3. lecke: Középértékek

Egy N elemű sokaság valamely X ismérv (változó) szerinti vizsgálatakor jó lenne, ha ezt a változót egyetlen olyan tipikus, közepes helyet elfoglaló értékkel tudnánk jellemezni, amelynek szemléletes tartalma van. A **középérték** olyan mutatószám, átlagos, közepes érték, amely a sokaság valamely tulajdonságát egy számmal fejezi ki. Mértékegysége azonos a vizsgált változóéval. A középérték a legkisebb és a legnagyobb ismérvérték között helyezkedik el.
A középértékeknek két csoportja van: a **számított középértékek (átlagok)** és a **helyzeti középértékek**.

1. Számított középértékek

A **számított középértékek (átlagok)** az ismérvértékekből számíthatók ki. **A négy legfontosabb átlag: a számtani (aritmetikai) átlag, a harmonikus átlag, a mértani (geometriai) átlag, a négyzetes (kvadratikus) átlag.** Az, hogy **melyik középértéket alkalmazzuk egy adatállományon, nem önkényes**. A választás attól függ, hogy a megadott adatokon milyen művelet elvégzésének van értelme.

A **számtani (aritmetikai) átlag** az a szám, melyet az egyes átlagolandó értékek helyére írva, azok összege nem változik.
A számtani átlag akkor használható, ha az ismérvértékek összegének van tárgyi értelme. Ez a leggyakrabban használt számított középérték. Az egyszerű számtani (aritmetikai) átlag a vizsgált változó x_i ismérvértékei összegének, azaz a teljes értékösszegnek és a sokaság elemszámának hányadosa.

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_N}{N} = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{S}{N}$$

Példa

Nyolc személy havi nettó keresete legyen rendre 50, 70, 60, 90, 80, 150, 120, 100 ezer forint. Ekkor a nyolc személy havi nettó keresetének átlaga:

$$\bar{x} = \frac{50+70+60+90+80+150+120+100}{8} = 90 \text{ ezer ft.}$$

Ha az egyes ismérvértékek többször is előfordulnak, akkor célszerűbb a súlyozott átlagformát használni. Ez gyakorlatilag azt jelenti, hogy az azonos tagokat összevonjuk. Például, 3,3,4,4,4 értékek esetében az ismérvértékek a számtani átlag összege felírható 3+3+4+4+4 alakban, de röviden csak 2*3+3*4 formában is. Ez a rövidített formátum egy súlyozott összeg. Általánosan, ha az egyes előforduló x_i ismérvértékek gyakoriságait f_i -vel jelöljük, akkor a súlyozott számtani átlag képlete az alábbi.

$$\bar{x} = \frac{f_1 \cdot x_1 + f_2 \cdot x_2 + \ldots + f_N \cdot x_N}{f_1 + f_2 + \ldots + f_N} = \frac{\sum_{i=1}^{k} f_i x_i}{\sum_{i=1}^{k} f_i} = \frac{\sum_{i=1}^{k} f_i x_i}{N} = \sum_{i=1}^{k} g_i x_i = \frac{\sum_{i=1}^{k} S_i}{\sum_{i=1}^{k} f_i} = \frac{S}{N}$$

A számított középértékek súlyozott formáiban a gyakoriságok helyett alkalmazhatjuk a relatív gyakoriságokat is. A számtani átlag esetében ez konkrétan az alábbi levezetést, illetve formulát jelenti.

$$\bar{x} = \frac{f_1 \cdot x_1 + f_2 \cdot x_2 + \ldots + f_N \cdot x_N}{N} = \frac{f_1 \cdot x_1}{N} + \frac{f_2 \cdot x_2}{N} + \ldots + \frac{f_N \cdot x_N}{N} =$$

$$= g_1 \cdot x_1 + g_2 \cdot x_2 + \ldots + g_N \cdot x_N = \sum_{i=1}^{k} g_i x_i$$

A súlyozott számtani átlag szemléltetésére nézzük a korábbi panelházas példát!

Példa

Egy lakótelepi panelház lakói a háztartás taglétszáma szerint

Taglétszám (fő) x_i	Háztartások száma (db) f_i	$S_i = f_i \cdot x_i$ (Lakók száma, fő)
1	5	5
2	10	20
3	7	21
4	25	100
5	2	10
6	1	6
Összesen	**50**	**162**

$$\bar{x} = \frac{\sum_{i=1}^{k} f_i x_i}{\sum_{i=1}^{k} f_i} = \frac{S}{N} = \frac{162}{50} = 3{,}24 \text{ fő/háztartás.}$$

Tehát a vizsgált panelházban a háztartások átlagos taglétszáma 3,24 fő, azaz 100 háztartásra átlagosan 324 lakó jut.

Amennyiben osztályközös gyakorisági sorból kell kiszámítanunk a számtani átlagot, akkor ennek értékét csak becsülni tudjuk, mivel ekkor csak annyit tudunk, hogy az egyes osztályközökbe hány ismérvérték esik (ezek a gyakoriságok), de az ismérvértékeket konkrétan nem ismerjük. Ekkor azt feltételeztük, hogy az egyes osztályközökbe eső sokasági elemek ismérvértékei az osztályközön belül egyenletesen oszlanak el, ezért azok helyettesíthetők az osztályközéppel. Ekkor a számtani átlag az osztályközepek – gyakorisággal – súlyozott átlagaként adódik. Ennek szemléltetésére nézzük a korábbi ABC-s példát!

Példa

Egy ABC forgalmára vonatkozó adatok

Vásárlások értéke (Ft)	Vásárlások száma f_i	X_i (osztályközép)	$\hat{S}_i = f_i \cdot x_i$
– 400	56	200 =(0+400)/2	11 200,00 $(= 56 \cdot 200)$
401 – 800	89	600 =(400+800)/2	53 400,00
801 – 1200	250	1000	250 000,00
1201 – 2000	260	1600	416 000,00
2001 – 2800	210	2400	504 000,00
2801 – 4000	84	3400	285 600,00
Összesen	949	–	1 520 200,00

$$\bar{x} = \frac{\sum_{i=1}^{k} f_i X_i}{\sum_{i=1}^{k} f_i} = \frac{\hat{S}}{N} = \frac{1520200}{949} = 1601{,}9 \text{ Ft/fő.}$$

Tehát a vásárlók átlagosan 1601,9 forint/fő értékben vásároltak, azaz az 1 vásárlóra jutó átlagos vásárlási érték 1601,9 Ft[1].

A számtani átlag fontosabb tulajdonságai:

1. Minden ismérvérték számtani átlaggal való helyettesítésekor elkövetett előjeles hibák kiegyenlítik egymást, vagyis az egyes ismérvértékek számtani átlagtól való eltéréseinek összege 0, azaz $\sum_{i=1}^{N}(x_i - \bar{x}) = 0$

2. Ha mindenegyes ismérvértékhez ugyanazt az *A* számot hozzáadjuk (kivonjuk), akkor ezek átlaga is *A*-val lesz nagyobb (kisebb) az ismérvértékek átlagánál.
   Például, ha mindenki kap 5000 forint fizetésemelést, akkor a fizetések átlaga is 5000 forinttal lesz nagyobb a korábbi értékéhez képest.

3. Ha mindenegyes ismérvértéket ugyanazzal a nullától különböző *A* számmal megszorozzuk, akkor ezek átlaga is A-szorosa lesz az ismérvértékek átlagának.

[1] Az értelmezések során a mértékegységet vagy Ft/fő alakban használjuk, vagy simán Ft alakot írunk, de ilyenkor a szövegbe beillesztjük az 1 főre jutó kifejezést.

Például, ha mindenki kap 5 százalék fizetésemelést, azaz mindenkinek 1,05-szorosára nő a fizetése, akkor a fizetések átlaga is 1,05-szorosára nő, azaz 5 százalékkal lesz nagyobb a korábbi értékéhez képest.

