STATISZTIKA I.

T.Nagy Judit

tnagy.judit@hjf.hu

Ajánlott irodalom:

- Korpás Attiláné (szerk.): Általános statisztika I.; Nemzeti Tankönyvkiadó, Budapest, 1997.
- Molnár Máténé - Tóth Mártonné: Általános statisztika példatár I.; Nemzeti Tankönyvkiadó, Budapest, 2001.

Követelmény: Gyakorlati jegy

- Ennek feltétele:
  - ZH megírása
  - Időpontja: 2008. május 10. szombat
    - Pótlás: 2008. május 16. péntek
  - (A zárthelyin csak kiadott képletgyűjtemény és számológép használható.)

A ponthatárok:

- -49% elégtelen
- 50-61% elégséges
- 62-73% közepes
- 74-85% jó
- 86-100% jeles

I. A STATISZTIKA ALAPFOGALMAI
A statisztika tárgya és szerepe

A **statisztika**

- tömegesen előforduló jelenségekre vonatkozó információk gyűjtése, feldolgozása, elemzése, közlése;
- Maguk az adatok;
- Tudományos módszertan.

Módszertan

Leíró statisztika
(deskriptív)
Jelenségek tömör,
átfogó jellemzése.

Matematikai (következtető)
statisztika
(inferencia)
Mintából való következtetés,
Általánosítás.

I./1Alapfogalmak

Sokaság: A megfigyelés tárgyát képező egyedek összessége.

Legkisebb részei az **egységek**, melyek lehetnek

- élőlények (pl. egy főiskola hallgatói),
- tárgyak (pl. kórházi ágyak),
- szervezetek (pl. magyar főiskolák),
- képzett egységek (pl. bruttó hazai termék).

Sokaság fajtái:

- az egységei szerint:
  - **Diszkrét:** egységei egyértelműen elkülönülő egységek (élőlények, tárgyak, szervezetek)
    pl. Magyarország népessége
  - **Folytonos:** egységei önkényesen határozhatók meg (képzett egységek)
    pl. Magyarország villamos energia fogyasztása 2007-ben

- a sokaság időbeli vizsgálatának lehetősége szerint:
  - **álló sokaság:** állapotot fejez ki, időpontra értelmezzük,

    pl. Magyarország népessége 2007. január 1-jén

  - **mozgó sokaság:** folyamatot fejez ki, időtartamra értelmezzük,

    pl. 2007-ben a születések száma Magyarországon

- az egységek száma szerint:
  - **véges** sokaság

    pl. a valós sokaságok

  - **végtelen** sokaság

    pl. tetszőlegesen sokszor megismételhető kísérletek eredményeinek halmaza

Példák:

A sokaság				
Megnevezése	**Egy egysége**	**Fajtája**		
Budapest népessége 2007. december 13-án				
Magyarország 2007. évi alkoholfogyasztása				
Felsőoktatási intézményben tanulók száma 2006/2007-ben, Magyarországon				
2006. december 31-én ittasan okozott balesetek száma Magyarországon				

Példák:

A sokaság				
Megnevezése	**Egy egysége**	**Fajtája**		
Budapest népessége 2007. január 1-jén	fő, ezer fő	Álló	Diszkrét	Véges
Magyarország 2007. évi alkoholfogyasztása	l, hl	Mozgó	Folytonos	Véges
Felsőoktatási intézményben tanulók száma 2006/2007-ben, Magyarországon	fő, ezer fő	Mozgó	Diszkrét	Véges
2006. december 31-én ittasan okozott balesetek száma Magyarországon	db	Álló	Diszkrét	Véges

Ismérv: Az a sokaság egységeit jellemző tulajdonság. Lehetséges kimenetelei az **ismérvváltozatok**.

Fajtái:

- a sokaság jellemzése szempontjából:
  - közös ismérv: a sokaság minden egységére jellemző, pl. az adott napon, adott iskolában érettségizett diákok sokaságában, érettségi időpontja, helye, …
  - megkülönböztető ismérv: amely szerint a sokaság egységei különböznek egymástól

    pl. diákok neve, neme, életkora, …

- tartalma szerint:
  - **Időbeli**: az egységek időbeli elhelyezésére szolgál, pl. születési idő
  - **Területi**: az egységek térbeli elhelyezésére szolgál, pl. születési hely
  - **Minőségi**: az egységek számszerűen nem mérhető tulajdonságai, pl. nem, iskolai végzettség
  - **Mennyiségi**: az egységek számszerűen mérhető tulajdonságai, pl. testsúly

I./2 Mérési szintek/skála

A mennyiségi ismérv változatai számadatok.

A többi ismérv változataihoz is rendelhetők számértékek.

A hozzárendelés szabálya alapján megkülönböztetett mérési skálák:

FEJLETTSÉG

- **névleges (nominális) skála**
  - csak azonosításra szolgál, közöttük relációk, számtani műveletek nem értelmezhetők, pl. ffi:0, nő:1; irányítószámok
- **sorrendi (ordinális) skála**
  - sorba rendezés közös tulajdonság alapján pl. végzettség: általános iskola: 1, középiskola: 2, főiskola: 3; országok hitelképességi sorrendje
- **intervallum (különbségi) skála**
  - nincs rögzített nullpont és mértékegység, meghatározásuk önkényes. Két pontnak csak a különbsége értelmezhető, az összege, aránya nem. Pl. naptári év (1998.+2007.=4005.???), tengerszint feletti magasság
- **arányskála**
  - valódi nullpont van (ami a tulajdonság hiányát jelzi), bármely két szintjének aránya független a mértékegységtől, elemzésekhez szükséges összes művelet elvégezhető. Pl. termelési érték, fogyasztás

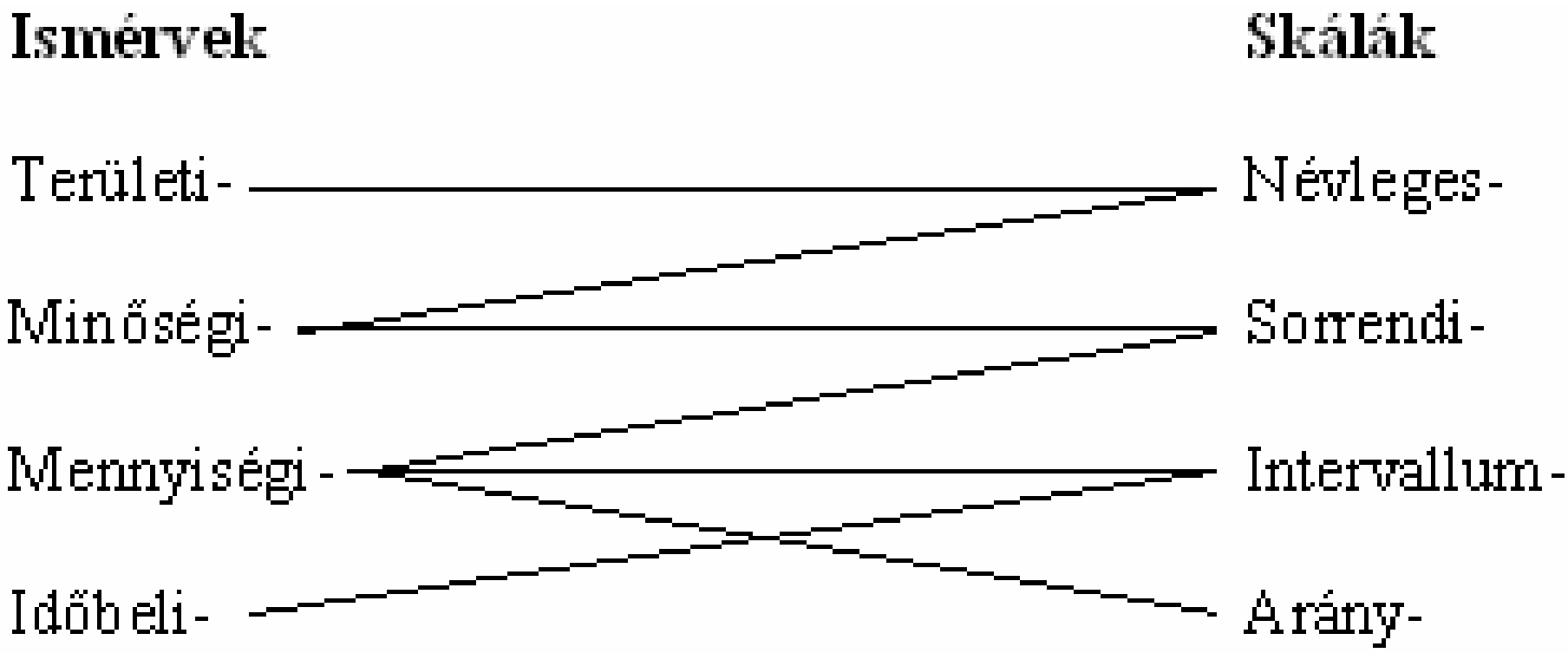
Példák:

Statisztikai ismérv	Skálája
Leckekönyvbe kerülő osztályzat	
Jövedelem	
Autók rendszáma	
Hőmérséklet	

Példák:

Statisztikai ismérv	Skálája
Leckekönyvbe kerülő osztályzat	sorrendi
Jövedelem	arány
Autók rendszáma	névleges
Hőmérséklet	intervallum

Ismérvek és mérési skáláik



Az információsűrítés eszközei

Csoportosítás

Valamilyen ismérv szerinti tagolás
Osztályok kialakítása átfedésmentes, teljes!

Összehasonlítás

Adatok egymáshoz való viszonyítása
Különbség, hányados képzés

Eredmény: statisztikai sor

Sorok rendszere: Statisztikai tábla

Példák:

- **Csoportosító sor:**

Egy cég dogozóinak iskolai végzettsége 2006. jan. 1-jén

Iskolai végzettség	Fő
Ált. isk.	34
Érettségi	75
Felsőfokú	38
Összesen	147

Összehasonlító sor

Egy főre jutó GDP néhány európai országban, 2004-ben

Ország	1 főre jutó GDP (USD)
Ausztria	27300
Belgium	26500
Ciprus	18200
Cseh Köztársaság	15700
Dánia	27200
Észtország	11300
Finnország	25600
Franciaország	24700

Statisztikai tábla

Egy főiskola hallgatóira vonatkozó adatok

A hallgatók		
Neme	**Száma (fő)**	
	1997	**2007**
Férfi	772	1312
Nő	745	1628
Összesen	**1517**	**2940**

I./3 Viszonyszámok

- Két egymással összefüggő statisztikai adat hányadosa.

$$V = \frac{A \leftarrow \textit{a viszonyítás tárgya}}{B \leftarrow \textit{a viszonyítás alapja}}$$

- Fajtái:
  - Megoszlási-
  - Koordinációs-
  - Intenzitási-
  - Dinamikus-

- **Megoszlási viszonyszám**

  Rész és egész aránya (%)

$$V_m = \frac{a}{A} \text{ vagy } \frac{x_i}{\Sigma x_i}$$

Pl. A nappalisok aránya a 2006/7-es tanévben:

$$Vm = \frac{\text{nappalisok száma}}{\text{hallgatók száma összesen}} = \frac{238674}{416348} = 0{,}5733 = 57{,}33\%$$

- **Koordinációs viszonyszám**

  a sokaság két részadatának hányadosa

$$\mathrm{Vk} = \frac{\mathrm{A}}{\mathrm{B}} \text{ vagy } \frac{\mathrm{x}}{\mathrm{y}}$$

Pl. Egy főiskolásra jutó egyetemisták száma a 2006/7-es tanévben:

$$Vk = \frac{egyetemisták\ száma}{föiskolásook\ száma} = \frac{121131}{163323} = 0{,}7417$$

Pl. Ezer főiskolásra jutó egyetemisták száma a 2006/7-es tanévben:

$$Vk = \frac{egyetemisták\ száma}{föiskolásook\ száma} * 1000 = \frac{121131}{163323} * 1000 = 741{,}7$$

- **Intenzitási viszonyszám**

  Két egymással kapcsolatban álló, de különböző fajta (akár különböző mértékegységű) adat hányadosa. Vi

  Pl. Egy intézményre jutó oktatók száma a 2006/7-es tanévben:

$$\mathrm{Vi} = \frac{\text{oktatók száma}}{\text{int ézmények száma}} = \frac{22076}{71} = 310{,}93$$

- **Dinamikus viszonyszám**
  Két időben különböző adat hányadosa (%)

  - **Lánc-:**
    előző évhez hasonlítom $$V_l = \frac{x_i}{x_{i-1}}$$

    Pl. $$V_l = \frac{\text{társasházak száma 2007 – ben}}{\text{társasházak száma 2006 – ban}}$$

  - **Bázis-:**
    bázis évhez hasonlítom $$V_b = \frac{x_i}{x_0}$$

    Pl. $$Vb = \frac{\text{társasházak száma 2007 – ben}}{\text{társasházak száma 2000 – ben}}$$

Dinamikus viszonyszámok értelmezése

- Példa a 2004/5. ill. 2005/6. tanévben a felsőoktatási intézmények számáról:

$$x_{2004/5} = 69$$

$$x_{2005/6} = 71$$

$$V_l = \frac{x_{2005/6}}{x_{2004/5}} = \frac{71}{69} = 1{,}029 = 102{,}9\%$$

  - A 2004/5-ös tanévről a 2005/6-os tanévre 1,029-szeresére nőtt a felsőoktatási intézmények száma.
  - A 2004/5-ös tanévről a 2005/6-os tanévre 102, 9%-ra nőtt a felsőoktatási intézmények száma.
  - A 2004/5-ös tanévről a 2005/6-os tanévre 2,9%-kal nőtt a felsőoktatási intézmények száma.

Dinamikus viszonyszámok értelmezése

- Példa a nappali tagozatos, főiskolai szintű oktatásról:

$$x_{2004/5} = 102380\,(fő)$$

$$x_{2005/6} = 96732\,(fő)$$

$$V_l = \frac{x_{2005/6}}{x_{2004/5}} = \frac{96732}{102380} = 0{,}9448 = 94{,}48\%$$

  - A 2004/5-ös tanévről a 2005/6-os tanévre 0,9448-szeresére csökkent a nappali tagozatos főiskolások száma.
  - A 2004/5-ös tanévről a 2005/6-os tanévre 94,48%-ra csökkent nappali tagozatos főiskolások száma.
  - A 2004/5-ös tanévről a 2005/6-os tanévre 5,52%-kal csökkent nappali tagozatos főiskolások száma.

Dinamikus viszonyszámok értelmezése

Feladat:

- Egy főiskola hallgatóinak létszáma 2006-ról 2007-re 13,25%-kal nőtt, így 2007-ben 2240 hallgatója volt. Hány hallgatója volt 2006-ban?

$$V = \frac{x_{2007}}{x_{2006}}$$

$$113{,}25\% = 1{,}1325 = \frac{2240}{x_{2006}} \quad \Rightarrow \quad x_{2006} = \frac{2240}{1{,}1325} = 1978$$

Megoszlási- és intenzitási viszonyszámok kapcsolata
Feladat:

- Egy főiskolán 2007-ben 100 férfi hallgatóra 145 nő hallgató jutott. Számítsa ki a férfi és a nő hallgatók arányát a főiskolán.

$$x : nök\ száma$$

$$y : férfiak\ száma$$

$$\frac{x}{y} \cdot 100 = 145 \Rightarrow x = 1{,}45y$$

$$nök\ aránya : \frac{x}{x+y} = ?$$

$$\frac{x}{x+y} = \frac{1{,}45y}{1{,}45y + y} = \frac{1{,}45y}{2{,}45y} = 0{,}5918 = 59{,}18\%$$

$$férfiak\ aránya : 100\% - 59{,}18\% = 40{,}82\%$$

Feladat:

Egy vállalat gépsorának termelési adatai 2006-ban és 2007-ben:

	2006	2007	Változás (%)
Termelés (ezer db)			
1 órára jutó termelés (ezer db/óra)			112,5
1 termékre fordított idő (óra/db)			
Összes munkaidő (óra)	5760		

Tudjuk, hogy a termelés 2006-ról 2007-re 43,7 ezer db-bal, azaz 19%-kal nőtt.

