

III. Modul Matematikai statisztika

6. lecke: Egyszerűbb statisztikai mérőszámok meghatározása

6.1. Számított középérték

A mintaelemek számtani középértékét **mintaátlagnak**, vagy **empirikus középnek** nevezzük és \hat{m}_n -pal vagy \bar{x} -al jelöljük.

A leggyakrabban alkalmazott középérték, jele: \bar{x} . A számtani átlag az a szám, amellyel az egyes átlagolandó értékeket helyettesítve azok összege változatlan marad.

Fajta az átlagolandó értékek előfordulása alapján:

- **Egyszerű számtani átlag:** minden átlagolandó érték egyszer fordul elő az adatsorban

$$\bar{x} = \frac{\sum x_i}{n}$$

1. bemutató feladat:

Egy focicsapat játékosainak cm-ben mért magasságai a következők:
189, 191, 185, 188, 190, 175, 180, 178, 185, 179, 184.

Megoldás:

$$\bar{x} = \frac{189 + 191 + 185 + 188 + 190 + 175 + 180 + 178 + 185 + 179 + 184}{11} = 184$$

Az átlagos testmagasság a csapatban 184 cm.

Gyakorisági sor: az ismerv előfordulásának gyakoriságát tüntetjük fel. A gyakoriság (f_i) megmutatja, hogy az egyes ismervváltozatok hányszor fordulnak elő a megfigyelt sokaságban. Ha az egyes gyakoriságokat azok összegéhez viszonyítjuk, akkor az adott ismervérték relatív gyakoriságát (g_i) kapjuk meg:

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

ahol: g_i : az i -edik ismervérték relatív gyakorisága
 f_i : az i -edik ismervérték gyakorisága
 $\sum f_i$: a sokaság elemeinek száma, azaz egyenlő n -nel.

Ha az ismervváltozatok száma nagy, akkor az adatokat rangsoroljuk, és ez megkönnyíti a változó osztályozását. Az osztályozás sűríti az információt. A legnagyobb és legkisebb ismérvek által adott intervallumot úgy osztjuk osztályokba, hogy az egyes osztályközökön belül a gyakoriságok közel egyenlő eloszlásúak legyenek, így az osztályközép alkalmas lesz az osztály jellemzésére.

Az osztályok olyan adatszoportok, ahol az egyes osztályok közötti mennyiségi változás minőségi változást takar.

Az osztályközök száma: $2^k > N$

Osztályköz hossza:

$$h = \frac{x_{\max} - x_{\min}}{k}$$

- **Súlyozott számtani átlag:** az átlagolandó értékek többször fordulnak elő az adatsorban:

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i}$$

Mivel $g_i = \frac{f_i}{\sum f_i}$, ezért a relatív gyakoriságokkal is kiszámolható az átlag:

$$\bar{x} = \sum g_i x_i$$

2. bemutató feladat:

Egy vidéki könyvtárban a kikölcsönzött könyvek száma és a könyvtárlagok száma:

Kölcsönzött könyvek száma (db)	f_i	$\bar{x} = \frac{14 \cdot 2 + 38 \cdot 3 + 22 \cdot 4 + 9 \cdot 5}{83} = 3,31$ <p>Egy könyvtárlag átlagosan 3,31 db könyvet kölcsönöz ki</p>
2	14	
3	38	
4	22	
5	9	
Összesen	83	

Osztályközös gyakorisági sor esetén az osztályközépek töltik be az átlagolandó értékek szerepét. Az osztályközép számítása:

$$x_i = \frac{x_{i(\text{alsóh.})} + x_{i(\text{felsőh.})}}{2}$$

ahol:

$x_{i(\text{alsóh.})}$: az i-edik osztályköz alsó határa,

$x_{i(\text{felsőh.})}$: az i-edik osztályköz felső határa.

A számításnál nem vesszük figyelembe az elkülönítést szolgáló utolsó számjegyet.

Pl.: 21-50 esetében:

$$x_i = \frac{20 + 50}{2} = 35$$

Ha az osztályközök alsó vagy felső határa nincs megadva, akkor nyitott osztályközről beszélünk. Az alsó határt úgy adjuk meg, hogy az osztályköz szélessége azonos legyen az őt követő osztályköz szélességével. A felső határ megadásakor pedig az előző osztályköz szélességét vesszük figyelembe.

Pl

.....-50
51-100
101-150
151-

Ebben az esetben az osztályköz hossza 50, tehát az első osztályköz 1-50 lesz, az utolsó pedig 151-200.

