STATISZTIKA

Mennyiségi ismérv szerinti elemzés (1)

Csoportosítás, helyzeti és számított középértékek, helyzetmutatók (kvantilisek)

A SOKASÁG ELEMZÉSE EGY ISMÉRV SZERINT

- egyedi és csoportosított adatok,
- helyzetmutatók,
- középértékek,
- szóródás,

- **A sokaság csoportosítása**

A csoportosítás a sokaság valamely ismérv szerinti tagolása. A csoportosításnak

- *átfedés mentesnek*
- és *teljesnek* kell lennie.
- Azaz: *minden adatnak* (a sokaság minden egységének, egyedének) *pontosan egy* csoportba kell tartoznia.

A *csoportosító ismérv* lehet

- **Minőségi ismérv:** pl. a hallgatók csoportosítása szakok szerint, a lakosság csoportosítása településtípus szerint.
- **Mennyiségi ismérv:**
  - Az egyszerűbb eset, amikor ***ismétlődő**, diszkrét* **ismérvértékek**ről van szó: pl. családok csoportosítása gyerekszám szerint, a hallgatók csoportosítása a Statisztika vizsgajegyük szerint, ami lehet 1, 2, …5.
  - ***Nem ismétlődő ismérvértékek*** esetén az érték szerint sorba rendezett sokaság intervallumokba, *osztályközökbe* sorolható. Pl. egy cég alkalmazottai kereseti intervallumok szerint csoportosíthatók.

- ***A csoportosított sokaság jellemzése.***
- A csoportosított sokaságról az első fontos információ, hogy az egyes csoportokban hány egyed, hány adat van. Ez a gyakoriság
  - Gyakoriság, relatív gyakoriság.
    - A ***gyakoriság* (f_i)** azt adja meg, hogy az *i*-edik csoportba a sokaságnak hány egysége tartozik.
    - A ***relatív gyakoriság* (g_i)** azt adja meg, hogy az *i*-edik csoportba a sokaságnak hányad része (hány százaléka) tartozik.

$$g_i = \frac{f_i}{\sum_{i=1}^{m} f_i} = \frac{f_i}{n}$$

  - ahol g_i az *i*-edik csoport relatív gyakorisága, *m* a csoportok száma, *n* a sokaság elemszáma.

- **Kumulált gyakoriság:** a kumulálás halmozott (göngyölített) összeadást jelent.

  - ***Kumulált gyakoriság*:** az első *k* csoportba a sokaságnak hány egysége tartozik.

$$f_i' = \sum_{i=1}^{k} f_i$$

  - ***Kumulált relatív gyakoriság:*** az *első k csoportba* a sokaságnak *hányad része* tartozik.

$$g_k' = \sum_{i=1}^{k} g_i = \frac{\sum_{i=1}^{k} f_i}{\sum_{i=1}^{m} f_i} = \frac{\sum_{i=1}^{k} f_i}{n}$$

- Mennyiségi ismérveknél megkülönböztethető – az értékek szerint – *felfelé és lefelé kumulált* gyakoriság. Ha magasabb *i* értékhez magasabb ismérvértékek tartoznak (pl. ha magasabb sorszámú osztályközbe magasabb keresetű dolgozók tartoznak, akkor:

  - a ***felfelé kumulált gyakoriságok*** (f_i') és relatív gyakoriságok (g_i') azt mutatják, hogy **<u>az első *i* osztályközben</u>** összesen hány adat, illetve összesen az adatok hányad része található.

  - a ***lefelé kumulált gyakoriságok*** (f_i'') és relatív gyakoriságok (g_i'') azt mutatják, hogy az **<u>*i*-edik és az azt követő osztályközökben</u>** összesen *hány adat*, illetve összesen az *adatok hányad része* található.

- ***Értékösszeg:*** Ha mennyiségi ismérvekről, adatokról van szó, akkor fontos számított adat az *értékösszeg*, **az egy csoportban szereplő értékek összege**. Jele: s_i. Ha azonos értékek szerepelnek egy csoportban, akkor az értékösszeg számítása:

$$s_i = f_i \cdot x_i$$

- ***Relatív értékösszeg:*** egy adott csoport értékösszegének a *teljes értékösszegen belüli részaránya:*

$$z_i = \frac{f_i \cdot x_i}{\sum_{i=1}^{m} f_i x_i}$$

- Mind az abszolút, mind a relatív értékösszegekből képezhetők *kumulált* sorok.

- ***statisztikai sor:*** *adott ismérvek szerinti* statisztikai adatok felsorolása. (Ezeket célszerű táblázatos formában megadni.)
- ***gyakorisági sor:*** *adott csoportok szerinti* gyakorisági adatok felsorolása.

Lehetnek még:

- ***relatív gyakorisági sor;***
- ***kumulált relatív gyakorisági sor;***
- ***relatív értékösszeg sor***, stb.

- **Példa.** Egy raktárban az alábbi táblázat szerint 200 doboz gyümölcslé van, négyféle űrtartalommal. (A térfogat deciliterben)

Értékek Térfogat (dl) x_i	**Gyakoriság f_i**	**Relatív gyakoriság g_i (%)**	**Értékösszeg s_i**	**Relatív értékösszeg z_i**	**Kumulált értékösszeg s'**	**Kumulált rel.ért.összeg z'**
20	15					
10	25					
3,3	50					
2	110					
összesen	**200**					

Középértékek és helyzetmutatók egyedi és csoportosított adatok esetén

Az adathalmazok, az empirikus eloszlások jellemzése:

***Helyzeti középértékek*:**

- *módusz (a leggyakrabban előforduló, a „legdivatosabb – a la mode”- érték);*
- *medián;*

Helyzetmutatók:

- *kvantilisek;*

Számított középértékek:

- számtani átlag (egyszerű, súlyozott);
- harmonikus átlag (egyszerű, súlyozott);
- mértani átlag (egyszerű, súlyozott);
- geometriai átlag (egyszerű, súlyozott);
- négyzetes átlag (egyszerű, súlyozott);
- kronologikus átlag (egyszerű, súlyozott);

***Helyzeti középértékek*:**

Módusz

***A módusz fogalma ismétlődő adatok* esetén:**

- **Diszkrét ismérv esetén:** a leggyakrabban előforduló elem;
- **Folytonos ismérv esetén:** a gyakorisági görbe maximuma;

Példa: Ha egy évfolyam Statisztika ZH jegyeit vesszük, és a 4-es jegyből van a legtöbb, akkor a vizsgajegyek módusza a 4-es.
A módusz nem mindig egyértelmű, lehet két vagy több módusza is egy adathalmaznak.