Feladat: Töltse ki a táblázat hiányzó adatait! Értelmezze a kapott eredményeket!

Tudjuk, hogy a termelés 2006-ról 2007-re 43,7 ezer db-bal, azaz 19%-kal nőtt.

A 2007-es termelést Y_1-gyel, a 2006-ost Y_0-lal jelölve:

I. $Y_1 - Y_0 = 43{,}7$

II. $Y_1 / Y_0 = 1{,}19 \quad \Rightarrow Y_1 = 1{,}19 Y_0$

I. $1{,}19 Y_0 - Y_0 = 43{,}7$

$0{,}19 Y_0 = 43{,}7$

$Y_0 = 230$

II. $Y_1 = 1{,}19 Y_0 = 1{,}19 * 230 = 273{,}7$

	2006	2007	Változás (%)
Termelés (ezer db)	230	273,70	119
1 órára jutó termelés (ezer db/óra)			112,5
1 termékre fordított idő (óra/db)			
Összes munkaidő (óra)	5760		

	2006	2007	Változás (%)
Termelés (ezer db)	230	273,70	119
1 órára jutó termelés (ezer db/óra)	0,040		112,5
1 termékre fordított idő (óra/db)	0,025		
Összes munkaidő (óra)	5760		

- 1 termékre fordított idő = Munkaidő / Termelés
- 1 termékre fordított idő = 5760 / 230 000 = 0,025 óra/db

- 1 órára jutó termelés = Termelés (ezer db) / Munkaidő
- 1 órára jutó termelés = 230/5760 = 0,04 ezer db / óra (= 40 db/óra)

	2006	2007	Változás (%)
Termelés (ezer db)	230	273,70	119
1 órára jutó termelés (ezer db/óra)	0,040	0,045	112,5
1 termékre fordított idő (óra/db)	0,025		
Összes munkaidő (óra)	5760		

A 2007-es 1 órára jutó termelést Y_1-gyel, a 2006-ost Y_0-lal jelölve:

$$Y_1/Y_0=1,125 \Rightarrow \quad Y_1=1,125Y_0$$

$$Y_1=1,125*0,04=0,045$$

	2006	2007	Változás (%)
Termelés (ezer db)	230	273,70	119
1 órára jutó termelés (ezer db/óra)	0,040	0,045	112,5
1 termékre fordított idő (óra/db)	0,025		
Összes munkaidő (óra)	5760	6082,22	

- 1 órára jutó termelés = Termelés (ezer db) / Munkaidő $\Rightarrow$

  Munkaidő = Termelés (ezer db) / 1 órára jutó termelés

  Munkaidő = 273,7 (ezer db) / 0,045 = 6082,22

	2006	2007	Változás (%)
Termelés (ezer db)	230	273,70	119
1 órára jutó termelés (ezer db/óra)	0,040	0,045	112,5
1 termékre fordított idő (óra/db)	0,025	0,022	
Összes munkaidő (óra)	5760	6082,22	

- 1 termékre fordított idő = Munkaidő / Termelés
- 1 termékre fordított idő = 6082,22 / 273 700 = 0,022 óra/db

	2006	2007	Változás (%)
Termelés (ezer db)	230	273,70	119
1 órára jutó termelés (ezer db/óra)	0,040	0,045	112,5
1 termékre fordított idő (óra/db)	0,025	0,022	88
Összes munkaidő (óra)	5760	6082,22	

A 2007-es 1 termékre fordított időt Y_1-gyel, a 2006-ost Y_0-lal jelölve:

$$Y_1/Y_0 = 0{,}022 / 0{,}025 = 0{,}88 = 88\%$$

	2006	2007	Változás (%)
Termelés (ezer db)	230	273,70	119
1 órára jutó termelés (ezer db/óra)	0,040	0,045	112,5
1 termékre fordított idő (óra/db)	0,025	0,022	88
Összes munkaidő (óra)	5760	6082,22	105,59

A 2007-es összes munkaidőt Y_1-gyel, a 2006-ost Y_0-lal jelölve:

$$Y_1/Y_0 = 6082,22/ 5760 = 1,0559 = 105,59\%$$

	2006	2007	Változás (%)
Termelés (ezer db)	230	273,70	119
1 órára jutó termelés (ezer db/óra)	0,040	0,045	112,5
1 termékre fordított idő (óra/db)	0,025	0,022	88
Összes munkaidő (óra)	5760	6082,22	105,59

Az 1 órára jutó termelés 2006-ról 2007-re óránként 5 db-bal (0,005 ezer db-bal), azaz 12,5%-kal nőtt.

Az 1 termékre fordított idő 2006-ról 2007-re termékenként 0,18 perccel (0,003 órával) azaz 18%-kal csökkent.

Az összes munkaidő 2006-ról 2007-re évi 322,22 órával, azaz 5,59%-kal nőtt.

II. Egy ismérv szerinti elemzés

II./1. Mennyiségi ismérv szerinti elemzés

Az alapdatok

- Ismérvváltozatok (ismérvértékek): $X_1, X_2, \ldots X_N$

1. Mintapélda:

Egy főiskola 24 fős gyakorlati csoportjának (26 pontos) statisztika zárthelyi dolgozatának elért pontszámai:

18, 17, 3, 13, 13, 20, 5, 23, 21, 20, 20, 11, 26, 13, 15, 26, 22, 13, 7, 16, 8, 21, 5, 9

Ezeket az ismérvváltozatokat X_1, X_2, …,X_N-nel jelöljük, ahol N a megfigyelt egységek száma.

N=24

- A sokaság további vizsgálata szempontjából kézenfekvő lehetőség az egységeinek (monoton növekvő) sorbarendezése, azaz a **rangsor** készítés: $X_1^*, X_2^*, \ldots X_N^*$

3, 5, 5, 7, 8, 9, 11, 13, 13, 13, 13, 15, 16, 17, 18, 20, 20, 20, 21, 21, 22, 23, 26, 26

A rangsor első eleme a **legkisebb** ismérvérték: $X_{min}=3$

A rangsor utolsó eleme a **legnagyobb** ismérvérték: $X_{max}=26$

3, 5, 5, 7, 8, 9, 11, 13, 13, 13, 13, 15, 16, 17, 18, 20, 20, 20, 21, 21, 22, 23, 26, 26

- **Helyzetmutatók**
  - Középértékek
    - Helyzeti-
      - Módusz
      - Medián
    - Számított-
      - Átlagok
  - Kvantilis értékek
    - Minimum
    - Alsó kvartilis
    - Medián
    - Felő kvartilis
    - Maximum

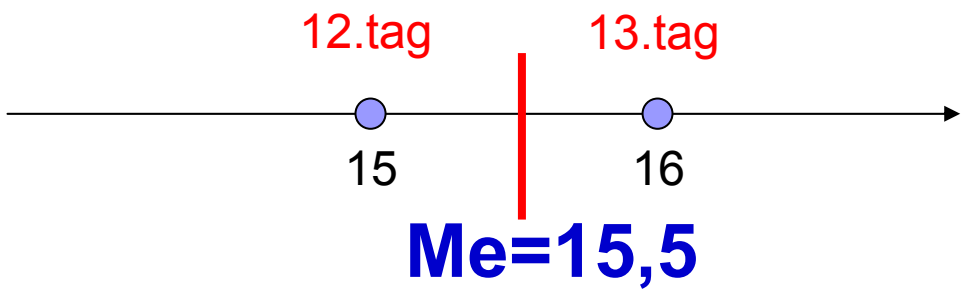
Keressük meg azt a pontszámot, amelynél a hallgatók fele rosszabb, (fele jobb) dolgozatot írt.

