

# Orvosi statisztika, informatika és telemedicina

## Leíró statisztika

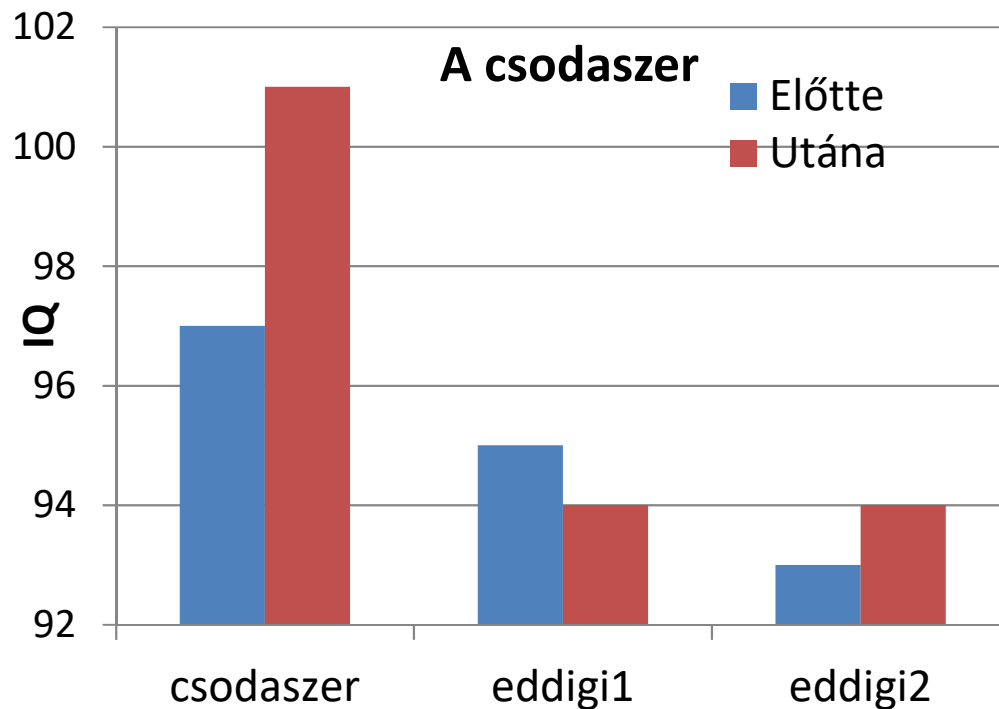
Adatok összegzése

készítette: VERES Dániel

2021. Szeptember 15.

# Miért tanuljunk Orvosi statisztikát?

„Azért, hogy el tudjuk dönteni, elhiggyünk-e valamit, amit olvasunk, vagy hogy észrevegyük, hol van benne a hiba, vagyis hogy ne dőlünk be olyan könnyen a statisztikai bűvészkedéseknek, műtermékeknek és tévedéseknek.”  
(Reiczigel-Harnos-Solymosi: Biostatisztika nem statisztikusoknak)



# Miért tanuljunk Orvosi statisztikát?

- „Azért, hogy saját vizsgálataink tervezését, kiértékelését és bemutatását ügyesebben el tudjuk végezni.” (diplomamunka...)
- „Azért, hogy könnyebben tudjunk tanácsot kérni - együtt dolgozni – statisztikussal, ha (amikor) elbizonytalanodnánk”

# VÁLTOZÉKONYSÁG



Megismeréséhez és leírásához:

Más **Gondolkozásmód**

új **Nevezéktan**

csekély **matematika**

A **statisztika** a véletlen tömegjelenségek leírója.



- Adatgyűjtés
- Adatok rendszerezése, áttekintése
- Adatok elemzése
- Következtetések levonása

**Leíró statisztika**

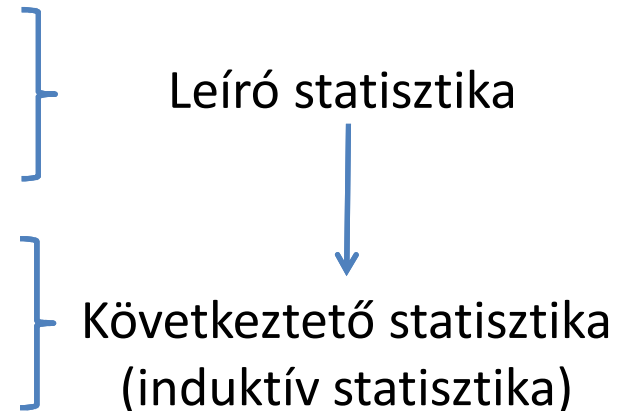


**Következtető statisztika**  
(induktív statisztika)



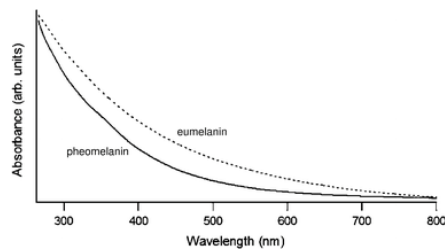
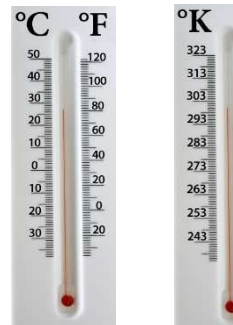


- Adatgyűjtés
- **Adatok rendszerezése, áttekintése**
- Adatok elemzése
- Következtetések levonása



# VÁLTOZÓK

Amit meg tudunk mérni vagy meg tudunk figyelni.



# VÁLTOZÓK típusai, mérési skálák

(KVÁZI-)  
DISZKRÉT

kvalitatív  
(kategoriális)

nominális  
„skála”

= ≠

természetes  
sorrend



ordinális  
skála

= ≠  
< >

ismert  
távolság



kvantitatív  
(számszerű)

