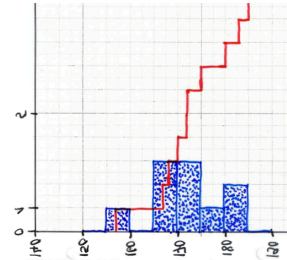


Biostatisztika és informatika alapjai



1. előadás: Bevezetés
2018. szeptember 13.
Agócs Gergely

Honnan készülünk fel?

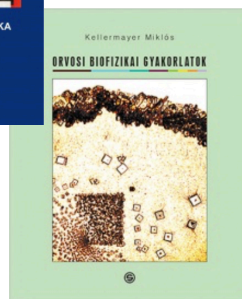
- egyetem = önálló tanulás
- források:
 - előadáson készített jegyzetek (csütörtök 15¹⁰–15⁵⁵; EOK Szent-Györgyi Albert előadó)
 - számítógépes laborgyakorlatokon készített jegyzetek (heti 1 alkalom, 90 perc; EOK első emelet „B” folyosó)
 - Herényi: „Statistika és Informatika” (Simmelweis Kiadó, 2016)
 - honlap: biofiz.simmelweis.hu
 - tantárgyi követelmények
 - előadásmatematika és diák
 - gyakorlati tematika
 - gyakorlófeladatok (házi feladatok)
 - korábbi évek anyagai



3

Honnan készülünk fel?

- egyetem = önálló tanulás
- források:
 - előadáson készített jegyzetek (csütörtök 15¹⁰–15⁵⁵; EOK Szent-Györgyi Albert előadó)
 - számítógépes laborgyakorlatokon készített jegyzetek (heti 1 alkalom, 90 perc; EOK első emelet „B” folyosó)
 - Herényi: „Statistika és Informatika” (Simmelweis Kiadó, 2016)



2

Tudomány és nemtudomány

Az ártatlanság vétele: „Minden gyanúsított személyt mindaddig ártatlannak kell vélelmezni, amíg bűnösségét a törvénynek megfelelően meg nem állapították.” AZ EURÓPAI UNIÓ ALAPJOGI CHARTÁJA, 48. cikk (1)

„A hatástalanság vétele”: Minden kezelést és szert mindaddig hatástalannak kell vélelmezni, amíg hatásosságát a tudományos követelményeknek megfelelően meg nem állapították.

Páciensek kezelése



Bizonyítékon alapuló orvoslás (tudományos orvoslás)



- 1) A döntéseket az egészségügy minden szintjén és ágában objektív bizonyítékokra alapozza. Az egészségügyi ellátás alapjául szolgáló bizonyítékoknak hozzáférhetőnek kell lenniük.
- 2) Az egészségügyben dolgozók képesek kell legyenek a tudományos közlemények színvonalának helyes megítélésére, kritikus olvasására és megértésére
- 3) Az egészségügy fejlesztéséhez szükséges a folyamatos kutatás

Alternatív vagy komplementer medicina (nem tudományos orvoslás, „kuruzslás”)

Nem bizonyítékokon, hanem hagyományon és hiten alapul, pl: hagyományos kínai orvoslás, akupunktúra, természetgyógyászat, homeopátia, íriszdiagnosztika, csontkovácsolás, köpölyözés, biorezonancia stb.



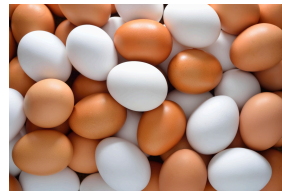
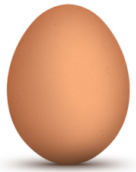
4

Miben segít nekünk a statisztika?

A **statisztika** az adatok gyűjtésével, rendszerezésével, elemzésével és következtetések levonásával foglalkozik

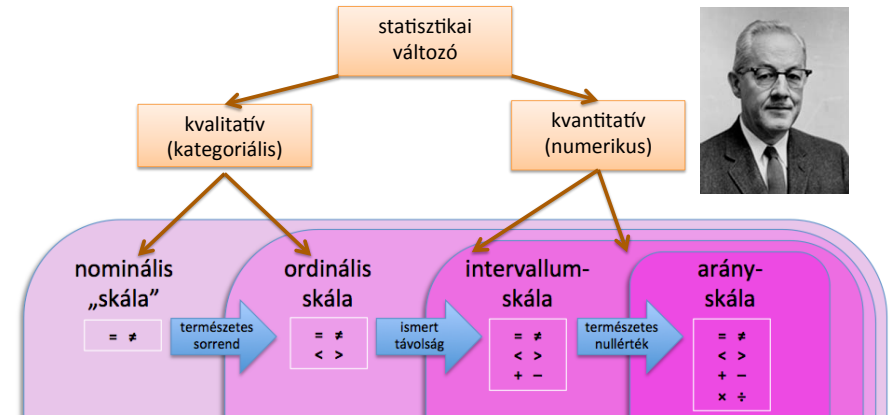
Leíró statisztika → Következtető statisztika

