

Adatfeldolgozási alapismeretek

2. rész

Helyzetparaméter becslése

Helyzetparaméter

- Leggyakrabban használt statisztikák:
 - átlag(ok)
 - medián (középső érték)
 - módusz (divatos, leggyakoribb érték)

Számtani átlag

- Adatok száma: n
- Adatok: x_1, x_2, \dots, x_n
- A számtani átlag:

$$\bar{x} = \text{ave}_{i=1}^n(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n x_i$$

Számítani átlag: Példa

$$\bar{x} = \text{ave}_{i=1}^5(\mathbf{x}) = \frac{67 + 109 + 28 + 312 + 84}{5} = 120$$

Medián

- A medián a *középső* érték
- Számolásmenet:
- Ha n páratlan
medián = a középső érték
- Ha n páros
medián = a két középső érték átlaga

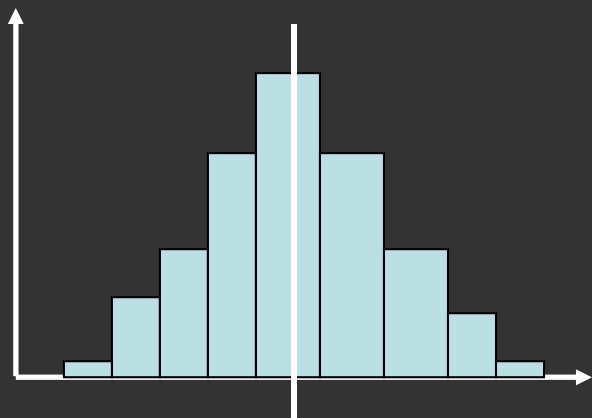
Példa

- Magok száma: 67, 109, 28, 412, 84
- Átlag: $(67+109+28+312+84)/5 = 120$
- Medián: 84

28 67 84 109 312

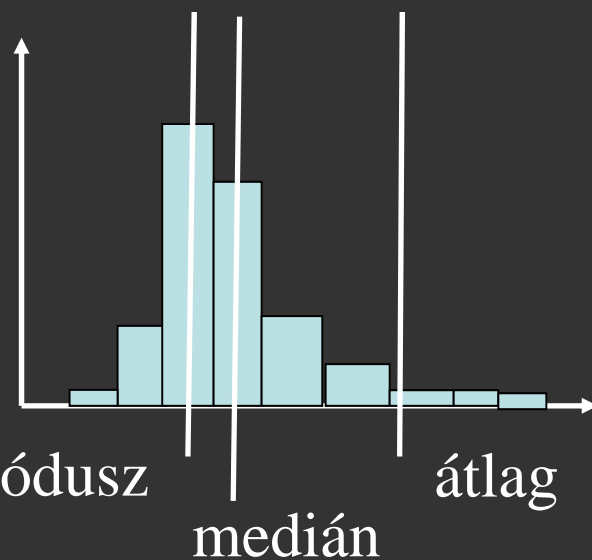
Módusz

- A leggyakoribb érték
- Nem mindig létezik
- Bimodális eloszlás esetén értéke nem egyértelmű