intervallum-  
skála

= ≠  
< >  
+ -

természetes  
nullérték



arány-  
skála

= ≠  
< >  
+ -  
× ÷



(KVÁZI-) FOLYTONOS



# Nominális változó jellemzése I.

## Analitikus

### Lista

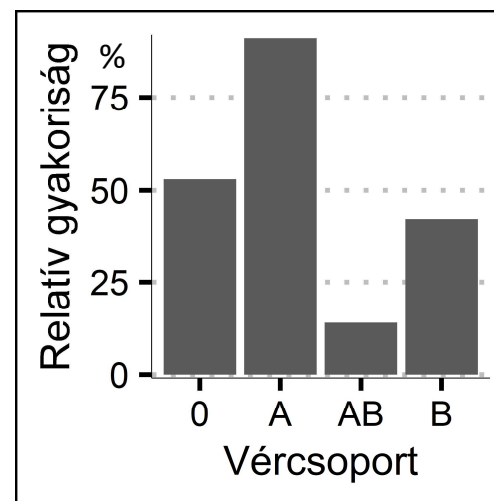
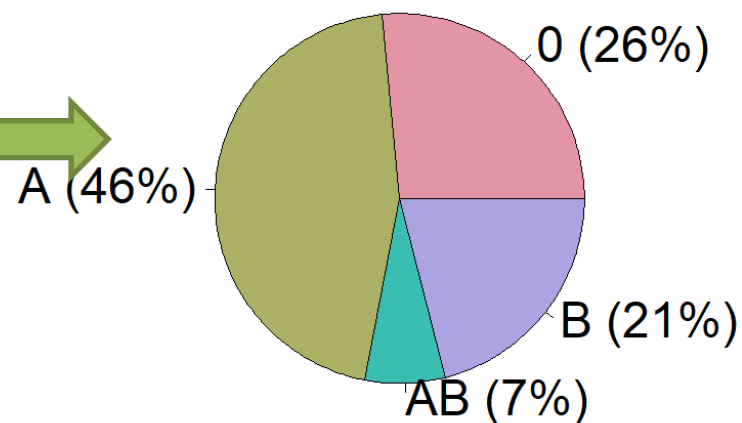
páciens sorszáma	vércsoport (ABO)	koleszterinszint (mg/dL)
1	0	149
2	A	125
3	A	127
4	B	159
5	A	134
6	A	153
7	0	164
8	0	140
9	A	139
10	B	150

### Gyakorisági sor (tábla)

vércsoport	(abszolút) gyakoriság	relatív gyakoriság
A	85	0.425
B	28	0.14
AB	10	0.05
0	77	0.385
$\Sigma$	200	1

## Grafikus

*Az agy és a józan ész*



# Nominális változó jellemzése II.

## Analitikus

Gyakorisági sor (tábla)

vércsoport	(abszolút) gyakoriság	relatív gyakoriság
A	85	0.425
B	28	0.14
AB	10	0.05
0	77	0.385
$\Sigma$	200	1

Tömörítés, de információvesztéssel.

Jellemző mutató („hasonlóságra”):

**Módusz:** leggyakoribb elem(ek)

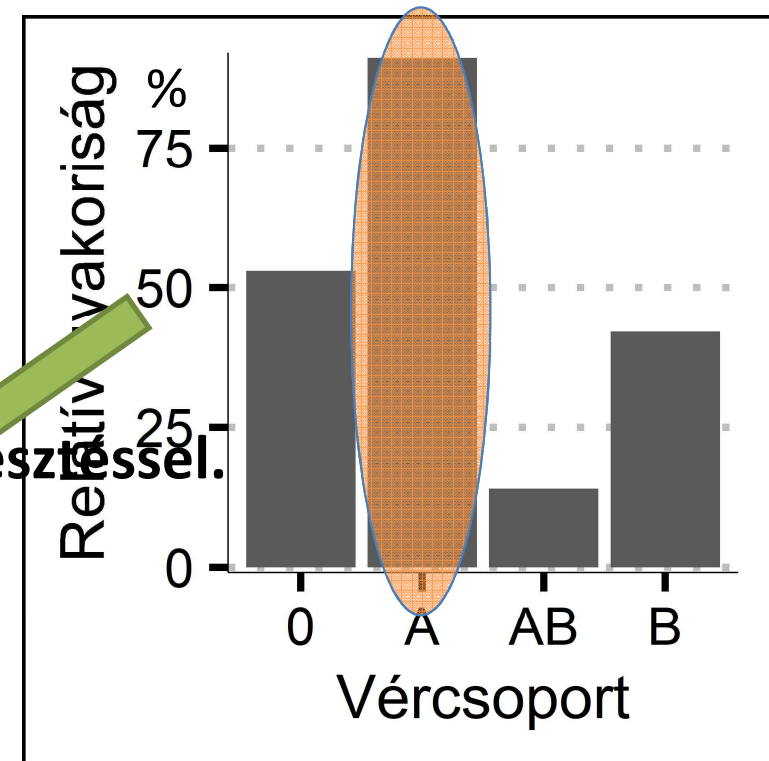
Jelölése:  $\text{Mod}$ ,  $x_{\text{mod}}$  ~~Átlag?!~~

Egyéb jellemzők: („változékonyságra”)

**Elemzés** ( $n$ ), kategóriák száma

## Grafikus

*Az agy és a józan ész*



# Ordinális változó jellemzése I.

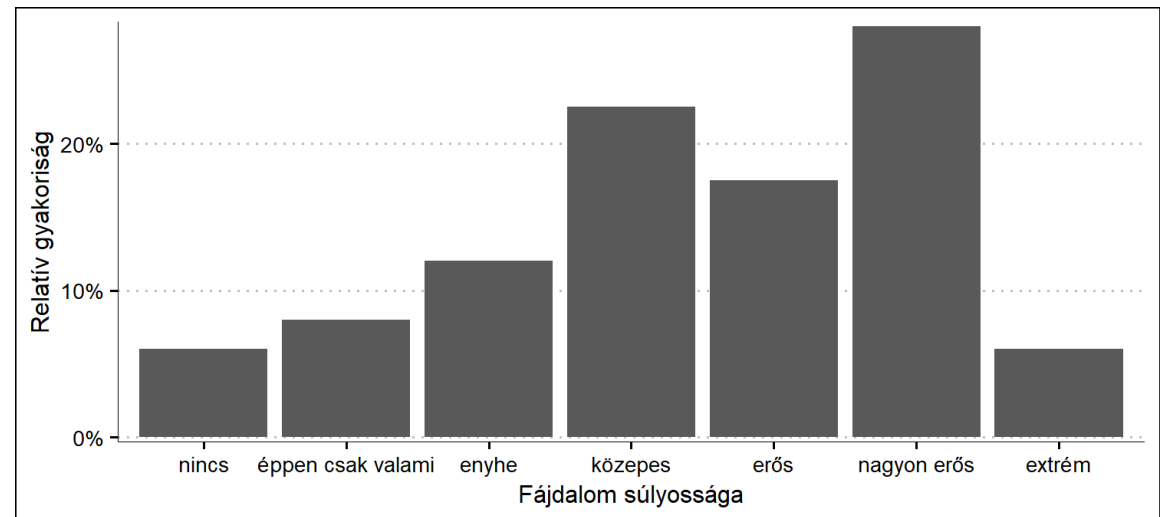
## Analitikus

## Grafikus

Adatrendezés információvesztés nélkül

Gyakorisági sor (tábla)

Fájdalom mértéke	Relatív gyakoriság
nincs	0,06
éppen csak valami	0,08
enyhe	0,12
közepes	0,225
erős	0,175
nagyon erős	0,28
extrém	0,06
$\Sigma$	1



