

## Egy ismerv szerinti elemzés

Az adatsor leggyakoribb értéke a [módusz](#). Hogyha például Bob matekjegyei ezek:

2, 3, 1, 4, 1, 2, 2, 3, 5, 2, 3, 2, 3, 2, 4, 3, 2, 4, 2, 4

Akkor egyszerűen meg kell számolni, hogy melyikből van a legtöbb, és az a matekjegy lesz a [módusz](#). Most 2-esből van a legtöbb, így Bob matekjegyeinek a módusza 2. A [módusz](#) jele  $Mo$  és így most  $Mo=2$ .

Léteznek olyan eloszlások is, amelyeknek több módusza van. Hogyha például Bob jegyei:

1, 2, 2, 3, 5, 3, 3, 4, 2

Itt 2-esből és 3-asból ugyanannyi van, mindkettőből 3 darab. Ez egy kétmódusú [eloszlás](#).

[Megnézem a kapcsolódó epizódot](#)

A [medián](#) a növekvő sorba rendezett adatsor középső értéke. Ha az adatsorban páros sok elem van, akkor nincs középső elem, ilyenkor a két középső elem átlagát vesszük.

Hogyha például Bob matekjegyei ezek:

2, 3, 1, 4, 1, 2, 2, 3, 5, 2, 3, 2, 3, 2, 4, 3, 2, 4, 2, 4

Akkor egyszerűen növekvő sorba kell rakni..

1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 4, 4, 4, 4, 5

És aztán meg kell keresni melyik a középső. Most nincsen középső, mert páros sok elem van, így ilyenkor a két középen lévő átlagoljuk:

1, 1, 2, 2, 2, 2, 2, 2, 2, **2, 3**, 3, 3, 3, 3, 4, 4, 4, 4, 5

Ezeknek az átlaga 2,5 vagyis a [medián](#) most 2,5. A [medián](#) jele  $Me$ , így  $Me=2,5$

[Megnézem a kapcsolódó epizódot](#)

Az átlagot úgy kapjuk meg, hogy az összes elemet összeadjuk, és aztán elosztjuk az elemek számával.

Jele:  $\bar{x}$

[Megnézem a kapcsolódó epizódot](#)

Az átlagtól való átlagos eltérés egyik legjobb mérőszáma a szórás. Hátránya, hogy egy kicsit ronda a szórás képlete. A szórást egy szigma nevű görög betűvel jelöljük.

$$\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}}$$

[Megnézem a kapcsolódó epizódot](#)

Az adatsor első felének a felezőpontja az alsó kvartilis.

Az alsó kvartilis jele:  $Q_1$

[Megnézem a kapcsolódó epizódot](#)

Az adatsor második felének a felezőpontja a felső kvartilis.

A felső kvartilis jele:  $Q_3$

[Megnézem a kapcsolódó epizódot](#)

A kvartilisek és a [medián](#) azt szemlélteti, hogyan oszlanak el az adatsorban szereplő adatok. Ezek segítségével készíthető el a doboz-ábra, vagy másnéven dobozdiagram. Szokás még sodrófa diagramnak is nevezni, és az angol elnevezést is gyakran használják, ami a box plot.

Egy sobarendezett adatsorban öt darab speciális negyedelőpontot fogunk használni. Az első az adatsor legkisebb értéke, ez a  $Q_0$ . Aztán a következő negyedelő az alsó kvartilis, ami  $Q_1$  utána jön a felezőpont vagyis a [medián](#), ezt  $Me$ -vel és  $Q_2$ -vel is jelöljük, végül a felső kvartilis, ami a  $Q_3$ . Az adatsor legnagyobb értéke pedig  $Q_4$ . A legnagyobb és a legkisebb érték különbsége a terjedelem, míg a két kvartilis különbségét félterjedelemnek vagy más néven interkvartilisnek hívjuk. Ezekből épül föl a doboz-ábra vagy másként dobozdiagram.

Előfordulhat, hogy az adatsorban kiugró értékek is szerepelnek. A kiugró érték az, ami az alsó kvartilisnél legalább a félterjedelem másfélszeresénél kisebb, vagy pedig a felső kvartilisnél legalább a félterjedelem másfélszeresénél nagyobb. Huh, ez elég bonyolult hangzik. De valójában nagyon egyszerű, csak nézd meg kapcsolódó epizódot és kiderül.

[Megnézem a kapcsolódó epizódot](#)

A relatív szórás azt mondja meg, hogy a szórás az átlagnak hány százaléka:

$$V = \frac{\sigma}{\bar{X}}$$

[Megnézem a kapcsolódó epizódot](#)

$$Mo = mo + \frac{k_1}{k_1 + k_2} \cdot h_{mo}$$

A képletben  $mo$  a nyers [módusz](#), ami a legnagyobb gyakoriságú osztály alsó határa. A  $k_1$ -et úgy kapjuk, ha ennek az osztályköznek a gyakoriságából kivonjuk az előtte lévő osztályköz gyakoriságát. A  $k_2$ -t pedig úgy kapjuk, ha ennek az osztályköznek a gyakoriságából az utána lévő osztályköz gyakoriságát vonjuk le. A  $h_{mo}$  pedig ennek az osztályköznek a hosszát jelöli.

