

STATOKOS JEGYZET: CSOPORTOK ÖSSZEHASONLÍTÁSA

StatOkos – Statisztikai és Módszertani Adatbázis

2019



Tárgymutató

- Populáció és minta
- Csoportok logikája
- Hány csoportot hasonlítunk össze?
- Milyen próbákat alkalmazzunk?
- Példa

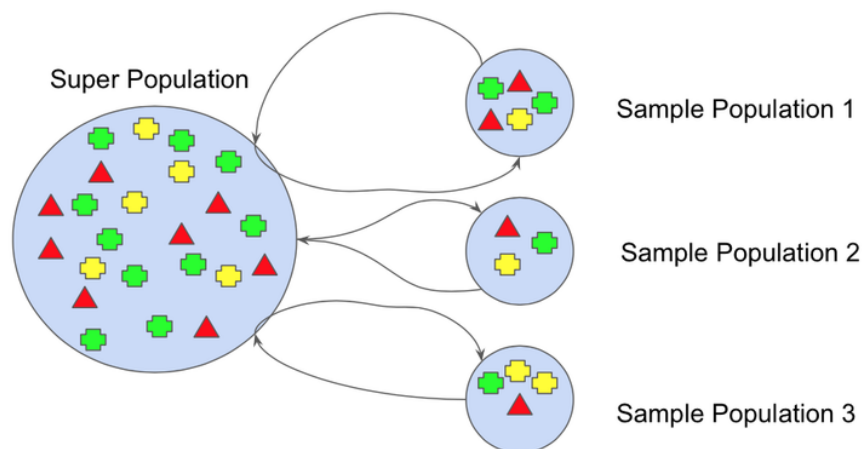
Populáció és minta

A gyakorlatban találkozhatunk az alábbi kifejezésekkel: populáció és minta. De miben különbözik a kettő?

- **A populáció rendszerint az az elméleti teljes egység, amit vizsgálni szeretnénk.**
- **A minta pedig a populációból vett "szelet", ami a lehető legjobban reprezentálja a teljes populációt.**

A minta értékeiből tudunk következtetni a teljes populációra. Ezért van az, hogy a legtöbb kísérlet nem vizsgál meg minden lehetséges esetet, hanem csak a teljes lehetséges esetek (populáció) egy bizonyos, számszerűen jól meghatározott hányadát (minta). A mintavételezés során általában két lehetséges vizsgálati módot választunk:

- **Kísérletezés**
- **Meglévő adatok vizsgálata (feltáró jelleggel)**



A nagy populációból vett (Super Population) vett minták (SP 1-2-3) közül melyik lehet a leginkább reprezentatív? Segítség: próbáljuk megállapítani a különböző színek darabszám-arányai alapján! (Kép forrása: wikipedia)

Csoportok logikája

A tudományos munkák általános eszköze a kísérletezés.

A kísérletezés során általában két különböző metodikai eljárást alkalmazunk:

- **Ugyanazt a jelenséget több csoporton (legalább kettő) vizsgáljuk.** Ez általában egy kísérleti csoport és egy kontrollcsoport összehasonlítása. Miért? Azért, mert a kísérleti csoporton (i) elvégezzük a szükséges beavatkozást vagy (ii) a kísérleti csoportnak van egy olyan tulajdonsága, ami eltér az átlagostól pl.: betegségben szenvednek. Az átlag itt a kontrollcsoportot jelenti.
- **Eltérő jelenségeket egy csoporton vizsgálunk.** Ez általában egy kísérleti feltétel és egy kontrollfeltétel összehasonlítása. Miért? Azért, mert a kísérleti feltétel (i) a terveink szerint másképp befolyásolja a csoportot, mint a kontrollfeltétel (ii), ami rendszerint azt jelenti, hogy nem alkalmazunk semmit. Vagyis azt várjuk, hogy a kísérleti feltétel egyéb hatást is képes kiváltani, mintha nem történt volna semmilyen beavatkozás.

Továbbá, az is lehetséges, hogy többváltozós statisztikai eljárásokkal, feltáró jelleggel próbálunk csoportba sorolni egy mintát. Erre alkalmas a klaszterelemzés. A kialakult klasztereket a későbbiekben csoportként kezelve tudjuk összehasonlítani.

