

# STATOKOS JEGYZET: TÖBBVÁLTOZÓS KÍSÉRLETI ELRENDEZÉSEK

StatOkos – Statisztikai és Módszertani Adatbázis

2019



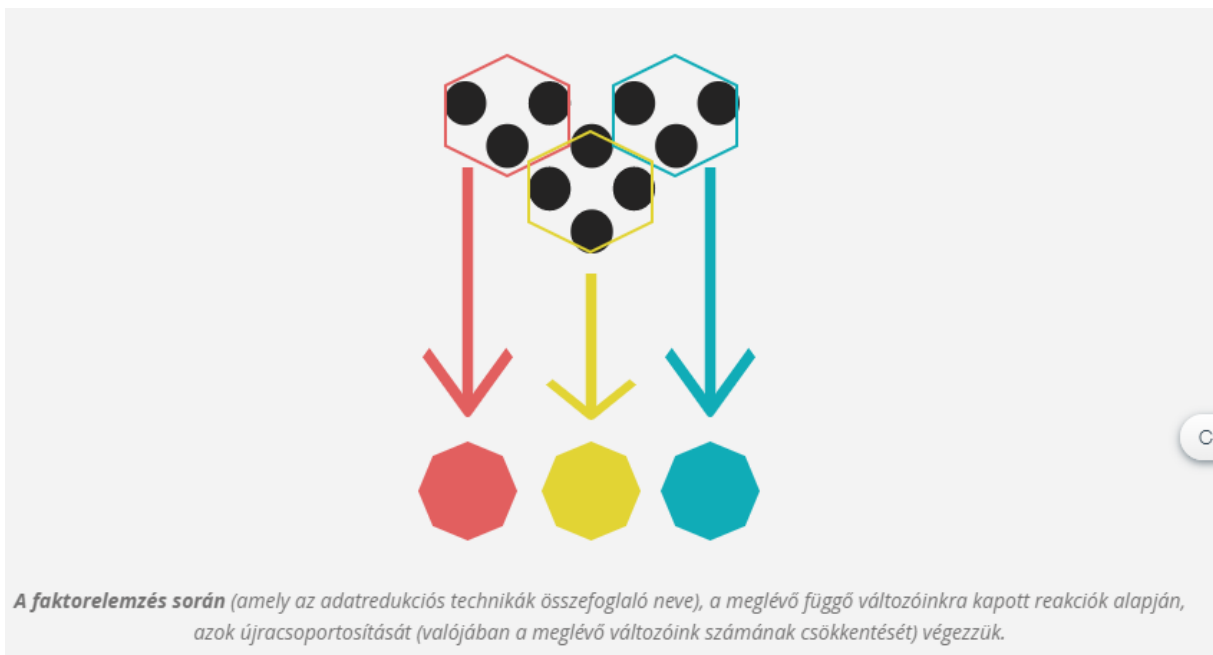
## Tárgymutató

- Milyen többváltozós modelleket alkalmazunk alapszinten?
- Milyen próbákat alkalmazzunk?
- Példa

## Többváltozós modellek

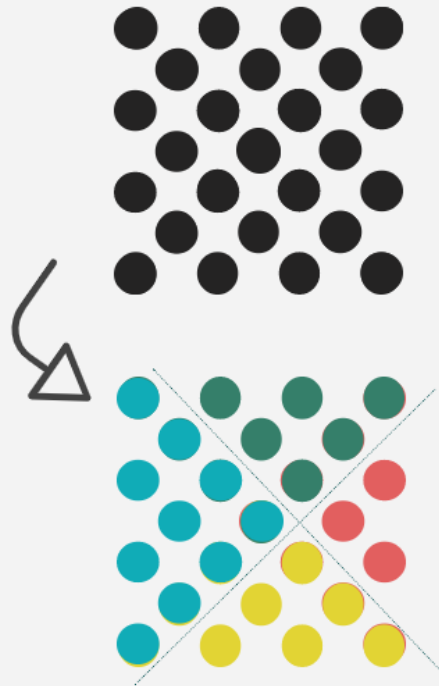
A legjellemzőbb, a többváltozós statisztikai eljárások esetében az, hogy:

- **Több függő és független változóval operálnak** (4-5 darabtól egészen 10-15 változóig is, vagy magasabb számban)
- A függő változók egy közös jelenséghez köthetők
- **A függő változók redukciója** azt jelenti, hogy például 10-15 változóból kevesebbet, akár 3 változót generálunk, amely valamilyen logika (vagy a program dönti el vagy mi) szerint értelmezhető
- **Az újracsoportosítás** (klaszterbe sorolás rendszerint) azt jelenti, hogy a vizsgálati személyek függő változó(k)ra adott reakciói alapján, hasonlóságot keresve egy csoportba sorolhatók

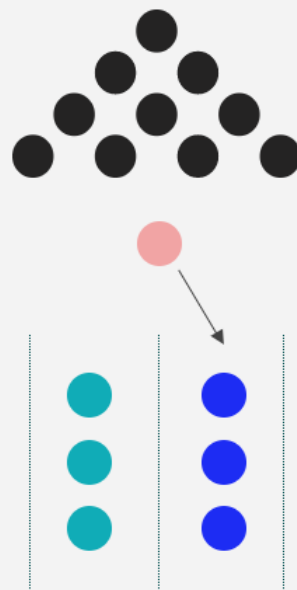


- **Ismeretesek azok az eljárások is** (logisztikus regresszió és diszkriminanciaanalízis, ROC-görbe), melyek segítségével a meglévő adatok alapján képesek vagyunk a vizsgálati személyeket

**csoportokba sorolni, elkülöníteni egymástól.** Ez a klaszterezéssel szemben nem csak keresztmetszeti (éppen aktuális állapot), hanem a jövőbeli esetekre vonatkozóan is **előrejelző funkcióval bír**, vagyis akár diagnosztikus célt is szolgálhat.