Nem ismétlődő egyedi adatok esetén, ha minden egyes adat különbözik (valamennyi érték gyakorisága 1), akkor nincs értelme móduszról beszélni.

A módusz jellemzői

<u>Előnyös tulajdonságok:</u>

- tipikus érték;
- valamennyi mérési skála esetén alkalmazható;
- nem érzékeny a szélső, kiugró értékekre;

<u>Hátrányos tulajdonságok:</u>

- nem minden esetben létezik, vagy előfordulhat, hogy több is van belőle;

A módusz fogalma osztályközös csoportosított adatok esetén:

az a *modális* osztályköz, amelyben az *adatok sűrűsége* a legnagyobb.

PÉLDA: Ha egy évfolyam ZH pontszámait elemezzük, ahol 100 pont volt a maximum, és a 60 és 80 pont közötti intervallumban helyezkednek el legsűrűbben az adatok, akkor ez a modális osztályköz.

Mivel előfordulhatnak azonos sűrűségű osztályközök, ezért *a modális osztályköz meghatározása nem mindig egyértelmű.*
A hisztogram segítségével jól szemléltethető a *modális* osztályköz: ez az az osztályköz, amelyhez a legmagasabb oszlop tartozik.

A módusz becslése:

A modális osztályközben ki szoktak jelölni egyetlen értéket, amelyről azt mondhatjuk, hogy – feltehetően – *e körül sűrűsödnek leginkább az adatok*. Az alapadatok ismerete híján ez természetesen csak becslés.

<u>A legegyszerűbb lehetőség, hogy módusznak a modális osztályköz közepét választjuk,</u> de a gyakorlati tapasztalatok alapján jobb becslésnek tűnik, ha a két *szomszédos intervallum sűrűségét* is figyelembe vesszük, úgy gondolván, hogy a módusz az intervallum azon széléhez áll közelebb, amelynek a sűrűsége közelebb van a modális sűrűséghez.

<u>A számítás ennek alapján: határozzuk meg, hogy a két szomszédos osztályköz adatsűrűsége mennyivel kisebb, mint a modális osztályköz sűrűsége.</u> Jelöljük ezeket a különbségeket k_1 és k_2-vel, és válasszuk az intervallumból azt az értéket módusznak, amely az intervallumot $k_1 : k_2$ arányban osztja ketté. Ennek alapján a módusz becsült értéke a következő képlettel határozható meg:

$$Mo = x_{i0} + \frac{k_1}{k_1 + k_2} \cdot h_i$$

ahol x_{i0} a modális osztályköz alsó határa, h_i pedig a modális osztályköz szélessége.

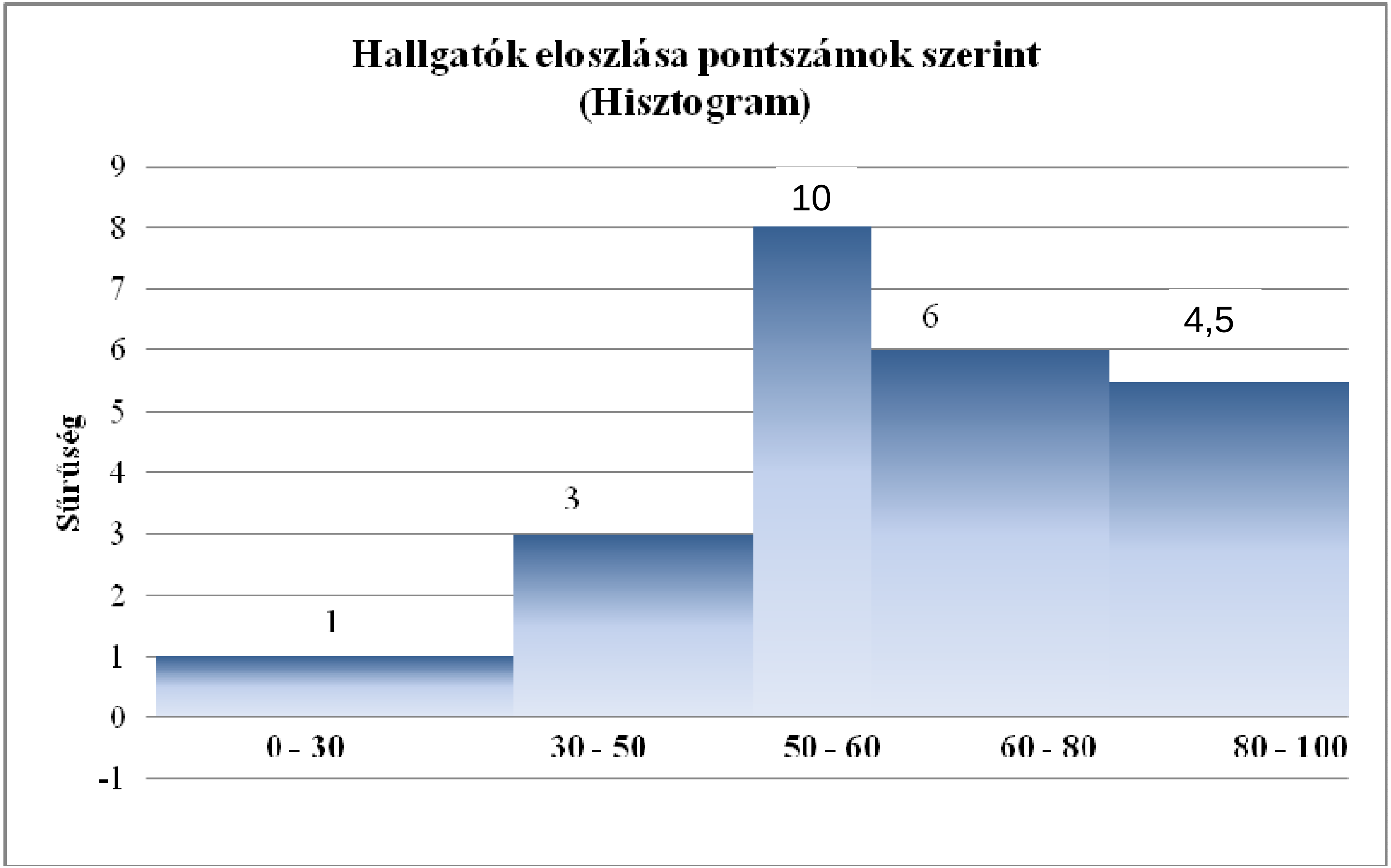
PÉLDA. Ha k_1 és k_2 és értéke 7 és 4, a modális osztályköz alsó határa pedig 50, az osztályköz szélessége 10, akkor a módusz becsült értéke:

