


Dr Ketskeméty László matematikus

<http://www.ketskemety.hu>
laszlo@ketskemety.hu
kela@szit.bme.hu

☎: +36/70/31 00 51 0



Megjelent az Eötvös Kiadó gondozásában
 Forgalmazza a Cytotech Kft
 Megjelenés éve: 2005
 Jellemzők: 459 o., 21×29.4 cm (A4)
 Nyelv: magyar
 ISBN: 963 463 823 6
 Ára: 7000 Ft ÁFA-val
 Megrendelhető:
<http://www.szit.bme.hu/~kela/spsskonvv.html>

**S
T
A
T
I
S
Z
T
I
K
A

T
A
N
F
O
L
Y
A
M**

Irodalom

<http://www.szit.bme.hu/~kela/ind2>
 Bolla-Krámlí: Statisztikai következtetések elmélete, Typotex Kiadó, Budapest, 2005.

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*,
 Academic Press, 1979.

S T A T I S Z T I K A T A N F O L Y A M	TÖBBVÁLTOZÓS STATISZTIKA
	• Bevezetés, alapfogalmak
	• Hipotéziselmélet
	• Korreláció- és regresszióelemzés
	• Adatszegmentálás
	• Adatredukció
	• Skálák konstrukciója, MDS
	• Idősorelemzés
	• Közlekedtetés

A L A P F O G A L M A K P É L D É K	Statistikai 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
	- Egyetemi, főiskolai hallgatók
	- Az autók halmaza
	- Egy adott termék vásárlóinak halmaza
	- Egy TV csatorna nézőinek halmaza

A L A P F O G A L M A K P É L D É K	Statistikai 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 felmérésbe bevont hallgatók halmaza
	- Adott biztosítóval szerződött autók halmaza
	- Egy adott napon megkérdezett vásárlók halmaza
	- Egy felmérésbe bevont TV nézők halmaza

A
L
A
P
F
O
G
A
L
M
A
K

P
É
L
D
Á
K

Mintavételezési eljárások

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

A minta elemszámának elég nagynek 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 egyedeket sorsolással választjuk ki.

A
L
A
P
F
O
G
A
L
M
A
K

P
É
L
D
Á
K

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

A
L
A
P
F
O
G
A
L
M
A
K

P
É
L
D
Á
K

Változó

A populáció egy mérhető jellemzője.

- Magyarország állampolgárai: fizetés; kor; nem; párt stb.

- Egyetemi, főiskolai hallgatók: kor; tanulmányi átlag; kezdés éve; neme stb.

- Az autók halmaza: gyorsulás; fogyasztás; lóerő; típus;...

- Egy adott termék vásárlóinak halmaza: ár; minőség;...

- Egy TV csatorna nézőinek halmaza: kor; nem; tetszési index; iskola; stb.

A
L
A
P
F
O
G
A
L
M
A
K

P
É
L
D
Á
K

Többváltozós analízis

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

Egyváltozós analízis

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

A
L
A
P
F
O
G
A
L
M
A
K

P
É
L
D
Á
K

Idősor

Időben egymást követő adatok sorozata.

Időlépték

Az idősorban egymást követő adatok időpontjainak különbsége

- Tőzszeindexek (pl. BUX)
- Lázgörbe
- Budapest szeptemberi közép-hőmérsékletei 1800 óta

A
L
A
P
F
O
G
A
L
M
A
K

P
É
L
D
Á
K

Összetartozó minták

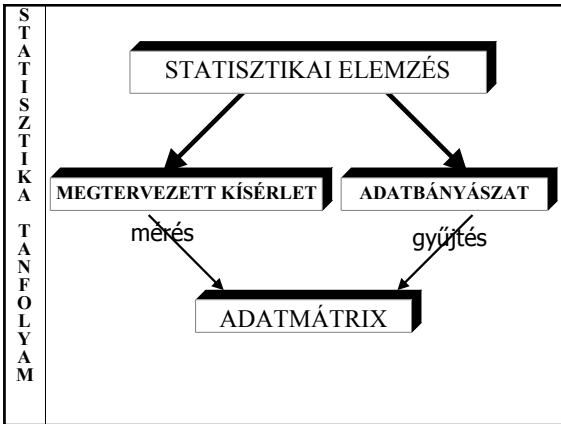
Időben szinkron mérési adatok összessége. (Több változó egy adatmátrixban)