Törvényszerűségek → Determinisztikus törvényszerűségek → tökéletes predikció
 Törvényszerűségek → Stochasztikus törvényszerűségek → bizonytalanság



5

A statisztikai változók típusai I. Mérési szintek (S. S. Stevens nyomán)



7

Milyen adatokkal foglalkozunk?

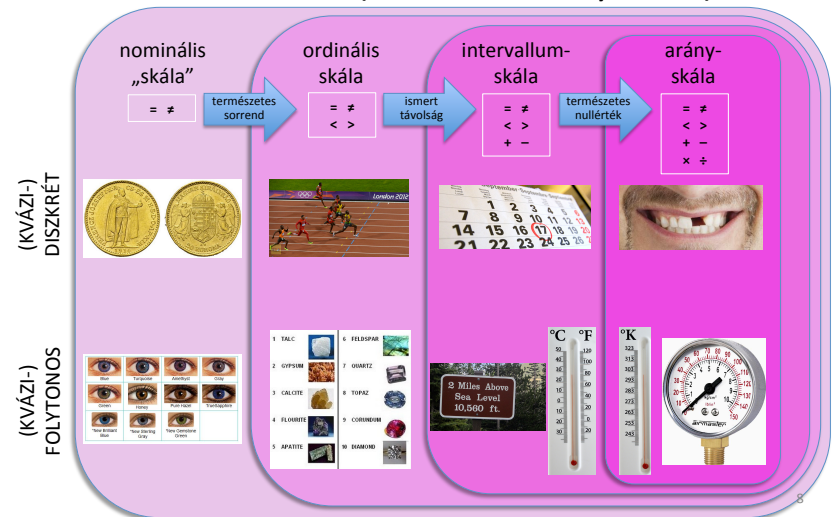
A feldolgozandó adatok rendkívül változatosak ...

LabCorp		LabCorp San Diego 13112 Freighting Court, 10th Fl., San Diego, CA 92121-4100 Phone: 619-468-3700
333-245-1455-0	22247223	333-245-1455-0
DONOR		Request: A Test, LTD
NONO		VART Verified
DOB: 28/07/13	17/16/83	DOB: 28/07/13
4897 THOMPSON DR	4897 THOMPSON DR	4897 THOMPSON DR
SAN ANTONIO, TX 78201	SAN ANTONIO, TX 78201	SAN ANTONIO, TX 78201
PHOTO ID REQUIRED		PHOTO ID REQUIRED
By and For: 13/28/10 10:52	By and For: 13/28/10 10:52	By and For: 13/28/10 10:52
13/28/10 10:52	13/28/10 10:52	13/28/10 10:52
CBC With Differential/Platelet		CBC With Differential/Platelet
WBC	5.5	x10E3/uL
HGB	14.94	g/dL
Hematocrit	46.2	%
HCT	94	%
HCTC	32.7	%
RDW	13.2	%
Platelets	205	x10E3/uL
Neutrophils	44	%
Lymphs	46	%
Monocytes	9	%

Mit mér a fizikus?	Mit mér az orvos?	Mit mér a hallgató?
hossz	testmagasság	vörösvérsejt átmérője (2)
frekvencia	pulzusszám	impulzusgyakoriság (22)
koncentráció	vércukorszint	vérplazma fehérjekonc. (4)
feszültség	EKG-jel	EKG-jel (27)
hangintenzitás	hallásküszöb	hallásküszöb (25)
impedancia	impedancia-pletizmográfia (térfogatmérés)	bőrimpedancia (24)
nyomás	vérvnyomás	vérvnyomás (28)
sebesség	véráramlás sebessége	

