

# Természettudományos rendszerek és komplexitása gyakorlat (ZTT1233L)

Irodalom:

Reiczigel J., Harnos A., Solymosi N. (2007): Biostatisztika nem statisztikusoknak, , Pars Kft

<http://biostatkonyv.hu/>

Kurzus információk, anyagok:

<http://zeus.nyf.hu/~szept/kurzusok.htm>

# Statisztika alapfogalmai

<https://e-akademia.hu/html/zold12/ch-statalap.html> *(2. fejezet)*

# Leíró statisztika

Mérőszámok, statisztikák

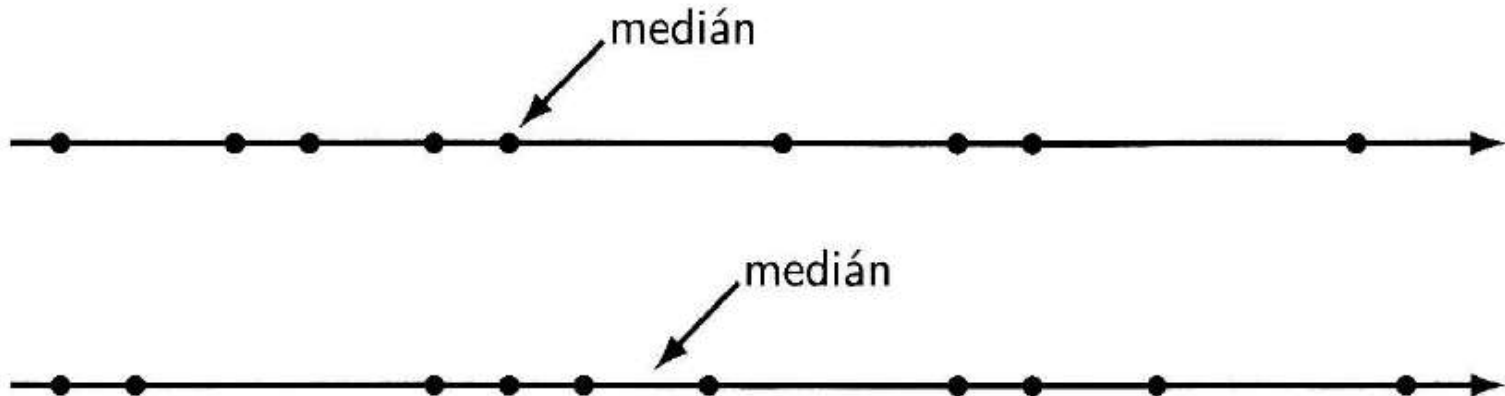
n: mintaszám

Helyzeti mutatók:

-Átlag

$$\bar{x} = \frac{\sum_{k=1}^n x_k}{n}.$$

-Medián

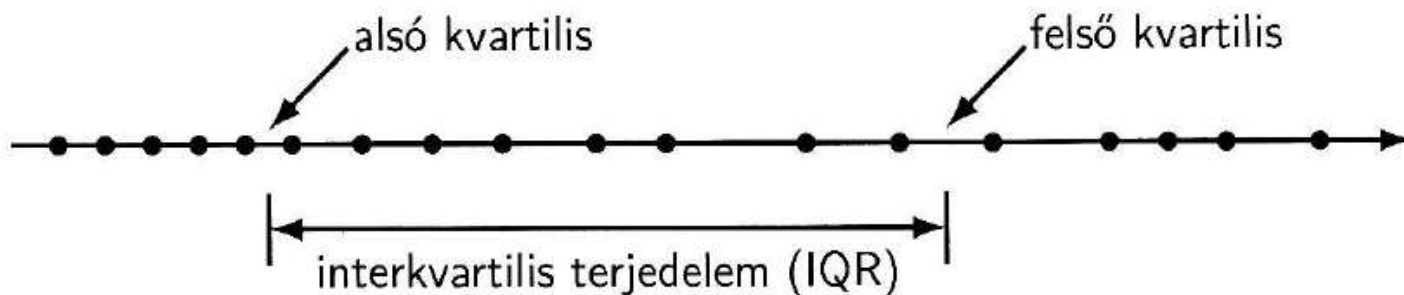


# Leíró statisztika

## Mérőszámok, statisztikák

### Szóródási mutatók:

- **Range** (minimum-maximum közötti különbség)
- **Alsó/Felső kvartilis**
- **Interkvartilis terjedelelem (IRQ)**



# Leíró statisztika

## Mérőszámok, statisztikák

Szóródási mutatók:

-Variancia/Szórásnégyzet

$$s^2 = \frac{\sum_{k=1}^n (x_k - \bar{x})^2}{n - 1}$$

-Szórás (SD)

$$s = \sqrt{s^2}$$

-Relatív szórás/variációs együttható (CV)

$$CV = 100 \frac{s}{\bar{x}} \%$$

# Leíró statisztika

## Változó ábrázolása

Nominális vagy ordinális (faktor) adatok :