4. Minden ismérvérték számtani átlaggal való helyettesítésekor elkövetett hibák négyzetösszege minimális lesz; és fordítva: a számtani átlag az a konstans, amely esetén a négyzetes hiba minimális. Ez az ún. négyzetes minimum tulajdonság: $\sum_{i=1}^{N}(x_i - A)^2 \to \min \Leftrightarrow A = \bar{x}$. Erre a tulajdonságra fog épülni a következő fejezetben tárgyalt szórás.

A **mértani átlag** az a szám, melyet az egyes átlagolandó értékek helyére írva, azok szorzata nem változik. A mértani átlagot akkor használhatjuk, ha az ismérvértékek (átlagolandó értékek) szorzatának van értelme. Az egyszerű mértani (geometriai) átlag a sokaság x_i ismérvértékei szorzatának N-edik gyöke.

$$x_g = \sqrt[N]{x_1 \cdot x_2 \cdot x_3 \cdot \ldots \cdot x_N}$$

Ha az egyes ismérvértékek többször is előfordulnak, akkor célszerűbb a súlyozott átlagformát használni. Ha az egyes előforduló x_i ismérvértékek gyakoriságait f_i -vel jelöljük, akkor a súlyozott mértani átlag képlete az alábbi.

$$x_g = \sqrt[N]{x_1^{f_1} \cdot x_2^{f_2} \cdot x_3^{f_3} \cdot \ldots \cdot x_N^{f_N}}$$

A mértani átlag leggyakoribb alkalmazása a fejlődés átlagos ütemének meghatározása. Ezt az idősorok vizsgálatánál fogjuk részletesen tárgyalni.

A **harmonikus átlag** az a szám, melyet az egyes átlagolandó értékek helyére írva, azok reciprokainak összege nem változik. Harmonikus átlagot akkor használunk, ha az átlagolandó értékek reciprokaiból számított összeg értelmezhető. Az egyszerű harmonikus átlag a sokaság elemszámának és az ismérvértékek reciprokai összegének hányadosa.

$$\bar{x}_h = \frac{N}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \ldots + \frac{1}{x_N}} = \frac{N}{\sum_{i=1}^{N} \frac{1}{x_i}}$$

Ha az egyes ismérvértékek többször is előfordulnak, akkor célszerűbb a súlyozott átlagformát használni. Ha az egyes előforduló x_i ismérvértékek gyakoriságait f_i-vel jelöljük, akkor a súlyozott harmonikus átlag képlete az alábbi.

$$\bar{x}_h = \frac{N}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \frac{f_3}{x_3} + \ldots + \frac{f_N}{x_N}} = \frac{\sum_{i=1}^{k} f_i}{\sum_{i=1}^{k} \frac{f_i}{x_i}} = \frac{1}{\sum_{i=1}^{k} \frac{g_i}{x_i}}$$

A harmonikus átlag számításának tipikus esete a részekre bontott sokaságokból történő átlagszámítás, melyet a következő szakaszban tárgyalunk.

A **négyzetes átlag** az a szám, melyet az egyes átlagolandó értékek helyére írva, azok négyzeteinek összege nem változik.
A négyzetes átlagot akkor használhatjuk, ha nem akarjuk figyelembe venni az átlagolandó értékek előjelét, és azt akarjuk, hogy az átlag a szélsőségesen nagy értékekre érzékenyen reagáljon. Az egyszerű négyzetes átlag az ismérvértékek négyzetösszegeinek és sokaság elemszáma hányadosának négyzetgyöke.

$$\bar{x}_q = \sqrt{\frac{x_1^2 + x_2^2 + x_3^2 + \ldots + x_N^2}{N}} = \sqrt{\frac{\sum_{i=1}^{N} x_i^2}{N}}$$

Ha az egyes ismérvértékek többször is előfordulnak, akkor célszerűbb a súlyozott átlagformát használni. Ha az egyes előforduló x_i ismérvértékek gyakoriságait f_i -vel jelöljük, akkor a súlyozott négyzetes átlag képlete az alábbi.

$$\bar{x}_q = \sqrt{\frac{f_1 \cdot x_1^2 + f_2 \cdot x_2^2 + f_3 \cdot x_3^2 + \ldots + f_N \cdot x_N^2}{N}} = \sqrt{\frac{\sum_{i=1}^{k} f_i x_i^2}{\sum_{i=1}^{k} f_i}} = \sqrt{\sum_{i=1}^{k} g_i x_i^2}$$

A négyzetes átlag tipikus alkalmazása a szórás számítása. Ezzel részletesen a következő leckében foglalkozunk.

A számított középértékek mindig a legkisebb és a legnagyobb ismérvérték közé esnek. Ugyanazon pozitív ismérvértékekből számított középértékek között az alábbi reláció áll fenn.

$$x_{\min} \le x_h \le x_g \le x \le x_q \le x_{\max}$$

Egyenlőség pontosan akkor áll fenn, ha minden egyes ismérvérték megegyezik. Ekkor az összes felsorolt, számított középérték átlaga megegyezik. Tehát ugyanazon adatokból számított középértékek matematikailag vagy mind megegyeznek, vagy mind különböznek.

A médiában gyakorta szerepelnek olyan állítások, hogy az átlagembernek hány gyermeke van, hány évig él, mekkora a havi nettó jövedelme, stb. Ez a megfogalmazás nem korrekt, ugyanis csak a változókból számolhatunk átlagokat, azaz statisztikailag átlagember nem létezik, csak átlag nettó jövedelem, várható átlagos élettartam stb.

2. Számított középértékek részekre bontott sokaság esetén

Ha egy sokaságot kettéválasztunk férfiakra és nőkre, akkor milyen összefüggést lehet felírni a nemek átlagkeresetei és az összes megfigyelt átlagkeresete között? Tehát adott egy vizsgált mennyiségi változó (kereset), melyből átlagot szeretnénk számítani, egy másik ismérv (nem), mely szerint részekre bontottuk a sokaságot. Ekkor a kapott diszjunkt részhalmazokat (férfiak, nők) **részcsoportoknak**, **részsokaságoknak** nevezzük. Ilyenkor egy adott részcsoporton belül számított átlagot **részátlagnak** (például férfiak átlagkeresete), míg az egész sokaság vonatkozásában számított átlagot **főátlagnak** (megfigyeltek átlagkeresete) nevezzük. Ekkor a főátlag a részátlagok súlyozott átlagaként számítható ki.

Ebben az esetben az ismérvértékeket x_{ij} szimbólummal jelöljük, melyben a j index arra utal, hogy az adott egyed melyik részcsoporthoz tartozik, az i index pedig az adott részcsoporton belüli i-edik elemet jelöli. A j-edik részcsoportba tartozó egyedek számának jelölése f_j, a részcsoporthoz tartozó értékösszegé pedig S_j. Ekkor a j-edik részcsoportban az ismérvértékek átlagát az

$$\bar{x}_j = \frac{\sum_{i=1}^{f_j} x_{ij}}{f_j} = \frac{S_j}{f_j}$$

formulával határozhatjuk meg.

