0$ -t	$H_0$ a valóságban	
	igaz ( $H_1$ nem igaz)	nem igaz ( $H_1$ igaz)
elvetjük	elsőfajú hiba ( $\alpha$ )	helyes döntés ( $1-\beta$ )
megtartjuk (nem vetjük el)	helyes döntés ( $1-\alpha$ )	másodfajú hiba ( $\beta$ )
$\Sigma$	(1)	(1)

Az első- és másodfajú hibával kapcsolatos fejtegetéseket azzal zárjuk, hogy  $H_0$  elvetése erős – „kemény” –,  $H_0$  megtartása (el nem vetése) azonban meglehetősen gyenge – „puha” – döntés. Ez azért van így, mert  $H_0$  elvetésekor minden további nélkül megadható és szükségképpen kontroll alá is vehető a hibás döntés esélye. Viszont  $H_0$  megtartása (el nem vetése) esetén ez nem tehető meg, mert a másodfajú hiba elkövetésének valószínűsége általában se nem ismert, se nem befolyásolható közvetlenül. A másodfajú hiba elkövetése ellen csak közvetetten lehet védekezni, elsősorban  $n$  növelésével.

<sup>6</sup> A próba(függvény) jelleggörbében azt a függvényt értjük, ami minden lehetséges egyszerű hipotézishez hozzárendeli azt a valószínűséget, amellyel a próbafüggvény az  $M$  tartományba esik. Ha  $H = H_0$ , ez a valószínűség  $1-\alpha$ , minden más esetben pedig  $\beta$ , azaz a másodfajú hiba elkövetésének valószínűsége. A megfelelő  $1-\beta$  komplementer valószínűségeket megadó függvényt a próba(függvény) erőfüggvényének nevezzük. Az erőfüggvény értéke a  $H = H_0$  esetben  $\alpha$ , minden más esetben pedig  $1-\beta$ . A gyakorlatban többnyire az erőfüggvényt használják a próbafüggvények viselkedésének minősítésére. Az erőfüggvény egy-egy értékét a próba(függvény) adott egyszerű hipotézishez tartozó erejének szokás nevezni.

Újabban szokásos még egy harmadik, az ún. harmadfajú hiba definiálása is. E hibának van egy teljesen általános<sup>7</sup> és egy, a hipotézisvizsgálathoz közvetlenül kötődő értelmezése is. A szűkebben értett harmadfajú hiba az, amikor egy kétoldali  $H_0$  helyes elvetése után hibás döntést hozunk a  $H_0$ -beli helyzettől való eltérés irányáról. Ennek elkövetési valószínűsége azonban a legtöbbször elhanyagolhatóan kicsi. (Lásd *Vargha* [2007] 156. old.)

## 7. Szignifikanciaszint-választás és hatáselemzés

A szignifikanciaszint megválasztását célszerű a kétféle hiba elkövetéséből adódó következmények, károk valamilyen együttes mérlegelésére támaszkodva megtenni. Ez nem mindig könnyű feladat.

A tudomány általános fejlődése szempontjából a kétféle hiba elkövetése más-más következményt von maga után. Míg az elsőfajú hiba túlzottan gyakori elkövetésével könnyen lejárátódik a tudomány, mert valamilyen téves megállapítást engedünk be annak tárházába, addig a másodfajú hiba „csak” azzal a – kevésbé káros – következménnyel jár, hogy az éppen elvégzett hipotézisvizsgálattal nem fedezünk fel valamilyen új, korábban nem ismert hatást vagy összefüggést, és azt legfeljebb később teszszük majd meg. (Lásd *Vargha* [2007] 161. old.) Emiatt  $H_0$  és  $H_1$  olyan módon történő megfogalmazására ajánlatos törekedni, hogy a kétféle hiba közül az elsőfajú hiba elkövetése legyen a kevésbé kívánatos számunkra, a szignifikanciaszint pedig ezzel összhangban minél kisebb legyen. Ekkor ugyanis kicsi a hibás döntés kockázata. Ez a stratégia azon alapszik, hogy az első- és másodfajú hiba konkrét tartalma attól függ, melyik hipotézist tekintjük  $H_0$ -nak és melyiket  $H_1$ -nek.<sup>8</sup> Ha nem lehet  $H_0$ -t úgy megfogalmazni, hogy a hipotézisvizsgálat végzőjének  $H_0$  elvetése álljon az érdekében, akkor viszonylag magas (10-20 százalékos vagy akár még magasabb) szignifikanciaszintet célszerű választani, és/vagy indokolt viszonylag nagy minta használatára törekedni.