### 4.1. példa. Kocák alomszáma – gyakoriságtáblázat és oszlopdiagram

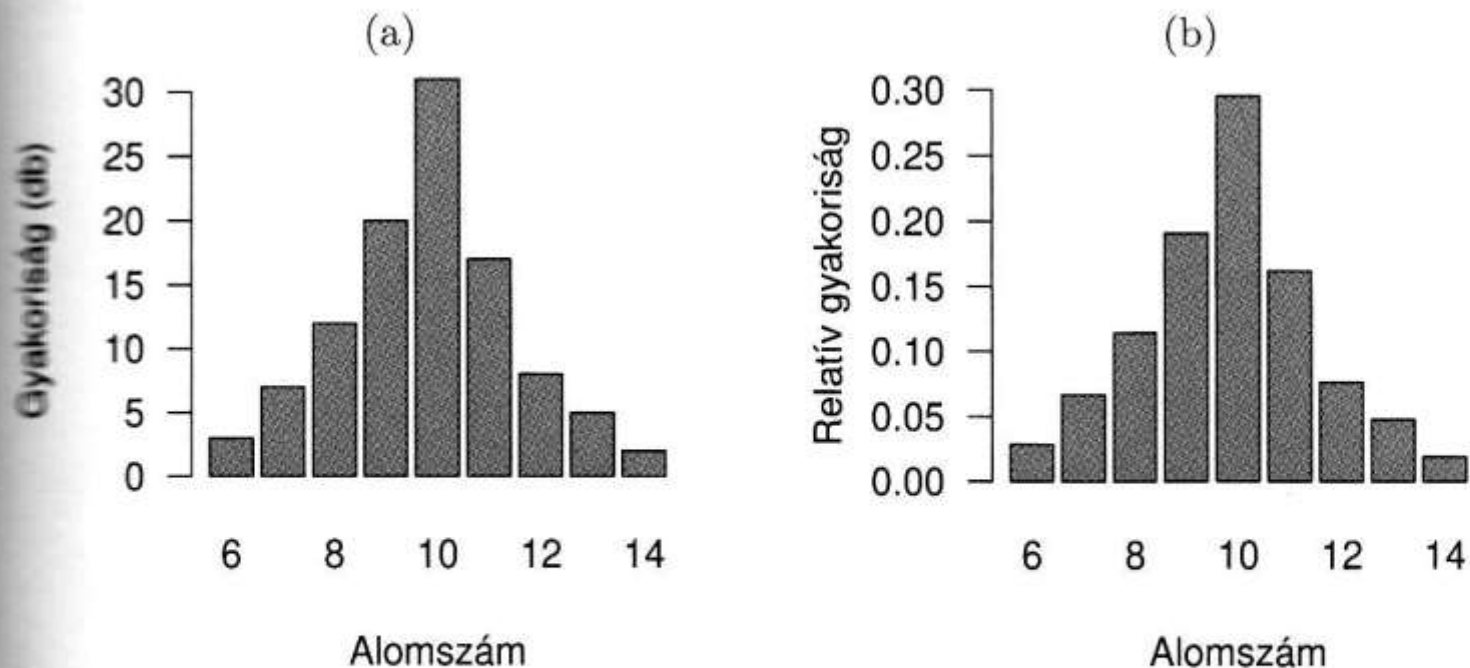
Anyakocák szaporaságát (születési alomszám) vizsgálták egy tenyészetben. Az egyes alomszám értékek gyakoriságát és relatív gyakoriságát tartalmazza a következő gyakoriságtáblázat:

alomszám születéskor	gyakoriság (db)	rel.gyak.	%
6	3	0.029	2.9
7	7	0.067	6.7
8	12	0.114	11.4
9	20	0.190	19.0
10	31	0.295	29.5
11	17	0.162	16.2
12	8	0.076	7.6
13	5	0.048	4.8
14	2	0.019	1.9
összesen:	105	1.000	100.0

# Leíró statisztika

## Változó ábrázolása

Nominális vagy ordinális (faktor) adatok :

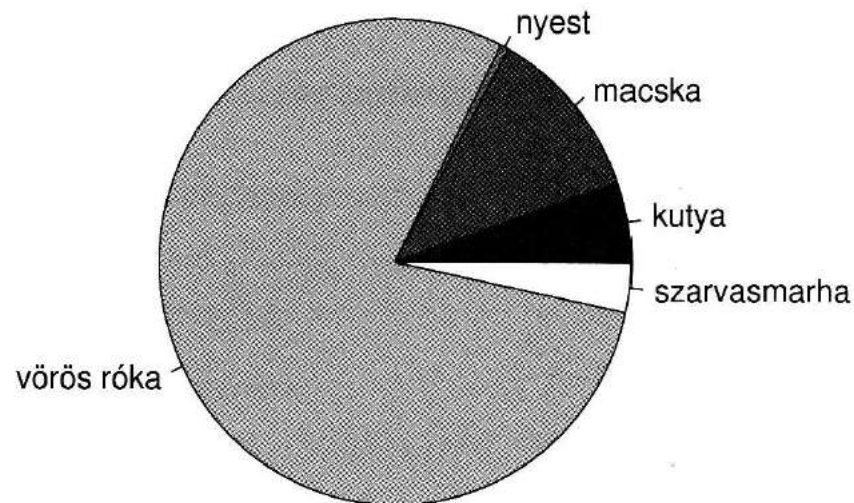


**41. ábra.** Anyakocák szaporaságának oszlopdiagramjai darabszámmal (a), illetve relatív gyakorisággal (b). Figyeljük meg, hogy a két oszlopdiagram egyforma, csak az  $y$  tengely skálázása különböző!