**A klaszterelemzés során** nem a változókat, hanem magukat a vizsgálati személyeket soroljuk új csoportokba (klaszterekbe) egy közös tulajdonság alapján. A feltáró kutatások egyik fontos eljárása a nagy elemszámú minta kisebb egységekre való bontása.



A predikciós funkcióval bíró többváltozós elemzések (logisztikus regresszió, diszkriminancia-analízis, ROC-görbe) azt hivatottak meghatározni, hogy a rendelkezésre álló adatok alapján a meglévő és a jövőbeni vizsgálati személyek (esetek) megbízhatóan besorolhatók-e egy-egy kategóriába; pl.: a meglévő vérnyomásértékek és nem alapján meghatározható-e, hogy valaki később stroke betegségben fog-e szenvedni.

## Milyen próbákat alkalmazunk?

Az előzőek alapján röviden tekintsük át, hogy milyen statisztikai próbát alkalmazunk a különböző esetekre.

- Abban az esetben, ha nagyon sok változónk van és szeretnénk a változók számát csökkenteni és valamilyen logika szerint rendszerezni, akkor válasszuk a faktorelemzést (faktorelemzést vagy főkomponenselemzést). A változóink szerencsésebb esetben metrikusak (paraméteresek).
- Abban az esetben, ha a vizsgálati személyeinket szeretnénk valamilyen közös tulajdonság alapján újracsoportosítani, használjunk klaszterelemzést.
- Abban az esetben, ha a vizsgálati személyeinket prediktív (előrejelző) céllal szeretnénk elkülöníteni, akkor használjuk a logisztikus regressziót vagy a diszkriminanciaanalízist.
- Abban az esetben, ha a meglévő tesztek vagy mérőeljárások diagnosztikus értékéről szeretnénk megbizonyosodni, használjunk ROC-görbét.

### Példa I.

Egy kutatás során egy falu lakosságának élettel való elégedettségét vizsgálták. A vizsgálatban különböző kérdőíveket kellett kitöltenie a lakóknak. A kutatók 13 darab eltérő kérdőívet vittek a helyszínre. A kitöltést követően, az adatok feldolgozásánál azt állapították meg, hogy túl sok változójuk van és összefüggéseket is szeretnének keresni a tesztek között.

**Független változó: A kitöltött teszteken kívül minden releváns információ (magasság, nem, tömeg, életkor, iskolázottság stb.)**

**Függő változó: A kitöltött tesztek**

**Csoportok száma: 1 (A falu lakossága)**

**Használt statisztikai próba: Faktorelemzés valamilyen típusa**

**Ha ismertek a tesztek és a várható eredmények is: Faktorelemzés**

**Ha feltáró jelleggel kutatnánk: Főkomponenselemzés**

### Példa II.

Egy kutatás során egy falu lakosságának élettel való elégedettségét vizsgálták. A vizsgálatban különböző kérdőíveket kellett kitöltenie a lakóknak. A kutatók 4 darab eltérő kérdőívet vittek a helyszínre, illetve felvették az alapvető demográfiai adatokat. A kitöltést követően, az adatok feldolgozásánál arra voltak kíváncsiak a kutatók, hogy a kapott eredmények alapján valamilyen módon csoportosítható-e a falu lakossága és ha igen, milyen különbségek vannak az emberek csoportjai közt.

**Független változó: A kitöltött teszteken kívül minden releváns információ (magasság, nem, tömeg, életkor, iskolázottság stb.)**

**Függő változó: A kitöltött tesztek + a független változók közül azok, amely alapján a csoportosítást végezni szeretnénk.**

**Csoportok száma: 1 (A falu lakossága)**

**Használt statisztikai próba: Klaszterelemzés**

### **Példa III.**

Járványos megbetegedés jelentkezett az egyik faluban. A Járványügyi Bizottság a helyszínen megvizsgálta a teljes lakosságot, hogy a betegség terjedésére vonatkozóan további információkat gyűjtsenek. Többek között (pl.: magasság, tömeg, vérnyomás, vércukorszint, nem) vérvizsgálattal meghatározták a különböző szervek működésére jellemző értékeket is. Megfigyelték, hogy a májfunkciókkal összefüggő értékek a tünetmentes időszakban valamilyen szintű növekedésnek indultak, azonban ez még nem volt elég a tudományos igényű bizonyossághoz. Azt is észrevették, hogy a nemek között eltérő a betegség megjelenése. A járvány későbbi megelőzése érdekében meg akarták állapítani, hogy a nem és a májfunkció értéke milyen rizikófaktorként jelenik meg, ugyanis így a veszélyeztetett személyeket még időben azonosítani tudják.

**Független változó: A véreredményekeken kívül minden releváns információ (magasság, nem, tömeg, életkor, iskolázottság stb.)**

**Függő változó: A véreredmények + a független változók közül azok, amely alapján a betegség bekövetkezését vizsgálni akarjuk.**

**Csoportok száma: 1 (A falu lakossága)**

**Használt statisztikai próba:**

**Ha van olyan függő változó, ami nem paraméteres (jelen esetben a nem változó), akkor: logisztikus regresszió**

**Ellenkező esetben: Diszkriminanciaelemzés**