A főátlag meghatározására három lehetőségünk van, attól függően, hogy a részcsoportokhoz tartozó ismérvértékek, részátlagok, értékösszegek közül mely mennyiségeket ismerjük. Ha ismerjük az ismérvértékeket, akkor a főátlagot megkaphatjuk az ismérvértékek összege, azaz a teljes értékösszeg és a sokaság elemszámának hányadosaként. Az ismérvértékek összegének meghatározásakor először a csoportokon belül határozzuk meg a $\sum_{i=1}^{f_j} x_{ij}$ (S_j) értékösszeget, majd a k darab csoport értékösszegét összeadjuk:

$$\bar{x} = \frac{\sum_{j=1}^{k}\sum_{i=1}^{f_j} x_{ij}}{\sum_{j=1}^{k} f_j} = \frac{\sum_{j=1}^{k}\sum_{i=1}^{f_j} x_{ij}}{N} = \frac{\sum_{j=1}^{k} S_j}{\sum_{j=1}^{k} f_j} = \frac{\sum_{j=1}^{k} S_j}{N} = \frac{S}{N}.$$

Ha a részátlagokat és a részcsoportok elemszámait ismerjük, akkor a főátlag megkapható a részátlagok – csoport elemszámokkal – súlyozott számtani átlagaként, azaz

$$\bar{x} = \frac{\sum_{j=1}^{k} f_j \cdot \bar{x}_j}{\sum_{j=1}^{k} f_j} = \sum_{j=1}^{k} g_j \cdot \bar{x}_j .$$

Ezt az alábbi levezetéssel igazolhatjuk.

$$\bar{x}_j = \frac{\sum_{i=1}^{f_j} x_{ij}}{f_j} \rightarrow \sum_{i=1}^{f_j} x_{ij} = f_j \cdot \bar{x}_j \Rightarrow \bar{x} = \frac{\sum_{j=1}^{k}\sum_{i=1}^{f_j} x_{ij}}{N} = \frac{\sum_{j=1}^{k} f_j \cdot \bar{x}_j}{N}$$

Példa

Három fagylaltárusító hely forgalmára és árára vonatkozóan az alábbi adatok ismertek.

Árus	Eladott mennyiség (gombóc)	Átlagár (Ft/gombóc)
1.	100	90
2.	80	110
3.	90	105

Ekkor a három árusító helyen átlagosan mennyibe kerül egy gombóc fagylalt?
A kérdés megválaszolásához azt kellene eldöntenünk, hogy milyen mennyiségek szerepelnek a táblázatban. Mivel az eladott mennyiség a részcsoportok elemszámait jelenti, az utolsó oszlopban pedig a részcsoportok részátlagait látjuk, ezért

$$\overline{x} = \frac{\sum_{j=1}^{k} f_j \cdot \overline{x}_j}{N} = \frac{100 \cdot 90 + 80 \cdot 110 + 90 \cdot 105}{100 + 80 + 90} = 100{,}9 \text{ Ft/gombóc.}$$

Ha a részátlagokat és a részcsoportokhoz tartozó értékösszegeket ismerjük, akkor a főátlag megkapható a részátlagok – csoport értékösszegekkel – súlyozott harmonikus átlagaként, azaz

$$\overline{x} = \frac{\sum_{j=1}^{k} S_j}{\sum_{j=1}^{k} \frac{S_j}{\overline{x}_j}} = \frac{1}{\sum_{j=1}^{k} \frac{Z_j}{\overline{x}_j}}\ .$$

Ezt az alábbi levezetéssel igazolhatjuk.

$$\overline{x}_j = \frac{S_j}{f_j} \to f_j = \frac{S_j}{\overline{x}_j} \Rightarrow \overline{x} = \frac{\sum_{j=1}^{k} S_j}{\sum_{j=1}^{k} f_j} = \frac{\sum_{j=1}^{k} S_j}{\sum_{j=1}^{k} \frac{S_j}{\overline{x}_j}}$$

Példa

Három fagylaltárusító hely forgalmára és árára vonatkozóan az alábbi adatok ismertek.

Árus	Napi forgalom (Ft)	Átlagár (Ft/gombóc)
1.	135 000	90
2.	264 000	110
3.	94 500	105

Ekkor a három árusító helyen átlagosan mennyibe kerül egy gombóc fagylalt?
A kérdés megválaszolásához azt kellene eldöntenünk, hogy milyen mennyiségek szerepelnek a táblázatban. Mivel a napi forgalom az egységár és az értékesített mennyiség szorzata, ezért a napi forgalom értékösszeg. Az utolsó oszlopban pedig a részcsoportok részátlagait látjuk. Ekkor

$$\bar{x} = \frac{\sum_{j=1}^{k} S_j}{\sum_{j=1}^{k} \frac{S_j}{\bar{x}_j}} = \frac{135000 + 264000 + 94500}{\frac{135000}{90} + \frac{264000}{110} + \frac{94500}{105}} = 102{,}8 \text{ Ft/gombóc.}$$

3. Helyzeti középértékek

A **helyzeti középértékek** az ismérvértékek közötti elhelyezkedésük alapján adnak valamiféle információt. **A két legismertebb helyzeti középérték a módusz és a medián.**

A **módusz** (jelölése: *Mo*) a tipikus, a divatos, a leginkább jellemző értéket mutatja. E körül sűrűsödnek, tömörülnek az ismérvértékek. *Diszkrét változó* esetén a módusz a leggyakrabban előforduló ismérvérték, tehát az az ismérvérték, amelyhez a legnagyobb gyakoriság, illetve relatív gyakoriság tartozik. *Folytonos változó* esetén a módusz a gyakorisági görbe maximumhelye.

Egy sokaság eloszlása lehet egy-, de akár többmóduszú is, attól függően, hogy hány módusz van. Előfordulhat az az eset is, hogy egy eloszlásnak nincs meghatározott módusza, például, ha mindegyik ismérvérték csak egyszer fordul elő.

Példa: módusz meghatározása rangsorból

Ismérvértékek rangsora: 10, 11, 12, 12, 13, 13, 13, 15. Ekkor Mo=13.
Ismérvértékek rangsora: 10, 11, 12, 12, 12, 13, 13, 13. Ekkor Mo_1=12 és Mo_2=13.
Ismérvértékek rangsora: 10, 10, 10, 10, 10, 10, 10, 10. Ekkor Mo=10.
Ismérvértékek rangsora: 10, 11, 12, 13, 14, 15, 16, 17. Ekkor Mo= –, azaz nem definiáljuk.

Példa: módusz meghatározása nem osztályközös gyakorisági sorból

Egy lakótelepi panelház háztartásai taglétszám szerint

Taglétszám (fő), x_i	Háztartások száma (db), f_i
1	5
2	10
3	7
4	25
5	2
6	1
Összesen	**50**

Forrás: fiktív adatok

Ekkor Mo=4; azaz a vizsgált panelházban a 4 fős háztartás a leggyakoribb.

Osztályközös gyakorisági sor esetén a módusz csak becsülhető, mivel az ismérvértékek konkrét értékét nem ismerjük. Ehhez **először** meg kell határoznunk, hogy melyik osztályban helyezkedik el a módusz. Ezt az osztályt **modális osztályköznek** nevezzük. Folytonos változó esetében a módusz a gyakorisági görbe maximumhelye, azaz azt mutatja meg, hogy hol sűrűsödnek legjobban az ismérvértékek. A módusz abban az osztályközben lesz, ahol a legnagyobb az $\frac{f_i}{h_i}$ adatsűrűség. A módusz egy durva becslése a modális osztályköz osztályközepe. Ezt az értéket **nyers módusznak** nevezzük. A **módusz becslésére osztályközös gyakorisági sorok esetében** az alábbi képletet használják.

$$\hat{Mo} = X_{Mo,0} + \frac{\frac{f_{Mo}}{h_{Mo}} - \frac{f_{Mo-1}}{h_{Mo-1}}}{\left(\frac{f_{Mo}}{h_{Mo}} - \frac{f_{Mo-1}}{h_{Mo-1}}\right) + \left(\frac{f_{Mo}}{h_{Mo}} - \frac{f_{Mo+1}}{h_{Mo+1}}\right)} \cdot h_{Mo} =$$

$$= X_{Mo,0} + \frac{\frac{g_{Mo}}{h_{Mo}} - \frac{g_{Mo-1}}{h_{Mo-1}}}{\left(\frac{g_{Mo}}{h_{Mo}} - \frac{g_{Mo-1}}{h_{Mo-1}}\right) + \left(\frac{g_{Mo}}{h_{Mo}} - \frac{g_{Mo+1}}{h_{Mo+1}}\right)} \cdot h_{Mo},$$

ahol
$X_{Mo,0}$: modális osztályköz alsó határa,
f_{Mo}: modális osztályköz gyakorisága,
h_{Mo}: modális osztályköz hossza,
f_{Mo-1}: modális osztályköz előtti osztályköz gyakorisága,
h_{Mo-1}: modális osztályköz előtti osztályköz hossza,
f_{Mo+1}: modális osztályköz utáni osztályköz gyakorisága,
h_{Mo+1}: modális osztályköz utáni osztályköz hossza,
g_{Mo}: modális osztályköz relatív gyakorisága,
g_{Mo-1}: modális osztályköz előtti osztályköz relatív gyakorisága,
g_{Mo+1}: modális osztályköz utáni osztályköz relatív gyakorisága.