# Leíró statisztika

## Változó ábrázolása

Nominális vagy ordinális (faktor) adatok :



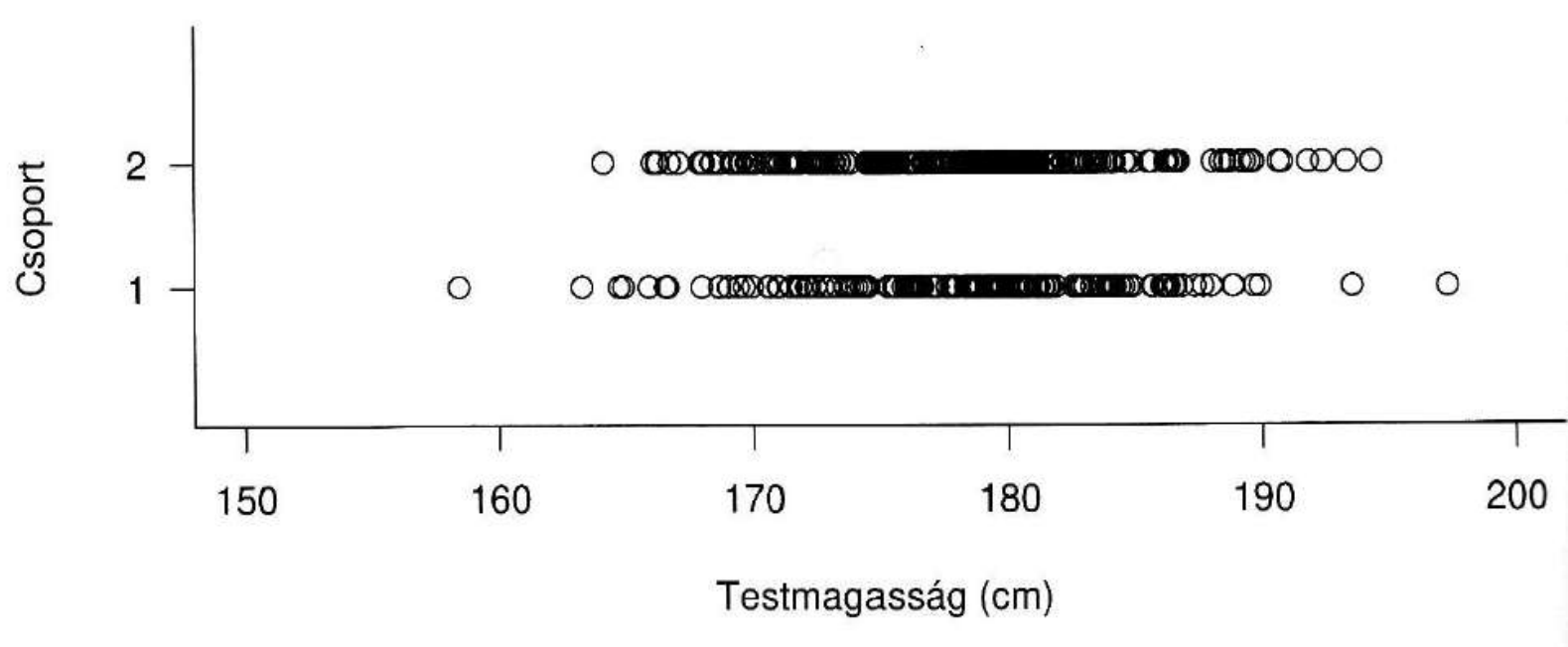
43. ábra. Hazánkban 1992-ben diagnosztizált veszettség esetek fajonkénti megoszlása



# Leíró statisztika

## Változó ábrázolása

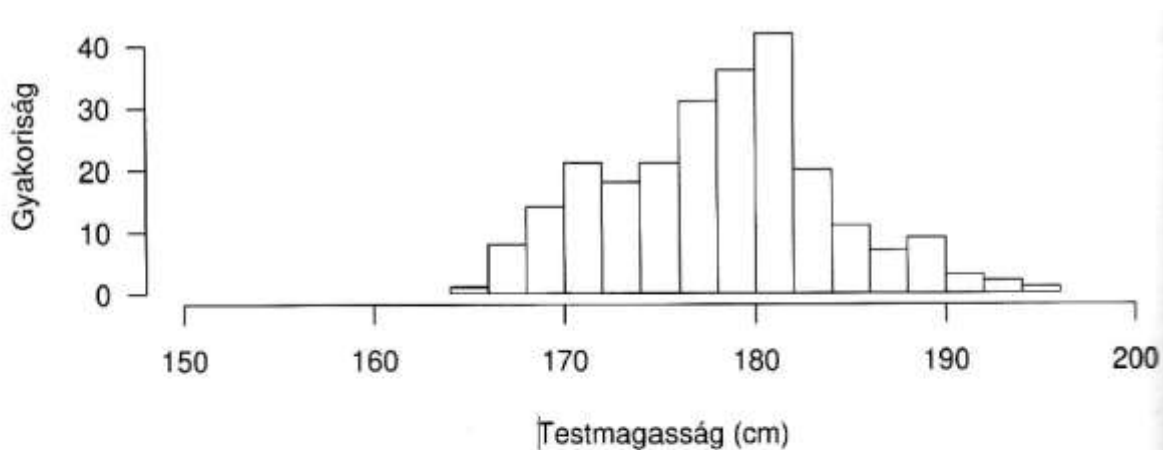
Számszerű (intervallum- vagy abszolút) adatok :