Jellemző mutató:

Módusz

Egyéb jellemzők:

Elemzés, kategóriák száma

# Ordinális változó jellemzése II.

## Analitikus

Gyakorisági sor (tábla)

Fájdalom mértéke	Kumulatív relatív gyakoriág
nincs	0,06
éppen csak	0,14
enyhe	0,26
közepes	0,485
erős	0,66
nagyon erős	0,94
extrém	1

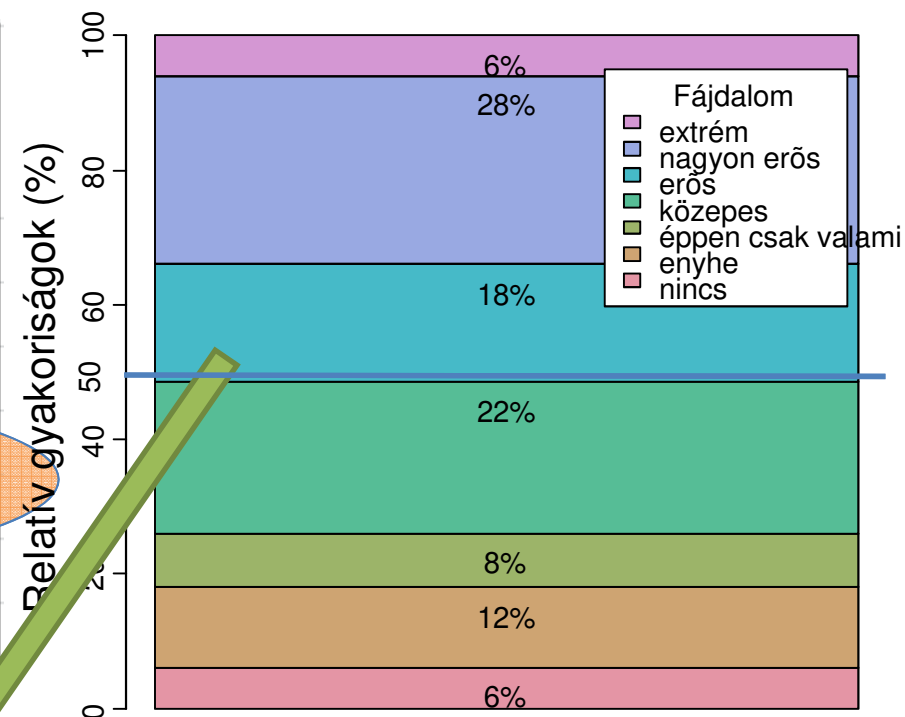


Új Jellemző („hasonlóságra”):

**Medián:** „középső” elem(ek)

Jelölése: Me, Med,  $x_{med}$

## Grafikus



Fájdalom



# Kvantitatív (számszerű) változó jellemzése I.

## Analitikus

### Gyakorisági sor (tábla)

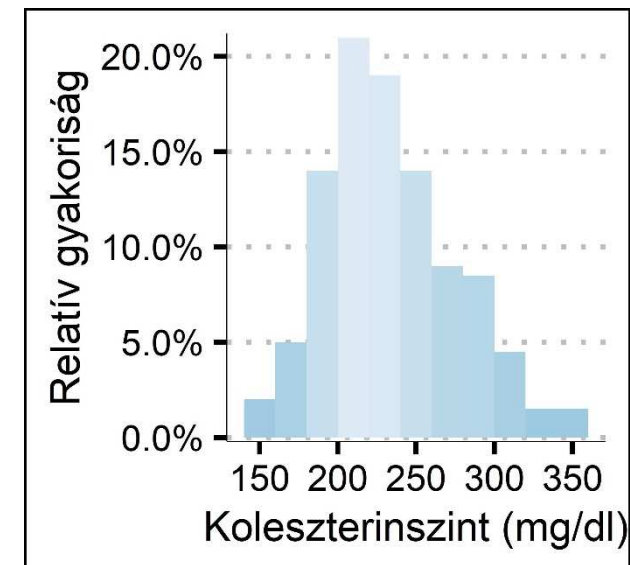
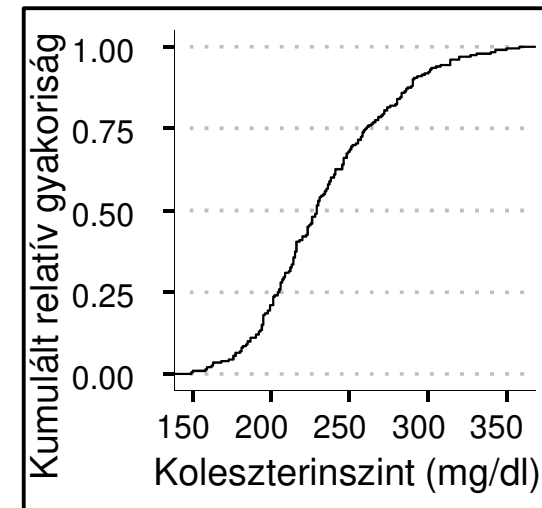
gyakorisági eloszlások (differenciáldiszkriminációs függvények)			
osztályok	osztályok felső (zárt) határa	(abszolút) gyakoriság (GYAKORISÁG)	(abszolút) gyakoriság (DARABT)
$x \leq 100$	100	0	
$100 < x \leq 110$	110	0	
$110 < x \leq 120$	120	2	
$120 < x \leq 130$	130	5	
$130 < x \leq 140$	140	22	
$140 < x \leq 150$	150	31	
$150 < x \leq 160$	160	48	
$160 < x \leq 170$	170	40	
$170 < x \leq 180$	180	22	

Adatrendezés **információvesztéssel** járhat.