Ez a **medián (**Me), ami a rangsor **(N+1)/2.** tagja. Esetünkben (N+1)/2=25/2=12,5. Tehát a medián a 12. és 13. tag között lesz félúton*:

Me

12.tag | 13.tag

3, 5, 5, 7, 8, 9, 11, 13, 13, 13, 13, 15, 16, 17, 18, 20, 20, 20, 21, 21, 22, 23, 26, 26



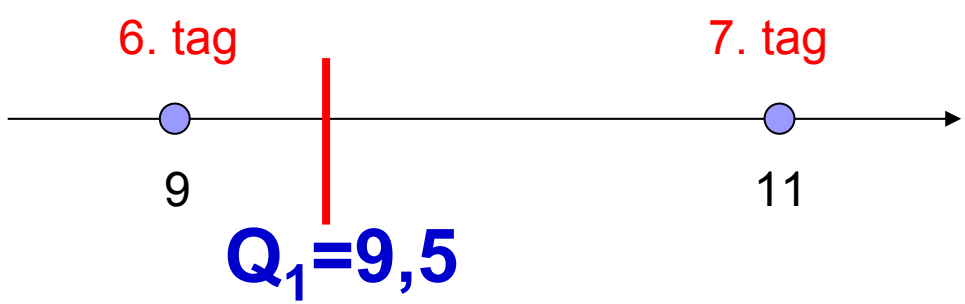
* A 12. és13. tag számtani átlaga.

Keressük meg azt a pontszámot, amelynél a hallgatók negyede rosszabb, (háromnegyede jobb) dolgozatot írt. Ez az **alsó kvartilis (Q_1)**, ami a rangsor **(N+1)/4.** tagja. Esetünkben (N+1)/4=25/4=6,25.

Tehát az alsó kvartilis a 6. és 7. tag között lesz,

pontosan a 6. és 7. tag közti távolság negyedénél:

$$Q_1 = 9 + 0{,}25 \cdot (11 - 9)$$

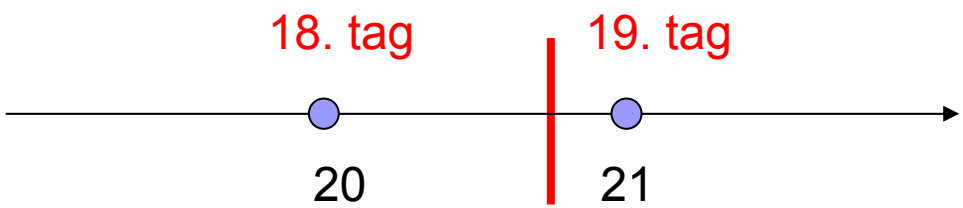


Keressük meg azt a pontszámot, amelynél a hallgatók háromnegyede rosszabb, (negyede jobb) dolgozatot írt. Ez a **felső kvartilis (Q_3)**, ami a rangsor **3(N+1)/4.** tagja. Esetünkben 3(N+1)/4=3*25/4=18,75.

Tehát a felső kvartilis a 18. és 19. tag között lesz,

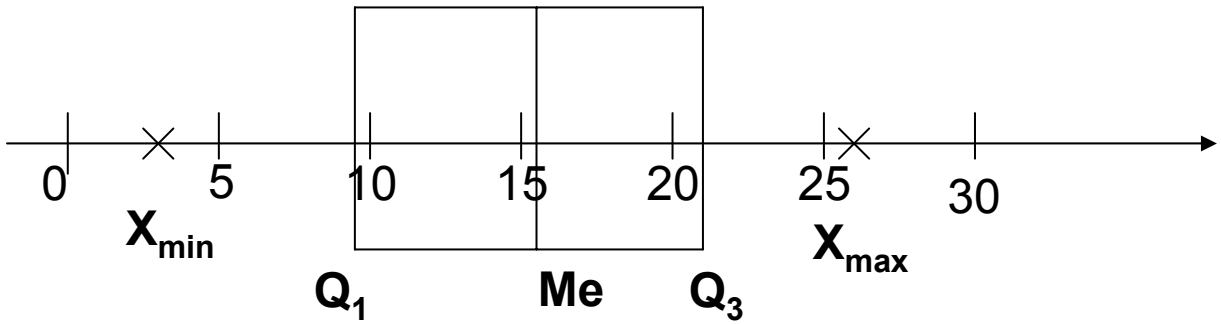
a 18. és 19. tag közti távolság háromnegyedénél:

$$Q_3 = 20 + 0{,}75 \cdot (21 - 20)$$



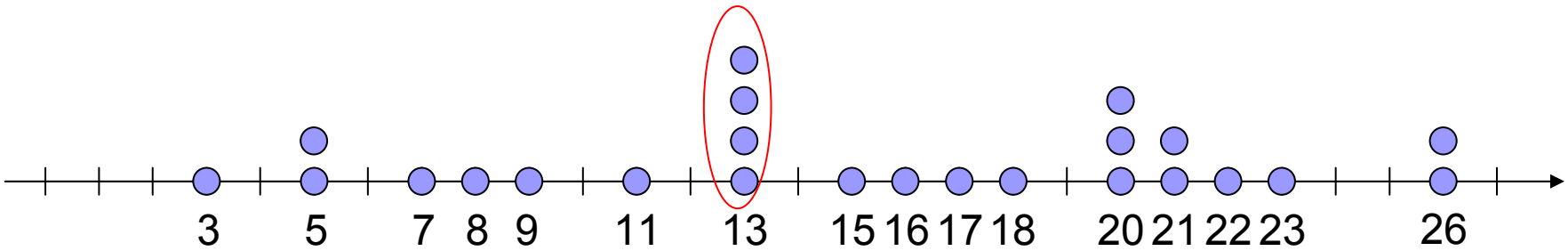
Q_3=20,75

Az előző helyzetmutatók segítségével elkészíthetjük a sokaság tömör jellemzésére szolgáló **ötszámos összegzést,** melyet a következő **doboz-ábrán** (box-and-whiskers) szemléltethetünk:



Keressük meg a tipikus pontszámot. Ez a **Módusz (Mo):** a leggyakoribb ismérvérték:

3, 5, 5, 7, 8, 9, 11, 13, 13, 13, 13, 15, 16, 17, 18, 20, 20, 20, 21, 21, 22, 23, 26, 26



Mo=13

Végül számítsuk ki a csoport átlagos pontszámát. Ez a **számtani átlag***:

$$\overline{X} = \frac{\sum_{i=1}^{N} X_i}{N} = \frac{3+5+5+7+\ldots+26}{24} = 15{,}21$$

3, 5, 5, 7, 8, 9, 11, 13, 13, 13, 13, 15, 16, 17, 18, 20, 20, 20, 21, 21, 22, 23, 26, 26

*további átlagokról később lesz szó.

- **Középértékek tulajdonságai:**

  Az előzőekben kiszámítottuk a sokaság helyzetének jellemzésére szolgáló **<u>középértékeket</u>**:
  - Helyzeti középértékek
    - Módusz
    - Medián
  - Számított középértékek
    - Átlagok

- Ezek eleget tesznek a középértékektől elvárt tulajdonságoknak:
  - pontosan definiált
  - közepes helyzetű
  - könnyen értelmezhető

- Szóródási mérőszámok
  - Terjedelem
  - Interkvartilis terjedelem
  - Szórás
  - Relatív szórás

Terjedelem: $T = X_{max} – X_{min}$

$$T = 26 – 3 = 23$$

A legjobb és legrosszabb dolgozat közti különbség 23 pont.

Interkvartilis terjedelelm: $IQT = Q_3 - Q_1$

$IQT = 20{,}75 - 9{,}5 = 11{,}25$

A hallgatók 50%-ának 11,25 pontnyi terjedelembe esik az eredménye.

Szórás:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}\left(X_i - \overline{X}\right)^2}{N}}$$

$$\sigma = \sqrt{\frac{(3-15{,}21)^2 + (5-15{,}21)^2 + \ldots + (26-15{,}21)^2}{24}} = 6{,}58$$

Az egyes hallgatók pontszámai átlagosan 6,58 ponttal különböznek a csoportátlagtól.

3, 5, 5, 7, 8, 9, 11, 13, 13, 13, 13, 15, 16, 17, 18, 20, 20, 20, 21, 21, 22, 23, 26, 26

Relatív szórás: Különböző mértékegységű adatok szórásának összehasonlítására szolgál.

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

$$V = \frac{6,58}{15,21} = 0,4326 = 43,26\%$$

Az egyes hallgatók pontszámai átlagosan 43,26%-kal különböznek az átlagtól.