6

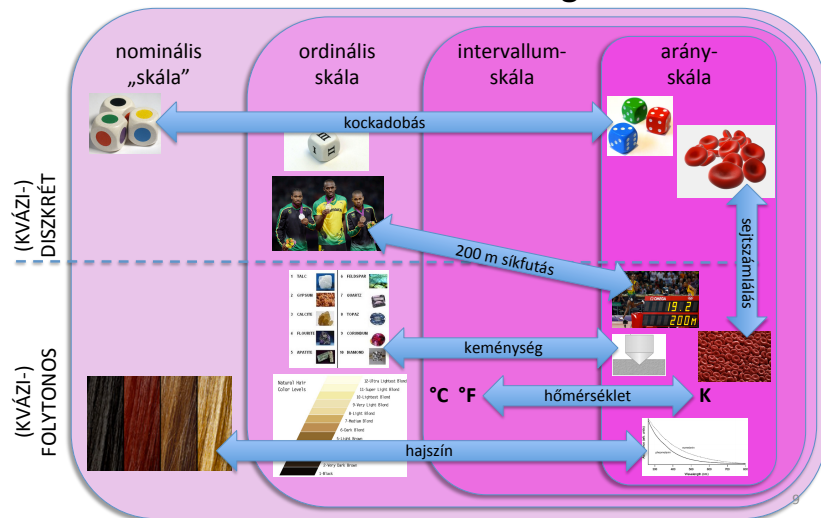
A statisztikai változók típusai I. Mérési szintek (S. S. Stevens nyomán)



(KVÁZI-) DISZKRÉT
(KVÁZI-) FOLYTÓNS

8

A statisztikai változók típusai I. A kontextus fontossága

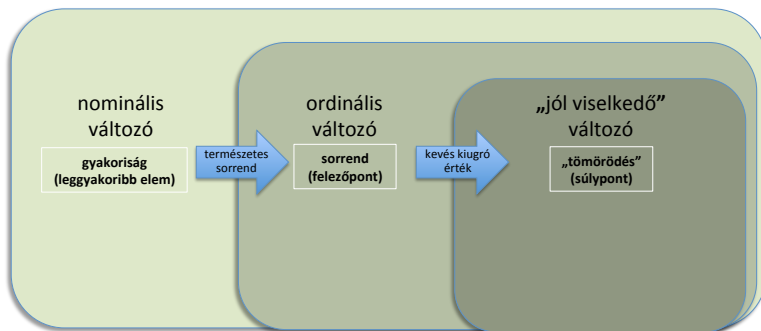


Alapsokaság és minta



11

A statisztikai változók típusai II. A statisztikai összehasonlítás alapja



10

Az adatok szemléltetése I.

- 1) Az adatok egyszerű listázása
- 2) A gyakoriságok táblázatos összefoglalása

- abszolút gyakoriság (Δn) és relatív gyakoriság ($\Delta n/n$)
- kvalitatív változók esetén a kategóriák adottak [Excel: =DARABTEL() vagy =COUNT() függvény]
- kvantitatív változók esetén a kategóriákat (osztályokat) magunk definiáljuk [Excel: =GYAKORISÁG() vagy =FREQUENCY() függvény is használható]
- gyakoriságsűrűség ($\Delta n/\Delta x$) és relatív gyakoriságsűrűség ($[\Delta n/n]/\Delta x$)

listázás: minden vizsgálat eredményét felsoroljuk

(abszolút) gyakoriság: az adott kimenetellel végződő vizsgálatok száma

relatív gyakoriság: az adott kimenetel adatokon belüli aránya

páciens sorszáma	vércsoport (AB0)	koleszterinszint (mg/dL)
1	O	149
2	A	130
3	A	130
4	B	159
5	A	134
6	A	153
7	O	147
8	O	14
9	A	139
10	A	139

