

Irodalom

<http://www.szit.bme.hu/~kela/ind2>

- Bolla-Krámlí: Statisztikai következtetések elmélete, Typotex, 2005
- Vargha A.: Matematikai statisztika, Pólya, 2000
- Bryman, A., Cramer, D. :
Quantitative Data Analysis with SPSS Release 10 for Windows
(Guide for Social Scientists), Routledge, 20001.
- Kinneer, P.R., Gray, C.D. :
SPSS for Windows Made Simple Release 10,
Psychology Press, 2000
- Neter, J. , Wasserman, W. , Kutner M.H. :
Applied Linear Statistical Models, IRWIN, 1985.
- Mardia, K.V., Kent, J.T., Bibby, J.M. : *Multivariate Analysis*,
Ac ademic Press, 1979.

Ketskeméty László - Izsó Lajos

Bevezetés az **SPSS** rogramrendszerbe

Módszertani útmutató és feladatgyűjtemény statisztikai
elemzésekhez



TÖBBVÁLTOZÓS STATISZTIKA

- Bevezetés, alapfogalmak
- Hipotéziselmélet
- Korreláció- és regresszióelemzés
- Adatszegmentálás
- Adatredukció
- Skálák konstrukciója, MDS
- CONJOINT-elemzés, korrespondencia-analízis
- Egyebek

Statisztikai populáció

**A vizsgálat tárgyát képező nagyszámú egyedek halmaza.
A halmaz egészének kevés adattal történő tömör jellemzése, és a populáció egyedeinek leírására bevezetett változók közötti kapcsolatok leírása a célunk.**

- Magyarország állampolgárai
- Harmadéves muszaki menedzser hallgatók
- Az autók halmaza
- Egy adott termék vásárlóinak halmaza
- Egy TV csatorna nézőinek halmaza

Statisztikai minta

A populáció egy kis elemszámú részhalmazára vonatkozó megfigyelések adatai. A minta úgy kell, hogy tükrözze a populáció tulajdonságait, ahogy a cseppben látjuk a tengert. Azaz a minta reprezentatív kell, hogy legyen.

- Egy felmérésbe bevont magyar állampolgárok halmaza
- Egy adott előadásra belátogatott hallgatók halmaza
- Adott biztosítóval szerződött autók halmaza
- Egy adott napon megkérdezett vásárlók halmaza
- Egy felmérésbe bevont TV nézők halmaza

Mintavételezési eljárások

A populáció minden egyes elemének ugyanakkora esélyt kell biztosítani a mintába kerüléshez.

A minta elemszámának elég nagynak kell lennie ahhoz, hogy a következtetéseink átvihetők lehessenek a populációra is.

Rétegzett mintavételezés: A populációt adott szempontok szerint csoportokba osztjuk, és a csoportok arányait a mintában is megtartjuk

Véletlen mintavételezés: A mintába kerülő egyedet sorsolással választjuk ki.

Változó

A populáció egy mérhető jellemzője.

- Magyarország állampolgárai: **fizetés; kor; nem; párt stb.**
- Harmadéves muszaki menedzser hallgatók: **B4 jegy; neptun-kód; nem stb.**
- Az autók halmaza: **gyorsulás; fogyasztás; lóero; típus;...**
- Egy adott termék vásárlóinak halmaza: **ár; minőség;...**
- Egy TV csatorna nézőinek halmaza: **kor; nem; tetszési index; iskola; stb.**

Eset

A minta egy eleme

Mintaelemszám

Az adott minta elemeinek száma. Egy adatmátrix sorainak száma.

Adatmátrix

n db eset és p db változó adatainak mátrixba rendezett alakzata

Többváltozós analízis

A populációt egynél több jellemzovel (változóval) írunk le, és az egyes változók közötti összefüggések feltárása az elsődleges cél. (Ld. valószínűségi változók együttes eloszlása.)

Egyváltozós analízis

Kiválasztunk egyetlen jellemzot (változót), és annak komplex statisztikai leírása a feladat, a többi változótól függetlenül. (Ld. valószínűségi változó eloszlása.)

Idosor

Idoben egymást követő adatok sorozata.