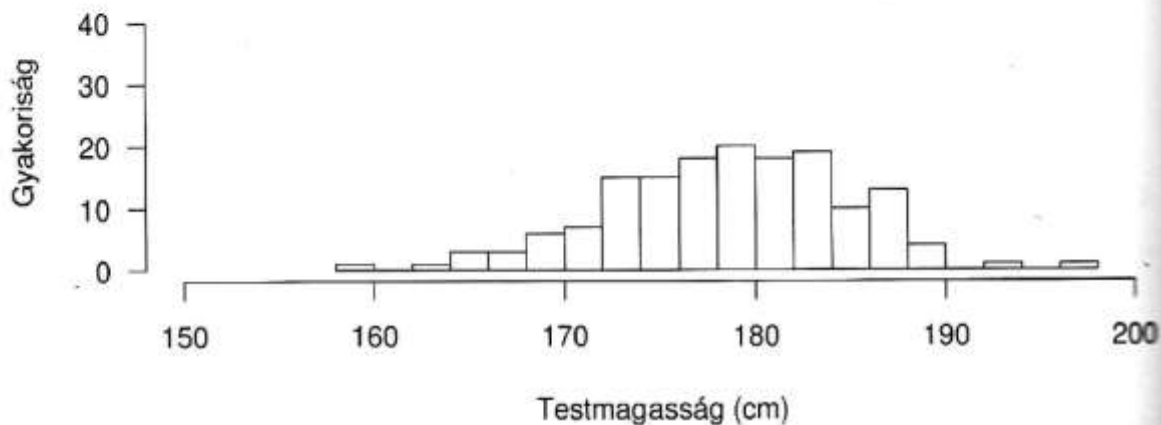
# Leíró statisztika

## Változó ábrázolása

Számszerű (intervallum- vagy abszolút) adatok :



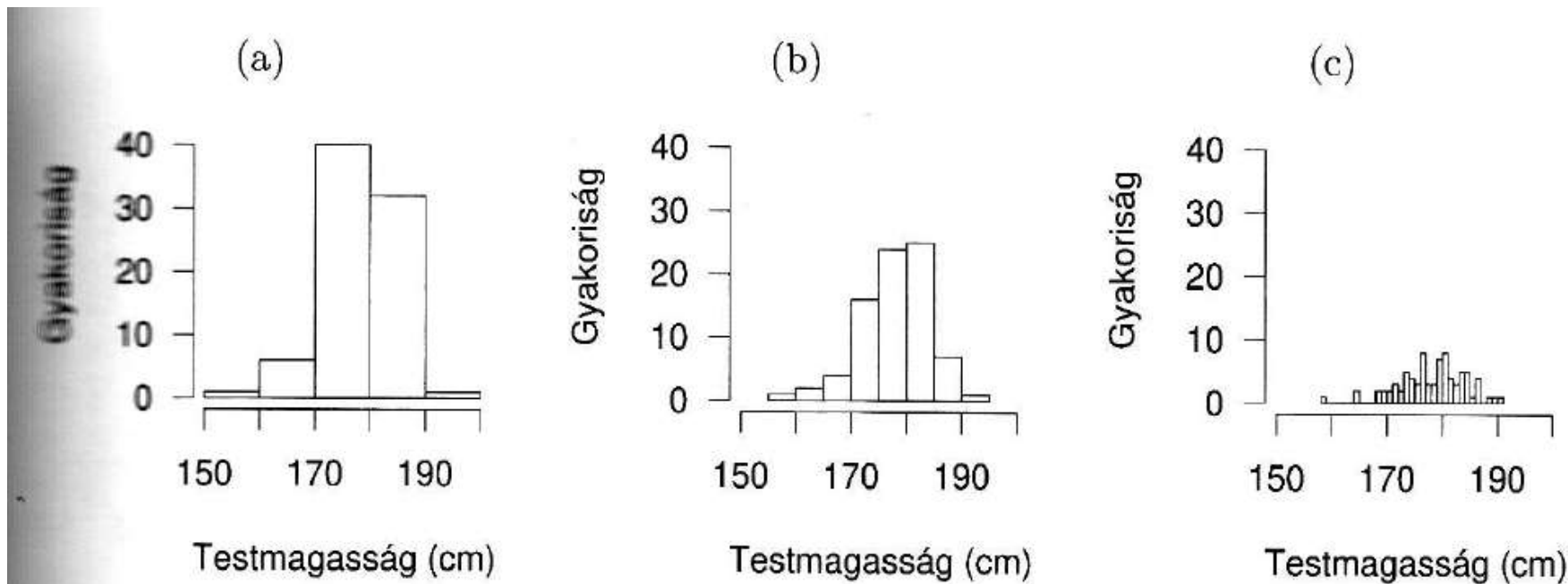
(c)