$$\mathrm{Mo} = 50 + \frac{7}{7+4} \cdot 10 \cong 56{,}4$$

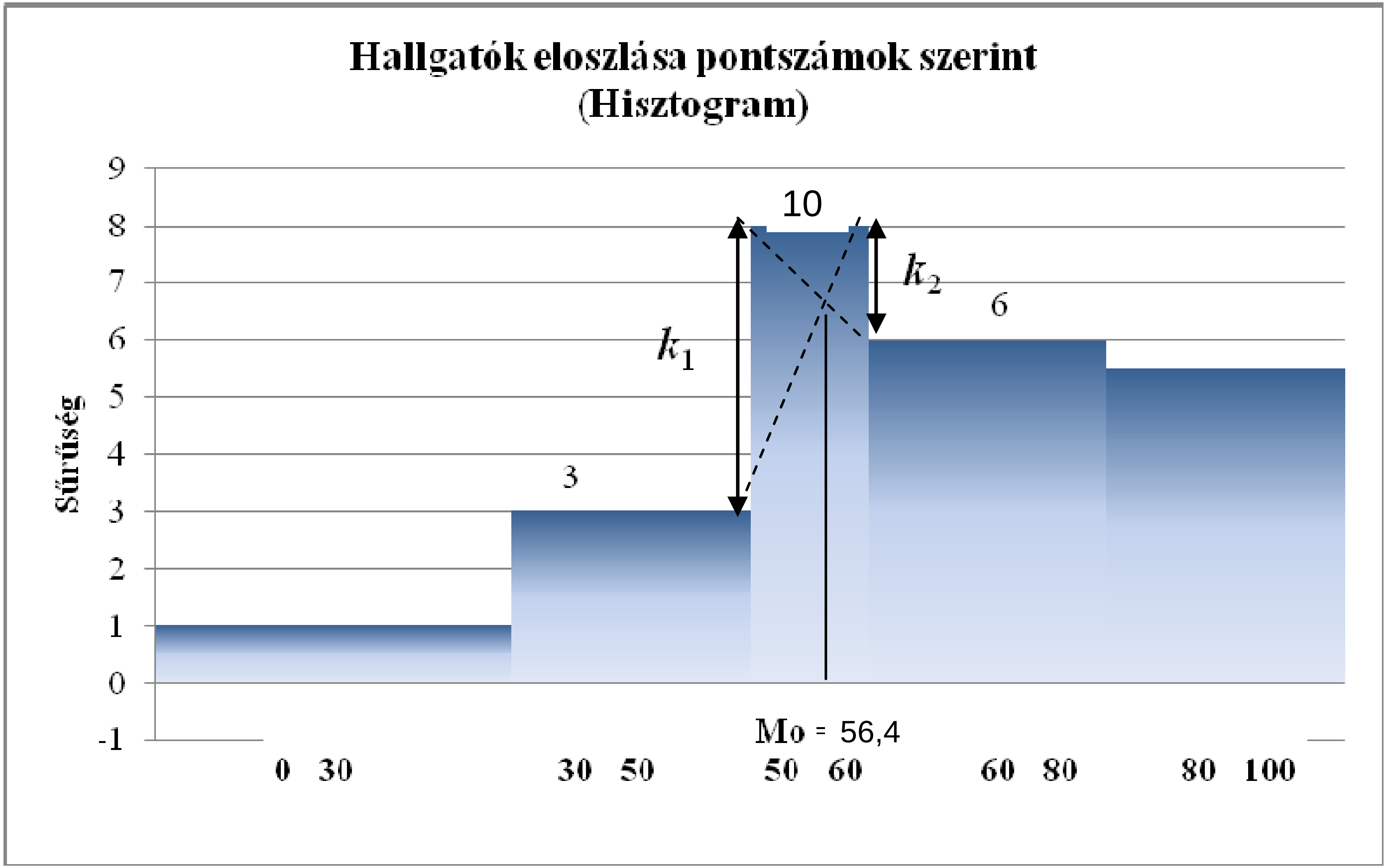
PÉLDA:

Hallgatók pontszám-intervallumok szerinti eloszlása:

Értékhatárok (Osztályközök)	Gyakoriság f_i	Osztályköz szélessége h_i	Sűrűség fi / hi
0 – 30	30	30	1
30 – 50	60	20	3
50 – 60	100	10	10
60 – 80	120	20	6
80 – 100	90	20	4,5
összesen	**400**		



A *legnagyobb gyakoriságú* osztályköz a negyedik,
a *legsűrűbb* viszont a harmadik.