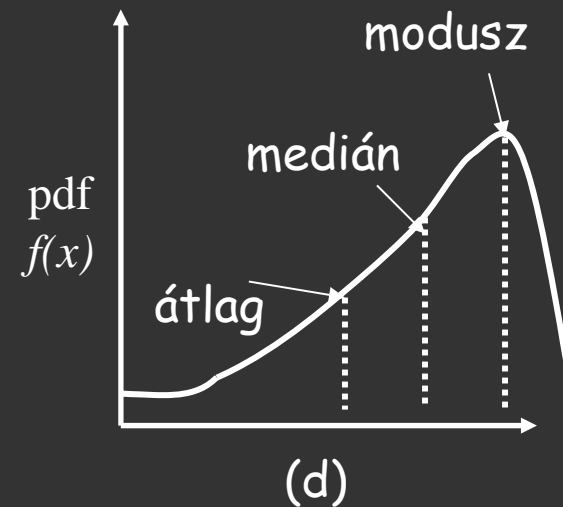
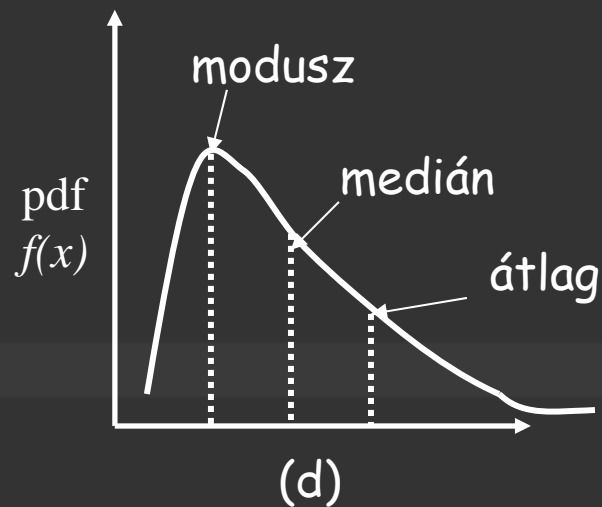
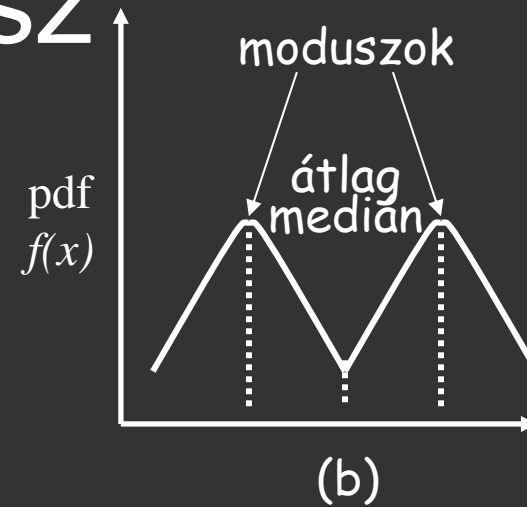
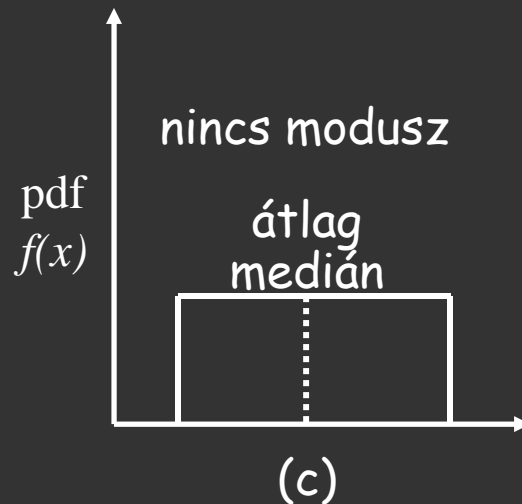
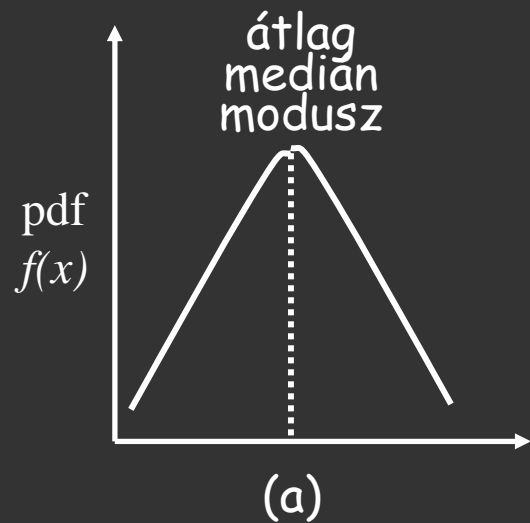
szimmetrikus eloszlás