2. Mintapélda:

Egyik budapesti valutaváltóban 2007. július 4-én 41-en váltottak valutát euróra. Az átváltott összegek rangsora (euró):

5, 20, 20, 26, 30, 35, 50, 60, 68, 76, 80, 84, 85, 85, 88, 93, 100, 100, 100, 108, 110, 113, 127, 130, 135, 145, 146, 147, 155, 158, 170, 175, 192, 200, 200, 252 276, 300, 350, 400, 480

N=41

Ha sok ismérvváltozat van ún. osztályközös gyakorisági sort készíthetünk:

5, 20, 20, 26, 30, 35, 50, 60, 68, 76, 80, 84, 85, 85, 88, 93, 100, 100, 100, 108, 110, 113, 127, 130, 135, 145, 146, 147, 155, 158, 170, 175, 192, 200, 200, 252, 276, 300, 350, 400, 480

- osztályközök kialakítása:
  - osztályközök száma: keressük azt a k-t, melyre $2^k>N$ (először $2^6>41$, tehát k=6 (db osztályköz lesz)
  - osztályközök hossza: h=(Xmax-Xmin)/k=(480-5)/6=79,16 (≈80
  - További kritériumok
    - Átfedésmentes
    - Teljes

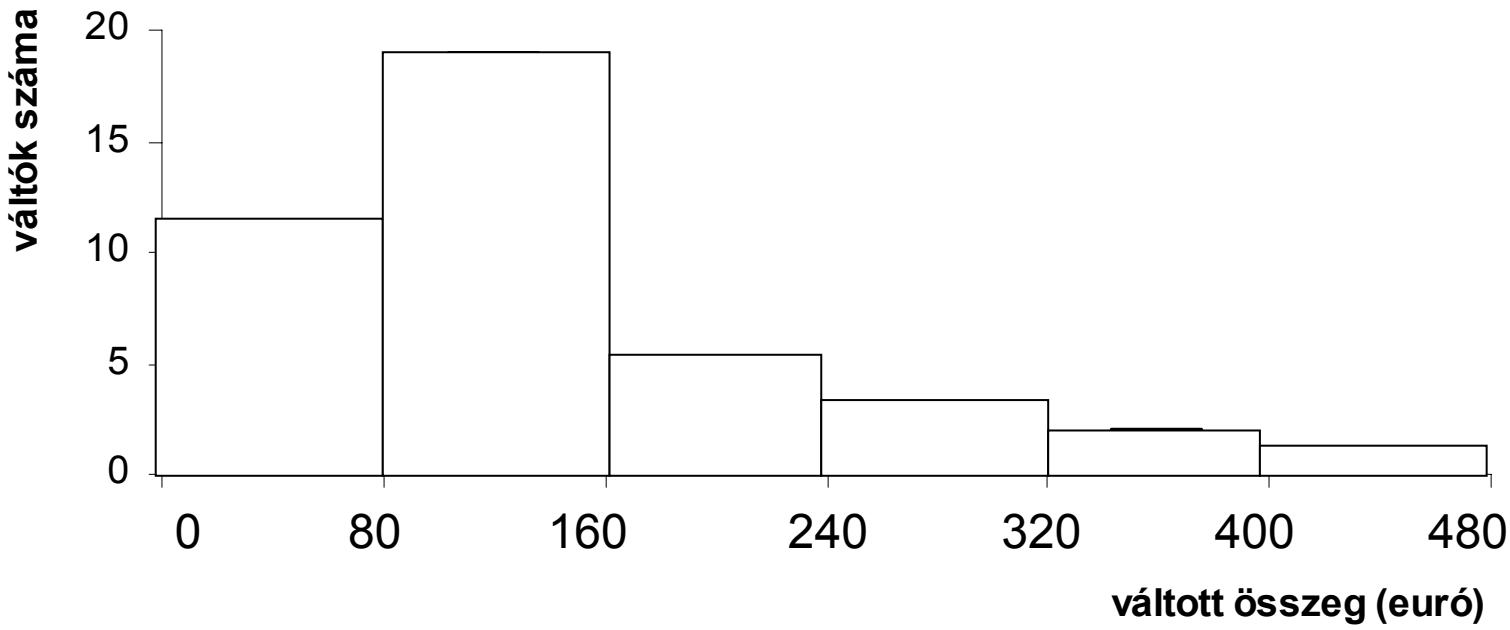
Ha sok ismérvváltozat van ún. osztályközös gyakorisági sort készíthetünk:

5, 20, 20, 26, 30, 35, 50, 60, 68, 76, 80, 84, 85, 85, 88, 93, 100, 100, 100, 108, 110, 113, 127, 130, 135, 145, 146, 147, 155, 158, 170, 175, 192, 200, 200, 252, 276, 300, 350, 400, 480

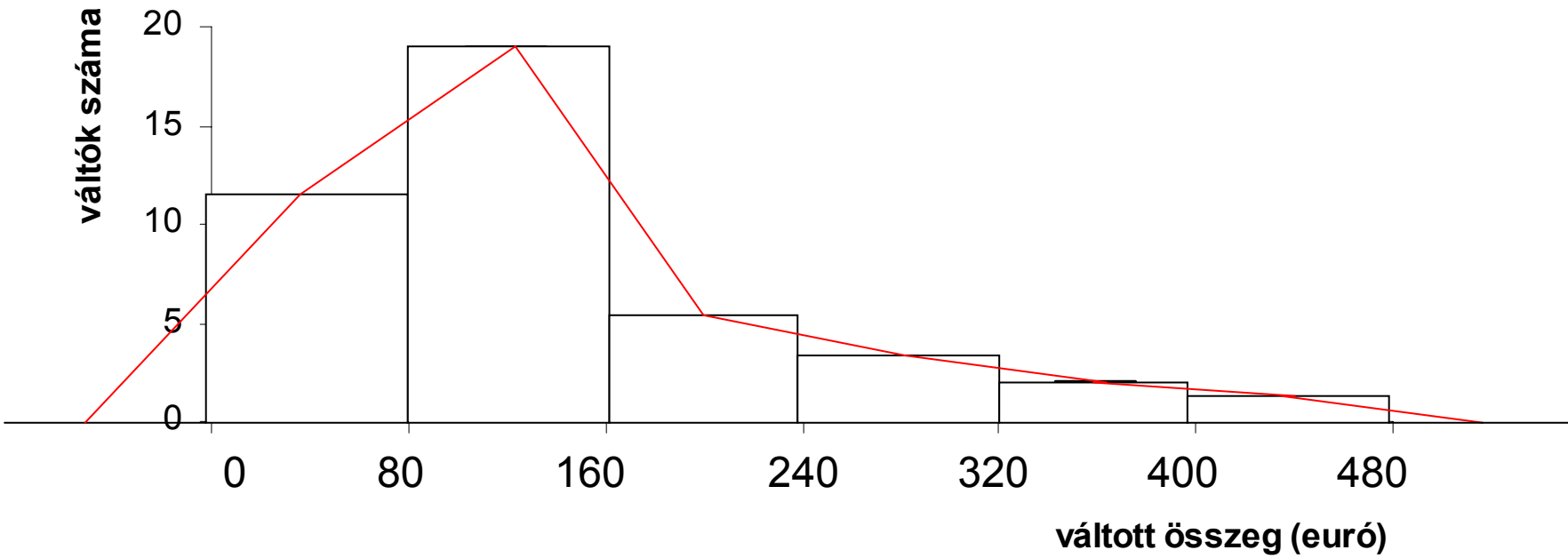
osztályközök kialakítása:

Váltott összeg (euró)	Váltók száma (fő)
0 – 80	11
81 – 160	19
161 – 240	5
241 – 320	3
321 – 400	2
401 - 480	1
Összesen:	41

- Osztályközös gyakorisági sor szemléltetése:
  - **Hisztogram**
  - Gyakorisági poligon



- Osztályközös gyakorisági sor szemléltetése:
  - **Hisztogram**
  - Gyakorisági poligon



Az osztályközök alsó határai helyett mindig az előző osztálykö felső határával számolunk!

Váltott összeg (euró)	Váltók száma (fő)
0 – 80	11
80 (81) – 160	19
160 (161) – 240	5
240 (241) – 320	3
320 (321) – 400	2
400(401) - 480	1
Összesen:	41

Meghatározzuk az osztályközepeket: az osztályközök felezőpontjait. Xi

A gyakoriságokat fi-vel jelöljük és gyakorisági sornak nevezzük.