# Leíró statisztika

## Változó ábrázolása

Számszerű (intervallum- vagy abszolút) adatok :

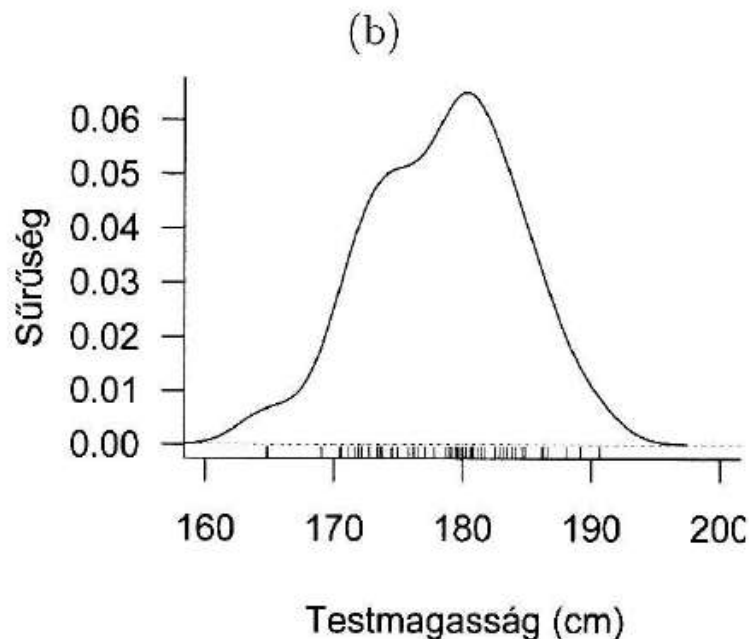
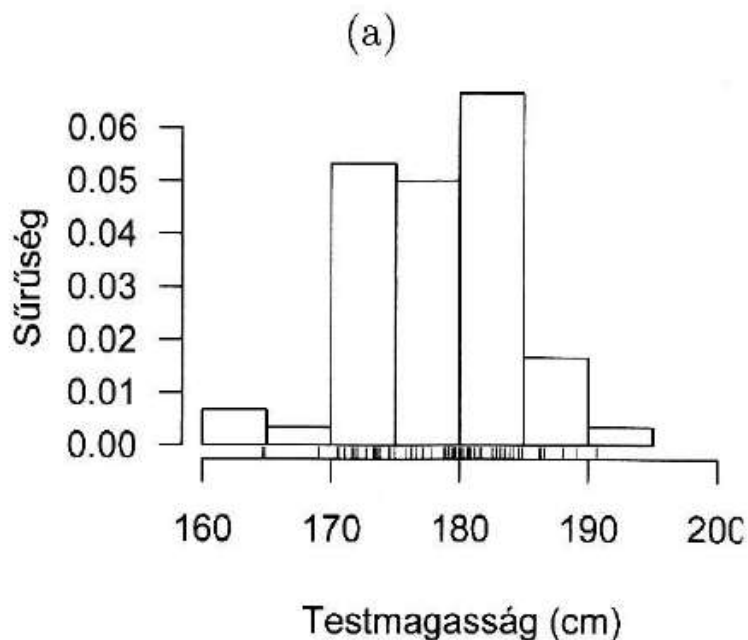


**4.8. ábra.** A hisztogramok ugyanazokat az adatokat (egyetemisták testmagassága) ábrázolják, csupán az osztályok száma különböző

# Leíró statisztika

## Változó ábrázolása

Számszerű (intervallum- vagy abszolút) adatok :