Értelmezés: a pontszámok az 56,4-es érték körül sűrűsödnek a legnagyobb mértékben.

Kvantilisek:

Az ismérvértékek elhelyezkedésének tömör leírását adják.

Elnevezésük:

Medián (M_{e2i}) ha **két** részre osztjuk;

Kvartilis (Q_{4i}) ha **négy** részre osztjuk;

Kvintilis (K_{5i}) ha **öt** részre osztjuk;

Decilis (D_{10i}) ha **tíz** részre osztjuk;

Pentilis (Pen_{20i}) ha **húsz** részre osztjuk;

Percentilis (Per_{100i}) ha **száz** részre osztjuk;

Meghatározásuk (diszkrét ismérvekre):

a) Rangsort készítünk (növekvő sorrend)

b) Meghatározzuk a kvantilis értékek sorszámát (osztópontját) (azaz melyik kvantilist szeretném meghatározni?)

$$s_j = \frac{j}{k}(n+1)$$

ahol:

n = az adatok száma;

k = az egyenlő részek (kvantilisek) száma;

j = 1,2,…k-1 az adott kvantilis értékeken belüli sorszám (azaz megadja, hogy hányadik sorszámú kvantilist keressük);

az s_j-edik sorszámú adat a kérdéses kvantilis

Példa: A havi bruttó fizetések kvantilisei adott n (=95) elemű minta alapján

Medián:

$$s = \frac{1}{2}(95+1) = 48$$

Me=201 ezer Ft

Kvartilisek: (Q_j)

$$s_1 = \frac{1}{4}(95+1) = 24$$

Q_1=159 ezer forint

$$s_2 = \frac{2}{4}(95+1) = 48$$

Q_2= 201 ezer forint

$$s_3 = \frac{3}{4}(95+1) = 72$$

Q_3= 269 ezer forint

***Medián*:**

A medián a *sorba rendezett adatok közül a középső* érték; vagy másképpen: a *medián* az az érték amely a sorba rendezett adatokat *két egyenlő részre* osztja.
Egyedi adatok esetén,

a) **ha az adatok *páratlan* számúak**, akkor az iménti meghatározás egyértelmű, mert akkor van egy középső adat, amely előtt ugyanannyi adat van, mint utána;

b) ***páros számú* adat esetén** két középső adat van, ez esetben a kettő közti bármelyik érték mediánnak tekinthető. A gyakorlatban a két érték számtani közepét szokták megadni.

$\Rightarrow$ ***A medián az az érték, amelynél az adatok legfeljebb 50%-a kisebb és legfeljebb 50 %-a nagyobb.***

Medián előnyös tulajdonságai

- egyértelműen meghatározható;
- nemcsak mennyiségi jellemzők esetén határozható meg, hanem rangsorba rendezhető minőségi ismérvek esetén is;
- értéke független a szélsőértékektől;

Medián hátrányos tulajdonságai

- csak rangsorba rendezett elemekből számítható;
- ha az egyedek jelentős hányada azonos ismérvértékkel rendelkezik, akkor nem célszerű használni;

A medián meghatározása osztályközös gyakorisági sor esetén:

Azt az értéket keressük, amely *a sorba rendezett adatokat két egyenlő részre osztja*. Ez is csak becslés, mivel nem ismerjük az alapadatokat, csak a gyakorisági sort.

Először meghatározzuk a mediánt tartalmazó osztályközt. Ez könnyen megtehető, ha figyelembe vesszük, hogy a mediánt tartalmazó osztályköz *előtti* osztályközökre és az azt követő osztályközökre is igaz, hogy azokban az adatoknak *kevesebb, mint a fele* található. **A mediánt tartalmazó osztályköz indexe legyen *i*.**

Hogy eljussunk az adatok feléig, az *i*-edik osztályköz adataiból még kell vennünk további adatokat, azaz annak valamekkora hányadát. Ezután az *i*-edik intervallumnak is vegyük ugyanekkora hányadát, és ezzel meg is határoztuk a mediánt.

A számítás menete a következő, ha már tudjuk, melyik osztályköz tartalmazza a mediánt (legyen ennek a sorszáma *i*):

- A mediánt tartalmazó *i*-edik osztályközt megelőző összes osztályközben összesen legyen f'_{i-1} db adat. Az *i*-edik osztályközből még $(n/2 - f'_{i-1})$ db adatot kell vennünk, hogy eljussunk az összes adat feléig.

- Nézzük meg, hogy az *i*-edik osztályközből vett adatok száma hányad része az *i*-edik osztályköz gyakoriságának. Az az érték, amely ugyanilyen arányban osztja ketté az *i*-edik osztályközt, az lesz a medián.

Azaz:

$$Me = x_{i0} + \frac{\frac{n}{2} - f'_{i-1}}{f_i} \cdot h_i$$

Ugyanez *relatív gyakoriságokkal*:

$$Me = x_{i0} + \frac{0,5 - g'_{i-1}}{g_i} \cdot h_i$$

Itt az x_{i0} a mediánt tartalmazó osztályköz alsó határa, h_i pedig ennek az osztályköznek a szélessége.

A medián rokonfogalmai, a kvantilisek:

kvartilisek, decilisek, pentilisek, percentilisek

Medián: a „középső érték”: a sorba rendezett adatok 50%-a ennél kisebb, a másik 50%-a ennél nagyobb.