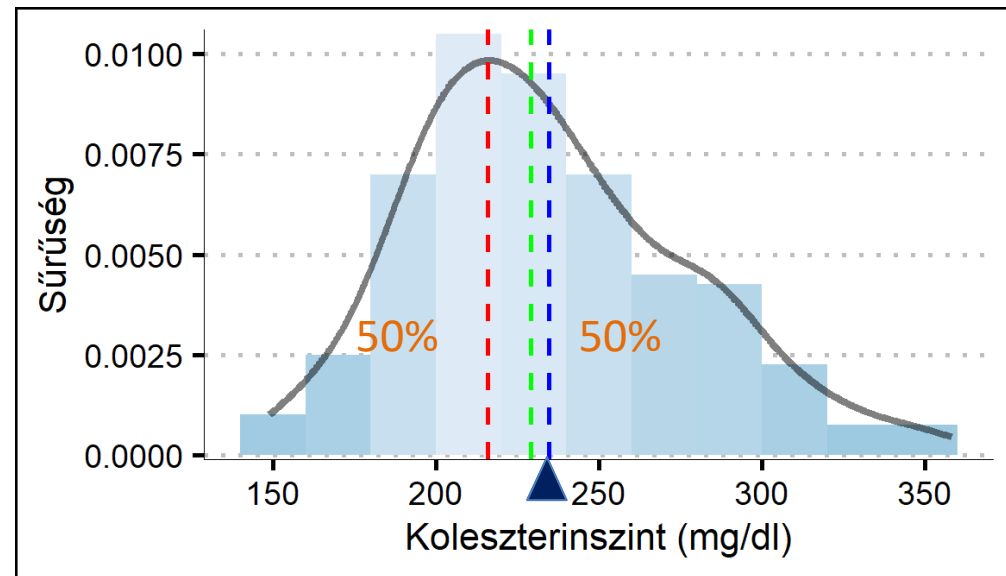
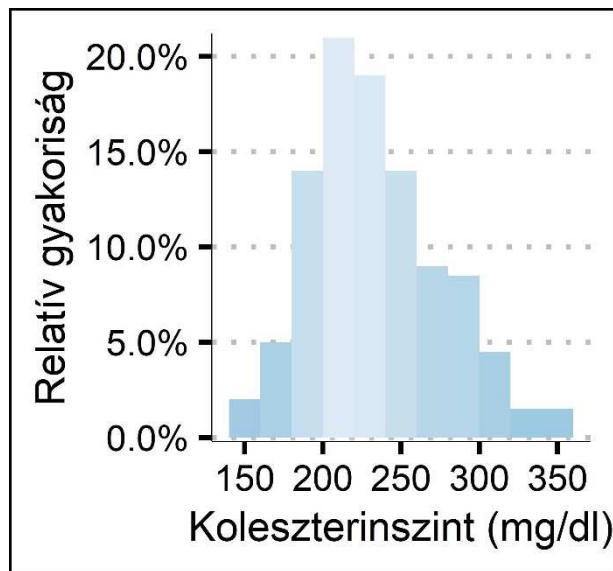
Osztályszélesség meghatározása:

- szakmai és esztétikai szempontok
- statisztikai szempontok alapján

## Grafikus



# Kvantitatív változó jellemzése II.



Jellemzők – **középértékek** (speciális **helyparaméterek**):

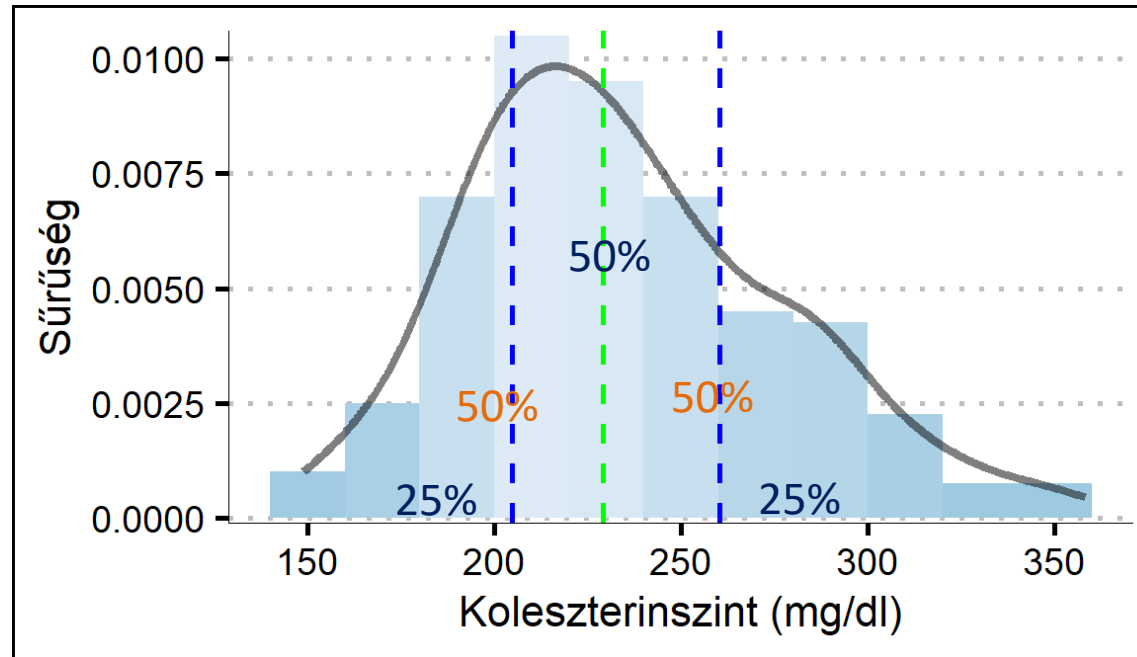
- **Módusz**(ok): leggyakoribb elem(ek) ?
- **Medián**: „középső” elem(ek?)
- **Átlag** (számtani közép): „súlypont” ,érzékeny a „kiszóró” adatokra ?!

Jelölése:  $x_{\text{átl}}$ ,  $\bar{x}$

Előny: tömörítés, **kevés adatból is számíthatóak**

Képletek: képlettárban

# Kvantilisek I.



Egyéb helyparaméterek:

- **Medián:** 50-50% ( $Q_2$ )
- **Kvantilisek** : alsó kvartilis ( $Q_1$ ): 25-75%; felső kvartilis ( $Q_3$ ): 75-25%

Általánosan

***p*-kvantilis(ek):** az adatrendszer *p*-kvantilisének nevezzük azt a számot, amelynél kisebb adatok darabszáma legfeljebb  $n \cdot p$  és amelynél nagyobb adatok darabszáma legfeljebb  $n \cdot (1 - p)$ , ahol  $p$  0 és 1 közötti szám

# Megjegyzés

Nap sorszáma	Várakozási idő			Nap sorszáma	Várakozási idő		
1	1,27	medián:	8,475	1	1,27	medián:	8,475
2	3,3	alsó kvartilis	3,59	2	3,3	alsó kvartilis	3,59
3	3,44	átlag	7,723333	3	3,44	átlag	9,141667
4	3,64			4	3,64		
5	6,33			5	6,33		
6	7,72			6	7,72		
7	9,23			7	9,23		
8	9,87			8	9,87		
9	10,31			9	10,31		
10	12,29			10	12,29		
11	12,3			11	12,3		
12	12,98			12	30		

**Medián**, kvantilisok elméletben és gyakorlatban eltérhetnek.  
 Átlag érzékeny a kiszóró adatokra, de kvantilisok nem érzékenyek.  
 Módusz?