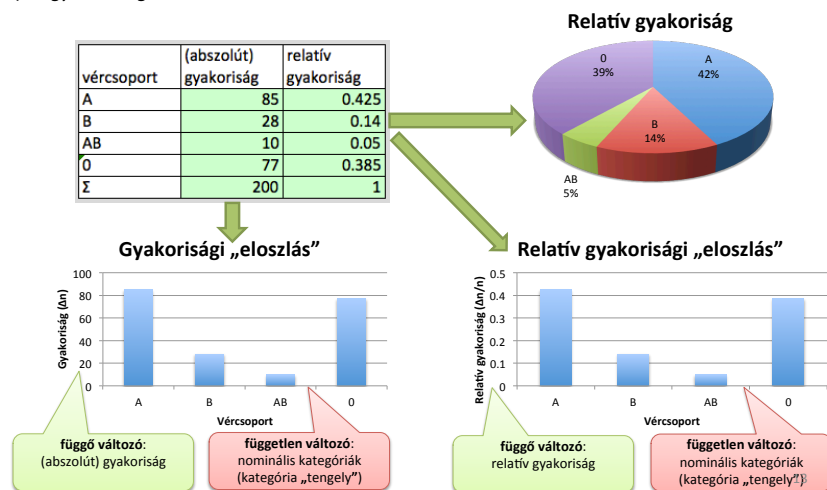
vércsoport	(abszolút) gyakoriság	relatív gyakoriság
A	85	0.425
B	28	0.14
AB	10	0.05
O	77	0.385
?	200	1

gyakorisági elosztások (differenciáldisztribúciós függvények)	osztályok felad (adati) határa	(abszolút) gyakoriság (GYAKORISÁG)	(abszolút) gyakoriság (DARABTEL)	relatív gyakoriság	(abszolút) gyakoriság-sűrűség	relatív gyakoriság-sűrűség
$x \leq 100$	100	0	0	0	0	0
$100 < x \leq 110$	110	0	0	0	0	0
$110 < x \leq 120$	120	2	2	0.01	0.2	0.001
$120 < x \leq 130$	130	5	5	0.025	0.5	0.0025
$130 < x \leq 140$	140	22	22	0.11	2.2	0.011
$140 < x \leq 150$	150	31	31	0.155	3.1	0.0155
$150 < x \leq 160$	160	48	48	0.24	4.8	0.024
$160 < x \leq 170$	170	40	40	0.2	4	0.02
$170 < x \leq 180$	180	32	32	0.16	3.2	0.016
$180 < x \leq 190$	190	10	10	0.05	1	0.005
$190 < x \leq 200$	200	9	9	0.045	0.9	0.0045
$200 < x \leq 210$	210	1	1	0.005	0.1	0.0005
$210 < x$		0	0			
összeg		200	200	1		

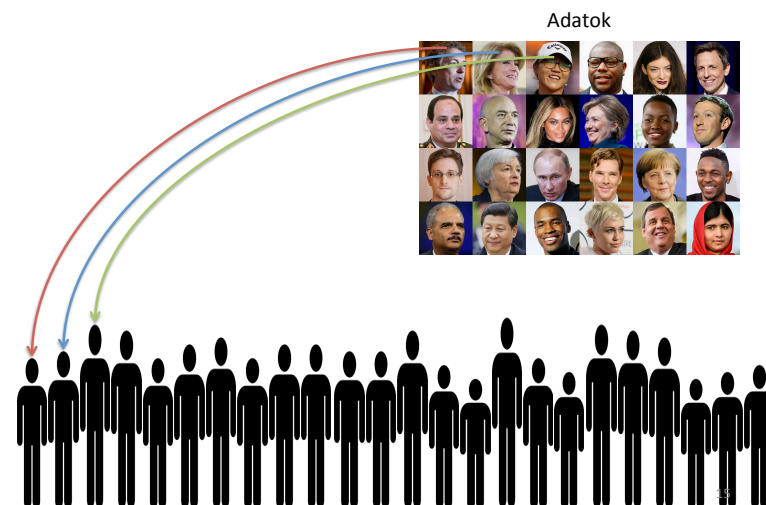
12

Az adatok szemléltetése II.