[Megnézem a kapcsolódó epizódot](#)

$$Me = me + \frac{\frac{N}{2} - f'_{me-1}}{f_{me}} \cdot h_{me}$$

Itt  $f'$  a kumulált gyakoriság.  $me$  a mediánt tartalmazó osztályköz eleje,  $f_{me}$  a mediánt tartalmazó osztályköz gyakorisága,  $f'_{me-1}$  a [medián](#) előtti osztályköz kumulált gyakorisága,  $h_{me}$  pedig a mediánt tartalmazó osztályköz hossza.

[Megnézem a kapcsolódó epizódot](#)

$$Q_{\frac{k}{m}} = a_i + \frac{\frac{k}{m} \cdot N - f'_{i-1}}{f_i} \cdot h_i$$

Az alsó kvartilis:

$$Q_{\frac{1}{4}} = a_i + \frac{\frac{1}{4} \cdot N - f'_{i-1}}{f_i} \cdot h_i$$

A felső kvartilis:

$$Q_{\frac{3}{4}} = a_i + \frac{\frac{3}{4} \cdot N - f'_{i-1}}{f_i} \cdot h_i$$

[Megnézem a kapcsolódó epizódot](#)

A relatív gyakoriság jele  $g_i$ , és úgy kell kiszámolni, hogy a gyakoriságot osztjuk az összes elemszámmal:

$$g_i = \frac{f_i}{N}$$

[Megnézem a kapcsolódó epizódot](#)

Az értékösszeg jele  $S_i$  és úgy kapjuk meg, hogy az osztályközepeket megszorozzuk a gyakorisággal.

$$S_i = X_i \cdot f_i$$

[Megnézem a kapcsolódó epizódot](#)

A Herfindahl-index egy eszköz a [koncentráció](#) vizsgálatára.

$$HI = \sum Z_i^2$$

A Herfindahl-index  $1/N$  és  $1$  között vesz fel értékeket és minél közelebb van az  $1$ -hez, annál nagyobb a [koncentráció](#).

[Megnézem a kapcsolódó epizódot](#)

A Lorenz-görbe egy eszköz a [koncentráció](#) vizsgálatára.

A Lorenz-görbe az úgynevezett koncentrációs területtel szemlélteti a [koncentráció](#) mértékét.

Minél nagyobb ez a terület, a [koncentráció](#) annál erősebb.

Olyankor pedig, amikor a [Lorenz görbe](#) egybeesik a négyzet átlójával, a [koncentráció](#) nulla.

[Megnézem a kapcsolódó epizódot](#)

Az [alakmutatók](#) arról szólnak, hogy az [eloszlás](#) mennyire asszimmetrikus.

Az egyik legegyszerűbb és leggyakrabban használt [alakmutatók](#), az úgynevezett Pearson-féle mérőszámok:

$$P = 3 \frac{\bar{X} - Me}{\sigma} \quad A = \frac{\bar{X} - Mo}{\sigma}$$

A negatív értékek jobb oldali asszimetriát jelentenek. A pozitív értékek esetén pedig bal oldali asszimetria van.

A Pearson-féle P és A mutatók általában -1 és 1 között tartózkodnak és csak extrém esetekben vesznek föl 1-nél nagyobb vagy -1-nél kisebb értéket.

[Megnézem a kapcsolódó epizódot](#)

Az [alakmutatók](#) arról szólnak, hogy az [eloszlás](#) mennyire asszimmetrikus.

Az egyik legegyszerűbb és leggyakrabban használt [alakmutatók](#), az úgynevezett Pearson-féle mérőszámok mellett az F-mutatók:

$$F_{0,25} = \frac{(Q_3 - Me) - (Me - Q_1)}{(Q_3 - Me) + (Me - Q_1)}$$

$$F_{0,1} = \frac{(D_9 - Me) - (Me - D_1)}{(D_9 - Me) + (Me - D_1)}$$

ahol  $D_1$  az első,  $D_9$  pedig a kilencedik decilist jelenti.

A negatív értékek jobb oldali asszimetriát jelentenek. A pozitív értékek esetén pedig bal oldali asszimetria van.

Az F mutató csak -1 és 1 között lehet.

[Megnézem a kapcsolódó epizódot](#)

A csúcosság azt jelenti, hogy az [eloszlás](#) görbéje mennyire csúcsosodik ki.

A csúcosság mérésére a következő mutató van forgalomban:

$$\alpha_4 = \frac{M_4(\bar{X})}{\sigma^4} - 3$$

Itt  $M_4(\bar{X})$  az úgynevezett negyedik momentum, és így számolható ki:

$$M_4(\bar{X}) = \frac{\sum (\bar{X} - x_i)^4}{N}$$

[Megnézem a kapcsolódó epizódot](#)