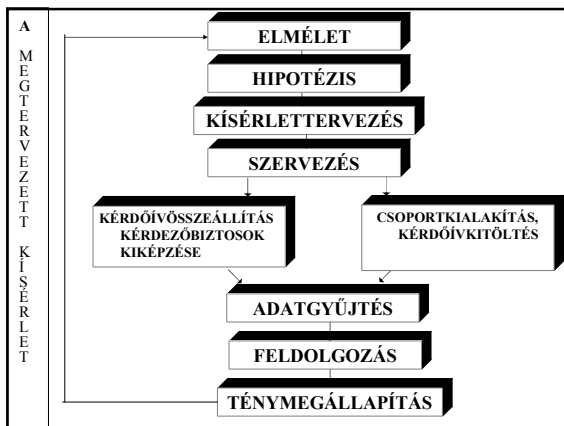
kezdőfizetés- jelenlegi fizetés
kezelés előtti súly - kezelés utáni súly
városok hőmérsékleti adatsorai 1800-2005

Független minták

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

női kezdőfizetés- férfi kezdőfizetés
kezelés előtti súly fiatalnál - kezelés előtti súly öregnél
Győr hőmérsékleti adatsora 1800-1900, 1901-2005





A
Z
A
D
A
T
B
Á
N
Y
Á
S
Z
A
T

- Ember számára emészthető, hasznos információk, rejtett összefüggések kinyerése nagy adathalmazokból.
- „Megfulladunk az információtól, miközben tudásra éhezünk.”

John Naisbitt

Alkalmazási területek:

marketing, orvosi biológia, genetika,
távközlés, csillagászat,
web-bányászat, szöveg-bányászat
...

A Z A D A T M Á T R I X

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

$\rightarrow \underline{x}^{(1)}$
 $\rightarrow \underline{x}^{(2)}$
 \vdots
 $\rightarrow \underline{x}^{(n)}$

1. eset
2. eset
n. eset

$\underline{x}^{(1)}$
1.
v
á
l
t
o
z
ó

$\underline{x}^{(2)}$
2.
v
á
l
t
o
z
ó

$\underline{x}^{(p)}$
p.
v
á
l
t
o
z
ó

eredet szerint: megjelenés szerint

- forrásadat
- numerikus
- képzett adat
- vesszős
- hiányzó adat
- pontos
- pótoltt adat
- 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, nő) 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


MÉRÉSI SZINTEK

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


nominális ⇔ **ordinális** ⇔ **intervallum** ⇔ **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*ek vagy *folytonos*ak


MÉRÉSI SZINTEK



Skálás: intervallum vagy aránytartó



Ordinális



Nominális

A HIÁNYZÓ ADAT

- 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/elvesztett
- A válasz hihetetlen/komolytalan/félreértett

A HIÁNYZÓ ADAT

- A rendszer által definiált

A feldolgozás során hiányzó adatként kezelendő esetek, melyeket a program definiál.
Változónként egységes. Jele: , illetve blank.

- A szűrt esetek helye az új változókban
- 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.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
5.	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
6.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

• 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.

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.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
5.	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
6.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

• 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 nő,
az esetek száma marad n=10.