Vannak néha olyan esetek is, amikor a kétféle hiba elkövetéséből adódó következmények, károk számszerűsíthetők valahogyan. Ha a kétféle hiba elkövetésének

<sup>7</sup> A teljesen általános, a statisztikai szaktanácsadáshoz kötődő értelmezés az, hogy statisztikai értelemben véve jó választ adunk egy szakmailag hibás kérdésre. Ez szinte mindig a megrendelő és a statisztikus közötti hibás kommunikáció következménye.

<sup>8</sup> Gondoljunk például arra, hogy mi az első- és másodfajú hiba konkrét tartalma az olyan bírósági ítékezés esetében, amikor a vádlott ártatlansága, illetve bűnössége a nullhipotézis!

van valamilyen költségkihatása, akkor ezt feltétlenül érdemes figyelembe venni, s a szignifikanciaszintet ezzel összhangban célszerű megválasztani. Például, ha az elsőfajú hiba elkövetése nagy anyagi veszteséggel jár, de a másodfajú hibáé nem okoz különösebb bajt, akkor a szignifikanciaszintet célszerű igen kicsire választani. (Ez lehet a helyzet például akkor, ha egy gyárban a minőségellenőrzés eredményére alapozzák annak eldöntését, hogy beavatkozzanak-e az adott folyamatba vagy sem, és a beavatkozás költsége magas. Ha ugyanis a nullhipotézis az, hogy a folyamat a technológiai előírásoknak megfelelően zajlik, akkor az elsőfajú hiba elkövetése a folyamatba való fölösleges beavatkozást jelenti.) Viszont, ha inkább a másodfajú hiba elkövetése ellen indokolt védekezni, akkor nyilván célszerű viszonylag magas szignifikanciaszintet használni, vagy ha lehetőség van rá, a hipotézisvizsgálat céljaira viszonylag nagy mintát venni. Mint korábban már megjegyeztük, mindkét lépés csökkenti a másodfajú hiba elkövetésének valószínűségét.

Világosan látnunk kell azonban, hogy adott mintanagyság és adott valóban fennálló egyszerű alternatíva mellett a kétféle hiba elkövetésének valószínűsége csakis egymás rovására változtatható. Ez többnyire a kétféle hiba következményeinek mérlegelésén alapuló valamilyen kompromisszum keresését igényli, ami sohasem egyszerű, és rendszerint nem is oldható meg teljesen objektív módon. Gondoljunk csak a bírósági ítélezéssel kapcsolatban megfogalmazott hibalehetőségekre, s a kétféle hiba összemérésének nehézségeire!

A gyakorlatban elterjedt 5 százalékos szignifikanciaszint használata egyfajta célszerű kompromisszumnak tekinthető  $\alpha$  és  $\beta$  nagysága tekintetében. Ehhez azonban természetesen nem szükséges és indokolt mereven ragaszkodni. Ha nem áll módunkban a hipotézisvizsgálatból származó hibák következményeinek, valamint a lehetséges alternatíváknak az áttekintése és mérlegelése, akkor  $\alpha$ -t célszerű oly módon megválasztani, hogy minél nehezebbé tegyük a számunkra kedvező eredmény fellépését. Ez úgy érhető el, hogy  $\alpha$ -t minél kisebbre vesszük, ha  $H_0$  elvetésében vagyunk érdekeltek, és minél nagyobbra, ha annak ellenkezőjében.