3) A gyakoriságok ábrázolása **kvalitatív** változó esetén

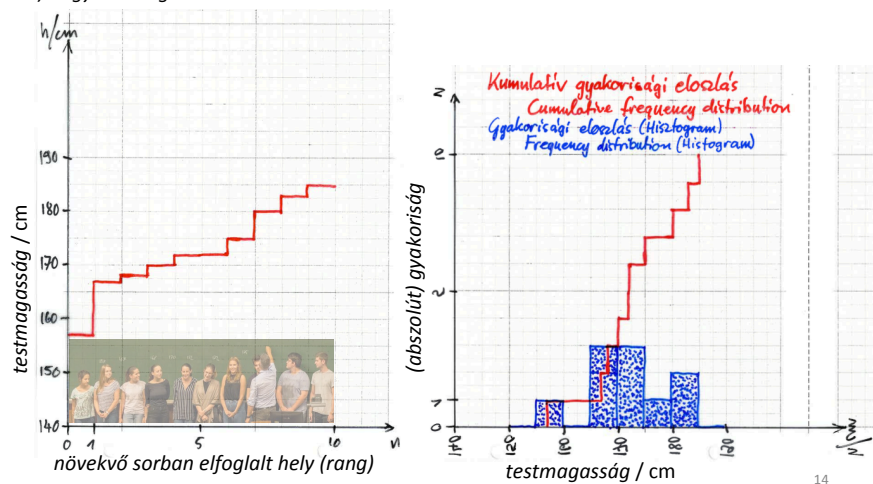


Az adatok összetétele

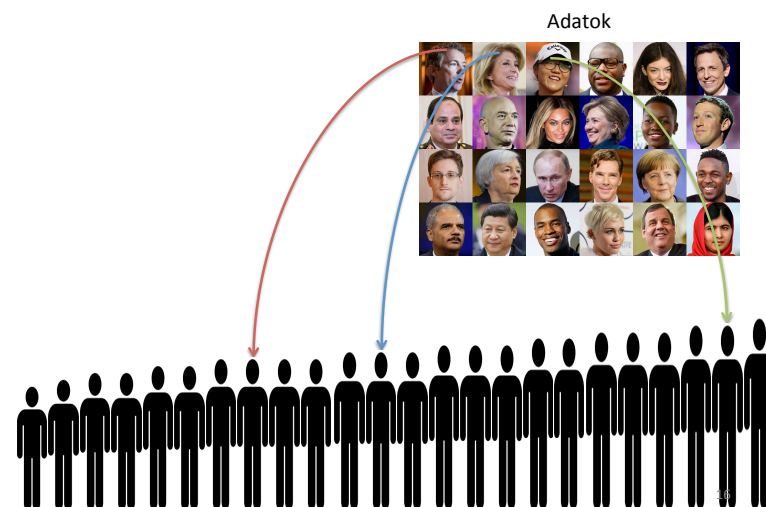


Az adatok szemléltetése III.

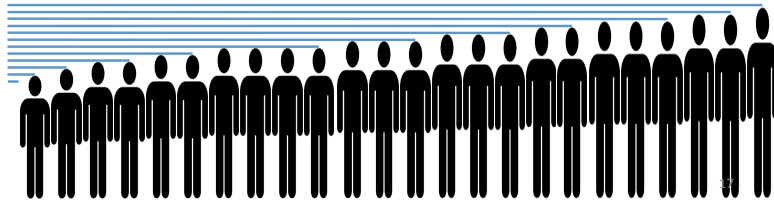
4) A gyakoriságok ábrázolása **kvantitatív** változó esetén



