

Matematika III. 7.

Helyzetmutatók, átlagok, kvantilisek

Prof. Dr. Závoti , József

Matematika III. 7. : Helyzetmutatók, átlagok, kvantilisek

Prof. Dr. Závoti , József

Lektor : Bischof , Annamária

Ez a modul a TÁMOP - 4.1.2-08/1/A-2009-0027 „Tananyagfejlesztéssel a GEO-ért” projekt keretében készült. A projektet az Európai Unió és a Magyar Állam 44 706 488 Ft összegben támogatta.

v 1.0

Publication date 2010

Szerzői jog © 2010 Nyugat-magyarországi Egyetem Geoinformatikai Kar

Kivonat

Ez a modul a statisztika elemi számítási módszereivel ismerteti meg az olvasót. Elsajátíthatja a középértékek, a medián, a módusz, a kvantilisek és egyéb átlagok gyakorlati számítási eszközeit.

Jelen szellemi terméket a szerzői jogról szóló 1999. évi LXXVI. törvény védi. Egészének vagy részeinek másolása, felhasználás kizárólag a szerző írásos engedélyével lehetséges.

Tartalom

7. Helyzetmutatók, átlagok, kvantilisek	1
1. 7.1 Bevezetés	1
2. 7.2 Középértékek	1
3. 7.3 Számítási közép (átlag)	1
4. 7.4 Medián (középső érték)	6
5. 7.5 Módusz	7
6. 7.6 Egyéb átlagfajták	7
7. 7.7 Kvantilisek	8
8. 7.8 Összefoglalás	9

7. fejezet - Helyzetmutatók, átlagok, kvantilisek

1. 7.1 Bevezetés

Jelen modul a Matematika III. tárgy hetedik fejezete, modulja. Az itt következő ismeretek megértéséhez javasoljuk, hogy olvassa el a Tárgy korábbi moduljainál írottakat. Amennyiben ez még nem lenne elég a megértéshez, akkor forduljon a szerzőhöz segítségért.

Jelen modul célja, hogy az Olvasó megismerkedjen a legfontosabb helyzetmutatókkal, és képessé váljon azok gyakorlati feladatok megoldásában való felhasználására.

A középértékek azonos fajta számszerű adatok centrumának közös jellemzői. Célunk használatukkal gyakorisági eloszlásokat kevés (1) adattal jellemezni.

2. 7.2 Középértékek

Fajtái:

1. **Számított középértékek vagy átlagok:** mindig számítással határozzuk meg őket. Értéküket minden egyes az átlagolásba bevont érték befolyásolja.

- a. számtani
- b. harmonikus
- i. mértani
- a. négyzetes

2. **Helyzeti középértékek:** az értékeknek egy bizonyos intervallumban való elhelyezkedése játszik szerepet. Az előforduló értékek egy része nem befolyásolja a középértékek nagyságát.

- a. módusz
- b. medián

Valamennyi középértéssel szemben támasztott **követelmény**, hogy közepes helyzetet foglaljon el, azaz a legkisebb x_{\min} és a legnagyobb x_{\max} értékek között helyezkedjen el. Fontos, hogy tipikus legyen, valamint könnyen értelmezhető, egyszerűen számítható.

3. 7.3 Számtani közép (átlag)

Definíció:

Adott: x_1, x_2, \dots, x_n n db alapsokaság metrikus skálán. Ekkor az x_i ismérvértékek **számtani átlaga**:

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

Példa:

Petiék biciklitáborban vesznek részt. Minden nap mérik a megtett távolságot, ami hétfőn 10 km, kedden 12, 7 km. Otthon kiszámolják, hogy ezen értékek számtani átlaga:

$$\bar{x} = \frac{10 + 12 + 16 + 12 + 17}{5} = \frac{67}{5} = 13,4$$

Azaz naponta átlag 13,4 kilométert tettek meg.

Megjegyzés:

Mivel ez az átlagfajta a legközismertebb, a mindennapokban gyakran elhagyják előle a számtani jelzöt.

Definíció:

Számtani átlagot nemcsak az egyenként ismert x_i adatokból, hanem **gyakorisági sorból** is számíthatunk. Ekkor:

$$\bar{x} = \frac{1}{n} \left(\underbrace{(x_1 + \dots + x_1)}_{f_1} + \underbrace{(x_2 + \dots + x_2)}_{f_2} + \dots + \underbrace{(x_k + \dots + x_k)}_{f_k} \right)$$

$$\bar{x} = \frac{1}{n} (x_1 f_1 + x_2 f_2 + \dots + x_k f_k)$$

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

ahol f_i az i -edik osztály gyakorisága, x_i az i -edik osztályhoz tartozó egyetlen ismérvérték.