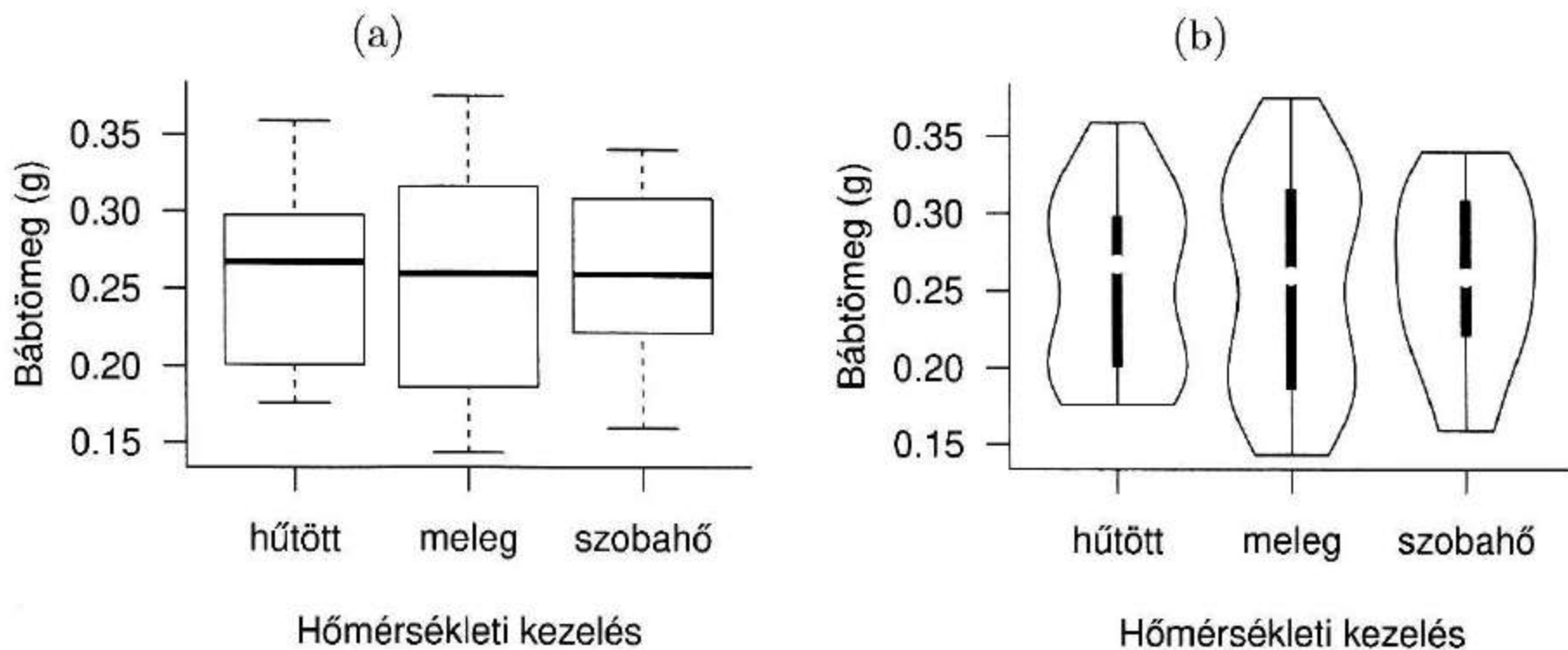


**4.7. ábra.** Hisztogram és simított hisztogram, az  $x$  tengely mentén az egyedi értékeket is ábrázolva a rug() függvénnyel

# Leíró statisztika

## Változó ábrázolása

Számszerű (intervallum- vagy abszolút) adatok :