Xi	Váltott összeg (euró)	Váltók száma (fő) fi
40	0 – 80	11
120	80 (81) – 160	19
200	160 (161) – 240	5
280	240 (241) – 320	3
360	320 (321) – 400	2
440	400(401) - 480	1
	Összesen:	41

Kumulált gyakorisági sor: fi’ az fi értékeket a megfelelő sorig összeadjuk (göngyölítjük).

Xi	Váltók száma (fő) fi	fi’
40	11	11
120	19	30
200	5	35
280	3	38
360	2	40
440	1	41
Összesen:	41	-

Relatív gyakorisági sor: gi =fi/N (%)

Xi	Váltók száma (fő) fi	fi'	gi (%)
40	11	11	26,83
120	19	30	46,34
200	5	35	12,20
280	3	38	7,32
360	2	40	4,88
440	1	41	2,44
Összesen:	41	-	100

Kumulált relatív gyakorisági sor: gi’ (%)

Xi	Váltók száma (fő) fi	fi’	gi (%)	gi’ (%)
40	11	11	26,83	26,83
120	19	30	46,34	73,17
200	5	35	12,20	85,37
280	3	38	7,32	92,68
360	2	40	4,88	97,56
440	1	41	2,44	100,00
Összesen:	41	-	100	-

Értékösszeg sor: Si=fi*Xi

Xi	Váltók száma (fő) fi	fi'	gi (%)	gi' (%)	Si
40	11	11	26,83	26,83	440
120	19	30	46,34	73,17	2280
200	5	35	12,20	85,37	1000
280	3	38	7,32	92,68	840
360	2	40	4,88	97,56	720
440	1	41	2,44	100,00	440
Összesen:	41	-	100	-	5720

Kumulált értékösszeg sor: Si’

Xi	Váltók száma (fő) fi	fi’	gi (%)	gi’ (%)	Si	Si’
40	11	11	26,83	26,83	440	440
120	19	30	46,34	73,17	2280	2720
200	5	35	12,20	85,37	1000	3720
280	3	38	7,32	92,68	840	4560
360	2	40	4,88	97,56	720	5280
440	1	41	2,44	100,00	440	5720
Összesen:	41	-	100	-	5720	-

Relatív Értékösszeg sor: Zi=Si/ΣSi (%)

Xi	Váltók száma (fő) fi	fi'	gi (%)	gi' (%)	Si	Si'	Zi (%)
40	11	11	26,83	26,83	440	440	7,69
120	19	30	46,34	73,17	2280	2720	39,86
200	5	35	12,20	85,37	1000	3720	17,48
280	3	38	7,32	92,68	840	4560	14,69
360	2	40	4,88	97,56	720	5280	12,59
440	1	41	2,44	100,00	440	5720	7,69
Összesen:	41	-	100	-	5720	-	100

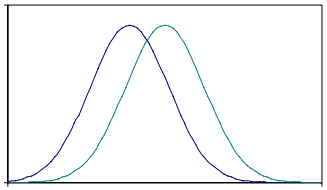
Kumulált relatív Értékösszeg sor: Zi’ (%)

Xi	Váltók száma (fő) fi	fi’	gi (%)	gi’ (%)	Si	Si’	Zi (%)	Zi’ (%)
40	11	11	26,83	26,83	440	440	7,69	7,69
120	19	30	46,34	73,17	2280	2720	39,86	47,55
200	5	35	12,20	85,37	1000	3720	17,48	65,03
280	3	38	7,32	92,68	840	4560	14,69	79,72
360	2	40	4,88	97,56	720	5280	12,59	92,31
440	1	41	2,44	100,00	440	5720	7,69	100
Összesen:	41	-	100	-	5720	-	100	-

Gyakorisági sorok jellegzetességei

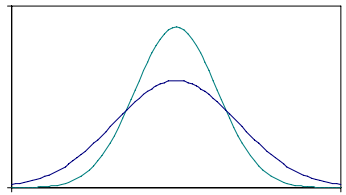
LEOLVASHATÓK vagy MÉRHETŐK

- Helyzetük



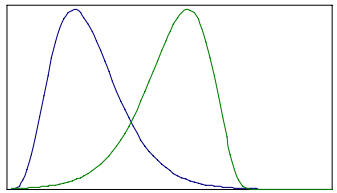
HELYZETMUTATÓK

- Szóródásuk



SZÓRÓDÁSI MUTATÓK

- Alakjuk



ASZIMMETRIA MUTATÓK

3. Mintapélda:

Egyik női ruhákat készítő cég nők testmagasságát vizsgálta, a következő véletlen minta alapján:

Testmagasság (cm)	Nők száma (fő)
– 150	7
151 – 160	29
161 – 170	48
171 – 180	22
181 –	4
Összesen:	110

Feladat

Becsülje meg és értelmezze a középértékeket!

Becsülje meg és értelmezze a kvartiliseket!

Becsülje meg és értelmezze a tanult szóródási mutatókat!

Ábrázolja poligonnal a testmagasság szerinti megoszlást!

Állapítsa meg az eloszlás aszimmetria tulajdonságát. Megállapítását aszimmetria mutatók kiszámításával támassza alá!

Az osztályközök előkészítése a további számoláshoz:

- **A két szélső (nyitott) osztályközt bezárjuk, ugyanakkorára, mint az ő szomszédos osztályközük.**
- **Az osztályközök alsó határai helyett az előző osztályköz felső határával számolunk.**

Testmagasság (cm)	Nők száma (fő)
140 – 150	7
150 ~~151~~ – 160	29
160 ~~161~~ – 170	48
170 ~~171~~ – 180	22
180 ~~181~~ – 190	4
Összesen:	110

További oszlopok elkészítése a számoláshoz:

- **Az osztályközép az osztályköz felezőpontja. Kiszámítása a következő képlettel történhet:**

$$X_i = \frac{X_{i_a} + X_{i_f}}{2}$$

Az i. osztályközép: az i. osztályköz alsó és felső határának számtani átlaga.

Osztályközép Xi	Testmagasság (cm)	Nők száma (fő) Gyakorisági sor fi
145	140 - 150	7
155	150 - 160	29
165	160 - 170	48
175	170 - 180	22
185	180 - 190	4
-	Összesen:	110

További oszlopok elkészítése a számoláshoz:

- **Kumulált gyakorisági sor (fi'): A gyakorisági sor tagjait göngyölítjük.**

Osztályközép Xi	Testmagasság (cm)	Nők száma (fő) Gyakorisági sor fi	Kumulált gyakorisági sor fi'
145	140 - 150	7	7
155	150 - 160	29	36
165	160 - 170	48	84
175	170 - 180	22	106
185	180 - 190	4	110
-	Összesen:	110	-

További oszlopok elkészítése a számoláshoz:

- **Értékösszeg sor (Si) tagjainak kiszámítása:**

$$S_i = f_i \cdot X_i$$

Osztály-közép Xi	Testmagasság (cm)	Nők száma (fő) Gyakorisági sor fi	Kumulált gyakorisági sor fi'	Értékösszeg sor Si=fiXi
145	140 - 150	7	7	1015
155	150 - 160	29	36	4495
165	160 - 170	48	84	7920
175	170 - 180	22	106	3850
185	180 - 190	4	110	740
-	Összesen:	110	-	18020

- Az **átlag** becsült értékének kiszámítása a súlyozott átlag képletével történik, hiszen osztályközös gyakorisági sorról van szó:

$$\overline{X} = \frac{\sum_{i=1}^{k} f_i \cdot X_i}{N}$$

- Ha rendelkezésünkre áll az értékösszeg sor, alkalmazható a következő képlet:

$$\overline{X} = \frac{\sum_{i=1}^{k} S_i}{N}$$

- $$\overline{X} = \frac{\sum_{i=1}^{k} S_i}{N} = \frac{18020}{110} = 163{,}82$$

- A vizsgált nők átlagos testmagassága 163,82 cm.