3. bemutató feladat

140 villanyégő égési idejét vizsgálva az alábbi adatokat kaptuk:

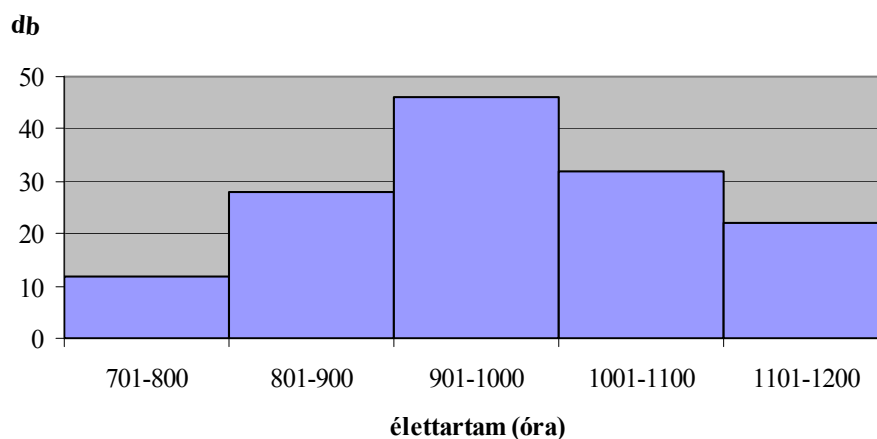
<i>Élettartam (óra)</i>	<i>f_i</i>	<i>Osztály- közép</i>
-800	12	750
801-900	28	850
901-1000	46	950
1001-1100	32	1050
1101-	22	1150
Összesen	140	

Számítsuk ki az égők átlagos élettartamát!

Megoldás:

<i>Élettartam (óra)</i>	<i>f_i</i>	<i>Osztály- közép</i>	$\bar{x} = \frac{12 \cdot 750 + 28 \cdot 850 + \dots + 22 \cdot 1150}{140} = 967,14$ <p>óra az átlagos élettartam</p>
701-800	12	750	
801-900	28	850	
901-1000	46	950	
1001-1100	32	1050	
1101-1200	22	1150	
Összesen	140		

Az adatokból elkészíthető a gyakorisági diagram, amelyet hisztogramnak nevezünk. Az egyes részintervallumok hossza 100 egység (egyenletes felosztás), a téglalapok magassága pedig az adott intervallumba eső elemek számával egyenlő.

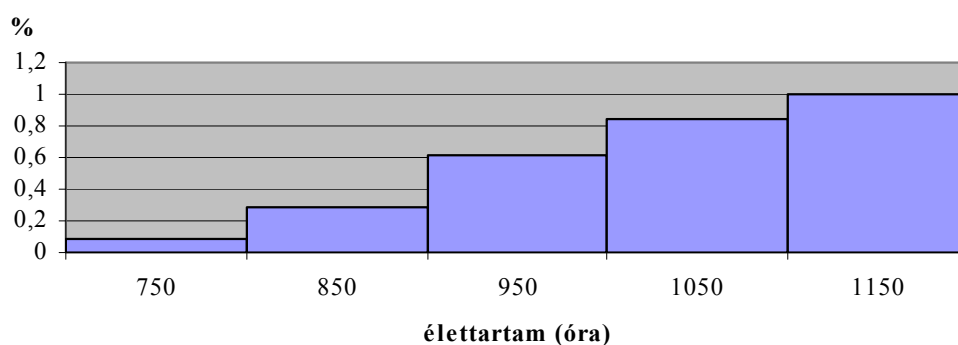


A tapasztalati eloszlásfüggvény:

Osztályokban (tól-ig) adott gyakoriságok esetén a tapasztalati eloszlásfüggvénynél az ugrásokat az osztályközepeknél jelöljük. Esetünkben az osztályok: [701;800), [801;900), [901;1000), [1001;1100) és [1101;1200) (óra), így az osztályközepek pedig 750, 850, 950, 1050, 1150 (óra). Először ki kell számolni a relatív gyakoriságot (g_i) és kumulált relatív gyakoriságot (G_i).

Élettartam (óra)	f_i	g_i	g_i'
701-800	12	0,09	0,09
801-900	28	0,20	0,29
901-1000	46	0,33	0,61
1001-1100	32	0,23	0,84
1101-1200	22	0,16	1,00
Összesen	140	1	-

A tapasztalati eloszlásfüggvény:



6.2. Helyzeti középértékek

A helyzeti középértékek a sokaságban elfoglalt helyzetüknél fogva jellemzik a vizsgált jelenséget vagy folyamatot. Ahhoz, hogy az egyedek sokaságon belüli elhelyezkedése jellemezhető legyen, az egyedeket valamilyen előre rögzített szabály szerint sorba kell rendezni. Általában növekvő sorrendbe rendezzük az ismértékeket, azaz rangsort képezünk.