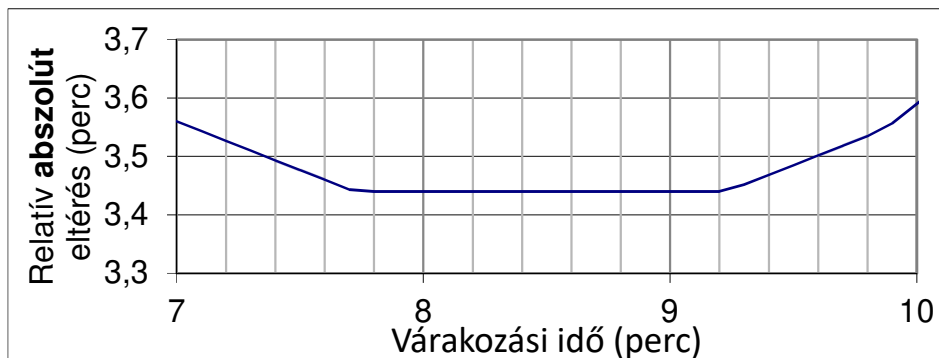
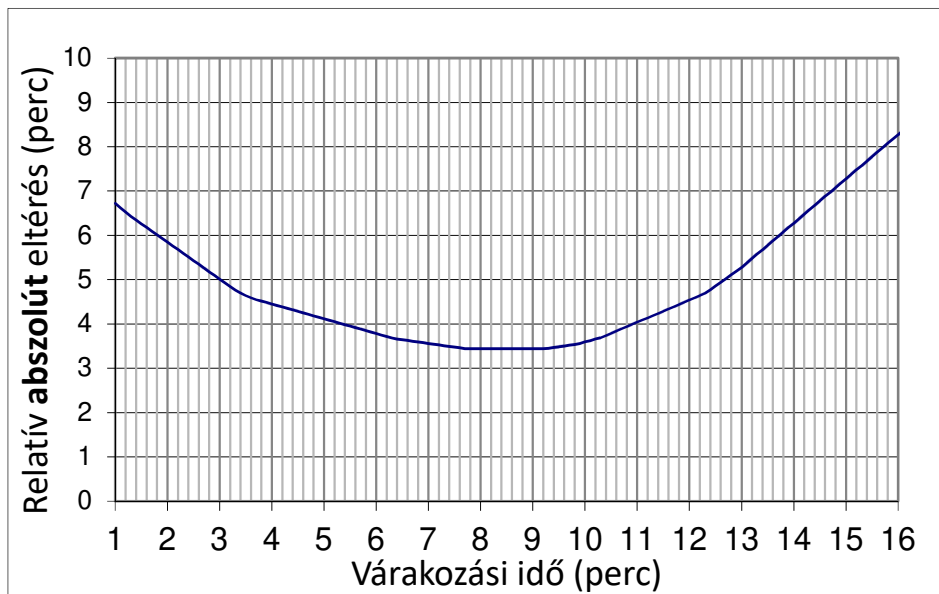


# Középérték - Távolságok

$$\frac{1}{n} \sum |x_i - x^*|$$

Minimális, ha:

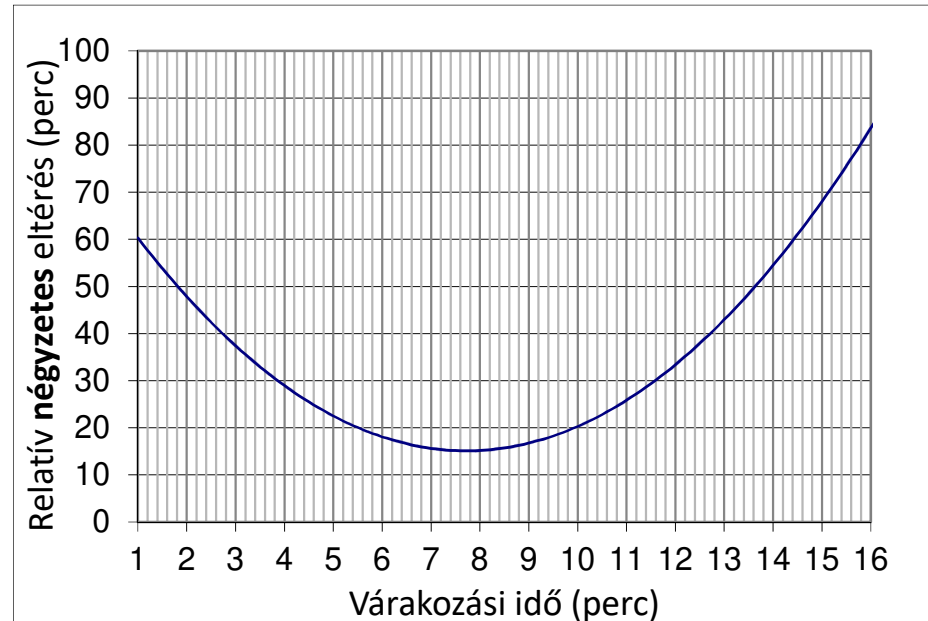
$$x^* = \textit{Medián}$$



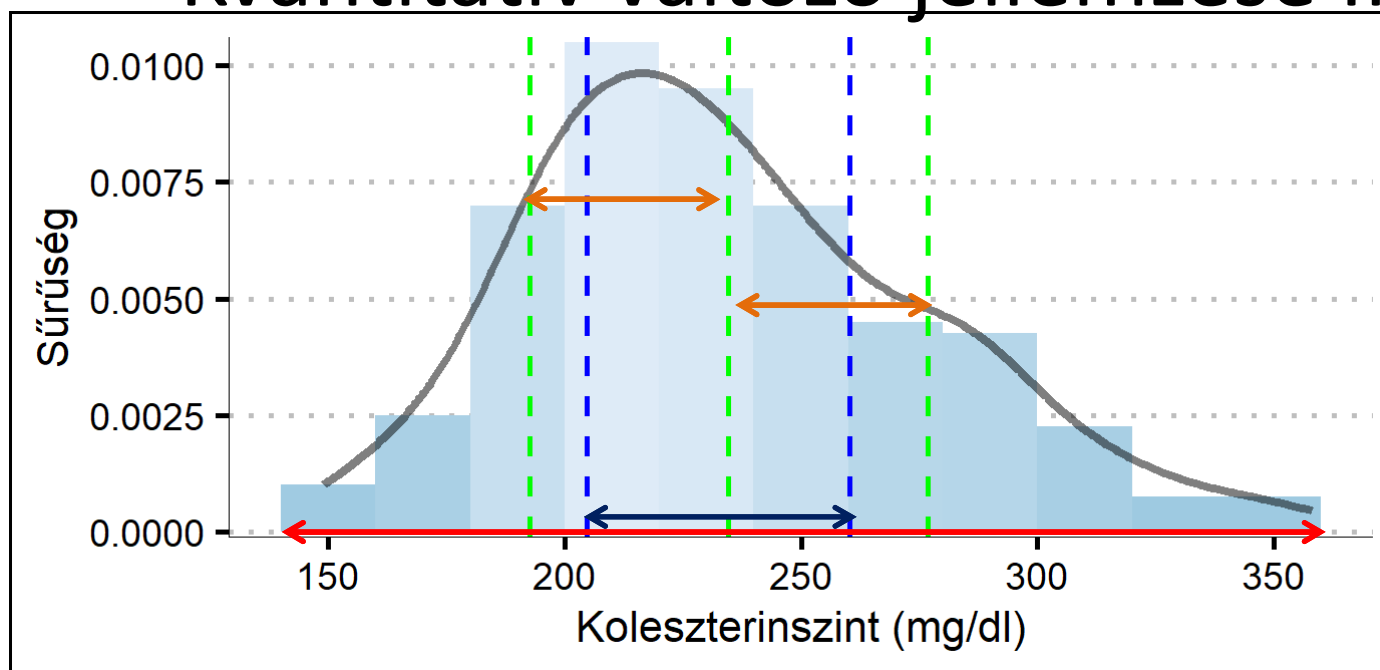
$$\frac{1}{n} \sum (x_i - x^*)^2$$

Minimális, ha:

$$x^* = \textit{Átlag}$$



## Kvantitatív változó jellemzése III.



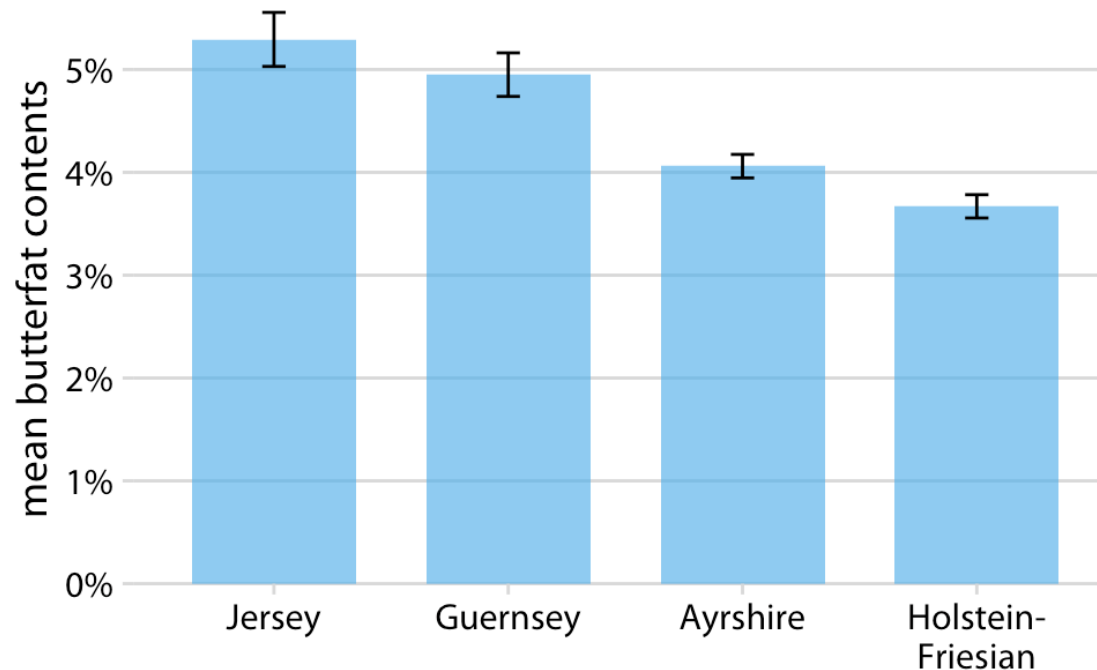
Jellemzők – szóródási paraméterek:

- **Terjedelem**: *maximális* érték és *minimális* érték különbsége
- **Variancia (szórásnégyzet,  $s^2$ )**: átlagtól vett átlagos négyzetes eltérés
- **Szórás ( $s$ )**: variancia négyzetgyöke – eloszlásgörbe „szélessége”
- **Interkvartilis távolság (IQR)** : felső és alsó kvartilis értékek különbsége, előnye: nem érzékeny a „kiszóró” pontokra

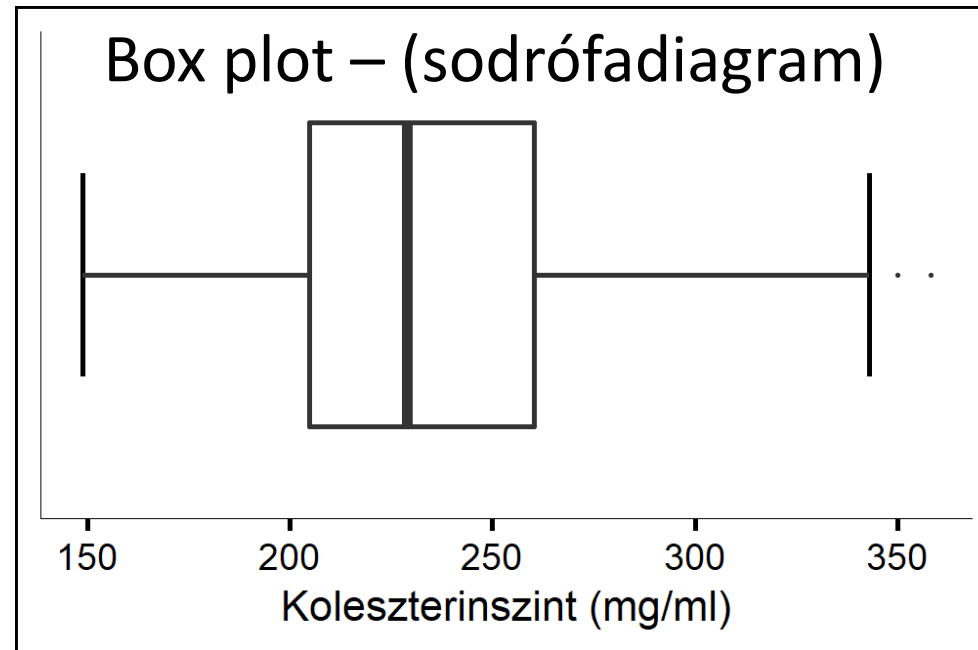
# Amit NE

**Oszlopdiaagram számszerű jellemzőkre: igazából a görbe alatti terület emeli ki az ábra,**

- de a „tetejét” nézzük nem is a görbe alatti területet...
- 0-nál kell(ene) kezdeni
- a szórás egy y-tartomány, az érték (átlag) egy terület
- és ha nem szimmetrikus a szórás? és ha nagyobb, mint az átlag?
- ha az átlag negatív?
- ...