- A **medián** becsült értékének kiszámítása:
  - Először meg kell határoznunk a mediánt tartalmazó osztályközt: ahol fi' először túllépi N/2-t
  - Majd alkalmazzuk a következő képletet:

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

ahol

- me: a mediánt tartalmazó osztályköz alsó határa
- f'me-1: a mediánt megelőző osztályköz kumulált gyakorisága
- fme: a mediánt tartalmazó osztályköz gyakorisága
- h: a mediánt tratalmazó osztályköz hossza

- A **medián** becsült értékének kiszámítása:
  - N/2=110/2=55
  - Mediánt tartalmazó osztályköz:

Osztály-közép Xi	Testmagasság (cm)	Nők száma (fő) Gyakorisági sor fi	Kumulált gyakorisági sor fi'	Értékösszeg sor Si=fiXi
145	140 - 150	7	7	1015
155	150 - 160	29	36	4495
165	160 - 170	48	**84**	7920
175	170 - 180	22	106	3850
185	180 - 190	4	110	740
-	Összesen:	110	-	18020

$$Me = me + \frac{\frac{N}{2} - f'_{me-1}}{f_{me}} \cdot h = 160 + \frac{55-36}{48} \cdot 10 = 163{,}96$$

- Me=163,96 cm

- A vizsgált nők fele 163,96 cm-nél alacsonyabb, fele ennél magasabb.

- A **módusz** becsült értékének kiszámítása:
  - Először meg kell határoznunk a móduszt tartalmazó (modális) osztálykőzt: ahol fi a legnagyobb
  - Majd alkalmazzuk a következő képletet:

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

ahol

    - mo: a modális osztálykőz alsó határa
    - k1: a modális és az azt megelőző osztálykőz (egységnyi osztálykőzhosszúságra jutó) gyakoriságainak különbsége
    - k2: a modális és az azután következő osztálykőz (egységnyi osztálykőzhosszúságra jutó) gyakoriságainak különbsége
    - h: a modális osztálykőz hossza

- A **módusz** becsült értékének kiszámítása:
  - A modális osztályköz

Osztály-közép Xi	Testmagasság (cm)	Nők száma (fő) Gyakorisági sor fi	Kumulált gyakorisági sor fi'	Értékösszeg sor Si=fiXi
145	140 - 150	7	7	1015
155	150 - 160	29	36	4495
165	160 - 170	**48**	84	7920
175	170 - 180	22	106	3850
185	180 - 190	4	110	740
-	Összesen:	110	-	18020

  - k1=48-29=19 k2=48-22=26

$$Mo = mo + \frac{k_1}{k_1 + k_2} \cdot h = 160 + \frac{19}{19 + 26} \cdot 10 = 164{,}22$$

- Mo=164,22 cm
- A vizsgált nők tipikus testmagassága 164,22 cm.

- Az **alsó kvartilis** becsült értékének kiszámítása, a mediánnál leírtakhoz hasonlóan történhet:
  - N/4=27,5
  - Alsó kvartilist tartalmazó osztályköz:

Osztály-közép Xi	Testmagasság (cm)	Nők száma (fő) Gyakorisági sor fi	Kumulált gyakorisági sor fi'	Értékösszeg sor Si=fiXi
145	140 - 150	7	7	1015
155	150 - 160	29	**36**	4495
165	160 - 170	48	84	7920
175	170 - 180	22	106	3850
185	180 - 190	4	110	740
-	Összesen:	110	-	18020

$$Q_1 = q_1 + \frac{\frac{N}{4} - f'_{q_1-1}}{f_{q_1}} \cdot h = 150 + \frac{27,5 - 7}{29} \cdot 10 = 157,07$$

- Q1=157,07 cm
- A vizsgált nők negyede 157,07cm-nél alacsonyabb, háromnegyede ennél magasabb.

- A felső **kvartilis** becsült értékének kiszámítása, a mediánnál leírtakhoz hasonlóan történhet:
  - 3N/4=82,5
  - Felső kvartilist tartalmazó osztályköz:

Osztály-közép Xi	Testmagasság (cm)	Nők száma (fő) Gyakorisági sor fi	Kumulált gyakorisági sor fi'	Értékösszeg sor Si=fiXi
145	140 - 150	7	7	1015
155	150 - 160	29	36	4495
165	160 - 170	48	**84**	7920
175	170 - 180	22	106	3850
185	180 - 190	4	110	740
-	Összesen:	110	-	18020

$$Q_3 = q_3 + \frac{\frac{3N}{4} - f'_{q_3-1}}{f_{q_3}} \cdot h = 160 + \frac{82,5 - 36}{48} \cdot 10 = 169,69$$

- Q3=169,69 cm
- A vizsgált nők háromnegyede 169,69 cm-nél alacsonyabb, negyede ennél magasabb.

- A **terjedelem** kiszámítása:
  - A terjedelem kiszámítása a T=Xmax-Xmin képlettel történik, ahol az első osztályköz alsó határát tekintjük x min-nek és az utolsó osztályköz felső határát Xmax-nak:

Osztály-közép Xi	Testmagasság (cm)	Nők száma (fő) Gyakorisági sor fi	Kumulált gyakorisági sor fi'	Értékösszeg sor Si=fiXi
145	140 - 150	7	7	1015
155	150 - 160	29	36	4495
165	160 - 170	48	84	7920
175	170 - 180	22	106	3850
185	180 - 190	4	110	740
-	Összesen:	110	-	18020

  - T= 190-140 =50

- T=50 cm.
- A legmagasabb és legalacsonyabb nő magasságkülönbsége 50 cm.

- Az **interkvartilis terjedelem** kiszámítása:
  - Az intervartilis terjedelem kiszámítása a IQT=Q3-Q1 képlettel történik.
  - IQT = 169,69 - 157,07 =12,62 cm
  - A vizsgált nők felének testmagssága 12,62 cm-nyi terjedelembe esik.

A **szórás** kiszámítása:

- A szórás kiszámítása a $\sigma = \sqrt{\frac{\sum_{i=1}^{k} f_i \left(X_i - \overline{X}\right)^2}{N}}$ képlettel történik, hiszen osztályközös gyakorisági sorról van szó.

- A számolás megkönnyítéséhez készíthetünk egy $\left(X_i - \overline{X}\right)^2$ értékeket tartalmazó segédoszlopot:

A **szórás** kiszámítása:

Osztály-közép Xi	Testmagasság (cm)	Nők száma (fő) Gyakorisági sor fi	Kumulált gyakorisági sor fi'	Értékösszeg sor Si=fiXi	$(X_i - \overline{X})^2$
145	140 - 150	7	7	1015	354,12
155	150 - 160	29	36	4495	77,76
165	160 - 170	48	84	7920	1,40
175	170 - 180	22	106	3850	125,03
185	180 - 190	4	110	740	448,67
-	Összesen:	110	-	18020	-

$$\sigma = \sqrt{\frac{\sum_{i=1}^{k} f_i (X_i - \overline{X})^2}{N}} = \sqrt{\frac{7 \cdot 354{,}12 + 29 \cdot 77{,}76 + 48 \cdot 1{,}4 + 22 \cdot 125{,}03 + 4 \cdot 448{,}67}{110}} = 9{,}22$$

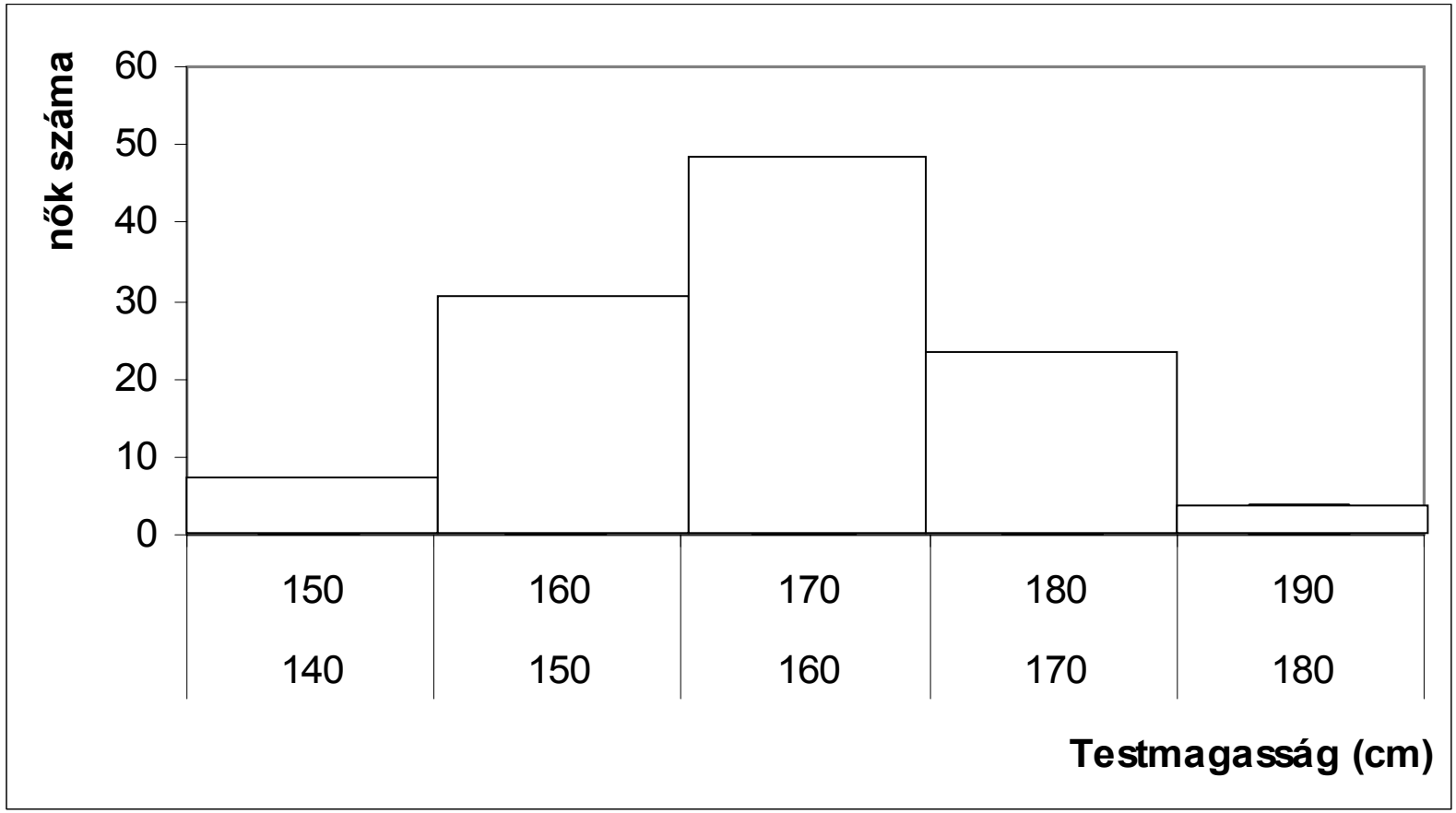
- σ =9,22 cm
- A vizsgált nők testmagasságának az átlagos (163,82 cm-es) testmagasságtól való eltérése 9,22 cm.

- A **relatív szórás** kiszámítása:

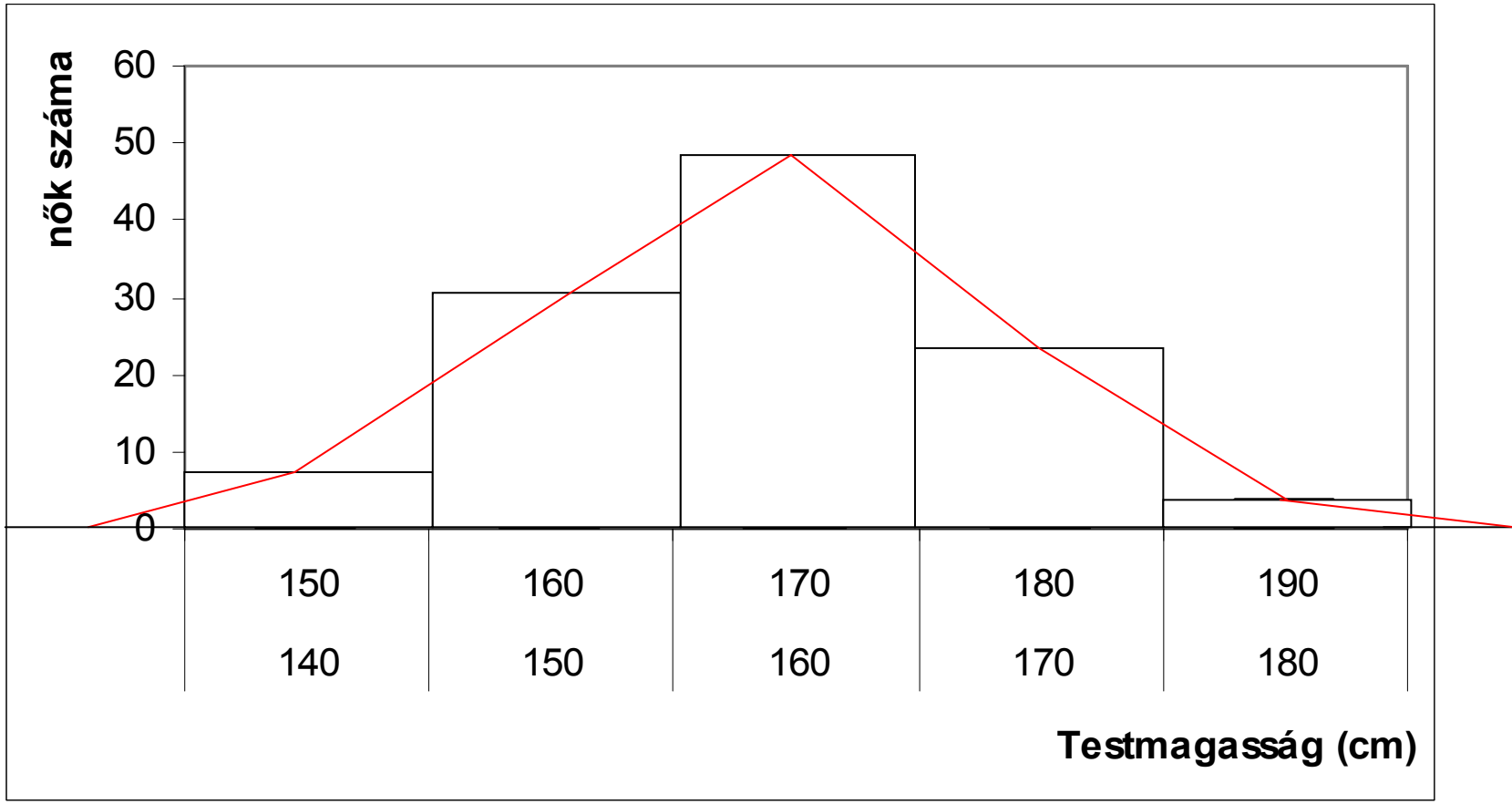
  - $$V = \frac{\sigma}{\overline{X}} = \frac{9{,}22}{163{,}82} = 0{,}0563 = 5{,}63\%$$

  - A vizsgált nők testmagasságának az átlagos (163,82 cm-es) testmagasságtól való eltérése 5,63%.

- Ábrázolás hisztogrammal:



- Ábrázolás gyakorisági poligonnal:



- Az aszimmetria számszerűsítése:
  - F mutatóval:
  - $$F = \frac{(Q_3 - Me) - (Me - Q_1)}{(Q_3 - Me) + (Me - Q_1)} = \frac{(169{,}69 - 163{,}82) - (163{,}82 - 157{,}07)}{(169{,}69 - 163{,}82) + (163{,}82 - 157{,}07)} = -0{,}07$$
  - Pearson-féle A mutatóval:
  - $$A = \frac{\overline{X} - Mo}{\sigma} = \frac{163{,}82 - 164{,}22}{9{,}22} = -0{,}04$$

Az aszimmetria számszerűsítése:

- Értelmezés:

- Ha A, F<0 jobboldali aszimmetria
- Ha A, F>0 baloldali aszimmetria
- A 0 körüli** érték szimmetriát jelez
- $-1 \leq F \leq 1$
- |A|-nak nem 1 a korlátja!
- Az |A| , |F| a szimmetria fokát jelzi (enyhe-közepes-erős)

- A mi esetünkben a 0-hoz közeli értékek szimmetriát jeleznek, mivel a fenti mutatókat osztályközös gyakorisági sorból becsültük.

Köszönöm a figyelmet!