Ha a modális osztályköz a legelső osztályköz, akkor $f_{Mo-1}= g_{Mo-1}=0$. Ha a modális osztályköz a legutolsó osztályköz, akkor $f_{Mo+1}= g_{Mo+1}=0$.

Észrevehető, hogy a módusz kiszámításához pontosan azokat a mennyiségeket használjuk, amelyeket a hisztogram készítésekor használtunk.

Példa: módusz becslése osztályközös gyakorisági sorból

Egy ABC vásárlóinak megoszlása vásárlási érték szerint

Vásárlások értéke (Ft)	Vásárlások száma, f_i	h_i	$\frac{f_i}{h_i}$
– 400	56	400	0,14
401 – 800	89	400	0,2225
801 – 1200	250	400	0,625
1201 – 2000	260	800	0,325
2001 – 2800	210	800	0,2625
2801 – 4000	84	1200	0,07
Összesen	949	-	-

$$\hat{Mo} = X_{Mo,0} + \frac{\frac{f_{Mo}}{h_{Mo}} - \frac{f_{Mo-1}}{h_{Mo-1}}}{\left(\frac{f_{Mo}}{h_{Mo}} - \frac{f_{Mo-1}}{h_{Mo-1}}\right) + \left(\frac{f_{Mo}}{h_{Mo}} - \frac{f_{Mo+1}}{h_{Mo+1}}\right)} \cdot h_{Mo} =$$

$$= 800 + \frac{\frac{250}{400} - \frac{89}{400}}{\left(\frac{250}{400} - \frac{89}{400}\right) + \left(\frac{250}{400} - \frac{260}{800}\right)} \cdot 400 =$$

$$= 800 + \frac{0{,}625 - 0{,}2225}{(0{,}625 - 0{,}2225) + (0{,}625 - 0{,}325)} \cdot 400 = 1029{,}2 \text{ Ft.}$$

Tehát a leggyakoribb (becsült) vásárlási érték 1029,2 Ft.

A **medián** (jelölése: *Me)* az az érték, aminél az összes előforduló ismérvérték legalább fele nem nagyobb, és legalább fele nem kisebb. Gyakorlatilag a medián két egyenlő részre osztja a sokaságot.
A medián meghatározásához **első lépésben rangsort** kell készítenünk. Amennyiben az ismérvértékek N száma páratlan, akkor a medián a rangsor középső, azaz az $\frac{N+1}{2}$-edik eleme lesz. Amennyiben az ismérvértékek N száma páros, akkor a medián a két középső, azaz az $\frac{N}{2}$ -edik és $\frac{N}{2}+1$-edik ismérvértékek átlaga:

$$Me = \frac{x_{N/2} + x_{(N/2)+1}}{2}.$$

Példa: medián meghatározása rangsorból

Ismérvértékek rangsora: 10, 11, 12, 12, 13, 13, 13. Ekkor *Me*=12 (N=7, azaz a medián a rangsor 4. (7+1)/2 eleme lesz).
Ismérvértékek rangsora: 10, 11, 12, 12, 13, 13, 13, 15. Ekkor *Me*=12,5. (N=8, azaz a medián a rangsor két középső elemének átlaga lesz).

Amennyiben nem osztályközös sorból kell meghatároznunk a mediánt, akkor is a fenti eljárást kell követnünk, hiszen egy ilyen mennyiségi sor a rangsor tömörített írásmódja. Így, a medián kiszámításában a felfelé kumulált gyakoriságok (relatív gyakoriságok) lehetnek a segítségünkre.

Példa: medián meghatározása nem osztályközös gyakorisági sorból

Egy lakótelepi panelház háztartásai taglétszám szerint

Taglétszám (fő), x_i	Háztartások száma (db), f_i	f'_i
1	5	5
2	10	15
3	7	22
4	25	47
5	2	49
6	1	50
Összesen	**50**	–

Forrás: fiktív adatok

Mivel 50, azaz páros elemünk van, ezért a medián a rangsor két középső elemének, azaz a 25. és a 26. elem átlaga lesz. A 25. és a 26. elem értékét a kumulált gyakoriságok alapján adhatjuk meg. Ugyanis a kumulált gyakoriságokat úgyis megragadhatjuk, hogy az első 5 elem egyes, a hatodiktól a tizenötödik elemig kettesek vannak; 16-22. elemek hármasok, míg 23-47. elemek 4-esek, azaz a 25. és a 26. elem is 4.

$$Me = \frac{x_{25.elem} + x_{26.elem}}{2} = \frac{4+4}{2} = 4\, fő.$$

Tehát a vizsgált panelházban a háztartások fele legfeljebb 4 fős, fele pedig legalább 4 fős.

Osztályközös gyakorisági sor esetén a medián értéke csak becsülhető. **Először** azt kell meghatároznunk, **hogy melyik osztályközbe** esik a medián. A medián definíciójából és kiszámításából adódóan a medián abban az osztályközben lesz, ahol a kumulált gyakoriságok értéke először lesz legalább $\frac{1}{2} \cdot N$,illetve a kumulált relatív gyakoriságok értéke először lesz legalább $\frac{1}{2}$. A medián durva becslésének tekinthető ennek az osztályköznek az osztályközepe, amelyet **nyers mediánnak** nevezzük. A medián értékét finomabban is lehet becsülni a

$$\hat{Me} = X_{Me,0} + \frac{\frac{1}{2} \cdot N - f'_{Me-1}}{f_{Me}} \cdot h_{Me} = X_{Me,0} + \frac{0{,}5 - g'_{Me-1}}{g_{Me}} \cdot h_{Me}$$

képlettel, ahol

$X_{Me,0}$: a mediánt tartalmazó osztályköz alsó határa,
f'_{Me-1}: a mediánt tartalmazó osztályköz előtti osztályköz kumulált gyakorisága,
f_{Me}: a mediánt tartalmazó osztályköz gyakorisága,
h_{Me}: a mediánt tartalmazó osztályköz hossza.
g'_{Me-1}: a mediánt tartalmazó osztályköz előtti osztályköz kumulált relatív gyakorisága,
g_{Me}: a mediánt tartalmazó osztályköz relatív gyakorisága.

Példa: medián becslése osztályközös gyakorisági sorból

Egy ABC vásárlóinak megoszlása vásárlási érték szerint

Vásárlások értéke (Ft)	Vásárlások száma, f_i	h_i	f_i'
– 400	56	400	56
401 – 800	89	400	145
801 – 1200	250	400	395
1201 – 2000	260	800	655
2001 – 2800	210	800	865
2801 – 4000	84	1200	949
Összesen	949	-	–

A medián abban az osztályközben van, ahol a felfelé kumulált gyakoriság először nagyobb, mint $\frac{1}{2} \cdot N = \frac{1}{2} \cdot 949 = 474{,}5$. Az első ilyen kumulált gyakoriság 655, tehát a medián az 1200-2000 tartományban helyezkedik el.

$$\hat{Me} = X_{Me,0} + \frac{\frac{1}{2} \cdot N - f'_{Me-1}}{f_{Me}} \cdot h_{Me} = 1200 + \frac{474{,}5 - 395}{260} \cdot 800 = 1444{,}615 \text{ Ft.}$$

Tehát a vásárlási értékek fele legfeljebb 1444,615 Ft.