A
L
A
P
S
T
A
T
I
S
Z
T
I
K
Á
K

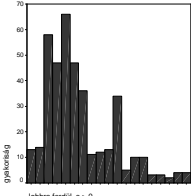
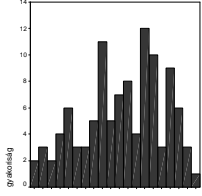
AZ ADATCENTRUMOT JELLEMZI

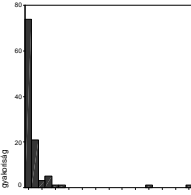
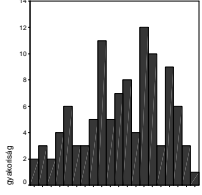
ÁTLAG (mean) $\frac{1}{n} \sum_{i=1}^n x_i$

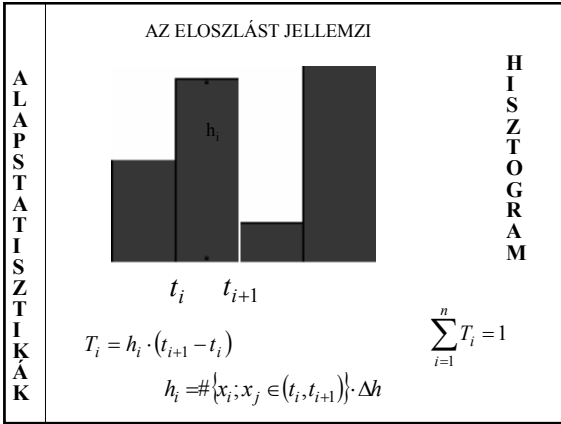
MEDIÁN (median) $\begin{cases} x_{n+1}^* & , \text{ ha } n \text{ páratlan} \\ \frac{x_n^* + x_{n+1}^*}{2} & , \text{ ha } n \text{ páros} \end{cases}$

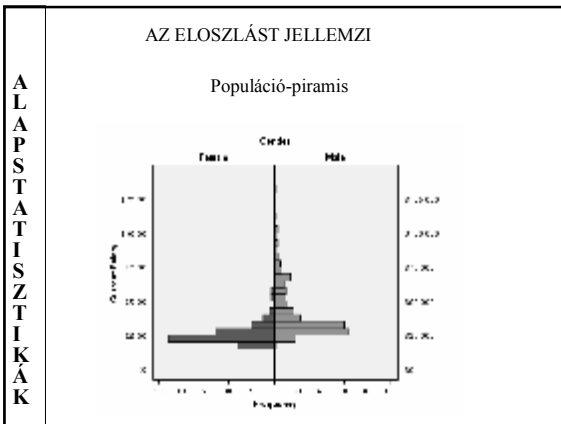
MÓDUSZ (mode) A leggyakrabban előforduló érték a mintában

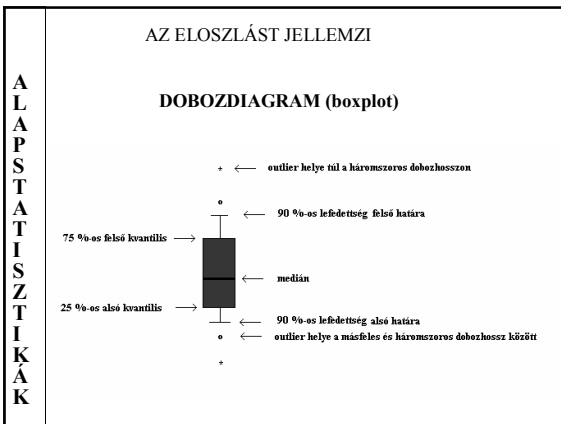
A L A P S T A T I S Z T I K Á K	A SZÓRÓDÁST JELEMLZI	
	STANDARD SZÓRÁS (deviation)	$\sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$
	VARIÁCIÓ (variance)	$\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$
	TERJEDELEM (range)	$x_n^* - x_1^*$
	STANDARD HIBA (s.e. mean)	$\frac{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}}{\sqrt{n}}$

A L A P S T A T I S Z T I K Á K	AZ ELOSZLÁST JELEMLZI	
	FERDESÉG (skewness)	$s = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left(\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \right)^3}$
		

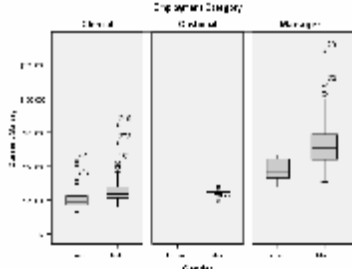
A L A P S T A T I S Z T I K Á K	AZ ELOSZLÁST JELEMLZI	
	LAPULTSÁG (curtosis)	$k = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left(\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \right)^4} - 3$
		