# Kvantitatív változó jellemzése IV.



**Sodrófa szeme:** *medián*

**Sodrófa teste:** *interkvartilis tartomány (IQR) (felső és alsó kvartilis)*

**Sodrófa szára:** *IQR 1,5-szerese...*

sodrófa szárán túl: **kiszóró pont**

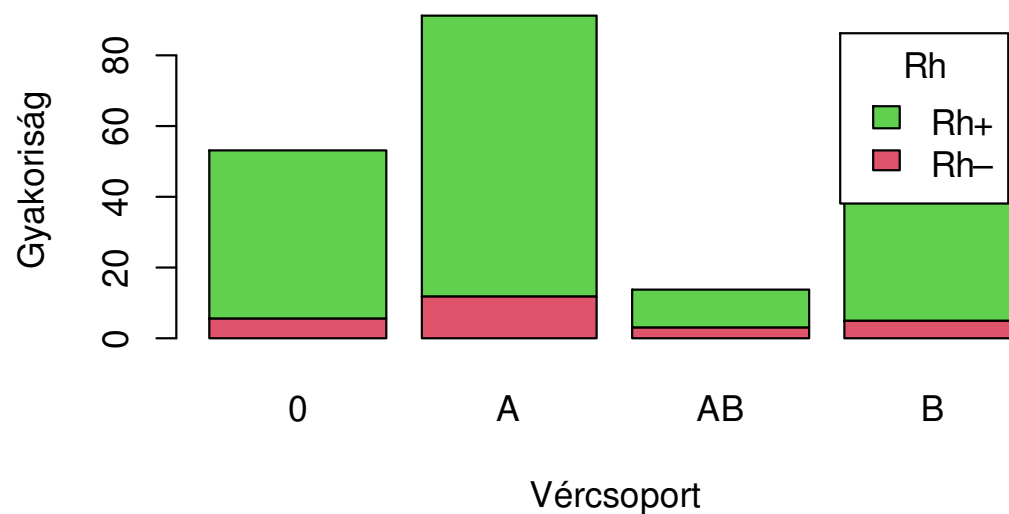


# Több kvalitatív változó jellemzése

Analitikus: ***kontingencia*** táblázat

	A	B	AB	0	$\Sigma$
Rh+	79	37	11	47	173
Rh-	12	5	3	6	27
$\Sigma$	39	115	38	8	200

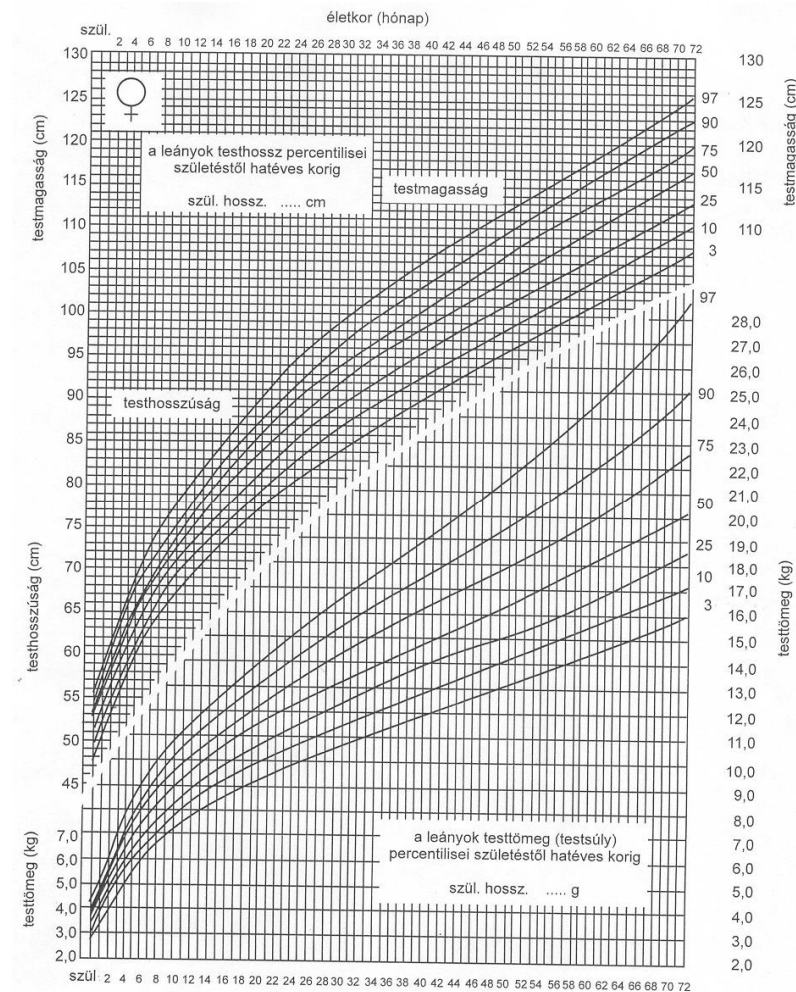
Grafikus: ***halmozott oszlopdiaagram***