A medián általánosításaként tekinthetőek a **kvantilisek**, amelyek a sokaságot *k* egyenlő részre osztják. Ahhoz, hogy egy sokaságot *k* részre daraboljunk fel, *k-1* osztópontra van szükségünk. A leggyakrabban használt kvantilisek az alábbiak.

Nevezetes kvantilisek

k	Neve	Osztópontok száma	Osztópontok jelölése
2	Medián	1	*Me*
3	Tercilis	2	T_1, T_2
4	Kvartilis	3	Q_1, Q_2, Q_3
10	Decilis	9	$D_1, D_2, \ldots, D_9$
100	Percentilis	99	P_1, P_2, P_{99}

A kvantiliseket, amelyek többek között egy változó eloszlásának jellemzésére is szolgálnak, általánosan $x_{i/k}$ szimbólummal is szokták jelölni, ahol *k* azt jelöli, hogy az adott kvantilisek hány részre osztják a sokaságot, *i* pedig azt mutatja meg, hogy az adott kvantilisen belül hányadik osztópontot keressük. Például Q_3 harmadik, vagy más néven felső kvartilis jelölése $x_{3/4}$, hiszen a kvartilisek négy részre osztják a sokaságot (*k*=4), ezen belül pedig a harmadik osztópontot (*i*=3) keressük.

Amennyiben rangsorból vagy nem osztályközös gyakorisági, relatív gyakorisági sorból kell meghatároznunk a keresett kvantilist, a medián kiszámításához hasonlóan a felfelé kumulált gyakoriságok (relatív gyakoriságok) lehetnek a segítségünkre.

Első lépésben azt kell meghatároznunk, hogy a keresett kvantilis hányadik elem a rangsorban. Ezt az

$$s_{i/k} = \frac{i}{k}(N+1)$$

képlettel tehetjük meg.

Ezután magát a kvantilist az

$$x_{i/k} = x_{[s_{i/k}]} + \{s_{i/k}\} \cdot \left(x_{[s_{i/k}]+1} - x_{[s_{i/k}]}\right)$$

formulával számíthatjuk ki, ahol a képletben szereplő [ ] szimbólum egy szám egész részét, míg a { } egy szám tört részét jelölik. Hogyan működik ez az eljárás a gyakorlatban?

Példa: kvantilisek meghatározása rangsorból

Ismérvértékek rangsora: 10, 11, 12, 12, 13, 13, 13.
Határozzuk meg T_2 értékét!
1. lépés: T_2 hányadik elem a rangsorban?

$s_{2/3} = \frac{2}{3}(7+1) = 5{,}33$-dik elem a második tercilis a rangsorban.

2. lépés: Határozzuk meg T_2 értékét!

$T_2 = x_{2/3} = 5.elem + 0{,}33 \cdot (6.elem - 5.elem) = 13 + 0{,}33 \cdot (13-13) = 13.$

Határozzuk meg Q_1 értékét!
1. lépés: Q_1 hányadik elem a rangsorban?

$s_{1/4} = \frac{1}{4}(7+1) = 2,0$ -dik elem az első kvartilis a rangsorban.

2. lépés: Határozzuk meg Q_1 értékét!

$Q_1 = x_{1/3} = 2.elem + 0,0 \cdot (3.elem - 2.elem) = 11 + 0,00 \cdot (12 - 11) = 11.$

Példa: kvantilisek meghatározása nem osztályközös gyakorisági sorból

Egy lakótelepi panelház háztartásai taglétszám szerint

Taglétszám (fő), x_i	Háztartások száma (db), f_i	f'_i
1	5	5
2	10	15
3	7	22
4	25	47
5	2	49
6	1	50
Összesen	**50**	–

Forrás: fiktív adatok

Határozzuk meg a T_1, Q_2, Q_3 értékét!

Határozzuk meg T_1 értékét!
1. lépés: T_1 hányadik elem a rangsorban?

$s_{1/3} = \frac{1}{3}(50+1) = 17,00$ -dik elem az első tercilis a rangsorban.

2. lépés: Határozzuk meg T_1 értékét!

$T_1 = x_{1/3} = 17.elem + 0,00 \cdot (18.elem - 17.elem) = 3 + 0,00 \cdot (3 - 3) = 3.$

Tehát a vizsgált panelházban a háztartások harmada legfeljebb 3 fős, kétharmada pedig legalább 3 fős.

Határozzuk meg Q_2 értékét!
A második kvartilis előtt az ismérvértékek 50 százaléka található, csak úgy, mint a medián előtt, mivel $Me=Q_2$.

Határozzuk meg Q_3 értékét!
1. lépés: Q_3 hányadik elem a rangsorban?

$s_{3/4} = \frac{3}{4}(50+1) = 38,25$ -dik elem a harmadik kvartilis a rangsorban.

2. lépés: Határozzuk meg Q_3 értékét!

$Q_3 = x_{3/4} = 38.elem + 0,25 \cdot (39.elem - 38.elem) = 4 + 0,25 \cdot (4 - 4) = 4.$

Tehát a vizsgált panelházban a háztartások háromnegyede legfeljebb 4 fős, negyede pedig legalább 4 fős.

Osztályközös gyakorisági sor esetén a kvantilisek értéke csak becsülhető. **Először** azt kell meghatároznunk, **hogy melyik osztályközbe** esik a keresett osztópont. Ez abban az osztályközben lesz, ahol a kumulált gyakoriságok értéke először lesz legalább $\frac{i}{k} \cdot N$,illetve a kumulált relatív gyakoriságok értéke először lesz legalább $\frac{i}{k}$. A medián becsléséhez analóg módon a kvantilisek értékét lehet becsülni az

$$\hat{x}_{i/k} = X_{i/k,0} + \frac{\frac{i}{k}N - f'_{(i/k)-1}}{f_{i/k}} \cdot h_{i/k} = X_{i/k,0} + \frac{0{,}5 - g'_{(i/k)-1}}{g_{i/k}} \cdot h_{i/k}$$

képlettel, ahol

$X_{i/k,0}$: az adott kvantilist (osztópontot) tartalmazó osztályköz alsó határa,

$f'_{i/k-1}$: az adott kvantilist (osztópontot) tartalmazó osztályköz előtti osztályköz kumulált gyakorisága,

$f_{i/k}$: az adott kvantilist (osztópontot) tartalmazó osztályköz gyakorisága,

$h_{i/k}$: az adott kvantilist (osztópontot) tartalmazó osztályköz hossza.

$g'_{i/k-1}$: az adott kvantilist (osztópontot) tartalmazó osztályköz előtti osztályköz kumulált relatív gyakorisága,

$g_{i/k}$: az adott kvantilist (osztópontot) tartalmazó osztályköz relatív gyakorisága.

Példa: kvantilisek becslése osztályközös gyakorisági sorból

Egy ABC vásárlóinak megoszlása vásárlási érték szerint

Vásárlások értéke (Ft)	Vásárlások száma, f_i	h_i	f_i'
– 400	56	400	56
401 – 800	89	400	145
801 – 1200	250	400	395
1201 – 2000	260	800	655
2001 – 2800	210	800	865
2801 – 4000	84	1200	949
Összesen	949	-	–

Határozzuk meg az első kvartilis, a harmadik kvartilis és a második decilis értékét.

Az első kvartilis abban az osztályközben van, hol a felfelé kumulált gyakoriság először nagyobb, mint $\frac{1}{4} \cdot N = \frac{1}{4} \cdot 949 = 237{,}25$. Az első ilyen kumulált gyakoriság 395, tehát a medián az 800-1200 tartományban helyezkedik el.

$$\hat{Q}_1 = \hat{x}_{1/4} = X_{1/4,0} + \frac{\frac{1}{4}N - f'_{(1/4)-1}}{f_{1/4}} \cdot h_{1/4} == 800 + \frac{949 \cdot \frac{1}{4} - 145}{250} \cdot 400 = 947{,}6Ft$$

Tehát a vásárlási értékek alacsonyabbik 25 százaléka (negyede) legfeljebb 947,6 Ft.