Napjainkban már az is elég gyakori, hogy a statisztikai próbákat előre megtervezetten hajtják végre. Ez azt jelenti, hogy igyekeznek mindkét fajta hiba elkövetési valószínűségét valamilyen elfogadható szinten tartani, az elsőfajúét például 5 százalékos, a másodfajúét pedig mondjuk 10-20 százalékos szinten. Ehhez az szükséges, hogy előre adjunk meg egy szakmailag már lényegesnek tekinthető minimális eltérést a  $H_0$ -beli helyzettől, válasszunk alkalmas szignifikanciaszintet, és keressük meg azt a legkisebb  $n$ -t, amin már teljesül a másodfajú hiba elkövetési valószínűségére vonatkozó elvárás. Erre ma már többféle számítógépes program is rendelkezésre áll. (Lásd *Vargha* [2007] 163. old.)

Az ún.  $p$ -érték használata – amit gyakran empirikus szignifikanciaszintnek neveznek – a szignifikanciaszint megválasztását valójában a hipotézisvizsgálat ered-

ményének felhasználójára bízta. A  $p$ -érték az a legkisebb szignifikanciaszint, amin  $H_0$  már éppen elvethető  $H_1$ -gyel szemben, ezért értéke úgy határozható meg, hogy a  $T$  próbafüggvénynek a hipotézisvizsgálathoz használt mintából nyert értékét annak előjelétől – egyes esetekben<sup>9</sup> nagyságától – függően alsó vagy felső kritikus értéknek tekintjük, és megállapítjuk a hozzá tartozó szignifikanciaszintet. Kétoldali alternatív hipotézis esetében ez az egyoldali szignifikanciaszint még kettővel szorzandó. Igen gyakori, hogy a  $p$ -értéknek csak a nagyságrendjét jelzik egy egyenlőtlenség megadásával. A  $p$ -érték ismeretében a hipotézisvizsgálat eredményének felhasználója mindig meghozhatja a saját igényeinek megfelelő döntést.

Sajnos az is elég gyakori, hogy a szignifikanciaszintet vagy  $p$ -értéket abszolutilizálják, és minél inkább szignifikáns eredmények elérését hajszojják, illetve méltányolják a kutatómunkában. Fontos világosan látni, hogy a statisztikai szignifikancia és a szakmai szignifikancia nem feltétlenül esik egybe. Különösen nagy mintaelemszámok esetében gyakran előfordul az, hogy az eredmény erősen szignifikáns, de e mögött csak a  $H_0$ -beli helyzettől való jelentéktelen, szakmailag teljesen érdektelen eltérés húzódik meg. Annak érdekében, hogy elkerüljük a statisztikai és a szakmai szignifikancia egymással való téves azonosítását – hacsak lehet – érdemes ún. hatásvizsgálatokat is végezni. Ez azt jelenti, hogy a „lehet-e ekkora” kérdés mellett mindig érdemes a „mekkora lehet” kérdésének is kellő figyelmet szentelni. A hatásvizsgálatokra még akkor is szükség van, ha valamely hipotézisvizsgálattal kapott két vagy több eredmény mindegyike erősen szignifikáns. Ezt kívánja szemléltetni a következő egyszerű példa.<sup>10</sup>

Valaki le akar fogyni, és két fogyasztótabletta között van módja választani. A kétféle tablettáról a következő információkkal rendelkezik.

A tabletták adott időszak – mondjuk egy hónap – alatti szedésével elérhető fogyás mindkét esetben normális eloszlású. A gyártók állítása szerint az  $A$  tablettá szedésével átlagosan 20 fontot lehet fogyni, az egy hónap alatti fogyások szórása 10 font, a  $B$  tablettá szedésével elérhető átlagos fogyás 5 font, a havi fogyások szórása pedig 0,5 font.

Megkér egy statisztikust, hogy segítsen neki eldönteni, melyik a hatásosabb tablettá; a kettő közül melyiket válassza.

A statisztikus a döntés megalapozása céljából egy-egy 16 elemű mintát vesz a kétféle tablettá vásárlói közül, és megkéri őket, hogy egy hónap múlva közöljék vele, mennyit fogytak. A két minta jellemzői a 2. táblázatban láthatók.

<sup>9</sup> Például a  $\chi^2$  -, illetve  $F$ -eloszlást követő próbafüggvények esetében.