Osztályozott gyakorisági eloszlások esetén:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k x'_i f_i$$

Ekkor x'_i az i -edik osztály közepe.

A számtani átlag számításához **relatív gyakoriságok** is használhatók:

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

$$\sum_{i=1}^n g_i = 1$$

g_i : abszolút gyakoriság

f_i : relatív gyakoriság

Ekkor a számtani átlag:

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

Példa:

Egy vállalatnál felmérték az alkalmazottak, összesen 250 ember éves keresetét.

Ezen ismerv alapján 10 osztályt alkottak, így számolták ki az átlagkeresetet.

osztály	Kereset (eFt/év)	f_i	x'_i	$x'_i \cdot f_i$	g_i	$x'_i \cdot g_i$
1.	500-799	6	650	3900	0,024	15,6
2.	800-1099	13	950	12350	0,052	49,4
3.	1100-1399	22	1250	27500	0,088	110
4.	1400-1699	32	1550	49600	0,128	198,4
5.	1700-1999	40	1850	74000	0,16	296
7.	2000-2299	42	2150	90300	0,168	361,2
7.	2300-2599	39	2450	95550	0,156	382,2
8.	2600-2899	31	2750	85250	0,124	341
9.	2900-3199	20	3050	61000	0,08	244

osztály	Kereset (eFt/év)	f_i	x'_i	$x'_i \cdot f_i$	g_i
1.	500-799	6	650	3900	0,024
2.	800-1099	13	950	12350	0,052
3.	1100-1399	22	1250	27500	0,088
4.	1400-1699	32	1550	49600	0,128
5.	1700-1999	40	1850	74000	0,16
7.	2000-2299	42	2150	90300	0,168
7.	2300-2599	39	2450	95550	0,156
8.	2600-2899	31	2750	85250	0,124
9.	2900-3199	20	3050	61000	0,08

Vagyis az egy dolgozóra jutó éves átlagkereset 2064,4 ezer Ft.

Definíció:

Súlyozott számtani középérték:

$$\bar{X} = \frac{w_1 X_1 + w_2 X_2 + \dots + w_n X_n}{w_1 + w_2 + \dots + w_k}$$

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

ahol: w_i -k az x_i értékekhez tartozó súlyszámok. Az x_i -k egymás közötti arányait szemléltetik.

Példa:

Egy vizsgán az írásbelin szerzett pontszámokat háromszoros, míg a szóbelin és a teszten elért pontokat egyszeres súlyozással veszik figyelembe. Az egyik tanuló írásbelin 85, szóbelin 70, míg a teszten 90 pontot szerzett. A végső jegynél az

$$\bar{x} = \frac{3 \cdot 85 + 1 \cdot 70 + 1 \cdot 90}{3 + 1 + 1} = 83$$

pontszámot veszik figyelembe.

Tétel:

A számtani átlag tulajdonságai:

1. Adott x_1, x_2, \dots, x_n adatok esetén a $d_i = x_i - \bar{x}$ előjeles hibák összességében kiegyenlítik egymást:

$$\sum_{i=1}^n (x_i - \bar{x}) = \sum_{i=1}^n x_i - n\bar{x} = n \frac{\sum_{i=1}^n x_i}{n} - n\bar{x} = n\bar{x} - n\bar{x} = 0$$

1. $(x_i - a)^2$ eltérésnégyzet-összeg akkor minimális, ha $a = \bar{x}$. Azaz fennáll az

$$\sum_{i=1}^n (x_i - \bar{x})^2 < \sum_{i=1}^n (x_i - a)^2$$

egyenlőtlenség, minden olyan esetben, amikor $a \neq \bar{x}$.

Bizonyítás:

Vegyük $f(a) = \sum_{i=1}^n (x_i - a)^2$ függvényt! Ennek 'a' szerinti első deriváltját nullával egyenlővé téve szélsőérték-helyet kapunk, ami pont \bar{x} :

$$f'(a) = -2 \sum_{i=1}^n (x_i - a) = 0$$

$$\sum_{i=1}^n x_i - n \cdot a = 0$$

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

A második derivált: $f''(a) = 2n > 0$, tehát a függvény az \bar{x} pontban veszi fel minimum-értékét.

1. Adottak x_i ismérértékek \bar{x} számtani átlaggal. Ekkor $x_i^* = \beta x_i + \alpha$, $i = 1, 2, \dots, n$ lineárisan transzformált ismérértékek számtani átlaga és az eredeti átlag között igazolható a következő összefüggés:

$$\bar{x}^* = \frac{\sum_{i=1}^n x_i^*}{n} = \frac{\sum_{i=1}^n (\beta x_i + \alpha)}{n} = \frac{\beta \sum_{i=1}^n x_i}{n} + \frac{n \cdot \alpha}{n} = \beta \bar{x} + \alpha$$

1. Ismerjük két részsokaság adatait:

${}_1x_i$ ($i = 1, 2, \dots, n_1$), átlaga ${}_1\bar{x}$

${}_2x_i$ ($i = 1, 2, \dots, n_2$), átlaga ${}_2\bar{x}$

Ekkor ${}_1x_i$ és ${}_2x_i$ elemekből álló egyesített sokaság átlaga:

$$\bar{x} = \frac{1}{n_1 + n_2} \left(\sum_{i=1}^{n_1} {}_1x_i + \sum_{i=1}^{n_2} {}_2x_i \right) = \frac{n_1 \cdot {}_1\bar{x} + n_2 \cdot {}_2\bar{x}}{n_1 + n_2}$$

Példa 1:

Átlag 250 km-t megyünk bérelt kocsinkkal naponta. Mennyibe kerül átlagosan az autókölcsönzés, ha az autókölcsönző naponta 4400 Ft fix díjat, valamint megtett km-enként 40 Ft-ot számol fel?

Az alapsokaság ekkor x_i : az egyes napokon megtett út.

$\alpha = 4400$ Ft (fix díj)

$\beta = 40$ Ft/km (benzinpénz)

$\bar{x} = 250$ km (átlag 250 km-t megyünk naponta)

A 3. tulajdonság alapján:

$$\bar{x}^* = 40 \frac{\text{Ft}}{\text{km}} \cdot 250 \text{ km} + 4400 \text{ Ft} = 14400 \text{ Ft}$$

. Ennyit fizetünk a kölcsönzőnek átlagosan naponta.

Példa 2:

Egy négyszáz fős üzemben az átlagkereset 29200 Ft. Egy másik üzemben 300 fő dolgozik, az ő átlagkeresetük 40100 Ft.

A 4. tulajdonság alapján együttesen a két üzemben dolgozók átlagosan

$$\bar{x} = \frac{400 \cdot 29200 + 300 \cdot 40120}{400 + 300} = 33880 \text{ Ft-ot keresnek.}$$

Megjegyzés:

A számtani közép nem mindig jó jellemzője egy sokaságnak, mivel nagyon érzékeny a kiugró értékekre. Például, ha egy 10 fős csoport 9 tagja 40000 Ft-ot keres, 1 pedig 400000 Ft-ot, a csoport átlagkeresete 76000 Ft:

$$\bar{x} = 40000 \cdot 0.9 + 400000 \cdot 0.1 = 76000$$

Ennek kiküszöbölésére alkalmazzák a **robosztus becslést** (trimmed mean), amikor a legkisebb és legnagyobb számot elhagyják az átlagolásnál.

4. 7.4 Medián (középső érték)

Definíció:

A **medián** rendezett mintában az a középső ismérérték, amelyiknél az összes adat fele kisebb, fele nagyobb.

Meghatározásának feltétele, hogy létezzon legalább ordinális skála. Ekkor rangsoroljuk az adatokat:

$$x_1^* \leq x_2^* \leq \dots \leq x_n^*$$

A medián értéke

$$m_e = x_{\left[\frac{n+1}{2}\right]}^*$$

ha n páratlan:

$$m_e = \frac{1}{2} \left(x_{\left[\frac{n}{2}\right]}^* + x_{\left[\frac{n}{2}+1\right]}^* \right)$$

ha n páros:

Példa:

$$m_e = x_{\left[\frac{5+1}{2}\right]}^* = x_3^* = 12$$

10, 12, 12, 16, 17 értékek esetén, ahol a minta nagysága páratlan (5):

Tétel:

$$\sum_{i=1}^n |x_i - m_e| < \sum_{i=1}^n |x_i - a|$$

Azaz, ha minden ismérértéket a mediánnal helyettesítenénk, akkor ezzel összességében a legkisebb hibát követnénk el.

Definíció:

Osztályozott adatokból kiindulva a mediánt a következő formulával becsülhetjük:

$$m_e = x_i^a + \frac{\frac{n}{2} - F_{i-1}}{f_i} \Delta x$$

ahol i azon legelső osztályköz sorszáma, melyre $F_i > \frac{n}{2}$

x_i^a : az i -edik osztályköz alsó határa

F_i : a kumulált gyakorisági sor i -edik eleme.

Megjegyzés:

A medián mindig egyértelműen meghatározható, mert bármilyenek is az ismérértékek, mindig található közöttük egy vagy több középső, ha azokat rangsorba rendezzük. Ha a rangsorban nagyon sok egyforma érték szerepel, akkor nem tanácsos használni, mert kevésbé illik rá a definíció.

5. 7.5 Módusz

Definíció:

A **módusz** a leggyakoribb érték a sokaságban.

Diszkrét ismérvték esetén a módusz a leggyakrabban előforduló ismérvték, folytonos ismérvték esetén pedig a gyakorisági görbe maximumhelye.

A módusz csak abban az esetben határozható meg, ha létezik legalább nominális skála.

Folytonos ismérvték esetén a módusz értéke csakis valamilyen osztályközös gyakorisági sorból kiindulva közelíthető, az alábbi formulával becsülhető:

$$m_o = x_i^* + \frac{f_i - f_{i-1}}{f_i - f_{i-1} + f_i - f_{i+1}} \cdot \Delta x$$

Ekkor i a leggyakoribb osztály sorszáma, Δx az osztályköz hosszúsága.

A meghatározáskor e képletet használva fontos, hogy az osztályközök azonos hosszúságúak legyenek.

Ha a módusz a legelső vagy a legfelső osztályközbe esik, akkor a képletbe $f_i - f_{i-1} = 0$ vagy $f_i - f_{i+1} = 0$ kerül.

A módusz nem mindig határozható meg egyértelműen, sőt nem is mindig létezik.

6. 7.6 Egyéb átlagfajták

1. Geometriai közép:

- diszkrét adatok esetén: $G = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}$
- gyakorisági adatok esetén: $G = \sqrt[n]{x_1^{f_1} \cdot x_2^{f_2} \cdot \dots \cdot x_k^{f_k}}$ $n = \sum_{i=1}^k f_i$

A geometriai átlag kiszámításának gyakorlati módja az alábbi:

$$\lg G = \frac{1}{n} (\lg x_1 + \lg x_2 + \dots + \lg x_n)$$

$$\lg G = \frac{1}{n} (f_1 \cdot \lg x_1 + f_2 \cdot \lg x_2 + \dots + f_k \cdot \lg x_k) = \sum_{i=1}^k f_i \cdot \lg x_i$$

1. Harmonikus átlag:

$$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}} \quad x_i \neq 0$$

$$\frac{1}{H} = \frac{1}{n} \left(\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n} \right)$$

1. Négyzetes átlag:

$$\text{RMS} = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n}}$$

A harmonikus, mértani és négyzetes átlag általában olyan esetekben használható, amikor nem az ismértékek összegének, hanem az azok négyzetösszegének, szorzatának, reciprokaiból képzett összegnek van valamilyen kézzelfogható értelme.

Például mértani átlagot könnyen számolunk láncviszonyozásokból, hiszen azok szorzata egy bázisviszonyozás.

Tétel:

A négy átlag közötti összefüggés a következő:

$$x_{\min} \leq H \leq G \leq \bar{x} \leq \text{RMS} \leq x_{\max}$$

Tétel:

A számtani átlag, a módusz és a medián közötti összefüggés: $|\bar{x} - m_o| = 3|\bar{x} - m_e|$

7. 7.7 Kvantilisek

Az osztályozás során nemcsak egyenlő hosszúságú, hanem egyenlő gyakoriságú osztályközök képzését is célul tűzhetjük ki.

Példa:

36 lakást ajánlottak az ingatlanközvetítő irodában. 6 napunk van a választásra, valamint minden nap 6 lakást akarunk megnézni, a könnyebb összehasonlítás kedvéért hasonló árúakat. Mondjuk, a legalacsonyabb árut tekintjük meg először és haladunk az egyre drágábbak felé.

Feltehetjük a következő kérdéseket:

Melyik a 6 legolcsóbb lakás?

A harmadik nap megnézett lakások árai milyen értékhatárokon belül mozognak?

Ekkor tulajdonképp a hatodrendű kvantiliseket adjuk meg.

Az osztályközök képzésénél a meghatározott osztópontokat p-ed rendű kvantilis értékeknek nevezzük.

Definíció:

A **p-ed rendű kvantilis** az a szám, amelynél az összes előforduló ismérték p-ed része nem nagyobb, (1-p)-ed

része nem kisebb. Például az $\frac{x_2}{5}$ kvantilis esetében az adatok 40%-a nem nagyobb, 60%-a nem kisebb a meghatározott kvantiliséknél.

Meghatározásánál fontos az adatok sorrendbe való rendezése.

Megkülönböztetett kvantilisek:

1. **Medián** (m_e): két egyenlő gyakoriságú részre osztja a sokaságot.
2. **Kvartilisek** (Q_1, Q_2, Q_3): a 3 kvantilis négy egyenlő gyakoriságú részre osztja a rendezett halmazt.
3. **Kvintilis** (K_1, K_2, K_3, K_4): öt egyenlő rész

4. **Decilis** ($D_1, D_2, \dots, D_8, D_9$): 10 egyenlő rész

5. **Percentilis** ($P_1, P_2, P_3, \dots, P_{98}, P_{99}$): 100 egyenlő rész

A definícióból egyértelműen következik, hogy például $Q_2 = m_e = D_5$, vagy például $P_{25} = Q_1$.

8. 7.8 Összefoglalás

1. Az Express újságban 1995. 10. 04.-én eladásra kínált 70 m² körüli lakások ára (mFt):

2.0, 4.0, 3.1, 3.4, 4.2, 6.0, 3.6, 3.1, 2.6, 3.3, 3.4, 3.5, 2.4, 3.2, 3.8, 3.1, 5.3, 2.5, 3.6, 3.0, 3.5, 3.5, 4.1.

a. Határozza meg az adatok számtani közepét, mediánját, móduszát!

b. Számítsa ki és értelmezze a ferdeségi mutatószámot, a csúcsossági mutatószámot!

1. 48 db eladásra kínált lakás megoszlása a kínálati ár szerint

Ár (mFt)	Lakások száma (db)
2.0-2.9	6
3.0-3.9	19
4.0-4.9	11
5.0-5.9	6
6.0-6.9	3
7.0-7.9	3
Összesen	48

Számítsa ki és értelmezze a helyzetmutatókat (átlag, módusz, medián)!

1. Egy iparág vállalataira vonatkozóan az alábbi adatokat ismerjük:

Létszám (fő)	Vállalatok száma
- 500	10
501 - 1000	15
1001 - 1500	21
1501 - 2000	9
2001 - 2500	3
2501 -	2

Számítsa ki és értelmezze a helyzetmutatókat (átlag, módusz, medián)

1. Egy közúti forgalom-ellenőrzés során 1000 személygépkocsi lépte túl a megengedett sebességet. A túllépés mértéke:

Létszám (fő)	Vállalatok száma
- 500	10
501 - 1000	15
1001 - 1500	21
1501 - 2000	9
2001 - 2500	3
2501 -	2

Számítsa ki és értelmezze a helyzetmutatókat (átlag, módusz, medián)!

1. Egy közkezdvelt gyorsétterem-hálózat egyik egységében megfigyelték a kiszolgálási időt (mp):

45	48	49	56	61	66	66	66	70	72	72	75	78	79	81	81	83	95	102	135
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----

a. Határozza meg az adatok átlagát, mediánját, móduszát!

b. Határozza meg ugyanezen értékeket osztályozással is!

1. A 18 éves fiúk körében kísérleti jelleggel intelligenciateszteket végeztek. A vizsgálathoz felkért 19 főnél az alábbi intelligencia-értékeket (IQ) mértek:

141	65	75	100	99	96	89	82	101	110	104	119	107	103	114	104	130	122	58
-----	----	----	-----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----

a. Határozza meg az adatok átlagát, mediánját, móduszát!

b. Határozza meg ugyanezen értékeket osztályozással is!

Irodalomjegyzék

Hunyadi-Vita: *Statisztika közgazdászoknak*, KSH, Budapest, 2002

Keresztély, Sugár, Szarvas: *Statisztika példatár közgazdászoknak*, BKE, Nemzeti Tankönyvkiadó, 2005

Korpás A.: *Általános statisztika I-II.*, Nemzeti Tankönyvkiadó, Budapest, 1996

Csanády V., Horváth R., Szalay L.: *Matematikai statisztika*, EFE Matematikai Intézet, Sopron, 1995

Závoti, Polgárné, Bischof: *Statisztikai képletgyűjtemény és táblázatok*, NYME Kiadó, Sopron, 2009

Reimann J. - Tóth J.: *Valószínűségszámítás és matematikai statisztika*, Tankönyvkiadó, Budapest, 1991