A harmadik kvartilis abban az osztályközben van, hol a felfelé kumulált gyakoriság először nagyobb, mint $\frac{3}{4} \cdot N = \frac{3}{4} \cdot 949 = 711{,}75$. Az első ilyen kumulált gyakoriság 865, tehát a medián az 2000-2800 tartományban helyezkedik el.

$$\hat{Q}_3 = \hat{x}_{3/4} = X_{3/4,0} + \frac{\frac{3}{4}N - f'_{(3/4)-1}}{f_{3/4}} \cdot h_{3/4} = 2000 + \frac{949 \cdot \frac{3}{4} - 655}{210} \cdot 800 = 2216{,}2Ft$$

Tehát a vásárlási értékek alacsonyabbik háromnegyede legfeljebb (magasabbik negyede legalább) 2216,6 Ft.

A második decilis abban az osztályközben van, hol a felfelé kumulált gyakoriság először nagyobb, mint $\frac{2}{10} \cdot N = \frac{2}{10} \cdot 949 = 189{,}8$. Az első ilyen kumulált gyakoriság 395, tehát a medián az 800-1200 tartományban helyezkedik el.

$$\hat{D}_2 = \hat{x}_{2/10} = X_{2/10,0} + \frac{\frac{2}{10}N - f'_{(2/10)-1}}{f_{2/10}} \cdot h_{2/10} = 800 + \frac{\frac{2}{10} \cdot 949 - 145}{250} \cdot 400 = 871{,}7Ft$$

Tehát a vásárlási értékek alacsonyabbik húsz százaléka (kéttizede) legfeljebb 817,7 Ft.

A helyzeti középértékek fontosabb tulajdonságai:

1. Ha mindenegyes ismérvértékhez ugyanazt az *A* számot hozzáadjuk, akkor a módusz, a medián és a kvantilisek is *A*-val nagyobbak lesznek a korábbi értéküknél.

2. Ha mindenegyes ismérvértéket ugyanazzal a nullától különböző *A* számmal megszorozzuk, akkor a módusz, a medián és a kvantilisek értéke is az A-szorosa lesz a korábbi értékének.

3. Minden ismérvérték mediánnal való helyettesítésekor elkövetett hibák abszolút értékének összege minimális lesz és fordítva: $\sum_{i=1}^{N} |x_i - A| \to \min \Leftrightarrow A = Me.$

4. Mintafeladatok

1. feladat

A feladat megoldásához készítsük el az alkalmazottak havi nettó bérének rangsorát.

Az alkalmazottak havi nettó bérének rangsora (ezer Ft)

40	43	49	52	53	54	56	56	57	58	59	62	63
65	65	66	66	66	67	67	68	68	68	68	69	69
70	70	71	72	72	72	73	73	73	73	74	75	75
76	76	77	77	77	78	78	79	79	79	80	80	80
81	82	82	82	83	83	83	84	84	85	85	86	86
87	87	87	87	87	87	88	88	88	89	90	91	91
92	92	92	92	92	92	92	92	93	93	93	93	93
93	93	93	94	94	94	94	94	94	94	95	95	95
95	95	95	95	95	96	96	96	96	97	98	98	98
98	99	99	99	99	99	99	100	100	100	100	100	101
102	102	102	102	103	103	103	104	104	104	104	105	105
105	106	106	106	106	107	107	108	108	108	109	109	110
110	110	111	112	112	112	112	113	113	113	113	114	114
114	114	114	115	117	117	117	117	118	118	119	119	120
120	120	120	121	122	122	122	122	123	124	124	125	125
126	126	127	127	127	127	127	128	129	129	129	130	130
130	130	131	131	131	133	134	134	134	134	134	135	136
137	137	138	138	138	139	139	139	140	141	141	141	144
144	145	147	148	148	149	151	151	153	153	154	154	154
154	156	156	159	160	160	162	167	173	173	175	179	180

A) A közölt adatok alapján számítsa ki és értelmezze az alkalmazottak havi átlagos nettó bérét!

Mivel az ismérvértékek összegének van tárgyi értelme, ezért számtani átlagot kell használnunk.

$$\bar{x} = \frac{\text{havi nettó bérek összege}}{\text{alkalmazottak száma}} = \frac{40 + 43 + 49 + \ldots + 180}{260} = 105{,}17 \text{ ezer Ft.}$$

Ezek szerint az alkalmazottak havi átlagos nettó bére 105,17 ezer Ft.

B) A közölt adatok alapján határozza meg és értelmezze a móduszt!

A módusz a leggyakrabban előforduló ismérvérték. Ezért meg kell határoznunk, hogy mely ismérvérték, esetleg ismérvértékek fordulnak elő a leggyakrabban.
Mivel három ismérvérték – 92, 93, 95 – fordul elő a leggyakrabban (nyolcszor), ezért a vizsgált sokaság eloszlása több móduszú.
Ezek szerint az alkalmazottak leggyakrabban előforduló havi nettó bérei 92, 93, 95 ezer forint.

C) A közölt adatok alapján számítsa ki és értelmezze a mediánt!

Mivel páros számú adatunk van, ezért a medián a rangsor két középső elemének átlaga lesz:

$$Me = \frac{130.\,\text{elem} + 131.\,\text{elem}}{2} = \frac{101+102}{2} = 101{,}5 \text{ ezer Ft.}$$

Ezek szerint a rangsorolt havi nettó bérek alacsonyabbik fele legfeljebb 101,5 ezer Ft, míg a magasabbik fele legalább 101,5 ezer Ft.

D) A közölt adatok alapján számítsa ki és értelmezze a kvartiliseket!

A kvantilisiek és ezért a kvartilisek is az

$$s_{i/k} = \frac{i}{k}(N+1);$$

$$x_{i/k} = x_{[s_{i/k}]} + \{s_{i/k}\} \cdot \left(x_{[s_{i/k}]+1} - x_{[s_{i/k}]}\right)$$

képletek alapján számíthatók ki.

A kvartilisek a sokaságot négy egyenlő részre osztják. Ebből következően, az alsó (első) kvartilis $s_{1/4} = \frac{1}{4}(260+1) = 65{,}25$ -ik eleme a rangsornak, ezért

$$Q_1 = x_{1/4} = 65.\text{ elem} + 0{,}25(66.\text{ elem} - 65.\text{ elem}) =$$

$$= 86 + 0{,}25 \cdot (87-86) = 86{,}25 \text{ ezer Ft.}$$

A középső (második) kvartilis $s_{2/4} = \frac{2}{4}(260+1) = 130{,}5$ -ik eleme a rangsornak, ezért

$$Q_2 = x_{2/4} = 130.\text{ elem} + 0{,}5(131.\text{ elem} - 130.\text{ elem}) =$$

$$= 101 + 0{,}5 \cdot (102-101) = 101{,}5 \text{ ezer Ft.}$$

A harmadik (felső) kvartilis $s_{3/4} = \frac{3}{4}(260+1) = 195{,}75$-ik eleme a rangsornak, ezért

$$Q_3 = x_{3/4} = 195.\text{ elem} + 0{,}75(196.\text{ elem} - 195.\text{ elem}) =$$
$$= 125 + 0{,}75 \cdot (126-125) = 125{,}75 \text{ ezer Ft.}$$

Tehát a havi nettó bérek legkisebb 25 százaléka legfeljebb 86,25 ezer forint, míg a legmagasabb 25 százaléka legalább 125,75 ezer forint. A havi nettó bérek legkisebb 50 százaléka legfeljebb 101,5 ezer forint.

E) A közölt adatok alapján számítsa ki és értelmezze a 8. decilist!

A feladat analóg az előző kérdéssel.

A 8. decilis $s_{8/10} = \frac{8}{10}(260+1) = 208{,}8$ -ik eleme a rangsornak, ezért

$$D_8 = x_{8/10} = 208.\ \text{elem} + 0{,}8(209.\ \text{elem} - 208.\ \text{elem}) =$$

$$= 130 + 0{,}8 \cdot (130 - 130) = 130\ \text{ezer Ft.}$$

Ezek szerint a havi nettó bérek legmagasabb 20 százaléka legalább 130 ezer Ft.