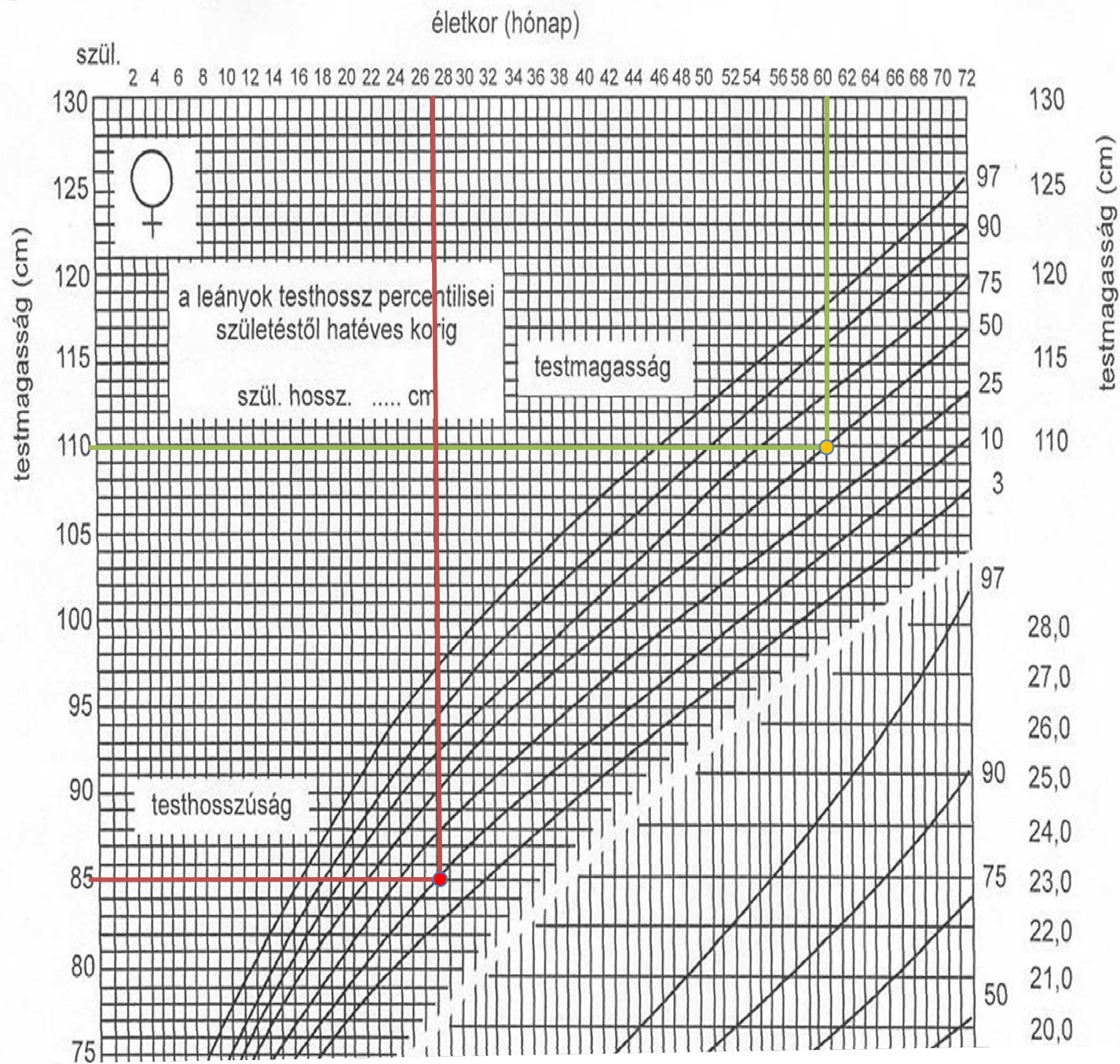


# Több kvantitatív változó jellemzése

Grafikus: **percentilis ábrák**

Percentilis: %-ban kifejezett kvantilis





# Színek

Milyet?

- **ha nincs sorrend:** kvalitatív színskála  
jól megkülönböztethető legyen  
ne mutasson sorrendet, „egyformák” legyenek:  
pl. ugyanolyan „világos” legyen mindegyik

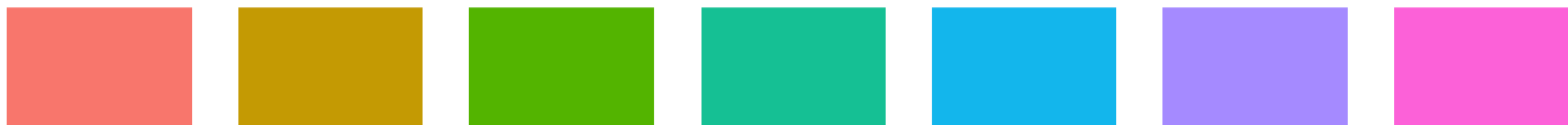
Okabe Ito



ColorBrewer Dark2



ggplot2 hue





# Színek

Milyet?

- **ha van sorrend:** szekvenciális színskálák  
mutasson sorrendet (kisebb-nagyobb)  
mutassa a különbség mértékét (mennyivel kisebb-nagyobb)

ColorBrewer Blues



Heat



Viridis

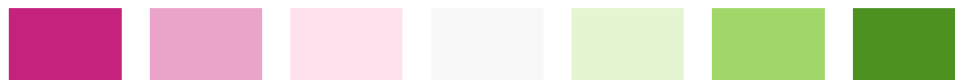


Ha szimmetrikus legyen

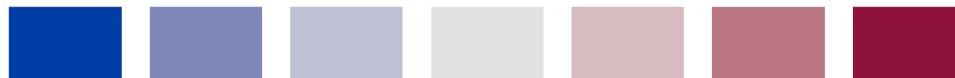
CARTO Earth



ColorBrewer PiYG



Blue-Red



# Színek

Milyet?

- **ha van sorrend:** szekvenciális színskálák  
mutasson sorrendet (kisebb-nagyobb)  
mutassa a különbség mértékét (mennyivel kisebb-nagyobb)

NEM ilyen: rainbow (hol „gyors”/”lassú”, szélein ugyanolyan)

rainbow scale



rainbow converted to grayscale



# Színek

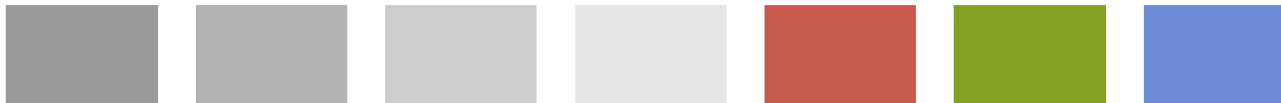
Milyet?

- **ha valamit ki kell emelni:**  
legyen más: sötétebb vagy telítettebb  
lehet több alapszín és néhány kiemelt

Okabe Ito Accent



Grays with accents



ColorBrewer Accent



# Színek

Milyet?

Gondoljunk pl. rám – **SZÍNTÉVESZTŐK** (régi projektor: piros-)  
(piros-zöld [...] ~ 8% férfiaknál, 0,5% nőknél)

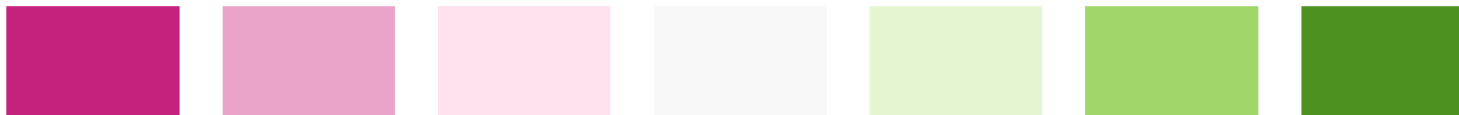
Jó lehet, pl:

Nincs sorrend



Szimmetrikus sorrend

ColorBrewer PiYG



Vannak szoftverek, aminél tesztelhető („szimuláljuk” a színtévesztőket)!

ImageJ-hez VischeckJ

Photoshopnál a CUD

# OMHV