<sup>10</sup> A példa kiinduló feltevései *Ziliak–McCloskey* [2008] írásából valók (43. old.), de az azokra alapozott mintaszimulálás, hipotézisvizsgálat és becslés a szerző munkája.

2. táblázat

*A kétféle tablettát szedők adatai*

Jellemző	<i>A</i>	<i>B</i>
	tablettát szedők	
Átlagos havi fogyás (font)	21,07	5,07
A havi fogyás korrigált szórása (font)	9,961	0,565
Mintanagyság	16	16

Az, hogy a két tablettát hatásos-e, a

$$H_0 : \mu \leq 0, \quad H_1 : \mu > 0$$

hipotézispár helyességének vizsgálatát igényli, ahol  $\mu$  az adott tablettával egy hónap alatt elérhető fogyás várható értéke. Az, hogy melyik a hatásosabb tablettát, attól függ, melyik minta adja a jobban szignifikáns eredményt. Erre a két  $p$ -érték meghatározásával és összehasonlításával lehet válaszolni. Elfogadva, hogy az egy hónap alatti fogyás eloszlása mindkét esetben normális, a havi fogyások szórása pedig annyi, amennyit a tablettát gyártói állítanak, mindkét hipotézispár a  $Z$ -próba segítségével vizsgálható. A próbafüggvény értéke az  $A$  tablettát szedők esetében

$$Z = \frac{21,07}{10/\sqrt{16}} \approx 8,427,$$

a  $B$  tablettát szedőket tekintve pedig ehhez hasonlóan 40,475. Minden számolás nélkül is nyilvánvaló, hogy a jobb oldali  $p$ -érték mindkét esetben 0, és az is, hogy a  $B$  tablettát szedők mintája adta az erősebben szignifikáns eredményt. Ezen az alapon tehát a statisztikus a  $B$  tablettát választását fogja javasolni. De biztosan jó ez a tanács?!

Ennek kiderítésére végezzünk most hatásvizsgálatot, és becsüljük a minták alapján mindkét tablettát átlagos fogyasztó hatását. Ennek érdekében határozzuk meg a 99 százalékos konfidenciaintervallumot mindkét esetben. Könnyen belátható, hogy ez az  $A$  tablettát szedőknél a [14,63; 33,95], a  $B$  tablettát szedőknél pedig a [4,75; 5,39] intervallum. Ennek alapján pedig nyilván az  $A$  tablettát érdemes választani, mert az jóval nagyobb átlagos fogyást valószínűsít, mint a  $B$  tablettát.

Ez az igen egyszerű példa is jól illusztrálja: az, hogy egy eredmény nagyobb mértékben szignifikáns, mint egy másik eredmény, önmagában még nem mond semmit a háttérben levő hatásokról.



Ezt a hatásvizsgálatot jelen esetben a kétmintás  $Z$ -próba használatával is el lehetett volna végezni a

$$H_0 : \mu_A \leq \mu_B, \quad H_1 : \mu_A > \mu_B$$

hipotézispár vizsgálata útján. Ennek eredménye

$$Z = \frac{21,07 - 5,07}{\sqrt{\frac{10^2}{16} + \frac{0,5^2}{16}}} \approx 6,390,$$

amelyhez 0-hoz nagyon közeli, jobb oldali  $p$ -érték tartozik. Ez is egyértelműen azt jelzi, hogy az  $A$  tablettát érdemes választani, azaz hogy az  $A$  tablettát hatásosabb, mint a  $B$ .

A  $p$ -érték használatával – mint már említettük – megkerülhető a szignifikanciaszint választásának problémája, illetve a hipotézisvizsgálat eredményének felhasználójára bízható az. Emellett a  $p$ -értékek használata még annak az etikai követelménynek is jól megfelel, mely szerint a statisztikai elemzés eredményeit lehetőleg mások által jól rekonstruálható – és esetleg felülbíráható – módon kívánatos közzétenni. A  $p$ -érték abszolútizálása azonban könnyen önkényes döntésekre vezethet. Ez ellen egy minimálisan elvárt szignifikanciaszint előzetes kikötésével lehet védekezni.