2. feladat

Ismertek egy település lakóira vonatkozó adatok: 890 családban nincs, 950 családban 1, 650 családban 2, 140 családban 3, 40 családban 4 gyermek van. Ötgyermekes családból 15, hatgyermekes családból 1 van.

A) Készítsen gyakorisági sort az adatok alapján!

A családok gyerekszám szerinti eloszlása a vizsgált településen

Gyerekszám (fő)	Családok száma (fő)
0	890
1	950
2	650
3	140
4	40
5	15
6	1
Összesen	2 686

Forrás: fiktív

B) A közölt adatok alapján számítsa ki és értelmezze a számtani átlagot!

A számtani átlag ebben az esetben azt jelenti, hogy 1 családban átlagosan hány gyermek van.

A feladat megoldásához az adatokból egy munkatáblát készítünk.

x_i	f_i	$x_i\ f_i$
0	890	0

1	950	950
2	650	1 300
3	140	420
4	40	160
5	15	75
6	1	6
Összesen	2 686	2 911

$$\bar{x}_a = \frac{\sum_{i=1}^{7} f_i x_i}{\sum_{i=1}^{7} f_i} = \frac{\sum_{i=1}^{7} f_i x_i}{N} = \frac{2911}{2686} = 1{,}08$$

Ez azt jelenti, hogy 100 családban átlagosan 108 gyermek él a vizsgált településen.

C) A közölt adatok alapján számítsa ki és értelmezze a móduszt!

A módusz a leggyakrabban előforduló ismérvérték, azaz *Mo*=1 gyermek. Tehát a vizsgált településen az 1 gyermekes család a leggyakoribb (950 család).

D) A közölt adatok alapján számítsa ki és értelmezze a mediánt!

A feladat megoldásához az adatokból egy munkatáblát készítünk.

x_i	f_i’
0	890
1	1 840
2	2 490
3	2 630
4	2 670
5	2 685
6	2 686
Összesen	-

Mivel páros számú elemű megfigyelésünk van (*N*=2686), ezért a medián a két középső elem átlaga, azaz 1343. és az 1344. elemek átlaga lesz. A kumulált gyakoriságok alapján mindkét elem 1, azaz *Me*=1 gyermek.

A családok fele legfeljebb 1 gyermekes, a másik fele legalább 1 gyermekes.

E) A közölt adatok alapján számítsa ki és értelmezze az alsó kvartilis és a felső kvartilis értékét!

Az alsó kvartilis az $s_{1/4} = \frac{1}{4}(2686+1) = 671{,}75$ -ik eleme a rangsornak, azaz

$$Q_1 = 671.\text{ elem} + 0{,}75(672.\text{ elem} - 671.\text{ elem}) =$$

$$= 0 + 0{,}75(0 - 0) = 0 \text{ gyermek;}$$

a felső kvartilis az $s_{3/4} = \frac{3}{4}(2686 + 1) = 2015{,}25$-ik eleme a rangsornak, azaz

$$Q_3 = 2015.\text{ elem} + 0{,}25(2016.\text{ elem} - 2015.\text{ elem}) =$$

$$= 2 + 0{,}25(2 - 2) = 2 \text{ gyermek.}$$

Ha a település összes családját a családonkénti gyermekek száma szerint rangsoroljuk, akkor a rangsor első negyedéhez tartozó családokban nincs gyermek, míg a családok legtöbb gyermekes 25%-a legalább 2 gyermekes.

3. feladat
Az 1. feladat adatai alapján
A) Készítsen az adatokból azonos hosszúságú osztályközös gyakorisági sort!

Az azonos hosszúságú osztályközös gyakorisági sor készítéséhez először meg kell határoznunk az osztályközök számát. Egy ajánlás – tehát nem kötelező szabály – szerint az N adatból készítendő osztályközök k száma az a legkisebb egész szám, amely teljesíti a $2^k > N$ összefüggést. Ez azt jelentené, hogy kilenc osztályközt kellene használnunk. Azonban, $2^8 = 256 < 260$ értékek közelsége miatt nem elvetendő gondolat az, hogy nyolc osztályközt vegyünk.

Ezek után határozzuk meg az osztályközök hosszát.
Ha kilenc osztállyal kalkulálunk, akkor

$$h = \frac{x_{\max} - x_{\min}}{k} = \frac{180 - 40}{9} = 15{,}\dot{5}\,;$$

viszont nyolc osztály használata esetén

$$h = \frac{x_{\max} - x_{\min}}{k} = \frac{180 - 40}{8} = 17{,}5\,.$$

Teljesen önkényesen, pusztán kényelmi okokból, használjunk nyolc osztályközt. Ezek alapján felírhatjuk az osztályközös gyakorisági sor szerkezetét a valódi határaival.

Havi nettó bérek (ezer Ft)	Alkalmazottak száma (fő)
40,0 – 57,5	
57,5 – 75,0	
75,0 – 92,5	
92,5 – 110,0	
110,0 – 127,5	

127,5 – 145,0	
145,0 – 162,5	
162,5 – 180,0	
Összesen	260

A közölt határokra egy lehetséges megoldás az alábbi.

Az alkalmazottak havi nettó bérének megoszlása

Havi nettó bérek (ezer Ft)	Alkalmazottak száma (fő)
40,0 – 57,4	9
57,5 – 74,9	30
75,0 – 92,4	47
92,5 – 109,9	72
110,0 – 127,4	44
127,5 – 144,9	34
145,0 – 162,4	18
162,5 – 180,0	6
Összesen	260

B) Az A) pont alapján számítsa ki és értelmezze az alkalmazottak havi átlagos nettó bérét!

A – súlyozott – számtani átlagot csak becsülni tudjuk, hiszen nem ismerjük az egyes ismérvértékek nagyságát. Mindegyik osztályból kiválasztunk egy reprezentánst, ez lesz az x_i osztályközép. Az egyes osztályokba tartozó ismérvértékeket az osztályközepükkel helyettesítjük. Az osztályközepek a valódi alsó és felső határok számtani átlagai.

Havi nettó bérek (ezer Ft)	f_i	x_i	$f_i x_i$
40,0 – 57,4	9	48,75	438,75
57,5 – 74,9	30	66,25	1 987,50
75,0 – 92,4	47	83,75	3 936,25
92,5 – 109,9	72	101,25	7 290,00
110,0 – 127,4	44	118,75	5 225,00
127,5 – 144,9	34	136,25	4 632,50
145,0 – 162,4	18	153,75	2 767,50
162,5 – 180,0	6	171,25	1 027,50
Összesen	260	–	27 305

$$\bar{x}_a = \frac{\sum_{i=1}^{8} f_i x_i}{\sum_{i=1}^{8} f_i} = \frac{27305}{260} \approx 105 \text{ ezer Ft.}$$

Ezek szerint az alkalmazottak havi nettó átlagbérének becsült értéke 105 ezer Ft.

C) Az A) pont alapján számítsa ki és értelmezze a mediánt!

A medián becslése az

$$\hat{M}e = X_{Me,0} + \frac{\frac{N}{2} - f'_{Me-1}}{f_{Me}} \cdot h_{Me}$$

képlettel történik. Az '*Me*' alsó index arra az osztályközre utal, amelyikben a medián elhelyezkedik. Ezt a kumulált gyakoriságokból állapíthatjuk meg.