Ehhez hasonlóan más *helyzetmutatókat* is definiálhatunk:
milyen érték található az adatok
- ✓ egynegyedénél,
- ✓ háromnegyedénél,
- ✓ egyharmadánál,
- ✓ valamilyen *p* hányadánál.

Az általános képlet a *p* hányadú kvantilis meghatározására:

$$Kvant(p) = x_{i0} + \frac{p - g'_{i-1}}{g_i} \cdot h_i$$

A gyakoribb hányadok, kvantilisek külön nevet kaptak, ezek közül néhány:

Alsó kvartilis: az adatok ¼ -e ennél kisebb, az adatok ¾-e ennél nagyobb:

$$Q_1 = x_{i0} + \frac{0,25 - g'_{i-1}}{g_i} \cdot h_i$$

***Középső kvartilis*:** azonos a mediánnal.
Felső kvartilis: adatok ¾-e ennél kisebb, az adatok ¼ -e ennél nagyobb:

$$Q_1 = x_{i0} + \frac{0,75 - g'_{i-1}}{g_i} \cdot h_i$$

Felső decilis: adatok 9/10-e ennél kisebb, az adatok 1/10-e ennél nagyobb.
Felső pentilis: az adatok 95%-a ennél kisebb, az adatok 5%-a ennél nagyobb.
Felső percentilis: adatok 99 %-a ennél kisebb, az adatok 1%-a ennél nagyobb.

KÖZÉPÉRTÉKEK

A sokaság/minta eloszlásának jellemzése

- tipikus értékek meghatározása;
- az adatok különbözőségének vizsgálata;
- a sokaság/minta eloszlásgörbéjének elemzése;

<table>
<tr><th>Középértékek</th><th>Szóródási mérőszámok</th></tr>
<tr><td>

- **helyzeti (medián, módusz)**
- **számított**

$\overline{X}$, $\overline{X}_h$, $\overline{X}_q$, $\overline{X}_g$, $\overline{X}_k$

</td><td>

- **szórás (σ)**
- **relatív szórás (V)**
- **terjedelem (R)**
- **interkvartilis terjedelem (IQR)**

</td></tr>
<tr><th>Aszimmetria mérése</th><th>Egyéb jellemzők</th></tr>
<tr><td>

- **Pearson féle mutató (A)**
- **F mutató**
- **β_1 mutató**
- **Grafikus ábrázolás**

</td><td>

- **koncentráció**
- **kvantilisek**
- **momentumok**
- **grafikus ábrák**

</td></tr>
</table>

Középértékekkel szembeni követelmények

- egyértelmű számítás;
- tipikus, jellemző értékek legyenek;
- szemléletes, jó értelmezhetőség;
- közepes helyzet $X_{min} \leq K \leq X_{max}$;

Középértékek jellemzői

- a mennyiségi ismérvet egyetlen számmal jellemzik;
- dimenzió: az ismérv mértékegysége;

$\bar{x}$

Középértékek :

Átlagok

- Számtani $\overline{x}$
- Harmonikus $\overline{x_h}$
- Mértani $\overline{x_g}$
- Négyzetes $\overline{x_q}$
- **Kronologikus** $\overline{x_k}$

Helyzeti középértékek

- Módusz (Mo)
- Medián (Me)

Számtani átlag

Az a szám, amelyet az átlagolandó értékek helyére téve azok összege változatlan marad.

Egyedi értékeknél:

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Súlyozott forma:

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i \cdot x_i}{\sum_{i=1}^{n} f_i}$$

Egyszerű átlag

Az értékek egyszer fordulnak elő:

Érdemjegy (x)	Hallgatók száma/fő (f)
5	1
4	1
3	1
2	1
1	1
Összesen	**5**

Az értékek többször, de ugyanannyiszor fordulnak elő:

Érdemjegy (x)	Hallgatók száma/fő (f)
5	2
4	2
3	2
2	2
1	2
Összesen	**10**

Súlyozott átlag

Az értékek többször, de nem ugyanannyiszor fordulnak elő:

Érdemjegy (x)	Hallgatók száma/fő (f)
5	3
4	8
3	6
2	2
1	1
Összesen	**20**

A számtani átlag matematikai tulajdonságai

- Az egyes elemek - átlagolandó értékek - átlagtól való eltéréseinek összege 0:

$$\sum_{i=1}^{n} \left(x_i - \overline{x}\right) = 0$$

- Ha minden egyes elemhez hozzáadunk egy "a" konstans értéket, az így kapott elemek számtani átlaga "a"-val tér el az eredeti elemek átlagától.

- Ha minden egyes elemet megszorzunk egy "b" konstans értékkel, akkor az így kapott elemek átlaga "b"-szerese lesz az eredeti elemek átlagának.

A számtani átlag matematikai tulajdonságai

- Ha az $x_1, x_2, \ldots, x_n$ elemek átlaga $\bar{x}$ és
- az $y_1, y_2, \ldots, y_n$ elemek átlaga $\bar{y}$
- akkor az $x_1 + y_1; x_2 + y_2; \ldots ; x_n + y_n$ átlaga $\bar{x} + \bar{y}$ lesz.
- Az elemek mindegyikéből egy tetszőleges "a" állandót levonva ezen eltérések négyzetösszege akkor lesz minimális, ha az "a" állandó éppen az $\bar{X}$, azaz

$$\sum_{i=1}^{n} (x_i - a)^2 \text{ minimális, ha } a = \bar{X}$$

Példa a számtani átlag tulajdonságaira

x_i	x_i+50	$x_i \cdot 1,1$	$Z=x_i+x_i \cdot 1,1$
100	150	110	210
150	200	165	315
210	260	231	441
240	290	264	504
300	350	330	630
Σ *1000*	*1250*	*1100*	*2100*
$\bar{x}$ **200**	**250**	**220**	**420**

A számtani átlag előnyös tulajdonságai

- Világos, érthető fogalom, számítása egyszerű.
- Minden adathalmaznak létezik számtani átlaga, s csak egy van belőle.
- Minden elem figyelembe vételével kerül kiszámításra.
- Kiszámításához nem szükséges az egyedi értékek ismerete, elegendő azok darabszámát és összegét tudni.