A  $p$ -érték használatával és értelmezésével kapcsolatban sajnos elég sok a félreértés, helytelen alkalmazás. Ezekről jó áttekintés ad a *Goodman* [2008] tanulmány. Az ott felsorolt téves értelmezések közül csak az általa legelterjedtebbnek és legveszélyesebbnek minősített félreértést emeljük ki, mely szerint az 5 százalékos  $p$ -érték azt jelenti, hogy  $H_0$  fennállásának esélye 5 százaléknyi. Ez természetesen nem lehet igaz, hiszen maga a  $p$ -érték meghatározása már eleve feltételezi  $H_0$  fennállását.  $H_0$  fennállásának valószínűségét csakis a bayesi statisztikán belül lehet meghatározni. A másik, általa ugyancsak fontosnak tartott téves értelmezésről a statisztikai és szakmai szignifikancia említése, illetve megkülönböztetése kapcsán korábban már szó esett.

## 8. Az alkalmazási feltételek sérülése

A próbák – szigorúan véve – csak akkor használhatók, ha alkalmazási feltételeik mindegyike pontosan teljesül. Egy próbát valamely alkalmazási feltétel szempontjára

ból akkor szokás robusztusnak nevezni, ha az, hogy az adott feltétel egyáltalán nem vagy nem pontosan teljesül, nem nagyon befolyásolja az első- és másodfajú hiba elkövetési valószínűségét adott  $n$ ,  $\alpha$ ,  $H_0$  és  $H_1$ , valamint ténylegesen fennálló egyszerű alternatíva mellett. A feltételek, amelyekről beszélünk igen változatosak és sokfélék lehetnek. Leggyakoribb feltétel a próbafüggvény eloszlására vonatkozik: a statisztika főként nagy minták esetén sűrűn él a normalitás feltevésével. A feltételek másik csoportja minták, változók függetlenségét mondja ki. Gyakoriak azok a feltételek, amelyek bizonyos paraméterek (például varianciák) egyezését írják elő, de feltétel lehet például a modell linearitása is, és fontos feltétel a mintavétel módja, amivel később kissé részletesebben is foglalkozunk. A feltételek teljesülését valójában a teszt tényleges végrehajtása előtt kell vizsgálni, amire nem mindig adódik lehetőség. Ezért a statisztika – főként manapság, a megnövekedett számítási lehetőségek birtokában – gyakran azt az utat választja, hogy szimulációs vizsgálatokkal elemzi, a feltételek esetleges megsértése milyen hatással van a próba eredményeire. Ezek a szimulációs vizsgálatok sok segítséget adnak az alkalmazásokhoz, ám érvényességük értelemszerűen korlátozott.

Ezen a helyen kell szót ejtenünk a paraméteres és nemparaméteres próbákról, bár részletes tárgyalásuk messze meghaladná a tanulmány kereteit. Nem egész pontos definíció szerint a paraméteres próbák valamely sokasági paraméter tesztelésére irányulnak, és általában feltételezéseket tartalmaznak a sokasági eloszlásra. Amennyiben nem paraméterek tesztelése (hanem például függetlenségvizsgálat) a próba célja, és/vagy az eloszlásbeli feltételezések nem indokoltak, ún. nemparaméteres próbákhoz fordulunk. A paraméteres próbáknak többnyire (de nem minden esetben) megvan a nemparaméteres párjuk. A kétféle próbatípus viszonya elég egyszerű: a paraméteres próbák több feltevést igényelnek, ennek fejében viszont erejük nagyobb, mint a hasonló feladatra alkalmazott nemparaméteres próbáké. Ezek után az már adott helyzetben a felhasználó döntése, hogy melyik próbatípushoz fordul.