6.2.1. Módusz

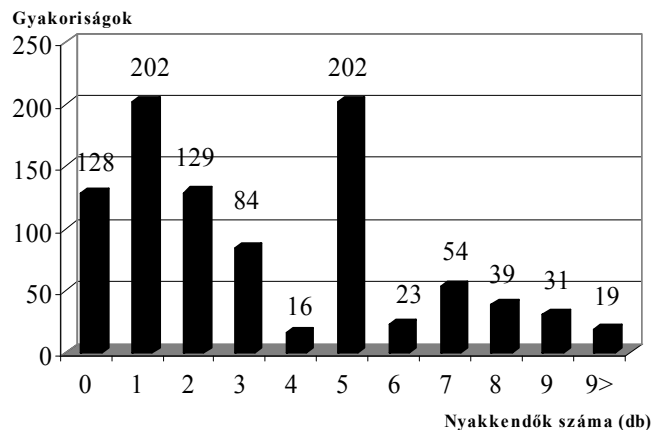
A legegyszerűbb helyzeti mutató a leggyakoribb érték, amely egyben a „legsűrűbb” érték, és **módusznak** nevezzük. A módusz egy gyakorisági eloszlásnak az az ismértékváltozata, amely a leggyakrabban fordul elő, azaz a legnagyobb gyakorisággal. Egy eloszlásnak több módusza is lehet.

4. bemutató feladat

927 férfit megkérdeztek a nyakkendőjük számáról:

Ny. száma	0	1	2	3	4	5	6	7	8	9	9>	Össz.
Gyak.	128	202	129	84	16	202	23	54	39	31	19	927

Diagramon ábrázolva:



$M_o=1$ és $M_o=5$

Azaz ennek a gyakorisági sornak két módusza van.

Folytonos gyakorisági sor esetén már nehezebb megállapítani a módusz konkrét értékét. Az osztályközös gyakorisággal megadott adatok esetében a statisztikai gyakorlat nem egységes. Kevés ismértértéket felvehető diszkrét jelenségnél adott osztályköz felső határát meghaladja a következő osztályköz alsó határa, és ekkor az osztályközepeknél az osztályközök tényleges és megadott határai megegyeznek egymással, és az osztályba sorolás egyértelmű. Ekkor az osztályközök megadott alsó és felső határaival kell számolni. Folytonos ismértértékeknel vagy sok ismértértékkel rendelkező diszkrét ismértértékeknel az osztályköz-határok kétféle megadása lehetséges. Első esetben az adott osztályköz felső határa és a következő osztályköz alsó határa megegyezik egymással, és az osztályköz-határral megegyező értékű elemeket az alacsonyabb osztályközbe soroljuk be. Második esetben közölt osztályköz-határokat adunk meg, amelyek csak arról tájékoztatnak, hogy az osztályköz-határokkal megegyező értékű elemeket az alacsonyabb osztályközbe soroljuk. Ekkor az osztályközepek kiszámításához és a módusz és medián becsléséhez az előző osztályköz felső határával számolunk.

$$M_o = m_0 + \frac{f_{m_o} - f_{m_{o-1}}}{(f_{m_o} - f_{m_{o-1}}) + (f_{m_o} - f_{m_{o+1}})} * h$$

m_o : a modális osztályköz alsó határa,

f_{m_o} : a modális osztályköz gyakorisága

$f_{m_{o-1}}$: a modális osztályközt megelőző osztályköz gyakorisága,

$f_{m_{o+1}}$: a modális osztályközt követő osztályköz gyakorisága

h : az osztályköz hossza

5. bemutató feladat

140 villanyégő égési idejét vizsgálva az alábbi adatokat kaptuk:

<i>Élettartam (óra)</i>	<i>f_i</i>	<i>Osztály- közép</i>
-800	12	750
801-900	28	850
901-1000	46	950
1001-1100	32	1050
1101-	22	1150
<i>Összesen</i>	<i>140</i>	

Számítsuk ki az égők élettartamának móduszát!

Megoldás:

A táblázatban jól látható, hogy a legtöbb adat 901-100 között van, azaz a leggyakoribb érték itt található.

$$M_o = m_o + \frac{f_{m_o} - f_{m_{o-1}}}{(f_{m_o} - f_{m_{o-1}}) + (f_{m_o} - f_{m_{o+1}})} * h = 900 + \frac{46 - 28}{(46 - 28) + (46 - 32)} 100 = 956,25$$

Az égők élettartama 956,25 óra körül sűrűsödi, azaz ez a leggyakoribb élettartam.

6.2.2. Medián

A másik leggyakrabban alkalmazott helyzeti középérték a medián. A nagyság szerint sorba rendezett értékek közül a középső. A medián az az érték, amelynél kisebb értékek gyakorisága azonos a nálánál nagyobb értékek gyakoriságával, azaz a medián a megfigyelt értékek rangsorát két egyenlő részre osztja.

Páratlan számú adat esetén:

$$Me = \frac{n+1}{2} \text{ -edik elem a medián.}$$

6. bemutató feladat

Egy vizsgán 17 hallgató az alábbi pontszámokat érte el:

8, 8, 9, 10, 10, 10, 11, 11, 12, 12, 13, 13, 13,
14, 14, 14, 14,

Megoldás:

A medián a 9. elem, így $Me=12$

Páros számú adat esetén: a medián sorszámát (s_{Me}) az alábbiak alapján számítjuk ki

$s_{Me} = \frac{n+1}{2} - 0,5$ és $\frac{n+1}{2} + 0,5$ -edik elem lesz a medián, a hozzátartozó értékeket átlagolni kell.