Idolépték

Az idosorban egymást követő adatok idopontjainak különbsége

- Tozsdeindexek (pl. BUX)
- Lázgörbe
- Budapest szeptemberi közép-homérsekeletei 1800 óta

Összetartozó minták

Idoben szinkron mérési adatok összessége. (Több változó egy adatmátrixban)

kezdofizetés- jelenlegi fizetés

kezelés előtti súly - kezelés utáni súly

városok homérsékleti adatsorai 1800-2005

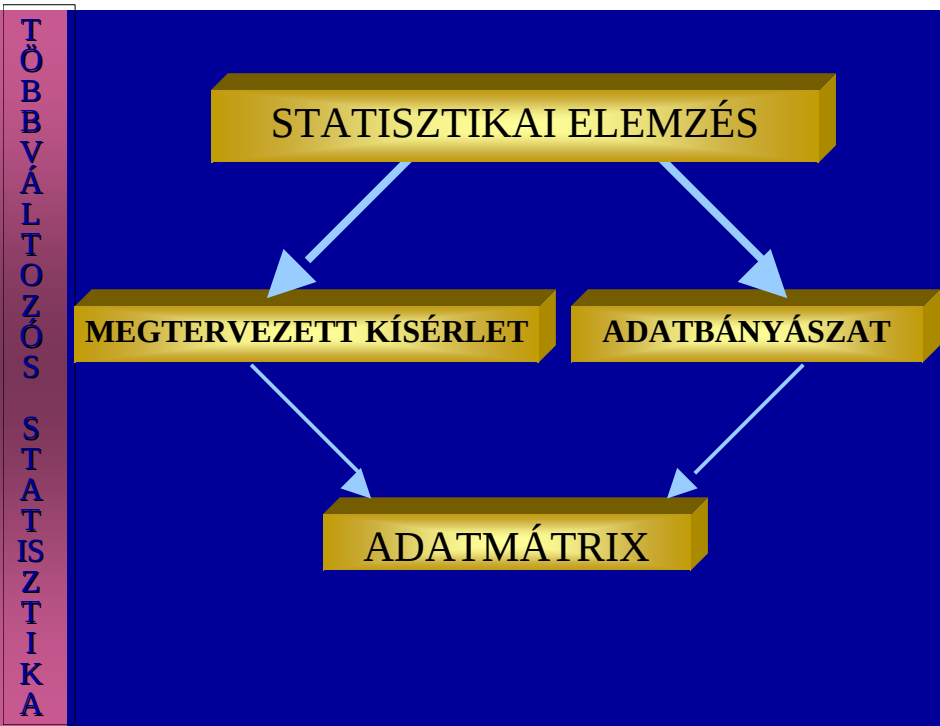
Független minták

Idoben és térben nem összekapcsolható, ugyanarra a változóhoz vett mérésekből származó adatok összessége

noi kezdofizetés- férfi kezdofizetés

kezelés előtti súly fiatalnál - kezelés előtti súly öregnél

Gyor homérsékleti adatsora 1800-1900, 1901-2005





Csoportok kialakítása



1. séma

G1: kísérlet → megfigyelés

G kontroll: nincs kísérlet → megfigyelés

2. séma

G1: megfigyelés → kísérlet 1 → megfigyelés

G2: megfigyelés → kísérlet 2 → megfigyelés

G kontroll: megfigyelés → nincs kísérlet → megfigyelés

3. séma



Nem	TV nézés hetente órában	Agresszív akció játék közben
fiú	9,50	9
lány	9,25	8
lány	8,75	7
fiú	8,25	7
fiú	8,00	6
lány	5,50	4
fiú	5,25	4
lány	4,75	5
fiú	4,50	3
lány	4,00	3

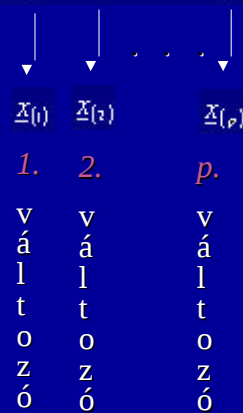
HIPOTÉZIS: A TV-nézés növeli a gyerek agresszivitását

KÍSÉRLET: Gyerekekkel agresszív akciófilmeket nézetnek

MEGFIGYELÉS: Játék közben hányszor mutatnak agressziót

AZ ADATMÁTRIX

$$\underline{\underline{X}} = \begin{pmatrix} x_1^1 & x_1^2 & \dots & x_1^p \\ x_2^1 & x_2^2 & \dots & x_2^p \\ \vdots & \vdots & \ddots & \vdots \\ x_n^1 & x_n^2 & \dots & x_n^p \end{pmatrix}$$



- $\underline{x}^{(1)}$ 1. eset
 → $\underline{x}^{(2)}$ 2. eset
 ⋮
 → $\underline{x}^{(n)}$ n. eset

$$x_j^i$$

eredet szerint:

- forrásadat
- képzett adat
- hiányzó adat
- pótolts adat

megjelenés szerint

- numerikus
 - vesszos
 - pontos
 - tudományos
 - dátum
 - dollár
- szöveg

MÉRÉSI SZINTEK	TÍPUS	LEÍRÁS	PÉLDA
	Nominális	Az objektumok diszkrét kategóriákba vannak osztva. A kategóriákat nem lehet rendezni.	nemzetiség vallás születési hely
	Dichotóm	Kétértékű változó. Lehet nominális, ordinális	nem (férfi, no) válasz (igen, nem)
	Ordinális	Az objektumok diszkrét kategóriákba vannak osztva. A kategóriákat sorba lehet rendezni. A távolság nem jellemző.	képzettség osztályzat magabiztosság
	Intervallum (a)	Az objektumok rendezhető kategóriákba csoportosíthatók. A szomszédos kategóriák távolsága egyenlő.	jövedelem életkor csapadék
	Intervallum (b)	Szigorúan véve ordinális, de a kategóriák száma nagy.	iskolai végzettség
	Arányítható	Értékei sorba rendezhetők, különbségük és arányuk is értelmezhető.	testsúly időtartam

A felsorolt skálatípusok a mérések négy növekvő szintjét képviselik a következő sorrendben:

nominális $\Rightarrow$ *ordinális* $\Rightarrow$ *intervallum* $\Rightarrow$ *arány*

- Az egyes skálatípusok ebben a sorrendben növekvő mennyiségű információt hordoznak
- Az egyes skálatípusok más-más statisztikai módszerekkel dolgozhatók fel
- A magasabb mérési szintű skálatípusok – információ-vesztés árán - átalakíthatók alacsonyabb szintűekké
- A változók egy más csoportosítás szerint lehetnek *diszkrét* vagy *folytonos*

É R V É N Y E S S É G

- **ARCULAT**

Azt mér, amit kell

- **IRÁNYULTSÁG**

Íránya egyezik más változókkal

- **IDOTÁLLÓSÁG**

Nem pillanatnyi állapotra vonatkozik

- **SZERKEZETI**

Kapcsolatos a hipotézissel

- **KONVERGENCIA**

Van koherencia más változókkal

- **KÜLÖNBÖZŐSÉG**

Mást mér, mint a többi változó

- Megismételhetőségi szempont
- Homogenitási szempont
- Belso megbízhatóság

egyértelműség : jól lehessen kódolni

legyen variancia: szubjektumok visszatükröződése

- Külso megbízhatóság

- A felhasználó által definiált

A feldolgozás során hiányzó adatként kezelendő
esetek definiálása változónként.

Define Missing Values

Missing Values for String Variables

- Megtagadott válasz
- Nincs vélemény/Nem tudja
- Forrásadat nincs/elveszett
- A válasz hihetetlen/komolytalan/félreértett

- A rendszer által definiált

A feldolgozás során hiányzó adatként kezelendő esetek, melyeket a program definiál.

Változónként egységes. Jele: , illetve *blank*.

- A szurt esetek helye az új változóknak
- Új változók azon esete, ahol a független változóban hiányzó eset volt
- Értelmezhetetlen művelet „eredménye”
pl. 0-val osztás, negatív logaritmus,
negatív négyzetgyök, érvénytelen függvény argumentum

**A
H
I
Á
N
Y
Z
Ó
É
R
T
É
K
E
K
K
E
Z
E
L
É
S
E**

	X	Y	Z
1.	☐	☐	☐
2.	☒	☐	☐
3.	☐	☐	☐
4.	☐	☐	☒
5.	☐	☒	☐
6.	☐	☐	☐
7.	☐	☐	☐
8.	☐	☐	☐
9.	☐	☐	☐
10.	☐	☐	☒

- Kizárás lista szerint

Csak az 1,3,6,7,8,9 esettel számolunk, n=6.

- Kizárás páronként

X-Y: Csak a 2. és 5. esetet zárjuk ki, n=8.

X-Z: Csak a 2.,4. és 10. esetet zárjuk ki, n=7.

Y-Z: Csak a 4.,5. és 10. esetet zárjuk ki, n=7.

- Átlagértékkel pótolunk

Mindegyik esettel számolunk, n=10.

	X	Y	Z
1.	☐	☐	☐
2.	☒	☐	☐
3.	☐	☐	☐
4.	☐	☐	☒
5.	☐	☒	☐
6.	☐	☐	☐
7.	☐	☐	☐
8.	☐	☐	☐
9.	☐	☐	☐
10.	☐	☐	☒

- Változónként egyedi elbánás
pl. Chi-négyzet próbánál

X: Csak a 2.-t zárjuk ki, n=9.

Y: Csak az 5.-et zárjuk ki, n=9.

Z: Csak a 4., 10.-et zárjuk ki, n=8.

- A hiányzó adat új kategória
pl. Explore, Chart

**X,Y,Z kategóriáinak száma eggyel no,
az esetek száma marad n=10.**