A számtani átlag hátrányos tulajdonságai

- A kiugró értékekre (ún. outlier-ekre) érzékeny (nyesett átlag – trimmed mean).
- Osztályközös gyakorisági sor alkalmazása esetén nem tudjuk figyelembe venni az egyedi értékeket.

Geometriai átlag

Geometriai átlag az a szám, amelyet az egyedi értékek helyére írva azok szorzata változatlan marad.

Egyedi értékek esetén:

$$\bar{x}_g = \sqrt[n]{\prod_{i=1}^{n} x_i}$$

Súlyozott átlagforma:

$$\bar{x}_g = \sqrt[n]{\prod_{i=1}^{n} x_i^{f_i}}$$

A GDP volumenindexének alakulása Magyarországon

Időszak	Előző negyedév=100%
2008. I. n.év	100,9
2008. II. n.év	99,8
2008. III. n.év	99,0
2008. IV. n.év	98,1

Forrás: KSH Gyorstájékoztató

A változás átlagos üteme:

$$\overline{x_g} = \sqrt[4]{1{,}009 \cdot 0{,}998 \cdot 0{,}99 \cdot 0{,}981} = \sqrt[4]{0{,}978} = 0{,}994 = 99{,}4\%$$

Harmonikus átlag:

Egyszerű harmonikus átlag: az átlagolandó értékek reciproka átlagának a reciproka.

Fordított intenzitási viszonyszámok átlagolására használható.

$$\bar{x}_h = \frac{1}{\dfrac{\dfrac{1}{x_1} + \dfrac{1}{x_2} + \ldots + \dfrac{1}{x_n}}{n}} \qquad \bar{X}_h = \frac{n}{\sum_{i=1}^{n} \dfrac{1}{x_i}}$$

100 munkadarab előállításának műszakóra szükséglete 4 géptípusnál

műszakóra/100 munkadarab

Típus	Teljesítmény
I.	45
II.	55
III.	40
IV.	43

$$\bar{x} = \frac{1}{\frac{\sum_{i=1}^{n} \frac{1}{x_i}}{n}} = \frac{1}{\frac{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}{n}} = \frac{1}{\frac{\frac{1}{45} + \frac{1}{55} + \frac{1}{40} + \frac{1}{43}}{4}} =$$

$$\bar{\mathrm{x}} = \frac{1}{\frac{0{,}088}{4}} = \frac{1}{0{,}022} = 45{,}45 \text{mű} / 100 \text{munkadarab}$$

- Súlyozott harmonikus átlag

Ha viszonyszámokat átlagolunk és súlyként a viszonyszámok számlálója van megadva.

$$\overline{X}_h = \frac{1}{\frac{f_1 \frac{1}{x_1} + f_2 \frac{1}{x_2} + \ldots + f_n \frac{1}{x_n}}{f_1 + f_2 + \ldots + f_n}} \qquad \overline{X}_h = \frac{\sum_{i=1}^{n} f_i}{\sum_{i=1}^{n} f_i \frac{1}{x_i}}$$

Egy termék előállított mennyisége és egy főre jutó átlaga területi egységenként, Észak-Magyarország, 2008

Területi egység	Előállított mennyiség (f_i) ezer t	Egy főre jutó átlag (x_i) t/fő	$\frac{f_i}{x_i}$
Borsod-Abaúj-Zemplén	**114,04**	**39,38**	2,896
Heves	**90,64**	**33,57**	2,700
Nógrád	**14,10**	**19,29**	0,731
Észak-Magyarország	*218,78*	-	*6,327*

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i}{\sum_{i=1}^{n} \frac{f_i}{x_i}} = \frac{114,04 + 90,64 + 14,10}{\frac{114,04}{39,38} + \frac{90,64}{33,57} + \frac{14,10}{19,29}} = \frac{218,78}{6,327} = 34,58t\ /\ fő$$

Harmonikus átlag

- **Csak akkor alkalmazható, ha az átlagolandó értékek reciprokainak van tárgyi értelme.**
- A gyakorlatban a súlyozott formája fordul elő gyakrabban:
  - átlagszámítás értékösszeg sor adataiból;
  - összetett viszonyszám számítása;

Mértani átlag:

A fejlődés átlagos ütemét mutatja.

A változás kifejezhető az abszolút (összegszerű) és a relatív (szorzatszerű) mértékben.

Nagyságát a két szélső érték dönti el.

Egyirányú tendenciával rendelkező értéksor esetében használható.

Ha a változás nem egyirányú, akkor a statisztikai sort szakaszokra kell bontani.

A statisztikai sor időbeni, vagy intenzitásbeli terjedelmére ad jellemző értéket.

Négyzetes átlag:

- a kiugró értékekre érzékeny;
- az átlagolandó értékek helyébe helyettesítve, azok négyzetösszege változatlan marad;

$$\overline{x}_q = \sqrt{\frac{\sum x_i^2}{n}} \qquad \overline{x}_q = \sqrt{\frac{\sum f_i x_i^2}{\sum f_i}}$$

Kronologikus átlag

A kronologikus átlag a számtani átlag egyik fajtája.
Alkalmazható állapot idősor átlagolására, ahol az adatok egyenlő időközökben állnak rendelkezésre.
Pl. számolhatunk kronologikus átlaggal átlagkészletet és átlagos létszámot.
Kronológikus átlagot úgy számolunk, hogy az első adatot (NyK) osztjuk kettővel, hozzáadjuk a többi adatot, majd az utolsó adat (ZK) felét, és elosztjuk eggyel kevesebbel, mint az összes elemszám (n).

$$\overline{X}_k = \frac{\frac{x_1}{2} + x_2 + \ldots + x_{n-1} + \frac{x_n}{2}}{n-1}$$

Mintafeladat:

Január 1-jén a készlet 40eFt = NyK (nyitókészlet)
Január 31-én a készlet 48eFt
Február 28-án a készlet 46eFt
Március 31-én a készlet 44eFt = ZK (zárókészlet)

Határozzuk meg a készlet kronologikus átlagát!