7. bemutató feladat

8 rúdugró az alábbi magasságokat ugrotta át:

4,6 4,8 4,8 4,9 5,0 5,0 5,1 5,3 (m)

Megoldás:

A medián a 4. és 5. elem átlaga lesz:

$$Me = \frac{1}{2} * (4,9 + 5,0) = 4,95 \text{ m}$$

Folytonos gyakorisági sor esetén hasonlóan a móduszhoz, nehezebb megállapítani a középső értéket.

$$Me = x_{me,a} + \frac{\frac{n}{2} - f'_{me-1}}{f_{me}} \cdot h_{me}$$

$n/2$: a medián sorszáma (középső érték)

me : a mediánt tartalmazó osztályköz alsó határa

f'_{me-1} : a mediánt tartalmazó osztályközt megelőző osztályköz kumulált gyakorisága

f_{me} : a mediánt tartalmazó osztályköz gyakorisága

h : az osztályköz hossza

8. bemutató feladat

140 villanyégő égési idejét vizsgálva az alábbi adatokat kaptuk:

<i>Élettartam (óra)</i>	<i>f_i</i>	<i>Osztály- közép</i>
-800	12	750
801-900	28	850
901-1000	46	950
1001-1100	32	1050
1101-	22	1150
Összesen	140	

Számítsuk ki az égők élettartamának mediánját!

Megoldás:

Először ki kell számolni a kumulált gyakoriságot:

<i>Élettartam (óra)</i>	<i>f_i</i>	<i>Osztály- közép</i>	<i>f'_i</i>
-800	12	750	12
801-900	28	850	40
901-1000	46	950	86
1001-1100	32	1050	118
1101-	22	1150	140
Összesen	140	-	-

$$Me = x_{me,a} + \frac{\frac{n}{2} - f'_{me-1}}{f_{me}} \cdot h_{me} = 900 + \frac{\frac{140}{2} - 40}{46} \cdot 100 = 965,22$$

Az égők felének élettartama 965,22 óránál kevesebb, felének pedig több.

6.2.3. Kvantilisek

A középértékek mellett fontos helyzeti mutatók a kvantilisek. Ha a rangsorba rendezett sokaságot egy x -ismérvérték $q:(1-q)$ arányban osztja ketté, akkor ezt az ismérvértéket q -ad rendű vagy q -adik kvantilisnek nevezzük. A kvantilisek meghatározása egyúttal a sokaság egy osztályozását jelenti. Ezen osztályozás során egyenlő gyakoriságú osztályközöket kapunk.

6.2.3.1. Kvartilisek

A sokaságot négy egyenlő elemszámú részsokaságra bontjuk.

Az alsó kvartilis: Q_1 , amely megmutatja, hogy a sokaság $1/4$ -része mely értéknél kisebb.

A középső kvartilis: Q_2 , amely megegyezik a mediánnal.

A felső kvartilis: Q_3 , amely megmutatja, hogy a sokaság $3/4$ -része mely értéknél kisebb, illetve $1/4$ része mely értéknél nagyobb.

Az alsó kvartilis sorszáma:

$$s = \frac{n+1}{4}$$

A felső kvartilis sorszáma pedig:

$$s = 3 * \frac{n+1}{4}$$

A képletek megegyeznek a medián képletével, természetesen a megfelelő fogalmak cseréje esetén.

Pl. a felső kvartilis kiszámítása:

$$Q_3 = q_3 + \frac{3 * \frac{n}{4} - f'_{q_{3-1}}}{f_{q_3}} * h$$

$3*n/4$: a felső kvartilis sorszáma

q_3 a felső kvartilist tartalmazó osztályköz alsó határa

$f'_{q_{3-1}}$: a felső kvartilist tartalmazó osztályköz megelőző osztályköz kumulált gyakorisága

f_{q_3} : a felső kvartilist tartalmazó osztályköz gyakorisága

h : az osztályköz hossza

9. bemutató feladat

140 villanyégő égési idejét vizsgálva az alábbi adatokat kaptuk:

<i>Élettartam (óra)</i>	<i>f_i</i>	<i>Osztály- közép</i>
-800	12	750
801-900	28	850
901-1000	46	950
1001-1100	32	1050
1101-	22	1150
Összesen	140	

Számítsuk ki az égők élettartamának alsó és felső kvartilisét!