A gyakorlatban talán a próbáknak a mintavétel módját illető alkalmazási feltétele sérül a legtöbbször, amelyek között ott szerepel az a megkötés is, hogy a hipotézisvizsgálat céljaira egy vagy több ún. FAE-minta<sup>11</sup> áll rendelkezésre. Sajnos a használt minták FAE-mintavételtől való eltéréseinek hatása szinte alig van feltérképezve. Ezért erről a problémáról csak annyit mondhatunk, hogy csekély kiválasztási arány esetén az egyszerű véletlen minták – sőt még azok valamilyen ismérv(ek) szerint utólag képzett részei is – jó közelítéssel FAE-mintáknak tekinthetők. Ugyanez érvényes a rétegzett mintán belüli rétegmintákra is, ha azokra nézve elég kicsik a kiválasztási arányok. Ilyenkor az egyes rétegekből vett minták még egymástól függetlenek is, ami jól kihasználható a két vagy több sokaságra vonat-

<sup>11</sup> Független, azonos eloszlású elemekből álló minta.

kozó nullhipotézisek helyességének vizsgálatakor. A kis kiválasztási arány ugyanis jelentősen korlátozza a mintaelemek egymástól való függőségét. Igazi problémát csak a csoportos minták használata okozhat, melyek elemeire a viszonylag erős egymástól való függőség jellemző. Ezért ilyen mintákat lehetőleg ne használjunk hipotézisvizsgálathoz.

A mintavételen alapuló hipotézisvizsgálat eredményeinek eddigi értelmezése kizárólag a mintavételi ingadozást és hibát veszi tekintetbe. A gyakorlatban azonban a mintavételi hiba mellett szinte mindig számolni kell kisebb-nagyobb nemmintavételi hibákkal is. A hipotézisvizsgálat eredményeinek értelmezése során feltétlenül szükséges figyelembe venni a nemmintavételi hibák nagyságával kapcsolatos információkat, tapasztalatokat is. A mintavételi hibához képest nagy nemmintavételi hibák esetében például úgy, hogy lemondunk a hipotézisvizsgálat használatáról.

## Irodalom

- BERGER, J. O. [2003]: Could Fisher, Jeffreys and Neyman Have Agreed on Testing? *Statistical Science*. Vol. 18. No. 1. pp 1–32.
- BLALOCK, H. M. [1972]: *Social Statistics*. McGraw-Hill Book Company. New York.
- CANAVOS, G. C. [1984]: *Applied Probability and Statistical Methods*. Little Brown Co. Boston.
- HAJTMAN B. [1968]: *Bevezetés a matematikai statisztikába pszichológusok számára*. Akadémiai Kiadó. Budapest.
- GOODMAN, S. [2008]: A Dirty Dozen: Twelve  $p$ -Value Misconceptions. *Seminars in Hematology*. Vol. 45. No. 3. pp. 135–140.
- HUNYADI L. [2001]: *Statisztikai következtetésemélet közgazdászoknak*. Központi Statisztikai Hivatal. Budapest.
- HUNYADI L. [2011]: Bayesi gondolkodás a statisztikában. *Statisztikai Szemle*. 89. évf. 10–11. sz. 1150–1172. old.
- HUNYADI L. – VITA L. [2004]: *Statisztika közgazdászoknak*. Központi Statisztikai Hivatal. Budapest.
- HUNYADI L. – VITA L. [2008]: *Statisztika II*. AULA Kiadó. Budapest.
- KÖVES P. – PÁRNICZKY G. [1981]: *Általános statisztika*. Közgazdasági és Jogi Könyvkiadó. Budapest.
- SERLIN, R. C. [1987]: Hypothesis Testing, Theory Building, and the Philosophy of Science. *Journal of Counseling Psychology*. Vol. 34. No. 4. pp. 365–371.
- VARGHA A. [2007]: *Matematikai statisztika*. Pólya Kiadó. Budapest.
- ZILIAK, S. T. – MCCLOSKEY, D. N. [2008]: *The Cult of Statistical Significance – How the Standard Error Costs Us Jobs, Justice, and Lives*. The University of Michigan Press. Ann Arbor.

## Summary

The paper considers the steps of testing hypotheses, discussing the role, logic, controversial points and obstacles of the succeeding steps. Then the possible errors of testing hypotheses and the problems and possibilities of selecting the significance level are touched upon. Related to this, reference is made to the most pervasive misconceptions of  $p$ -value, and also to the importance of making clear distinction between statistical and substantial significance. In this connection the importance of examining effect size in addition to statistical significance is emphasized. This is also illustrated by a simple example.