Átlagok

	Súlyozatlan	Súlyozott
Számtani $\bar{x}$	$\bar{x} = \frac{\sum x_i}{n}$	$\bar{x} = \frac{\sum f_i x_i}{n}$
Harmonikus $\bar{x}_h$	$\bar{x} = \frac{n}{\sum \frac{1}{x_i}}$	$\bar{x} = \frac{\sum f_i}{\sum \frac{f_i}{x_i}}$
Mértani $\bar{x}_g$	$\bar{x} = \sqrt[n]{\prod x_i}$	$\bar{x} = \sqrt[n]{\prod x_i^{f_i}}$
Négyzetes $\bar{x}_q$	$\bar{x} = \sqrt{\frac{\sum x_i^2}{n}}$	$\bar{x} = \sqrt{\frac{\sum f_i x_i^2}{\sum f_i}}$

Ugyanazon pozitív értékekből számított átlagok nagyságrendje

$$x_{\min} \le \bar{x}_h \le \bar{x}_g \le \bar{x} \le \bar{x}_q \le x_{\max}$$

$\bar{x}_h$ és $\bar{x}_g$ érzékeny a kiugróan alacsony értékekre

$\bar{x}$ és $\bar{x}_q$ érzékeny a kiugróan magas értékekre

Példa/1: (egyszerű/súlyozatlan átlagok – az értékek csak egyszer fordulnak elő (**egyedi értékek**) vagy ugyanannyiszor)

Az átlagolandó értékek: 3, 4, 5, 8 – az értékek egyszer fordulnak elő

(vagy: 3, 3, 4 ,4, 5, 5, 8, 8 – az értékek többször, de ugyanannyiszor fordulnak elő)

<u>Feladat</u>

a) Számítsa ki a számtani, a harmonikus, a mértani és a négyzetes átlagot!

b) Hasonlítsa össze a kapott eredményeket!

c) Állapítsa meg ugyanazon pozitív számokból számolt átlagok sorrendjét!

d) Amennyiben az átlagolandó értékek között szerepelne még egy kiugróan alacsony érték (pl. 1), akkor mely átlagok reagálnának rá érzékenyen?

e) Mely átlagok értékét befolyásolja jobban, ha az átlagolandó értékek között még egy kiugróan magas érték (pl. 32) is található?

Megoldás

Számtani átlag:

$$\bar{x} = \frac{3+4+5+8}{4} = 5$$

Mértani átlag:

$$\bar{x}_g = \sqrt[4]{3 \cdot 4 \cdot 5 \cdot 8} = 4.681$$

Harmonikus átlag:

$$\bar{x}_h = \frac{4}{\frac{1}{3}+\frac{1}{4}+\frac{1}{5}+\frac{1}{8}} = 4.404$$

Négyzetes átlag:

$$\bar{x}_q = \sqrt{\frac{3^2+4^2+5^2+8^2}{4}} = \sqrt{\frac{114}{4}} = \sqrt{28,5} = 5.339$$

Példa/2: (**súlyozott átlag** – az értékek többször fordulnak elő és nem ugyanannyiszor)

Az átlagolandó értékek és a hozzájuk tartozó súlyok:

(x_i) adatok: 3, 4, 5, 8

(f_i) gyakoriság: 4, 4, 1, 1

Feladat:

a) Számítsa ki a számtani, a harmonikus, a mértani és a négyzetes átlagot!

Megoldás

Számtani átlag

$$\bar{x} = \frac{4\cdot 3 + 4\cdot 4 + 1\cdot 5 + 1\cdot 8}{10} = 4.1$$

Mértani átlag:

$$\bar{x}_g = \sqrt[10]{3^4 \cdot 4^4 \cdot 5^1 \cdot 8^1} = 3.907$$

Harmonikus átlag:

$$\bar{x}_h = \frac{10}{\frac{4}{3} + \frac{4}{4} + \frac{1}{5} + \frac{1}{8}} = \frac{10}{2.658} = 3.762$$

Négyzetes átlag:

$$\bar{x}_q = \sqrt{\frac{4\cdot 3^2 + 4\cdot 4^2 + 1\cdot 5^2 + 1\cdot 8^2}{10}} = 4.347$$

Példa/1. (egyedi értékek)

Egy bp.-i lakóparkban télen megkérdezték a 3 szobás lakások tulajdonosait, hogy mennyi volt az előző havi rezsiköltségük. Az alábbi adatokat kapták ezer Ft-ban:

75, 64, 69, 80, 76, 77, 86, 79, 65, 72, 73, 75, 75, 70

<u>Feladat:</u>
Jellemezzük a 3 szobás lakástulajdonosok előző havi rezsiköltségét az adott esetben felhasználható középértékekkel! (átlag, módusz, medián)

Megoldás

Számtani átlag:

$$\overline{X} = \frac{75 + \ldots + 70}{14} = 74$$

A lakástulajdonosok előző havi átlagos rezsiköltsége 74 ezer Ft.

Rangsor készítése:

64, 65, 69, 70, 72, 73, **75, 75, 75**, 76, 77, 79, 80, 86

Medián:

$$\frac{n+1}{2} = \frac{15}{2} = 7{,}5$$

Me=75 ezer Ft →A lakástulajdonosok felének 75 ezer Ft-nál kevesebb (a lakástulajdonosok másik felének pedig 75 ezer Ft-nál nagyobb) volt az előző havi rezsiköltsége.