Havi nettó bérek (ezer Ft)	Alkalmazottak száma (fő)	f'_i
40,0 – 57,4	9	9
57,5 – 74,9	30	39
75,0 – 92,4	47	86
92,5 – 109,9	72	158
110,0 – 127,4	44	202
127,5 – 144,9	34	236
145,0 – 162,4	18	254
162,5 – 180,0	6	260
Összesen	260	–

A medián becslése a 130. elem lesz. Kérdés, hogy mely osztályközben van a 130. elem. A kumulált gyakoriságok alapján tudjuk, hogy ez a 92,5-109,9 tartomány.
Tehát

$$\hat{M}e = 92{,}5 + \frac{\frac{260}{2} - 86}{72} \cdot 17{,}5 = 103{,}2 \text{ ezer Ft.}$$

A havi nettó bérek alacsonyabbik fele legfeljebb 103,2 ezer Ft.

D) Az A) pont alapján számítsa ki és értelmezze a kvartiliseket!

Abban az esetben, ha a kvantilisek becsült értékét kell kiszámítanunk, az alábbi képletet használhatjuk.

$$\hat{x}_{i/k} = X_{i/k,0} + \frac{\frac{i}{k}N - f'_{(i/k)-1}}{f_{i/k}} \cdot h_{i/k}$$

Az alsó kvartilis becslése a 65. elem lesz. Kérdés, hogy mely osztályközben van a 65. elem. A kumulált gyakoriságok alapján tudjuk, hogy ez a 75,0-92,4 tartomány.
Tehát

$$\hat{Q}_1 = \hat{x}_{1/4} = 75{,}0 + \frac{\frac{1}{4}260 - 39}{47} \cdot 17{,}5 = 84{,}7 \text{ ezer Ft.}$$

A havi nettó bérek legalacsonyabb negyede legfeljebb 84,7 ezer Ft.

A középső kvartilis értéke megegyezik a mediánnal. Ezt pedig az előző kérdésben vizsgáltuk.

A felső kvartilis becslése a 195. elem lesz. Kérdés, hogy mely osztályközben van a 195. elem. A kumulált gyakoriságok alapján tudjuk, hogy ez a 110,0-127,4 tartomány.
Tehát

$$\hat{Q}_3 = \hat{x}_{3/4} = 110{,}0 + \frac{\frac{3}{4}260 - 158}{44} \cdot 17{,}5 = 124{,}7 \text{ ezer Ft.}$$

A havi nettó bérek legmagasabb negyede legalább 124,7 ezer Ft.

E) Az A) pont alapján számítsa ki és értelmezze a 8. decilist!

A 8. decilis becslése a 208. elem lesz. Kérdés, hogy mely osztályközben van a 208. elem. A kumulált gyakoriságok alapján tudjuk, hogy ez a 127,5-144,9 tartomány.
Tehát

$$\hat{D}_8 = \hat{x}_{8/10} = 127{,}5 + \frac{\frac{8}{10}260 - 202}{34} \cdot 17{,}5 = 130{,}6 \text{ ezer Ft.}$$

A havi nettó bérek legmagasabb 20 százaléka legalább 130,6 ezer Ft.

F) Készítse el az adatok kvartiliseloszlását!

A kvartiliseloszlás elkészítéséhez számítsuk ki a rangsorból a kvartilisek értékét.

$s_{1/4} = \frac{1}{4}(260+1) = 65{,}25 \quad \rightarrow Q_1 = x_{1/4} = 65.$ elem + 0,25(66. elem – 65. elem)=
$= 86 + 0{,}25 \cdot (87 - 86) = 86{,}25$ ezer Ft.

$s_{2/4} = \frac{2}{4}(260+1) = 130{,}5 \rightarrow Q_2 = x_{2/4} = 130.$ elem + 0,5(131. elem – 130. elem)= $= 101 + 0{,}5 \cdot (102 - 101) = 101{,}5$ ezer Ft.

$s_{3/4} = \frac{3}{4}(260+1) = 195{,}75 \rightarrow Q_3 = x_{3/4} = 195.$ elem + 0,75(196. elem – 195. elem)= $= 125 + 0{,}75 \cdot (126 - 125) = 125{,}75$ ezer Ft.

Az alkalmazottak havi nettó bérének
kvartiliseloszlása

Havi nettó bérek (ezer Ft)	Alkalmazottak száma (fő)
40,00 – 86,25	65
86,25 – 101,50	65
101,50 – 125,75	65
125,75 – 180,00	65
Összesen	260

G) Az F) pont alapján számítsa ki és értelmezze az alkalmazottak havi átlagos nettó bérét!

A számtani átlag becslése mindegyik osztályközös gyakorisági sor esetén ugyanarra – a már tárgyalt – eljárásra épül.

Havi nettó bérek (ezer Ft)	f_i	x_i	$f_i x_i$
40,00 – 86,25	65	63,125	4 103,125
86,25 – 101,50	65	93,875	6 101,875
101,50 – 125,75	65	113,625	7 385,625
125,75 – 180,00	65	152,875	9 936,875
Összesen	260	–	27 527,500

Tehát

$$\bar{x}_a = \frac{\sum_{i=1}^{4} f_i x_i}{\sum_{i=1}^{4} f_i} = \frac{27527{,}5}{260} = 105{,}875 \text{ ezer Ft.}$$

Ezek szerint az alkalmazottak havi nettó átlagkeresetének becsült értéke 105,875 ezer Ft.

H) Az F) pont alapján határozza meg és értelmezze a móduszt!

A módusz becsléséhez azt kellene tudnunk, hogy melyik osztályközbe esik. A módusz abban az osztályközben van, amelyikben a legjobban sűrűsödnek az ismérvértékek. Különböző hosszúságú osztályközök esetén az adatok sűrűségét a gyakoriságok és az osztályköz hosszok hányadosa adja.

Havi nettó bérek (ezer Ft)	f_i	h_i	$\frac{f_i}{h_i}$
40,00 – 86,25	65	46,25	1,405
86,25 – 101,50	65	15,25	4,262
101,50 – 125,75	65	24,25	2,680
125,75 – 180,00	65	54,25	1,198
Összesen	260	–	–

Ezek szerint

$$\hat{Mo} = 86{,}25 + \frac{4{,}262 - 1{,}405}{(4{,}262 - 1{,}405) + (4{,}262 - 2{,}68)} \cdot 15{,}25 = 96{,}1 \text{ ezer Ft.}$$

Tehát az alkalmazottak leggyakoribb havi nettó bérének becsült értéke 96,1 ezer Ft.

5. Ellenőrző kérdések

1. Csoportosítsa a középértékeket és adjon példát mindegyik típusra!
2. Melyik számított középértéket mikor alkalmazhatjuk?
3. Definiálja a számtani átlagot! Hogyan lehet kiszámítani?
4. Sorolja fel a számtani átlag tulajdonságait!
5. Definiálja a mértani átlagot! Hogyan lehet kiszámítani?
6. Definiálja a harmonikus átlagot! Hogyan lehet kiszámítani?
7. Definiálja a négyzetes átlagot! Hogyan lehet kiszámítani?
8. Milyen reláció áll fenn a számított középértékek és az ismérvértékek között?
9. Részekre bontott sokaság esetében hogy számítható ki a főátlag?
10. Mi a módusz? Hogyan lehet kiszámítani?
11. Mi a modális osztályköz?
12. Mi a nyers módusz?
13. Mi a medián?
14. Hogyan lehet a mediánt kiszámítani?
15. Mi a kvantilis? Sorolja fel a nevezetes kvantiliseket!
16. Hogyan értelmezné a kvartiliseket?
17. Tegye növekvő sorrendbe az alábbi kvantiliseket: D_8, T_2, P_5, Me, Q_1!
18. Sorolja fel a helyzeti középértékek fontosabb tulajdonságait!
19. Mi a kvantiliseloszlás?

JELEN TANANYAG
A SZEGEDI TUDOMÁNYEGYETEMEN KÉSZÜLT
AZ EURÓPAI UNIÓ TÁMOGATÁSÁVAL.
PROJEKT AZONOSÍTÓ: EFOP-3.4.3-16-2016-00014

SZÉCHENYI 2020
MAGYARORSZÁG KORMÁNYA
Európai Unió
Európai Szociális Alap
BEFEKTETÉS A JÖVŐBE