Az adatok összetétele

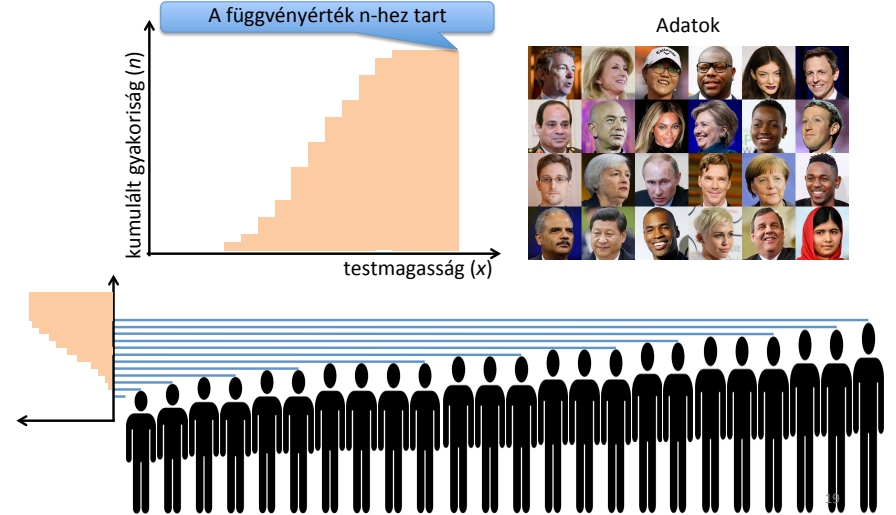


Az adatok összetétele



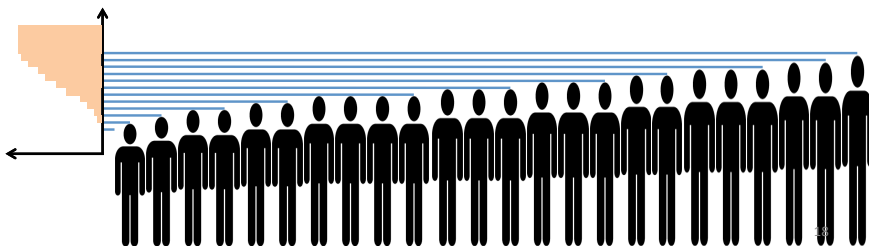
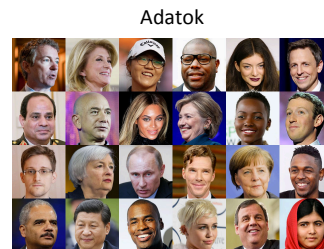
Kumulatív gyakorisági eloszlás

Hány elem **kisebbsége** mint egy adott x érték?



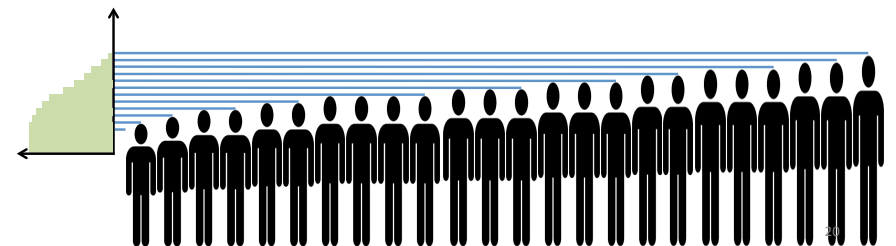
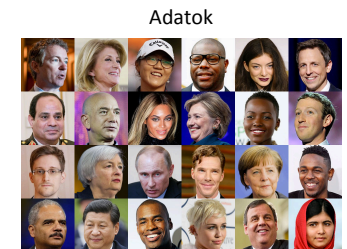
Kumulatív gyakorisági eloszlás

Hány elem **kisebbsége** mint egy adott x érték?



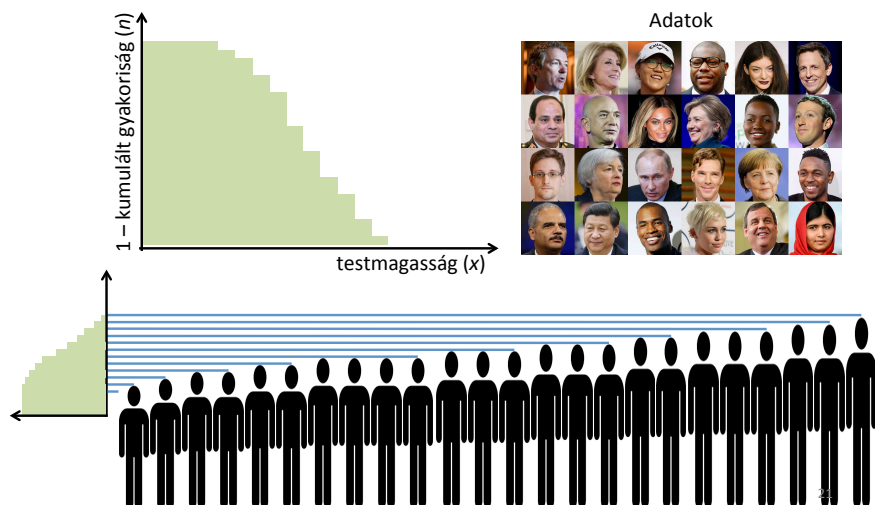
Integráldiszkriminációs eloszlás

Hány elem **nagyobb** mint egy adott x érték?