Megoldás:

<i>Élettartam (óra)</i>	<i>f_i</i>	<i>Osztály- közép</i>	<i>f'_i</i>
-800	12	750	12
801-900	28	850	40
901-1000	46	950	86
1001-1100	32	1050	118
1101-	22	1150	140
Összesen	140	-	-

Az alsó kvartilis $140/4=35$, azaz a 35. elem, ez a 801-900-as intervallumban található:

$$Q_1 = q_1 + \frac{\frac{n}{4} - f'_{q_{1-1}}}{f_{q_1}} * h = 800 + \frac{\frac{140}{4} - 12}{28} * 100 = 882,14$$

Az felső kvartilis $3*(140/4)=105$, azaz a 105. elem, ez a 1001-1100-as intervallumban található:

$$Q_3 = q_3 + \frac{3 * \frac{n}{4} - f'_{q_{3-1}}}{f_{q_3}} * h = 1000 + \frac{3 * \frac{140}{4} - 86}{32} * 100 = 1059,375$$

6.2.3.2. Tercilisek

Ha két osztópont segítségével három egyenlő részre osztjuk a sokaságot, akkor terciliseket kapunk (T_1 és T_2). Az alsó tercilis (T_1) a sokaság első harmadolópontja, amely megmutatja, hogy mely értéknél kisebb a sokaság 1/3-ad része, illetve mely értéknél nagyobb a sokaság 2/3-ad része. A felső tercilis (T_2) a sokaság felső harmadolópontja, amely megmutatja, hogy mely értéknél kisebb a sokaság 2/3-ad része, illetve mely értéknél nagyobb a sokaság 1/3-ad része.

Kiszámításuk a mediánhoz hasonló.

Az alsó tercilis képletet:

$$T_1 = t_1 + \frac{\frac{n}{3} - f'_{t_{1-1}}}{f_{t_1}} * h$$

A felső tercilis képlete:

$$T_2 = t_2 + \frac{2 * \frac{n}{3} - f'_{t_{3-1}}}{f_{t_3}} * h$$

10. bemutató feladat

140 villanyégő égési idejét vizsgálva az alábbi adatokat kaptuk:

<i>Élettartam (óra)</i>	<i>f_i</i>	<i>Osztály- közép</i>
-800	12	750
801-900	28	850
901-1000	46	950
1001-1100	32	1050
1101-	22	1150
<i>Összesen</i>	<i>140</i>	

Számítsuk ki az égők élettartamának alsó és felső tercilisét!

Megoldás:

<i>Élettartam (óra)</i>	<i>f_i</i>	<i>Osztály- közép</i>	<i>f'_i</i>
-800	12	750	12
801-900	28	850	40
901-1000	46	950	86
1001-1100	32	1050	118
1101-	22	1150	140
Összesen	140	-	-

Az alsó tercilis $140/3=46,67$, azaz a 46-47. elem, ez a 901-1000-as intervallumban található:

$$T_1 = t_1 + \frac{\frac{n}{3} - f'_{t_{1-1}}}{f_{t_1}} * h = 900 + \frac{\frac{140}{3} - 40}{46} * 100 = 914,49$$

Az felső tercilis $2*(140/3)=93,33$, azaz a 93-94. elem, ez a 1001-1100-as intervallumban található:

$$T_2 = t_2 + \frac{2 * \frac{n}{3} - f'_{t_{3-1}}}{f_{t_3}} * h = 1000 + \frac{2 * \frac{140}{3} - 86}{32} * 100 = 1022,92$$

6.2.3.3. Decilisek

A decilisek számítása során 9 osztópont segítségével tíz egyenlő részre bontjuk a sokaságot, azaz tíz egyenlő gyakoriságú osztályközöt hozunk létre. Az első decilis (D_1) megmutatja, hogy a sokaság $1/10$ -d része mely értéknél kisebb.

Kiszámítása a kvartilisekhez hasonló, természetesen a megfelelő jelölések változtatásával. Például a 7. decilis értéke az alábbi képlettel számítható ki:

$$D_7 = d_7 + \frac{7 * \frac{n}{10} - f'_{d_{7-1}}}{f_{d_7}} * h$$

11. bemutató feladat

140 villanyégő égési idejét vizsgálva az alábbi adatokat kaptuk:

<i>Élettartam (óra)</i>	<i>f_i</i>	<i>Osztály- közép</i>
-800	12	750
801-900	28	850
901-1000	46	950
1001-1100	32	1050
1101-	22	1150
Összesen	140	

Számítsuk ki az égők élettartamának 4. és 7. decilisét!