Módusz:

Mo=75 ezer Ft

A legtöbb lakástulajdonos előző havi rezsije 75 ezer Ft.

Példa/2. (egyenlő osztályközök)

Egy benzinkútnál a napi eladott mennyiség szerint
a személygépkocsik megoszlása a következő volt:

Értékesített benzin mennyisége (liter)	Gépkocsik száma
10 – 19	10
20 – 29	28
30 – 39	42
40 – 49	15
50 – 59	5
Összesen	100

Feladat:

Számítsa ki és értelmezze az átlagot!

Becsülje meg a mediánt és a móduszt, és írja le jelentésüket!

Megoldás

Értékesített benzin mennyisége (liter)	Gépkocsik száma	Osztály-közép	Kumulált gyakoriság
10 – 19	10	15	10
20 – 29	28	25	38
30 – 39	42	35	80
40 – 49	15	45	95
50 –59	5	55	100
Összesen	100	---	---

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{10 \cdot 15 + 28 \cdot 25 + \ldots + 5 \cdot 55}{100} = \text{32,7 liter}$$

A gépkocsik átlagosan 32,7 litert tankoltak a benzinkútnál az adott napon.

Megoldás

Medián:

$s_{me} = \frac{n}{2} = \frac{100}{2} = 50$ **és** $f' \geq 50 \rightarrow Me$ a 3. osztályközben van

$$Me = x_{me,0} + \frac{\frac{n}{2} - f'_{me-1}}{f_{me}} h_{me} = 30 + \frac{50-38}{42} \cdot 10 =$$ 32,86 liter

A gépkocsik fele 32,86 liter benzinnél kevesebbet tankolt, a gépkocsik másik fele pedig ennél többet az adott napon.

Módusz: 3. osztályközben van

$$Mo = x_{mo,0} + \frac{k_1}{k_1 + k_2} \cdot h_{mo} = 30 + \frac{(42-28)}{(42-28)+(42-15)} \cdot 10 =$$ 33,41 liter

A legtöbb kocsi 33,41 liter benzin körüli mennyiséget tankolt az adott napon.

Példa/3. (nem egyenlő osztályközök)

1999-ben az átlagkeresetek alakulása egy vállalatnál

Keresetek (ezer Ft)	Létszám
40 – 50	12
50 – 60	20
60 – 80	34
80 – 100	32
100 – 150	14
150 – 200	3
Összesen	115

Feladat:

Számítsa ki és értelmezze az átlagot!

Becsülje meg a mediánt, a móduszt és a kvartiliseket és írja le jelentésüket!

Csak a MÓDUSZHOZ!

Megoldás

Keresetek (ezer Ft) (x)	Létszám (f)	Osztály-közép (x)	Kumulált gyakoriság (f')	f* (új oszt.köz= 20e Ft)
40 – 50	12	45	12	24
50 – 60 (Q1),(Mo)	20	55	32	40
60 – 80 (Me)	34	70	66	34
80 – 100 (Q3)	32	90	98	32
100 – 150	14	125	112	5,6
150 – 200	3	175	115	1,2
Összesen	115	---	---	---

$$\bar{x}=\frac{\sum f_i x_i}{\sum f_i}=\frac{45\cdot 12+55\cdot 20+\ldots+175\cdot 3}{115}=75{,}1eFt$$

$$f^*=\frac{gyakoriság}{eredeti\ oszt.\ köz\ h.}\cdot új\ oszt.\ köz\ h.$$

A vállalatnál a dolgozók átlagosan 75,1 ezer Ft-ot keresnek.

Megoldás

Medián:

$s_{me} = \frac{n}{2} = \frac{115}{2} = 57{,}5$ (A Me 3. osztályközben van.)

$$Me = x_{me,0} + \frac{\frac{n}{2} - f'_{me-1}}{f_{me}} h_{me} = 60 + \frac{57{,}5 - 32}{34} \cdot 20 = \text{ 75 ezer Ft}$$

A dolgozók fele 75 ezer Ft-nál kevesebbet keresett, (a másik fele pedig ennél többet) az adott évben.

Alsó kvartilis: $\frac{n}{4} = \frac{115}{4} = 28{,}75$ (A Q1 a 2. osztályközben van.)

$$Q_1 = x_{q1,0} + \frac{\frac{n}{4} - f'_{q1-1}}{f_{q1}} \cdot h_{q1} = 50 + \frac{28{,}75 - 12}{20} \cdot 10 = \text{ 58,375 ezer Ft}$$

A dolgozók negyede 58,4 ezer Ft-nál kevesebbet keresett, (három negyede pedig ennél többet) az adott évben.

Megoldás

Felső kvartilis:

$s_{q3} = \frac{3n}{4} = \frac{3 \cdot 115}{4} = 86{,}25$ (A Q3 4. osztályközben van.)

$$Q_3 = x_{q3,0} + \frac{\frac{3 \cdot n}{4} - f'_{q3-1}}{f_{q3}} h_{q3} = 80 + \frac{86{,}25 - 66}{32} \cdot 20 = 92{,}65 eFt$$

A dolgozók negyede 92,65 ezer Ft-nál többet keresett, (a három negyede pedig ennél kevesebbet) az adott évben.

Módusz: (A Mo a 2. osztályközben van.)

$$Mo = x_{mo,0} + \frac{k_1}{k_1 + k_2} \cdot h_{mo} = 50 + \frac{(40-24)}{(40-24)+(40-34)} \cdot 10 = 57{,}27 eFt$$

A dolgozók legtöbbje 57,27 ezer Ft-ot keresett az adott évben.

Nézzük mindig a dolgok napos oldalát!

Mára befejeztük, viszontlátásra!