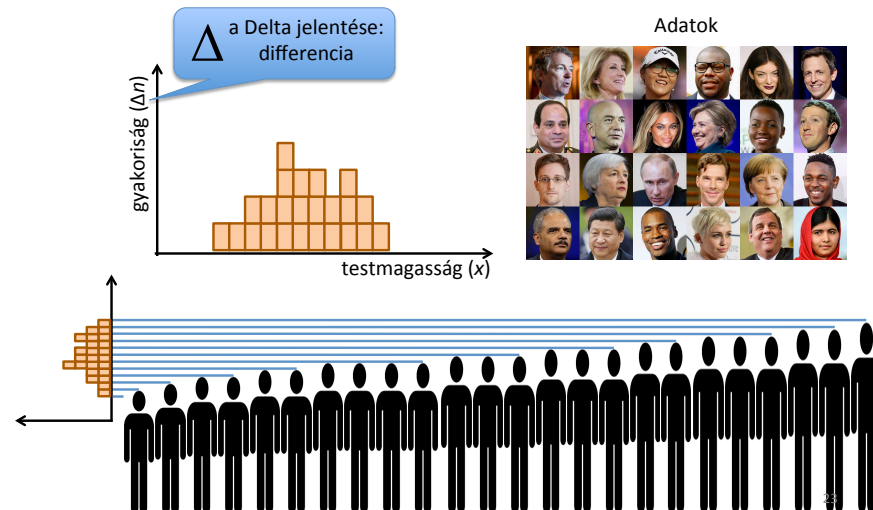
Integráldiszkriminációs eloszlás

Hány elem **nagyobb** mint egy adott x érték?



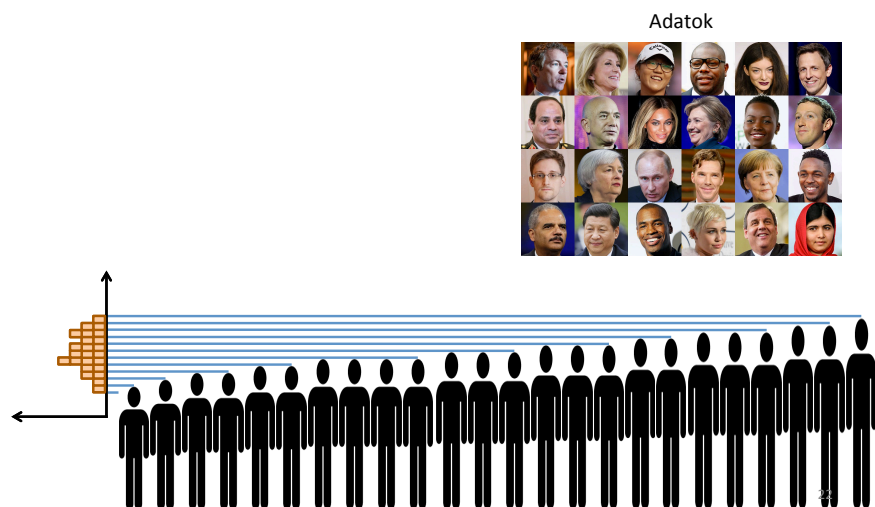
Gyakorisági eloszlás

Hány elem esik egy Δx szélességű osztályba?



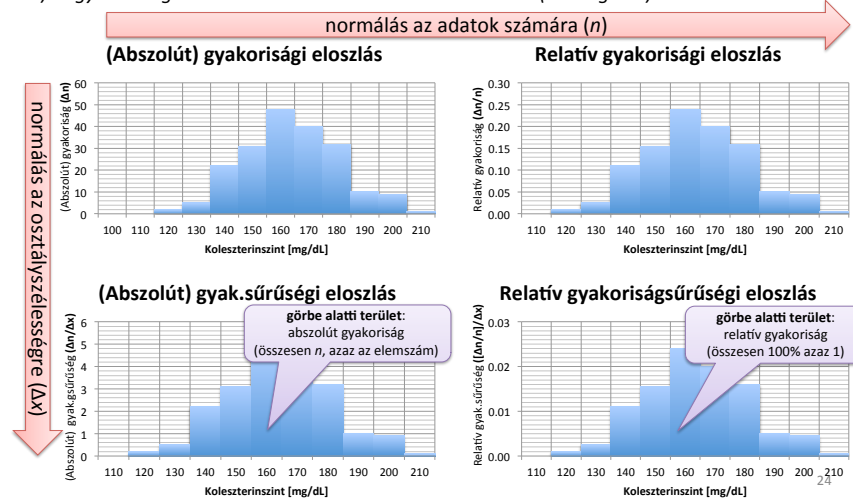
Gyakorisági eloszlás

Hány elem esik egy Δx szélességű osztályba?



Az adatok szemléltetése III.