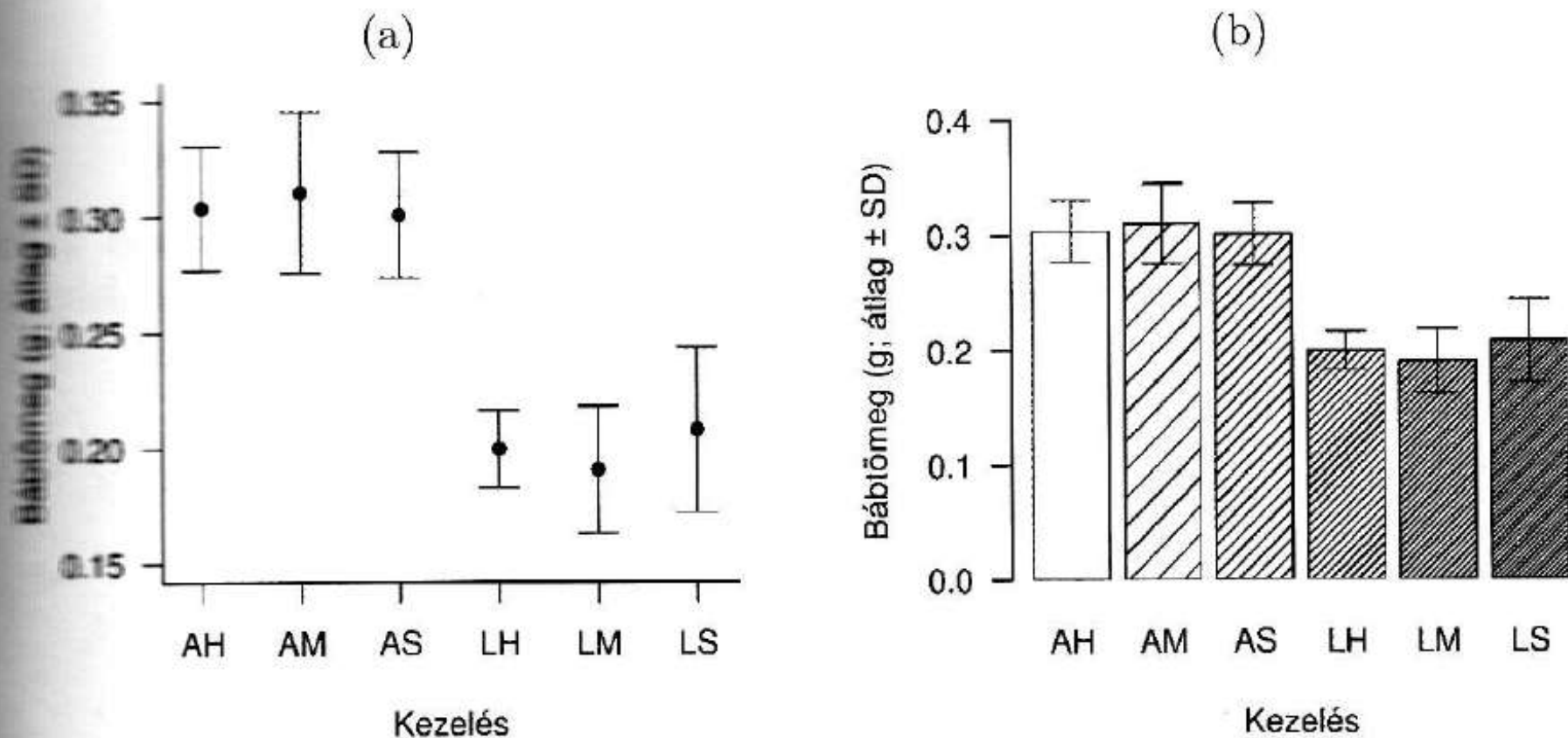


4.9. ábra. Boxplot (a) és hegedűábra (b) 3 csoportra

# Leíró statisztika

## Változó ábrázolása

Számszerű (intervallum- vagy abszolút) adatok :



10. ábra. Lepke bábtömegek átlaga és szórása a különböző kezelések esetén

# Becslés

## Alapfogalmak

Pontbecslés:

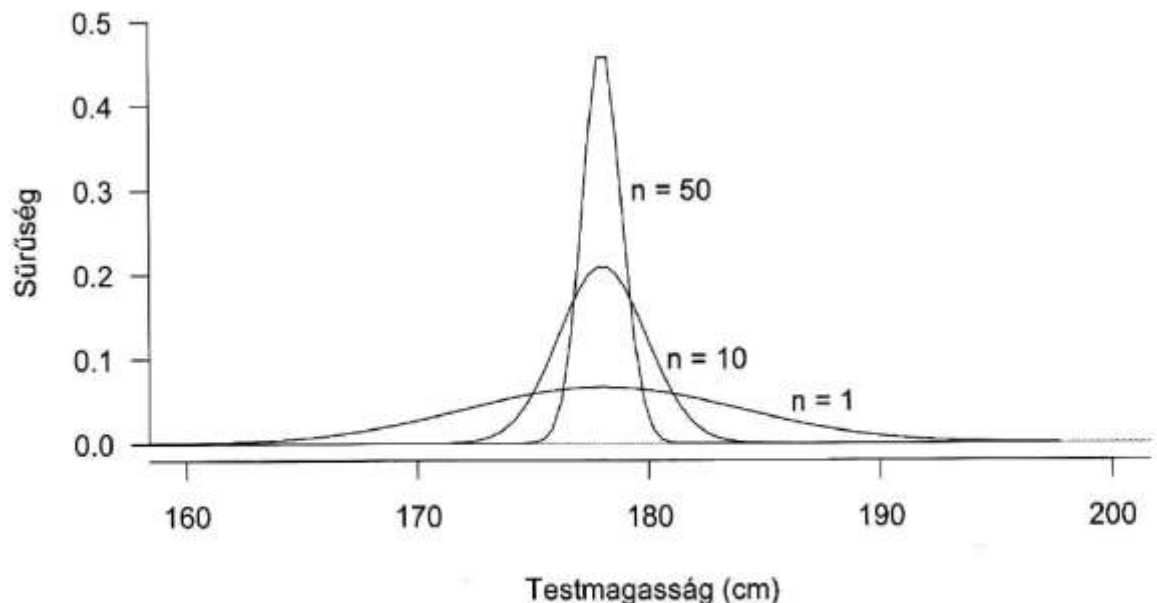
-Átlag

-Standard hiba (SE)

Átlag (SE)

*Átlag  $\pm$  (SD)*

$$SD / \sqrt{n}$$



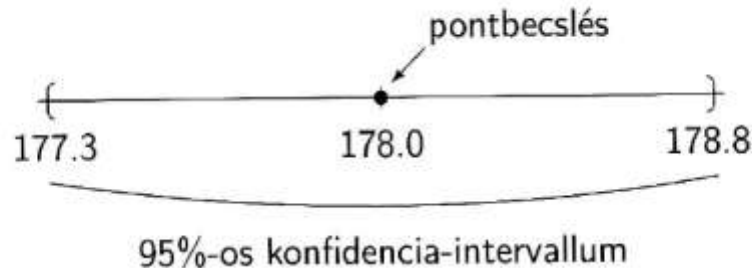
5.2. ábra. Különböző elemszámú véletlen mintákból számolt átlagos testmagasságok sűrűségfüggvénye ( $n = 1$ -re a testmagasság populációbeli sűrűségfüggvényét kapjuk)

# Becslés

## Alapfogalmak

Intervallumbecslés:

- Konfidencia-intervallum (95%)



**5.3. ábra.** Pontbecslés és 95%-os konfidencia-intervallum a fenti mintából a testmagasság populációbeli átlagára



**5.5. ábra.** Pontbecslés és konfidencia-intervallum az esélyhányadosra ugyanabból a mintából (az ábra csak illusztráció, nem kapcsolódik egyik ismertetett példához sem). A pontbecslés  $OR = 3.71$ , a 95%-os konfidencia-intervallum pedig: (0.35, 188.7). Figyeljük meg, hogy a végpontok nem a pontbecslésre szimmetrikusan helyezkednek el!