Hány csoportot hasonlítunk össze?

Az, hogy hány csoportot hasonlítunk össze, értelemszerűen függ attól, hogy a kontrollcsoporthoz képest milyen feltételekkel szeretnénk dolgozni. Azért vesszük alapvetőnek a kontrollcsoport meglétét, mert különböző kísérleti csoportok összehasonlításánál így nehezen tudjuk megmondani, miben térnek el az átlag populációtól és mi jellemző a kísérleti csoportra, csoportokra.

A legegyszerűbb megoldás, ha kezdetnek egy kontroll csoport és egy kísérleti csoport tulajdonságait hasonlítjuk össze. Ennél magasabb szintén már gondolkodhatunk több csoportban, de ha van lehetőségünk, akkor kerüljük a kontroll nélküli összehasonlításokat (bekeretezett rész). Ellenkező esetben, legyenek megfelelő indokaink ennek elhagyására!

Milyen próbákat alkalmazunk?

A csoportok összehasonlításánál szerencsénk van a statisztikai próbák kiválasztását illetően, ugyanis a leggyakrabban használt eljárások valamilyen csoportokat hasonlítanak össze. Mielőtt megállapítanánk, hogy milyen eljárást használjunk, gondoljuk végig, hogy az általunk vizsgált változó (függő változó) milyen típusú:

- nominális (diszkrét)
- ordinális (diszkrét)
- metrikus (folytonos)

Hány darab függő változónk van:

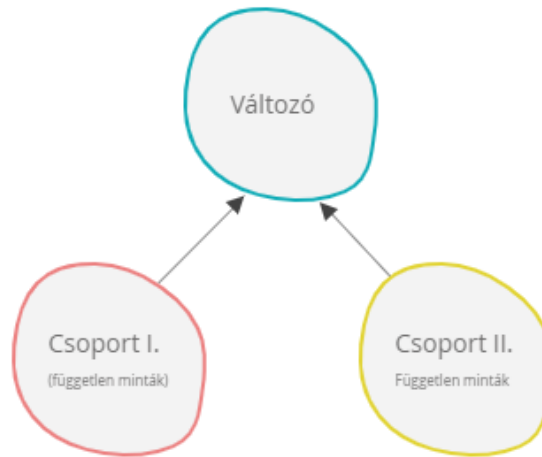
- Egy darab
- Két darab
- Három vagy több

Hogy hány csoportunk van:

- két csoport
- három vagy több csoport

a mérések száma is lényeges, ugyanis jelen esetben egy mérést vizsgálunk, több csoport és több mérés esetén Kevert ANOVA modelleket alkalmazunk. Továbbá, az egy csoporton történő egyszeri mérés is más kategóriába esik ebben az esetben.

Elrendezések I.

