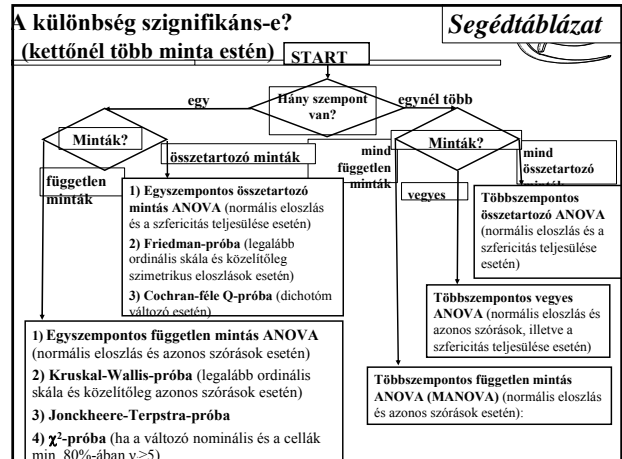
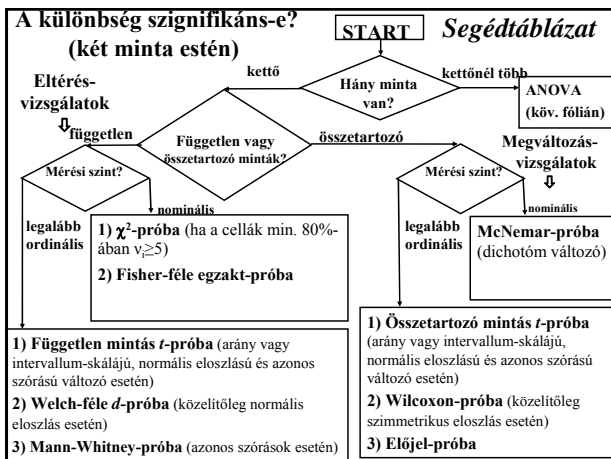
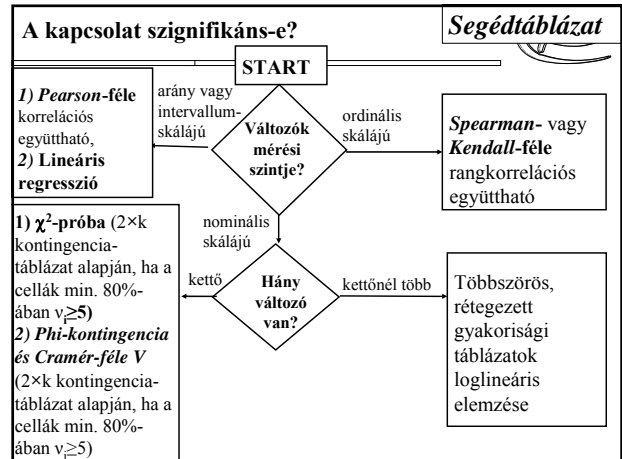


A STATISZTIKAI PRÓBÁK KIVÁLASZTÁSA



A STATISZTIKAI PRÓBÁK KIVÁLASZTÁSA

Segéd táblázatok a próbák kiválasztásához
A statisztikai eljárások alkalmazási feltételei




A statisztikai eljárások alkalmazási feltételei

A statisztikai hipotéziseknek két alapvető típusa van:

Paraméteres próbát végzünk, ha az elemzett minta eloszlását ismerjük, csupán az eloszlás paramétereit kell ellenőrizni. Olyan hipotézisvizsgálat, amelynél a nullhipotézis esetén érvényes próbastatisztika eloszlása függ a minta eloszlásától és ezért arra nézve feltételezést igényel (amely feltételezést a próbában már nem vizsgáljuk).


Nemparaméteres próbát végzünk, ha a minta eloszlását nem ismerjük. Olyan hipotézisvizsgálat, amelynél a nullhipotézis esetén érvényes próbastatisztika eloszlása a nullhipotézis esetén független a minta eloszlásától.

A statisztikai eljárások alkalmazási feltételei 

Régi, de újabban is gyakran visszatérő kérdés annak eldöntése, hogy paraméteres vagy nemparaméteres eljárásokat használjunk konkrét adataink feldolgozásához.

Ez különösen azoknak okoz gondot, akik elsősorban ordinális skálájú adatokkal dolgoznak (pszichológusok, pedagógusok, szociológusok, közvéleménykutatók, piackutatók stb.).


Ennek a problémának a megítélése sajátos hullámzást mutatott az idők folyamán.

A statisztikai eljárások alkalmazási feltételei 

Ezek alapján a kutatók azt az óvatosságot vonták le, hogy a szigorú normalitást általában nem kell megkövetelni az intervallum- vagy arányskálájú változóktól, sőt, gyakran még ordinális skálájú adatok is feldolgozhatók így.


Kizáró oknak csak a markánsan nem szimmetrikus eloszlásokat tekintették, amelyek egyébként megfelelő transzformációval gyakran normálissá tehetők.

A varianciára vonatkozó tapasztalatok már akkor is némileg rosszabbak voltak: ha sem az adatok varianciája az egyes változóknál, sem pedig az adatminták mérete nem tekinthető még közelítőleg sem azonosnak, akkor paraméteres eljárások még ezen szerzők szerint sem alkalmazhatók.