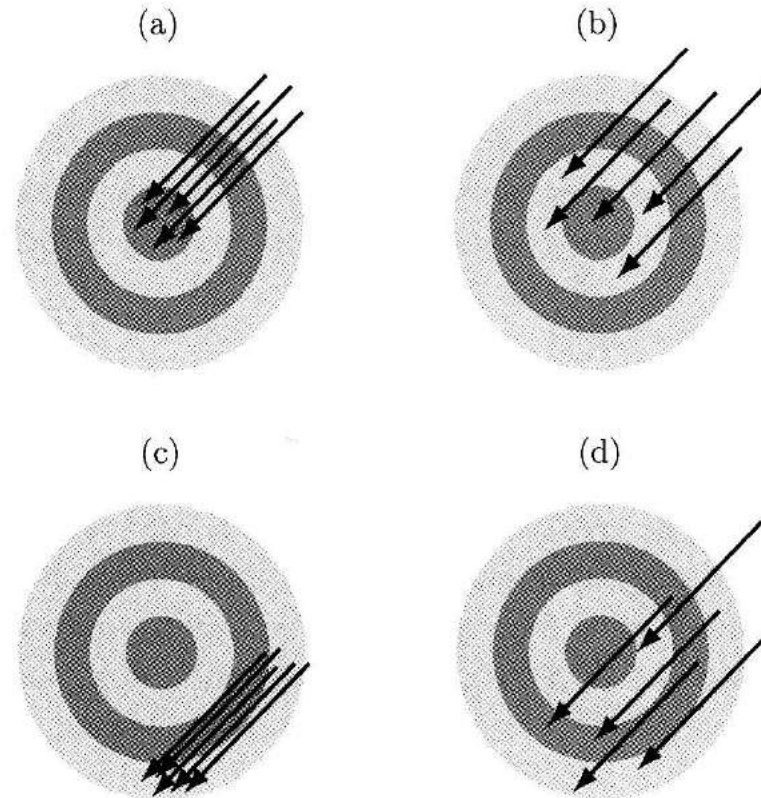
# Becslés

## Alapfogalmak

Becslés pontossága:

-Torzítás (unbiased)

-Szórás



**5.9. ábra.** Pontosság, szórás és torzítás, becslés helyett lövésekkel szemléltetve. A becslendő paraméternek (a valódi értéknek) a céltábla középpontja felel meg. Egy becslés lehet kis szórású és torzítatlan (a), torzítatlan, de nagy szórású (b), kis szórású, de torzított (c), nagy szórású és torzított (d). Pontosnak csak az elsőt nevezhetjük, de a pontatlanokról is jó tudni, hogy melyik típusba tartoznak


# R statisztikai programcsomag

## R commander (Rcmdr) „Grafikus Felhasználói Felület” alkalmazása

Installálni kell az R statisztikai programcsomagot a PC-re:

<https://www.r-project.org/>

a



**The R Project for Statistical Computing**

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To **download R**, please choose your preferred CRAN mirror.

[Download CRAN](#)

The Comprehensive R Archive Network is available at the

If you want to host a new mirror at your institution, please

0-Cloud  
<https://cloud.r-project.org/>

Argentina  
<http://mirror.fcaglp.unlp.edu.ar/CRAN/>

Australia

**Download and Install R**

Precompiled binary distributions of the base system and contribu  
of R:

- [Download R for Linux](#) ([Debian](#), [Fedora/Redhat](#), [Ubuntu](#))
- [Download R for macOS](#)
- [Download R for Windows](#)

b

c

# R statisztikai programcsomag

R commander (Rcmdr) „Grafikus Felhasználói Felület” alkalmazása

Rcmdr csomagot installálni kell az R-ben:



az R indítása után ki kell adni az alábbi parancsot:

```
install.packages("Rcmdr")
```

R Commander indítása az alábbi paranccsal:

```
library(Rcmdr)
```

# Első gyakorlat során végzett feladatok

Vizsgált adatok: elseeves.txt, lile.txt

-Adatok beolvasása:

Data->Import data->from text file...

-Adat átkódolás:

Data->Manage variables in active data set-> Convert numeric variables to factors

-Számított változók számítása:

Data->Manage variables in active data set->Compute New Variable

Ábrázolás:

Graphs

Histogram

Density estimate

Boxplot

Scatterplot

# Gyakorlat során végzett feladatok

Statisztika elemzések:

Statistics

Summaries

Numerical summaries

Shapiro-Wilk test of normality

Means

Independent samples t-test

Variances

Two-variances F-test

Nonparametric tests

Two-sample Wilcoxon test