átlag, medián, módusz egybeesik



aszimmetrikus eloszlás

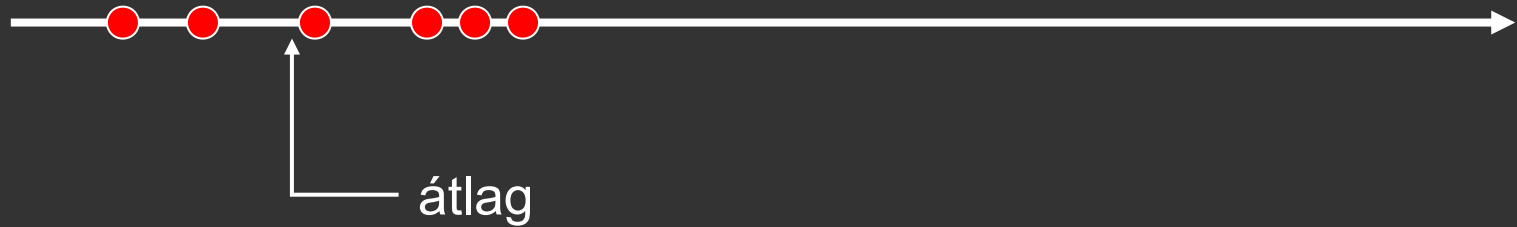
Átlag, medián és módusz



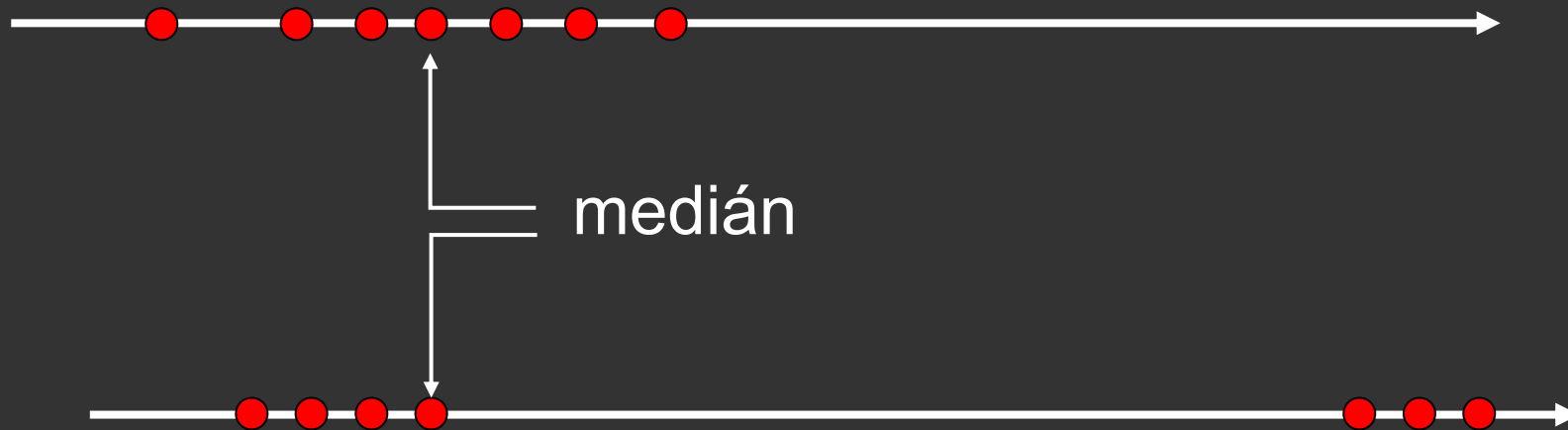
Átlag

- Egyetlen kilógó érték teljesen tönkretelheti az eredményt
- Ha egyetlen adatot növelése minden határon túl növeli az átlag értékét is
- Ferde (vagy bimodális) eloszlások esetén az átlag teljesen félrevezető lehet

Az átlag szenzitivitása a kilógó adatokra



A medián „vak” a többi adat tényleges értékét illetően





Még további átlagok!

- Számítani
- Harmonikus
- Geometriai
- Négyzetes



Számtani átlag

$$\overline{x}_A = \frac{1}{n} \sum_{i=1}^n x_i$$

Harmonikus átlag

$$\overline{x}_H = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

Geometriai átlag

$$\begin{aligned}\overline{x}_G &= \sqrt[n]{x_1 x_2 \cdots x_i \cdots x_n} \\ &= \left(\prod_{i=1}^n x_i \right)^{1/n}\end{aligned}$$

Súlyozott átlagok

$$\sum_{i=1}^n w_i = 1$$

$$\bar{x}_A = \sum_{i=1}^n w_i x_i$$

$$\bar{x}_H = \frac{1}{\sum_{i=1}^n \frac{w_i}{x_i}}$$

- Az összes adat egyformán fontos
- Súlyok az egyes adatok relatív fontosságát jelentik

Határozókulcs

- Kategóriális adataink vannak?
 - IGEN, használjuk a *módot*
 - példa: cipőméret
- Ferde az eloszlás
 - IGEN, használjuk a *mediánt*
 - Fajok egyedszámai a közösségben
- A fentiek egyike sem áll fenn
 - használjuk az *átlagot*