DOBOZDIAGRAM (boxplot)



KAPCSOLATOT JELLEMEZ

KONTINGENCIA TÁBLÁZAT

Két diszkrét változó közötti összefüggés jellemzésére használt gyakoriság-táblázat

1. VÁLT. \ 2. VÁLT.	x_1	x_n	SORÖSSZESEN
y_1	v_{11}		v_{1n}	$s_{1.}$
.	.		.	.
.	.		.	.
y_m	v_{m1}		v_{mn}	$s_{m.}$
OSZLOP ÖSSZESEN	$o_{.1}$		$o_{.n}$	N

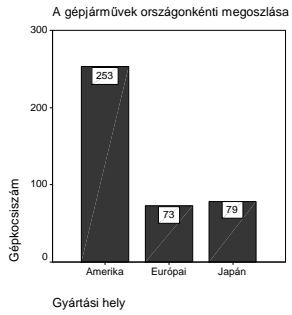
KONTINGENCIA TÁBLÁZAT

Number of Cylinders * Country of Origin Crosstabulation

Number of Cylinders		Country of Origin			Total
		American	European	Japanese	
3 Cylinders	Count	0	0	4	4
	% of Total	.0%	.0%	1.0%	1.0%
4 Cylinders	Count	72	66	69	207
	% of Total	17.8%	16.3%	17.0%	51.1%
5 Cylinders	Count	0	3	0	3
	% of Total	.0%	.7%	.0%	.7%
6 Cylinders	Count	74	4	6	84
	% of Total	18.3%	1.0%	1.5%	20.7%
8 Cylinders	Count	107	0	0	107
	% of Total	26.4%	.0%	.0%	26.4%
Total	Count	253	73	79	405
	% of Total	62.5%	18.0%	19.5%	100.0%

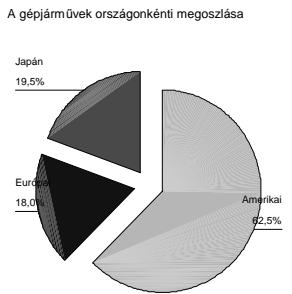
GRAFIKONOK

Oszlopdiaagram



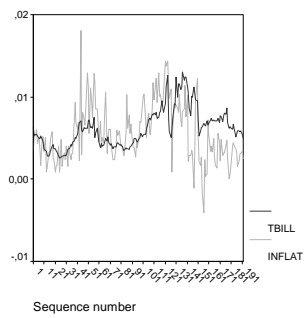
GRAFIKONOK

Torta-diaagram



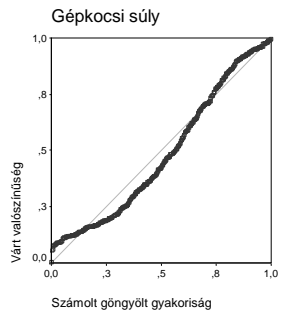
GRAFIKONOK

Vonal-diaagram



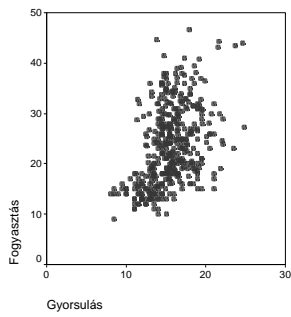
GRAFIKONOK

P-P plot (illeszkedésábra)



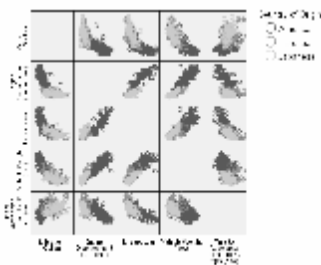
GRAFIKONOK

Szóródás-ábra (scatter)



GRAFIKONOK

Szóródás-ábra (matrix scatter)



Szóródás-ábra (3D scatter)