Megoldás:

<i>Élettartam (óra)</i>	<i>f_i</i>	<i>Osztály- közép</i>	<i>f_i'</i>
-800	12	750	12
801-900	28	850	40
901-1000	46	950	86
1001-1100	32	1050	118
1101-	22	1150	140
Összesen	140	-	-

A 4. decilis $4 \cdot (140/10) = 56$, azaz az 56. elem, ez a 901-1000-as intervallumban található:

$$D_4 = d_4 + \frac{4 \cdot \frac{n}{10} - f'_{d_4-1}}{f_{d_4}} \cdot h = 900 + \frac{4 \cdot \frac{140}{10} - 40}{46} \cdot 100 = 934,78$$

A 7. decilis $7 \cdot (140/10) = 98$, azaz az 98. elem, ez a 1001-1100-as intervallumban található:

$$D_7 = d_7 + \frac{7 \cdot \frac{n}{10} - f'_{d_7-1}}{f_{d_7}} \cdot h = 1000 + \frac{7 \cdot \frac{140}{10} - 86}{32} \cdot 100 = 1037,5$$

6.3. A statisztikai minta további jellemzői

A $\xi_1, \xi_2, \dots, \xi_n$ minta eloszlásfüggvénye (empirikus eloszlásfüggvény)

$$F_n(x) = \begin{cases} 0, & \text{ha } x \leq \xi_1^*, \\ \frac{k}{n}, & \text{ha } \xi_k^* < x < \xi_{k+1}^*, \text{ ha } k = 1, 2, \dots, n-1, \\ 1, & \text{ha } x \geq \xi_n^* \end{cases}$$

ahol ξ_i^* a minta növekvő nagyság szerint rendezett elemei közül az i -edik.

Legyen $\xi_1, \xi_2, \dots, \xi_n$ egy adott n elemű minta, az a, b számokra pedig teljesüljön az $a \leq \xi_1^*$ és a $\xi_n^* < b$ feltétel. Osszuk fel az $[a; b]$ intervallumot m részintervallumra (osztályra) az $a = x_0 < x_1 < \dots < x_{m-1} < x_m = b$ osztópontok segítségével. Az egyes $[x_{i-1}, x_i)$ részintervallum-ba eső mintaelemek számát jelöljük k_i -vel ($i = 1, 2, \dots, m$).

A **gyakorisági hisztogram**ot úgy kapjuk, hogy az $[x_{i-1}, x_i)$ intervallumra $\frac{k}{x_i - x_{i-1}}$ magasságú téglalapot rajzolunk ($i = 1, 2, \dots, m$).

A **sűrűség hisztogram**ot úgy kapjuk, hogy az $[x_{i-1}, x_i)$ intervallumra $\frac{k}{n \cdot (x_i - x_{i-1})}$ magasságú téglalapot rajzolunk ($i = 1, 2, \dots, m$).

6.4. Szóródási mutatók

A középértékek csak egyetlen tulajdonságát rögzítik az eloszlásnak, ezért a statisztikai sokaság jellemzésére általában nem elegendőek. Egy sokaság eloszlása nagyon sokféle lehet, így az eloszlás formája is különböző lehet.

Az ismértértékek szóródásának mérésére és mérőszámok képzésére különböző lehetőségek vannak.

Szóródáson azonos fajta számszerű adatok (általában egy mennyiségi ismértértékének) különbözőségét értjük. A szóródás mérése az ismértértékek valamely középértéktől vett eltérései vagy egymás közötti különbségei alapján történik.

Ezen eltérések, különbségek alapján számított mérőszámok a **szóródás abszolút mutatói**, amelyek mértékegysége megegyezik a megfigyelt ismért mértékegységével.

A szóródás **relatív mutatói** elvonatkoztatnak az ismértértékek mértékegységétől, nagyságrendjétől, a szóródás térbeli vagy időbeli összehasonlítására szolgálnak.

A szóródás legfontosabb mutatószámai:

- terjedelem (R),
- szórás (σ , s),
- relatív szórás (V, CV)

6.4.1. A szóródás terjedelme (R)

Azt mutatja meg, hogy a sokaság elemei milyen értékintervallumban helyezkednek el.

A **szóródás terjedelme** alatt az előforduló legkisebb és legnagyobb érték különbségét értjük:

$$R = x_{\max} - x_{\min}$$

A szóródás terjedelme egy nagyon egyszerű és csak közelítő mérőszáma a szóródásnak. Az eloszlásnak csak a legkisebb és a legnagyobb értékét veszi figyelembe, és a két szélső érték közötti többi értéket nem. Arról, hogy a többi érték hol helyezkedik el, nem mond semmit.

6.4.2. A szórás

A legismertebb és a leggyakrabban használt mutató a szórás. A szórás az egyes értékek számtani átlagtól való eltéréseinek négyzetes átlaga, azaz megmutatja, hogy az egyes ismértértékek átlagosan mennyivel térnek el az átlagtól.

Tapasztalati szórás: az egész sokaságra vonatkozik.

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} \quad \sigma = \sqrt{\frac{\sum f_i (x_i - \bar{x})^2}{\sum f_i}} \quad \sigma = \sqrt{\sum g_i (x_i - \bar{x})^2}$$

elméleti szórás (korrigált tapasztalati szórás): a mintából számoljuk ki.

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}} \quad s = \sqrt{\frac{\sum f_i (x_i - \bar{x})^2}{\sum f_i - 1}}$$

Tulajdonságai:

- Ha az ismértértékhez hozzáadunk egy állandót (A), a szórás értéke nem változik, mivel ilyenkor a számtani átlag is pont ezzel az állandóval lesz nagyobb.
- Ha az ismértértéket megszorozzuk egy állandóval (B), akkor a szórás $|B|$ -szeresére változik, mivel ebben az esetben a számtani átlag értéke „B”-szer nagyobb lesz.
- Ha $x_i > 0$, akkor a szórásra az alábbi összefüggést írhatjuk fel:

$$0 \leq \sigma \leq \bar{x} * \sqrt{n - 1}$$

Az alsó korlát a $\sigma=0$ minden olyan esetben fennáll, amikor $x_i = \bar{x}$, ($i=1,2,\dots,n$). A felső korlát $\sigma = \bar{x} * \sqrt{n-1}$ csak akkor igaz, ha $x_i=0$ és $x_n = n * \bar{x}$.

12. bemutató feladat

140 villanyégő égési idejét vizsgálva az alábbi adatokat kaptuk:

Élettartam (óra)	f_i	Osztály- közép
-800	12	750
801-900	28	850
901-1000	46	950
1001-1100	32	1050
1101-	22	1150
Összesen	140	

Számítsuk ki az égők átlagos élettartamának tapasztalati és elméleti szórását!

Megoldás:

$$\sigma = \sqrt{\frac{\sum f_i (x_i - \bar{x})^2}{\sum f_i}} = \sqrt{\frac{12 * (750 - 967,14)^2 + 28 * (850 - 967,14)^2 + \dots + 22 * (1150 - 967,14)^2}{140}} =$$

$$= 117,0073$$

$$s = \sqrt{\frac{\sum f_i (x_i - \bar{x})^2}{\sum f_i - 1}} = \sqrt{\frac{12 * (750 - 967,14)^2 + 28 * (850 - 967,14)^2 + \dots + 22 * (1150 - 967,14)^2}{139}} =$$

$$= 117,494$$

A szórást leggyorsabban a zsebszámológépek statisztikai üzemmódjával lehet meghatározni

$\sigma=117,0073$ óra

$s=117,494$ óra

6.4.2.1. Variancia

A variancia, vagy szórásnégyzet önálló mutatóként is használatos. Nagy jelentősége például a varianciaanalízisben van, amely több középérték összehasonlítására szolgál. Többféle jelölést találunk a szakirodalomban: s^2 , SS, MQ

6.4.2.2. A szórásnégyzet és a szórás felbontása

Ha a sokaság részekre, azaz részsokaságokra bontható, akkor a szórásnégyzetet is fel lehet bontani két részre:

- Belső szórásnégyzetre: σ_B^2
- Külső szórásnégyzetre σ_K^2

A belső- és külső szórásnégyzet összege a teljes szórásnégyzet:

$$\sigma_T^2 = \sigma_B^2 + \sigma_K^2$$

és a csoport főátlaga között értelmezett különbségekből meghatározott) szórásnégyzeteinek az átlaga.

$$\sigma_j^2 = \frac{\sum_{i=1}^j (x_{ij} - \bar{x}_j)^2}{n_j}$$

$$\sigma_B^2 = \frac{\sum_{j=1}^n n_j * \sigma_j^2}{n}, \quad \sigma_B^2 = \frac{\sum_{j=1}^n \sum_{i=1}^j (x_{ij} - \bar{x}_j)^2}{n}$$

ahol j : a csoportosító ismérv változatainak a száma
 σ_j^2 : a j -edik részsokaság szórásnégyzete
 n_j : a j -edik részsokaság elemszáma
 n : a sokaság elemszáma

Belső szórás:

$$\sigma_B = \sqrt{\sigma_B^2}$$

A **külső szórásnégyzet** az egyes részsokaságok átlagai (részátlag) és a teljes sokasági átlag (főátlag) eltérés négyzetösszegének az átlaga (a részsokaságok nagyságával súlyozva).

$$\sigma_K^2 = \frac{\sum_{j=1}^n n_j * (\bar{x}_j - \bar{x})^2}{n}, \quad \sigma_K^2 = \frac{\sum_{j=1}^n \sum_{i=1}^j (x_{ij} - \bar{x})^2}{n}$$

ahol \bar{x}_j : a részsokaságok átlaga
 \bar{x} : a teljes sokaság átlaga (főátlag)

Külső szórás:

$$\sigma_K = \sqrt{\sigma_K^2}$$

A **teljes szórás** (σ_T) azt fejezi ki, hogy a vizsgált sokaságban az egyes értékek átlagosan mennyivel térnek el a főátlagtól.

A **belső szórás** (σ_B) azt fejezi ki, hogy a részsokaságok egyes értékei átlagosan mennyivel térnek el a saját részátlaguktól.

A **külső szórás** (σ_K) azt fejezi ki, hogy a vizsgált sokaságban az egyes részsokaságok átlagai átlagosan mennyivel térnek el a sokasági főátlagtól.

13. bemutató feladat

A hallgatók szakok szerinti csoportosításban az alábbi eredményt érték el egy vizsgán. A maximálisan elérhető pontszám 100 pont volt.

Az egyes szakokon tanuló hallgatók vizsgaeredménye

Szak	Hallgatók száma (fő)	Az elért eredményük átlagpont-száma	Az eredmények szórás
Közgazdász	20	63	5,6
Jogász	30	60	5,9
Mérnök	25	50	6,4

Számítsa ki az adatok alapján a belső-, a külső és a teljes szórást, majd értelmezze azokat!

Megoldás:

A vizsga átlagpontszáma:

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{20 * 63 + 30 * 60 + 25 * 50}{20 + 30 + 25} = 57,47$$

A belső szórásnégyzet:

$$\sigma_B^2 = \frac{\sum n_j * \sigma_j^2}{n} = \frac{20 * 5,6^2 + 30 * 5,9^2 + 25 * 6,4^2}{20 + 30 + 25} = 35,94$$

A belső szórás:

$$\sigma_B = \sqrt{\sigma_B^2} = \sqrt{35,94} = 5,99$$

Az egyes szakokon belül a hallgatók eredménye átlagosan 5,99 ponttal tér el az adott szakon elért átlagos pontszámtól.

A külső szórásnégyzet:

$$\sigma_K^2 = \frac{\sum n_j * (\bar{x}_j - \bar{x})^2}{n} = \frac{20 * (63 - 57,47)^2 + 30 * (60 - 57,47)^2 + 25 * (50 - 57,47)^2}{20 + 30 + 25} = 29,32$$

A külső szórás:

$$\sigma_K = \sqrt{\sigma_K^2} = \sqrt{29,32} = 5,41$$

Az egyes szakok átlagos vizsgaeredménye 5,41 ponttal tér el az összes hallgató által elért átlageredménytől.

A teljes szórásnégyzet:

$$\sigma_T^2 = \sigma_B^2 + \sigma_K^2 = 35,94 + 29,32 = 65,26$$

A teljes szórás:

$$\sigma = \sqrt{\sigma^2} = \sqrt{65,26} = 8,08,$$

A vizsgán megjelent hallgatók eredménye átlagosan 8,08 ponttal tér el az összes hallgató által elért átlageredménytől.

6.4.3. Relatív szórás (V, CV)

Sok esetben a szóródás vizsgálatára a szórás, mint átlagos szóródási paraméter nem elegendő. Továbbá bizonyos esetekben szükség lehet arra, hogy az ismértékek nagyságrendjétől, illetve mértékegységétől elvonatkoztatott „tisztá szám” jellegű mérőszámmal mérjünk, és összehasonlíthatóvá tegyük a szórást. E célból bármelyik eddig bemutatott mérőszámot viszonyíthatjuk az átlaghoz. A legismertebb ilyen szóródási mérőszám a relatív szórás vagy variációs koefficiens, amely kifejezi, hogy a szórás az átlag hányad része. A relatív szórás az ismértékek átlagtól vett átlagos relatív eltérését fejezi ki. Mértékegység nélküli mutató, általában százalékban kifejezve adjuk meg.

$$V = \frac{\sigma}{\bar{x}}, \text{ illetve } V = \frac{s}{\bar{x}}$$

Kifejezi, hogy a sokaság egyes egyedeinek értéke átlagosan hány százalékkal tér el az átlagtól.

Mivel mértékegység nélküli mutató, ezért különböző mértékegységű ismértékek szóródásának összehasonlítására is alkalmas. A szóráshoz hasonlóan megadhatjuk az alsó és a felső korlátját:

$$0 \leq V \leq \sqrt{n-1}$$

A szóródási együtthatót százalékosan kifejezve a gazdasági gyakorlatban a változékonyságot az alábbiak szerint minősíthetjük:

- 0-10%: állandóságot (homogenitást)
- 10-20%: közepes változékonyságot,
- 20-30%: erős változékonyságot
- 30% felett szélsőséges ingadozást fejez ki.

14. bemutató feladat

140 villanyégő égési idejét vizsgálva az alábbi adatokat kaptuk:

<i>Élettartam (óra)</i>	<i>f_i</i>	<i>Osztály- közép</i>
-800	12	750
801-900	28	850
901-1000	46	950
1001-1100	32	1050
1101-	22	1150
Összesen	140	

Számítsuk ki az égők átlagos élettartamának relatív tapasztalati és relatív elméleti szórását!

Megoldás:

Ha a szórás az alapsokaságra vonatkozna:

$$V = \frac{\sigma}{\bar{x}} = \frac{117,073}{967,14} = 0,1211 = 12,11\%$$

A villanyégők élettartama átlagosan 12,11%-kal tér el a sokaság átlagától

Ha a szórás a mintára vonatkozik

$$V = \frac{s}{\bar{x}} = \frac{117,494}{967,14} = 0,1215 = 12,15\%$$

A villanyégők élettartama átlagosan 12,15%-kal tér el a sokaság átlagától