A statisztikai eljárások alkalmazási feltételei 

A kezdetben uralkodó szigorúbb felfogás szerint paraméteres eljárásokat (elsősorban a *t*-próbákat és varianciaanalíziseket) csak abban az esetben jogos használni, ha a kidolgozásuk során tett következő feltételek teljesülnek:


- 1) a mintába minden elem egymástól függetlenül került,
- 2) az adatok mérési szintje magasabb az ordinálisnál (legalább intervallum-skálájú),
- 3) az adatok normális eloszlásúak,
- 4) az adatok varianciája az egyes változóknál, illetve mintákban közelítőleg azonos,
- 5) az adatok száma az egyes mintákban közelítőleg azonos.

A statisztikai eljárások alkalmazási feltételei 

Később azonban, ahogy gyűltek a tapasztalatok, ismét kezdtek túlsúlyba kerülni a feltételek szigorúbb kezelését megkövetelő eredmények.

Wilcox (1987) például úgy találta, hogy eltérő varianciájú mintákra még egyenlő mintaméretű esetekben alkalmazhatók bizonyos paraméteres módszerek.


A nemparaméteres próbák ereje ugyanakkor – ha az adott problémára kidolgozva ilyenek egyáltalán léteznek – gyakran alig marad el a paraméteresekétől, és mivel így azokkal többnyire egyező eredményeket adnak, ilyen problémás esetekben ezeket biztonságosabban használni.

A statisztikai eljárások alkalmazási feltételei 

Számos összehasonlító vizsgálat azzal az eredménnyel járt, hogy bizonyos esetekben egyes paraméteres eljárások nem nagyon érzékenyek egyes követelmények (elsősorban a normalitás és a szórásazonosság) megsértésére, szakkifejezéssel élve „*robustusnak*”.

Egy statisztikai próbát akkor nevezünk „*robustusnak*” valamely alkalmazási feltételre vonatkozólag, ha a próba elsőfajú hibája akkor sem tér el lényegesen a választott névleges szinttől – pl. a szokásos 0,05-től – ha az adott feltétel nem teljesül.

Scheffé (1959) például kimutatta, hogy az egyik leggyakrabban használt paraméteres próba, a kétmintás *t*-próba egyenlő elemszámok esetén meglehetősen *robustus* a szórások mérsékelt különbségére.

A statisztikai eljárások alkalmazási feltételei 

Amikor ilyen próbák nem állnak rendelkezésre – pl. az SPSS nem tartalmazza az ANOVA-hoz kapcsolódó „*Contrast*” vagy a „*Post Hoc*” analízis nemparaméteres megfelelőit – akkor megfelelő óvatossággal kénytelenek vagyunk a paraméteres próbákhoz nyúlni.

A statisztikai eljárások



alkalmazási feltételei

Az uralkodó vélemények változása a nemparaméteres próbákkal kapcsolatban is megfigyelhető volt. A korábbi optimistább nézetek szerint például a *Mann-Whitney* próba a kétmintás *t*-próba olyan nemparaméteres alternatívája, amely nem követeli meg sem a normalitást, sem pedig a szórások egyenlőségét. Később viszont többen úgy találták, hogy ez a próba csak akkor végezhető el jogosan, ha az adott változó eloszlása a két populációban azonos alakú (azonos szórású, ferdeségű, csúcsosságú, stb.).

A statisztikai eljárások



alkalmazási feltételei

Vargha András (1996, 2000, 2003b) egymintás *t*-próbával végzett szimulációs vizsgálatának eredményei

(2) A próba kétoldalú ellenhipotézis melletti érvényességét illetően az állapítható meg, hogy a próba

- $n = 80$ esetén a 0–2 ferdeségi tartományban elfogadható;
- $n = 40$ esetén a 0–1 ferdeségi tartományban elfogadható, de 1,5 fölött már mérsékelten emelkedett, jóllehet a torzítás mértéke még nem drasztikus;
- $n = 20$ esetén az 1,5–2 ferdeségi tartományban már jelentős, 40–60%-os torzulás is felléphet;
- $n < 20$ esetén már mérsékelt ferdeség esetén is komoly torzulást szenvedhet, különösen alacsony csúcsosság mellett.

A statisztikai eljárások



alkalmazási feltételei

Vargha András (1996, 2000, 2003b) egymintás *t*-próbával végzett szimulációs vizsgálatának eredményei

A szimulációs elemzésekre a ferdeségi és a csúcsossági értékek széles tartományát átfogó lambda és kevert normális eloszláscsaláddal került sor 5 és 80 közötti elemszámokkal.

A beállított ferdeségi értékek tartománya 0–3, a csúcsossági értékek pedig 1,75–15,8 volt.

A statisztikai eljárások



alkalmazási feltételei

Vargha András (1996, 2000, 2003b) egymintás *t*-próbával végzett szimulációs vizsgálatának eredményei

(3) Egyoldalú ellenhipotézis alkalmazása esetén - különösen erősen ferde eloszlásoknál, és ha az elemszám igen kicsi ($n \leq 20$) - az egymintás *t*-próba igen érzékeny a normalitás megsértésére, az elsőfajú hiba torzulása jelentős.

(4) A próba robusztussága a normalitás megsértésével szemben $\alpha = 0,01$ szinten az 5%-os szinthez ($\alpha = 0,05$) viszonyítva számottevően kisebb, $\alpha = 0,10$ szinten pedig érezhetően nagyobb.

A statisztikai eljárások



alkalmazási feltételei

Vargha András (1996, 2000, 2003b) egymintás *t*-próbával végzett szimulációs vizsgálatának eredményei

(1) A próba elsőfajú hibája az eloszlás ferdeségének erősödésével egyre nő, viszont a csúcsosság növekedésével párhuzamosan szisztematikusan csökken:

