

SZÁMSOKASÁGOK STATISZTIKAI JELLEMZŐI

ELMÉLET

Egy véges számsokaság legfontosabb **statisztikai jellemzői**:

	Leírása
Átlag	a sokaság elemeinek az összegét elosztjuk az elemek számával
Módusz	a leggyakrabban előforduló elem (lehet több módusz is)
Medián	a nagyság szerinti felsorolásban páratlan elemszám esetén a középen álló szám, páros elemszám esetén a középen álló két szám átlaga
Terjedelem	a legnagyobb és a legkisebb szám különbsége

Ha az adatok gyakorisága ismert, akkor az átlag kiszámolható úgy, hogy a számokat megszorozzuk a gyakoriságukkal, ezeknek a szorzatoknak az összegét elosztjuk a gyakoriságok összegével. (Ezt az eljárást nevezik **súlyozott számtani közép** számolásának.)

BEVEZETŐ FELADAT:

Egy cégnél 2001-ben 20-an dolgoztak és 5-féle fizetést kaptak. Az igazgató 300 ezer forintot, 5 főmunkatárs 250 ezret, 5 tisztviselő 200 ezret, 6 adminisztrátor 120 ezret, 3 részmunkaidőben dolgozó takarító 80 ezret keresett havonta. Mit olvashatunk ki ezekből az adatokból?

Megoldás

A fizetések jegyzéke 20 adatot tartalmaz.

Ezek az adatok (ezer forintokban kifejezve) növekvő sorrendben a következők:

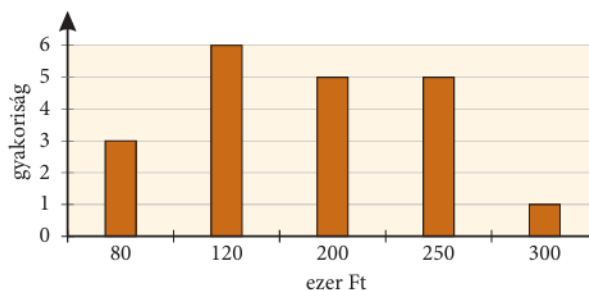
80, 80, 80, 120, 120, 120, 120, 120, 120, 120, 200, 200, 200, 200, 200, 250, 250, 250, 250, 300.

Leggyakrabban a 120 fordul elő, ez a számsokaság **módusza**.

A felsorolásban középen, a 9. és a 10. helyen 200 áll, ezeknek az átlaga szintén 200, ami a számsokaság **mediánja**.

A fizetéseket gyakorisági grafikonon (diagramon) is ábrázolhatjuk.

A legmagasabb oszlop a 120-hoz tartozik, a legalacsonyabb a 300-hoz.



A számsokaság **terjedelme** a legnagyobb és a legkisebb szám különbsége: $300 - 80 = 220$.

A terjedelem a grafikon vízszintes tengelyén olvasható le.

Számítsuk ki a fizetések **átlagát**! Az átlagot úgy kapjuk meg, hogy a 20 szám összegét elosztjuk 20-szal. Mivel sok egyenlő adat van, érdemes súlyozott átlagot számolni: $(300 + 5 \cdot 250 + 5 \cdot 200 + 6 \cdot 120 + 3 \cdot 80) : 20 = 3510 : 20 = 175,5$.

A 20 dolgozó fizetésének átlaga 175 500 forint. 11-en ennél többet keresnek, 9-en ennél kevesebbet.

Az átlagtól kevéssel tér el az 5 tisztviselő fizetése, de az igazgató és a takarítók fizetése nagyon messze van az átlagtól. Azt is mondhatjuk, hogy nagy a fizetések **szóródása**.

1. Add meg a következő számsokaságok főbb statisztikai jellemzőit!

a) 8, 9, 3, 5, 9, 4

b) 80, 90, 30, 50, 90, 40

a) Számsokaság: 8, 9, 3, 5, 9, 4

Először érdemes sorba rendezni a számokat: **3, 4, 5, 8, 9, 9**

- **Átlag ($\bar{x}$):** Az összeget elosztjuk a darabszámmal (6).

$$\frac{3 + 4 + 5 + 8 + 9 + 9}{6} = \frac{38}{6} \approx \mathbf{6,33}$$

- **Módusz:** A leggyakrabban előforduló elem.
Ez a **9** (kétszer szerepel).
- **Medián:** A sorba rendezett adatok középső értéke. Mivel páros számú elemünk van, a két középső (5 és 8) átlaga:

$$\frac{5 + 8}{2} = \mathbf{6,5}$$

- **Terjedelem:** A legnagyobb és a legkisebb érték különbsége.
 $9 - 3 = \mathbf{6}$

b) Számsokaság: 80, 90, 30, 50, 90, 40

Sorba rendezve: **30, 40, 50, 80, 90, 90**

Megfigyelhető, hogy ezek pontosan az "a" feladat számainak tízszeresei, így a jellemzők is eszerint alakulnak:

- **Átlag ($\bar{x}$):**

$$\frac{30 + 40 + 50 + 80 + 90 + 90}{6} = \frac{380}{6} \approx \mathbf{63,33}$$

- **Módusz:** A leggyakoribb a **90**.
- **Medián:** Az 50 és a 80 átlaga:

$$\frac{50 + 80}{2} = \mathbf{65}$$

- **Terjedelem:**
 $90 - 30 = \mathbf{60}$

2 Egy 30 számból álló adathalmaznak ismerjük a terjedelmét, átlagát, mediánját és móduszát. A négy közül melyik az a statisztikai jellemző, amelyiknek az értéke biztosan szerepel az adathalmaz elemei között?

1. Módusz (Ami biztosan benne van)

A módusz definíció szerint a számsokaság **leggyakrabban előforduló eleme**. Ahhoz, hogy valami a leggyakoribb legyen, elengedhetetlen, hogy legalább egyszer (de inkább többször) szerepeljen az adatok között. Így a módusz értéke 100%-os biztonsággal eleme a halmaznak.

2. Átlag (Ami nem biztos)

Az átlag egy számított érték ($\frac{\text{összeg}}{\text{darabszám}}$). Gyakran előfordul, hogy az eredmény egy olyan tört szám, amely egyáltalán nem szerepel az adatok között.

- *Példa:* Ha van két számunk, a 2 és a 4, az átlaguk 3, de a 3 nincs benne az adathalmazban.

3. Medián (Ami nem biztos)

A medián a sorba rendezett adatok középső értéke.

- Ha az adatok száma **páratlan**, akkor a medián a középső elem, tehát benne van a halmazban.
- Mivel azonban a feladatban **30 számról** (páros számú adat) van szó, a mediánt a két középső elem **átlagaként** számoljuk ki. Ez az átlag pedig lehet olyan szám is, ami nem szerepel az eredeti listában.

4. Terjedelem (Ami nem biztos)

A terjedelem a legnagyobb és a legkisebb érték **különbsége** ($X_{max} - X_{min}$). Ez egy távolságot jelöl, nem pedig egy konkrét adatpontot.

- *Példa:* Ha a legkisebb szám 10, a legnagyobb 15, a terjedelem 5. Nem biztos, hogy az 5-ös szám szerepel az adatok között.

3  Adott négy szám: 5, -2 ; -1 ; 0. Adj meg egy ötödik számot úgy, hogy

- a)** az öt szám mediánja 0 legyen!
- b)** az öt szám átlaga 0 legyen!
- c)** az öt szám mediánja és egyetlen módusza megegyezzen!
- d)** az öt szám átlaga és mediánja megegyezzen!
- e)** az öt szám átlaga és egyetlen módusza megegyezzen!

Jelöljük az ötödik, ismeretlen számot x -szel.

a) Az öt szám mediánja 0 legyen

Mivel öt számunk van (páratlan darabszám), a medián a sorba rendezett adatok pontosan **középső (3.) eleme** lesz.

A jelenlegi sorrend: -2 ; -1 ; **0**; 5. Ahhoz, hogy a 0 maradjon középen, az x értékének a 0-val megegyezőnek vagy annál nagyobbabbnak kell lennie.

- **Megoldás:** Bármely $x \geq 0$ szám megfelel. Például: $x = 1$. (A sorrend így: -2 ; -1 ; 0; 1; 5 $\rightarrow$ a középső a 0.)

b) Az öt szám átlaga 0 legyen

Az átlag akkor 0, ha a számok **összege is 0**.

Adjuk össze a meglévőket: $-2 + (-1) + 0 + 5 = 2$.

Ahhoz, hogy az összeg 0 legyen, $2 + x = 0$ kell, hogy teljesüljön.

- **Megoldás:** $x = -2$.

c) Az öt szám mediánja és egyetlen módusza megegyezzen

Itt két feltételnek kell egyszerre megfelelni. Próbáljuk meg elérni, hogy mindkettő 0 legyen!

1. **Módusz:** Ahhoz, hogy a 0 legyen az egyetlen módusz (a leggyakoribb elem), az x értéke legyen **0**. Így a 0 kétszer szerepel, a többi csak egyszer.
2. **Medián:** Ha $x = 0$, a rendezett sor: -2 ; -1 ; 0; 0; 5. A középső elem a 0.

- **Megoldás:** $x = 0$.

d) Az öt szám átlaga és mediánja megegyezzen

Ez egy kicsit trükkösebb, mert x helyzete változtatja a mediánt.

Próbáljuk megint a 0-t megcélozni közös értékként!

1. Láttuk a b) pontban, hogy az átlag akkor 0, ha $x = -2$.
2. Ellenőrizzük a mediánt, ha $x = -2$: A sorrend $-2; -2; -1; 0; 5$. A középső elem a **-1**. (Ez nem jó, $0 \neq -1$).

Próbáljuk meg, hogy a közös érték a **-1** legyen!

1. Átlag legyen -1: $\frac{2+x}{5} = -1 \rightarrow 2+x = -5 \rightarrow x = -7$.
2. Ellenőrizzük a mediánt, ha $x = -7$: A sorrend $-7; -2; -1; 0; 5$. A középső elem a **-1**. (Ez tökéletes!)

- **Megoldás:** $x = -7$.

d) nincs megoldás

Vegyük sorra a lehetőségeket az átlag képletével: $\text{Átlag} = \frac{-2+(-1)+0+5+x}{5} = \frac{2+x}{5}$

1. **Ha a módusz -2:** Akkor $x = -2$. Az átlag ekkor: $\frac{2+(-2)}{5} = 0$. (A $0 \neq -2$, tehát nem jó).
2. **Ha a módusz -1:** Akkor $x = -1$. Az átlag ekkor: $\frac{2+(-1)}{5} = 0,2$. (A $0,2 \neq -1$, nem jó).
3. **Ha a módusz 0:** Akkor $x = 0$. Az átlag ekkor: $\frac{2+0}{5} = 0,4$. (A $0,4 \neq 0$, nem jó).
4. **Ha a módusz 5:** Akkor $x = 5$. Az átlag ekkor: $\frac{2+5}{5} = 1,4$. (Az $1,4 \neq 5$, nem jó).

Mivel az x -nek a módusznak **kell** lennie, és az átlagnak is **meg kell egyeznie** vele, az egyetlen matematikai megoldás az lenne, ha $\frac{2+x}{5} = x$ teljesülne. Ebből (ahogy az előbb is számoltuk) $x = 0,5$ jön ki. De ha $x = 0,5$, akkor minden szám csak egyszer szerepel, tehát **nincs módusz**.

Matematikai értelemben, ha ragaszkodunk az eredeti négy számhoz ($-2, -1, 0, 5$), akkor az **e) feladatra nincs megoldás**.

Ennek az az oka, hogy az "egyetlen módusz" feltétel kényszerpályára teszi az átlagot. Az átlag csak akkor lehetne egyenlő az egyetlen módusszal, ha az ötödik szám $0,5$ lenne, de ekkor (mivel a $0,5$ nem szerepelt korábban) nem alakul ki módusz.

Összegezve:

Ennél a konkrét számsornál az **átlag és az egyetlen módusz nem tud megegyezni**, bármilyen számot is választasz ötödiknek.

4. Tizenegy tanuló olyan tesztoldozatot írt, amelyben 100 pontot lehetett elérni.

A tanulók elért pontszámai: 100; 100; 100; 63; 62; 60; 12; 12; 6; 2; 0.

- a) Add meg az elért pontszámok főbb statisztikai jellemzőit!
- b) A csoportban az a vélemény van többségben, hogy nehéz volt a teszt. Tudsz-e érvelni a többség véleménye mellett a statisztikai mutatókra támaszkodva?

a) Főbb statisztikai jellemzők

- **Darabszám (n):** 11
- **Módusz:** 100 (ez szerepel a legtöbbször, háromszor).
- **Medián:** A sorba rendezett adatok középső (6.) eleme.
100, 100, 100, 63, 62, **60**, 12, 12, 6, 2, 0 $\rightarrow$ **60**
- **Terjedelem:** $100 - 0 = 100$
- **Átlag ($\bar{x}$):** Az összeg osztva 11-gyel.
$$\frac{100+100+100+63+62+60+12+12+6+2+0}{11} = \frac{517}{11} = \mathbf{47}$$

b) Érvelés a "nehéz volt a teszt" vélemény mellett

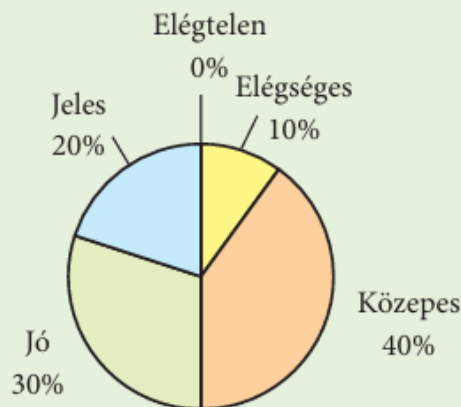
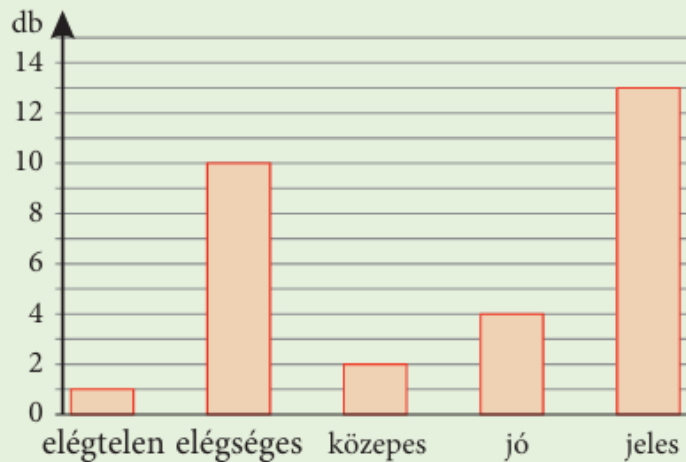
Bár a **módusz 100**, ami azt sugallná, hogy a teszt könnyű volt (hiszen a legtöbben maximumot írtak), a statisztika más mutatói a diákok véleményét támogatják:

1. **Az Átlag (47 pont):** Az elért átlagpontszám még a maximum (100) felét sem éri el.
Statisztikai szempontból egy olyan teszt, ahol a csoport átlaga "elégtelen" vagy gyenge "elégséges" szinten van, nehéznek minősül.
2. **A Medián (60 pont):** Ez azt mutatja, hogy a tanulók több mint fele (6 fő a 11-ből) 60 pontot vagy annál kevesebbet ért el.
3. **Kiemelkedő adatok (Outliers):** A három darab 100 pontos eredmény "húzza fel" az átlagot. Ha ezt a három zseniális vagy nagyon jól felkészült tanulót kivesszük, a maradék 8 diák átlaga drasztikusan lecsökken:

$$\frac{517-300}{8} = \frac{217}{8} = \mathbf{27,125}$$

Ez azt jelenti, hogy a közösség nagy része (közel 73%-a) nagyon alacsony pontszámot kapott.

5 A két diagram a félév végi kémiaosztályzatokat mutatja egy iskola 10.A, illetve 10.B osztályában.



- a)** Határozd meg, hogy mennyi a félévi kémiaosztályzatok terjedelme, módusza, mediánja és átlaga az egyes osztályokban!
- b)** Fogalmazd meg, hogy az egyes statisztikai jellemzők milyen módon olvashatók le a diagramokról!
- c)** Szerinted melyik osztályban lettek jobbak az eredmények? Milyen érvek szólnak az egyik, milyen érvek a másik osztály mellett?

a) Statisztikai jellemzők meghatározása**10.A osztály (Oszlopdiagram)**

Először olvassuk le a darabszámokat: Elégtelen (1), Elégséges (10), Közepes (2), Jó (4), Jeles (13).

Összesen: $1 + 10 + 2 + 4 + 13 = 30$ tanuló.

- **Terjedelem:** A legjobb és legrosszabb osztályzat különbsége: $5 - 1 = 4$.
- **Módusz:** A legmagasabb oszlop a jeles (13 fő), tehát a módusz az **5**.
- **Medián:** 30 tanuló esetén a 15. és 16. tanuló átlaga (sorba rendezve). Az első 11 tanuló 1-es vagy 2-es, a következő 2 közepes (13-ig), a következő 4 jó (17-ig). Tehát a 15. és 16. is **4-es (jó)**.
- **Átlag:** $\frac{(1 \cdot 1) + (10 \cdot 2) + (2 \cdot 3) + (4 \cdot 4) + (13 \cdot 5)}{30} = \frac{1 + 20 + 6 + 16 + 65}{30} = \frac{108}{30} = 3,6$.

10.B osztály (Kördiagram)

Itt csak százalékokat látunk, de mivel a statisztikai jellemzők (átlag, medián, módusz) relatívak, a konkrét létszám ismerete nélkül is kiszámíthatóak.

- **Terjedelem:** Van jeles (5) és elégséges (2) is, de elégtelen (1) nincs (0%). Így $5 - 2 = 3$.
- **Módusz:** A legnagyobb szelet a közepes (40%), tehát a módusz a **3**.
- **Medián:** Ha sorba rakjuk a százalékokat (0% elégtelen, 10% elégséges, 40% közepes...), az 50%-nál pont a **közepes (3)** és a **jó (4)** határán vagyunk. Statisztikailag ilyenkor a két érték átlagát, a **3,5**-öt vesszük.
- **Átlag:** $0,1 \cdot 2 + 0,4 \cdot 3 + 0,3 \cdot 4 + 0,2 \cdot 5 = 0,2 + 1,2 + 1,2 + 1,0 = 3,6$.

b) Hogyan olvashatók le a jellemzők?

- **Módusz:** Oszlopdiagramon a legmagasabb oszlop, kördiagramon a legnagyobb területű cikkely.
- **Terjedelem:** A skála két szélső értéke közötti különbség (oszlopnál ahol van oszlop, körnél ahol nem 0%).
- **Medián:** Oszlopnál a gyakoriságok összegzésével keressük a középső elemet; körnél a százalékok halmozott összegével nézzük meg, hová esik az 50%.
- **Átlag:** Súlyozott átlaggal számoljuk (érték szorozva a darabszámmal vagy százalékkal).

c) Melyik osztály eredménye a jobb?

Ez nézőpont kérdése, mindkét osztály mellett szólhatnak érvek:

10.A mellett szól:

- **Kiemelkedő teljesítmény:** Sokkal több a jeles (13 fő vs. 20%).
- **Magasabb medián:** A diákok fele legalább jó (4) lett, míg a B-ben a medián alacsonyabb.
- **Átlag:** Hajszálnyival (vagy kerekítve ugyanúgy) hozza a szintet, de több a "húzóerő".

10.B mellett szól:

- **Biztonság:** Senki nem bukott meg (0% elégtelen), míg az A-ban van bukás.
- **Egyenletesség:** Kevesebb a nagyon gyenge (elégséges) tanuló (csak 10%, míg az A-ban az osztály harmada az).
- **Kisebb terjedelem:** Kisebb a különbség a legjobb és leggyengébb tanuló között, homogénebb a csoport.

Összegzés: Ha a kiválóság a cél, a **10.A** a jobb, ha a bukás elkerülése és a kiegyensúlyozott középmezőny, akkor a **10.B**.