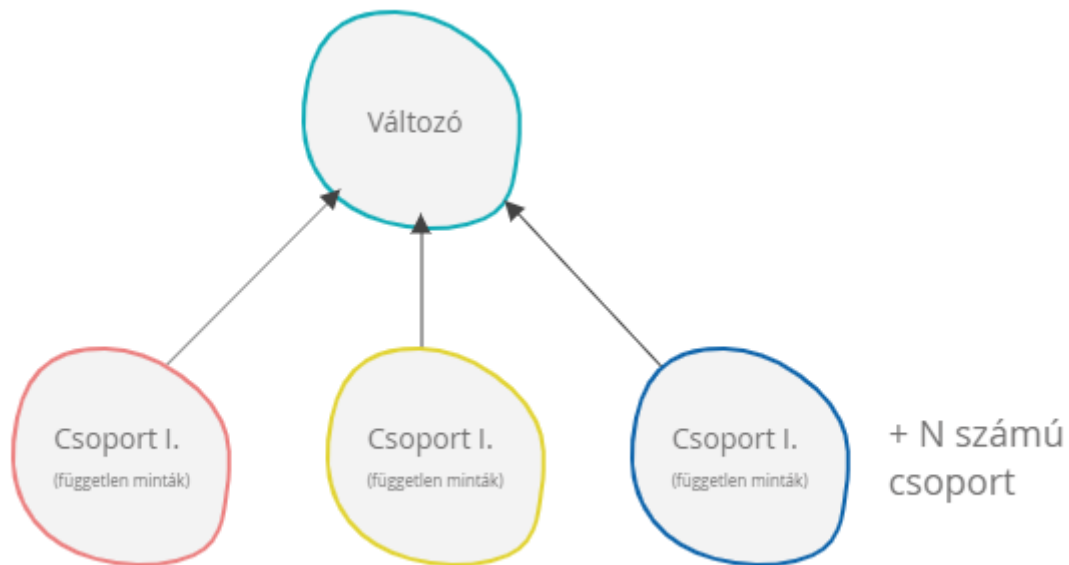


Ebben az elrendezésben két csoportot hasonlítunk össze egy változó alapján. Ez azt jelenti, hogy a két csoportból kivett minta ugyanahhoz a függő változóhoz tartozik, pl.: reakcióidő, életkor stb.

Ha ez az egy változónk:

- **nominális** (kettő, vagy legfeljebb néhány értéket vehet fel, pl.: nemek, hajszínek, stb.): khí-négyzet próba > ennek magyarázata, hogy a változó csoportok közötti megoszlását teszteli.
- **ordinális (vagy metrikus, de nem tapasztalunk normál eloszlást és a varianciák sem egyeznek meg a két csoport között)** (néhány, sorbarendeázhető értéket vesz fel, pl.: osztályzatok, korcsoportok stb.): Mann-Whitney U Próba > mivel nem tapasztalható normál eloszlás, ezért a rangsorolt adatokat hasonlítja össze.
- **metrikus** (nagyon sok különböző értéket vesz fel, aminek lehet/van sorrendje, pl.: magasság, tömeg, stb.): független-mintás t-próba, abban az esetben, ha normalitás van, de a varianciák eltérnek: Welch-próba

Elrendezések II.



Ebben az elrendezésben három vagy több csoportot hasonlítunk össze egy változó alapján. Ez azt jelenti, hogy a több csoportból kivett minta ugyanahhoz a függő változóhoz tartozik, pl.: reakcióidő, életkor stb.

Ha ez az egy változónk:

- **nominális** (kettő, vagy legfeljebb néhány értéket vehet fel, pl.: nemek, hajszínek, stb.): chí-négyzet próba > ennek magyarázata, hogy a változó csoportok közötti megoszlását teszteli.
- **ordinális (vagy metrikus, de nem tapasztalunk normál eloszlást és a varianciák sem egyeznek meg a csoportok között)** (néhány, sorbarendezhető értéket vesz fel, pl.: osztályzatok, korcsoportok stb.): Kruskal-Wallis Próba > mivel nem tapasztalható normál eloszlás, ezért a rangsorolt adatokat hasonlítja össze.
- **metrikus** (nagyon sok különböző értéket vesz fel, aminek lehet/van sorrendje, pl.: magasság, tömeg, stb.): Egyszempontos Varianciaanalízis (Egyszempontos ANOVA)

Példa

Egy kutatás arra keresi a választ, hogy a dohányzók, nemdohányzók és E-cigarettát használó személyek testtömegindexe eltérő-e a 18-35 év közötti egyetemisták körében. A vizsgált csoportok testtömegindexe normál eloszlást mutat, a varianciák megegyeznek a csoportok között. Kérdés: a három csoport közül találunk-e legalább egy olyat, amely különbözik a többitől?

Független változó: csoportokosító változó (dohányzók, nemdohányzók, E-cigarettát használók)

Függő változó: testtömegindex

Csoportok száma: 3 (dohányzók, nemdohányzók, E-cigarettát használók)

Paraméteres adatsor: Igen, normáloszlás van és metrikus változót használunk (testtömegindex)

Variációk megegyeznek: Igen

Használt statisztikai próba: három csoport, egy metrikus független változó, normál eloszlást tapasztalunk, variációk között nincs eltérés csoportszinten > **Egyszempontos ANOVA**