4) A gyakoriságok ábrázolása **kvantitatív** változó esetén (hisztogram)



Ellenőrző kérdések I.

- Mik a tudomány legfontosabb jellemzői?
- Mi jellemző a tudományos gondolkodásra?
- Mi különbözteti meg a tudományos orvoslást a kuruzslástól?
- Hogyan győződhetünk meg arról, hogy egy adott orvosi eljárás tudományos bizonyítékokon alapul?
- Mondj példát nem tudományos gyógymódokra!
- Kinek kell viselnie a bizonyítás terhét: Aki egy módszerről azt állítja, hogy hatásos, vagy aki szerint az nem hatásos?
- Mi a statisztika?
- Mi a statisztika célja?
- A matematika mely területeit használja fel a statisztika?
- Mi a modellalkotás célja?
- Mi a kapcsolat a modell és a valóság között?
- Mi a probléma a következő állítással: „Senki sem bizonyította, hogy ez a módszer hatástalan, ezért igazságtalan lenne korlátozni a használatát.”
- Hogyan lehet csoportosítani egy beteg megfigyelése során nyert statisztikai változókat?
- Nevez meg nominális változókat.
- Nevez meg ordinális változókat.
- Nevez meg intervallumskálán mért változókat.
- Nevez meg arányskálán mért változókat.
- Nevez meg olyan kísérletet, amelynek során többféle típusba is sorolható a megfigyelt változó.
- Nevez meg „jól viselkedő” statisztikai változókat.
- Mit jelent az, hogy egy statisztikai változó „jól viselkedik”?
- Miért lehet félrevezető a „nominális skála” kifejezésben a skála szó?

25

Ellenőrző kérdések III.

- Hogyan tehetők különböző méretű adathalmazokból nyert gyakoriságok közvetlenül összehasonlíthatóvá?
- Hogyan tehetők a gyakoriságok közvetlenül összehasonlíthatóvá, ha az osztályszélességek eltérnek?
- A relatív gyakoriság mitől „relatív”?
- A gyakoriságsűrűség kifejezésben mire utal a „sűrűség”?
- Mit és hogyan lehet leolvasni egy gyakorisági eloszlásról?
- Mit és hogyan lehet leolvasni egy relatív gyakorisági eloszlásról?
- Mit és hogyan lehet leolvasni egy gyakoriságsűrűség eloszlásról?
- Mit és hogyan lehet leolvasni egy relatív gyakoriságsűrűség eloszlásról?
- Mekkora a teljes görbe alatti területe egy gyakoriságsűrűség eloszlásnak?
- Mekkora a teljes görbe alatti területe egyrelatív gyakoriságsűrűség eloszlásnak?
- Mit és hogyan lehet leolvasni egy kumulatív gyakorisági eloszlásról?
- Mit és hogyan lehet leolvasni egy integráldiszkriminációs függvényről?
- Mi a kumulatív gyakorisági eloszlás előnye a hisztogrammal szemben?
- Mi a hisztogram előnye a kumulatív gyakorisági eloszlással szemben?

27

Ellenőrző kérdések II.

- Mi az alapvető különbség a nominális és az ordinális skála között?
- Mi az alapvető különbség az ordinális és az intervallumskála között?
- Mi az alapvető különbség az intervallum- és az arányskála között?
- Miért annyira fontos a statisztikai változók pontos definíciója?
- Összehasonlítást akarunk végezni. Hogyan célszerű a statisztikai változókat kategorizálni?
- Egy változót intervallumskálán mérünk. Ez azt is jelenti, hogy jól viselkedő?
- Mi az alapsokaság (populáció)?
- Mi a minta?
- Hogyan vegyünk mintát az alapsokaságból?
- Miért nem szokta a minta az alapsokaságot tökéletesen jellemezni?
- Mi az oka annak, hogy az alapsokaságról a minta alapján levont következtetéseinket bizonytalanság terheli?
- Hogyan lehet egy mintát leírni?
- Mi a „gyakoriság” jelentése?
- Ha csak azt mondjuk „gyakoriság”, akkor az abszolút vagy relatív gyakoriságot jelent?
- Hogyan lehet Excelben egy adathalmazhoz tartozó gyakorisági értékeket meghatározni kvalitatív változó esetén?
- Hogyan lehet Excelben egy adathalmazhoz tartozó gyakorisági értékeket meghatározni kvantitatív változó esetén?

26