KÖZÉPÉRTÉKEK A GYAKORLATBAN[1]

SUGÁR ANDRÁS
Budapesti Gazdasági Egyetem – Budapesti Corvinus Egyetem

1 Bevezetés

Időről időre (főleg a gyakorlati alkalmazásokból leszűrhető tapasztalatok alapján) érdemes áttekinteni még az egyszerűbb statisztikai módszerek elvi alapjait, gyakorlati alkalmazhatóságát is. Ebben a leírásban ezt tesszük a középértékek esetében. A középértékek a leggyakrabban használt statisztikai mutatók/módszerek közé tartoznak. A tankönyvek is eléggé hasonló módon, szerkezetben tárgyalják őket. Ebben a cikkben arra vállalkozunk, hogy röviden összefoglaljuk, mit is értünk középértéken, ezen belül főleg az átlagokkal foglalkozunk. Áttekintjük, milyen fajtái vannak, melyek a középértékek/átlagok tulajdonságai a szakirodalom (Pl. a gazdasági jellegű magyar felsőoktatás tananyagai, amelyek közül kiemeljük Köves–Párniczky (1981), Hunyadi–Vita (2003), Kerékgyártóné et al. (2009), Kovács (2011), Sándorné et al. (2013) tankönyveket/jegyzeteket, illetve Statisztikai Szemle cikkek alapján. Másrészt néhány újszerű szemponttal bővítjük az átlagokról szóló diskurzust, főleg az átlagok fajtái, azok gyakorlati használata, valamint a leggyakrabban használt számtani átlag általánosításai, variációi kapcsán.

2 A középértékekről

A középérték az információsűrítés egy szélsőséges (de nagyon hasznos) eszköze, egyetlen értékben testesítjük meg, tömörítjük az információt, amit fontosnak tartunk. (Ennek természetesen az az ára, hogy az eloszlás jellegére vonatkozó másfajta információt, pl. a szóródásra, vagy az eloszlás alakjára vonatkozókat a középérték nem mutatja meg, ezért a középértékek mellett egyéb mutatók számítása is fontos.) A középértékektől elvárjuk, hogy a) közepes helyzetűek, b) tipikusak, c) egyértelműen meghatározhatók, d) lehetőleg könnyen értelmezhetők legyenek. A gyakorlatban használt középértékek nem egyforma mértékben tesznek eleget ezeknek a követelményeknek, az átlagok az a) és c) követelményeknek megfelelnek, b) és d) követelményeknek nem feltétlenül. (A gyakorlatban általában feltételezzük, hogy a megfigyelt értékek nem mind ugyanazok, ebben a szélsőséges esetben ugyanis nem érdemes középértéket számolni.)

[1] Beérkezett 2024. február 27. DOI: https://doi.org/10.15170/SZIGMA.55.1241. E-mail: `sugar.andras@uni-bge.hu`.

A középértékek két fajtája a helyzeti és a számított középérték (utóbbiakat nevezzük átlagoknak). A helyzeti középértékek valamilyen jellegzetes helyzetüknél fogva (pl. szó szerint felezik a sokaságot, vagy tipikusak) jellemzik a sokaságot, a gyakorlatban is gyakran használjuk a mediánt és a móduszt. Mielőtt az átlagok részletesebb tárgyalásába kezdenénk, röviden jellemezzük a két helyzeti középértéket.

A medián a középen elhelyezkedő érték, az a speciális kvantilis (a kvantilis egy meghatározott arányú osztópont, itt speciálisan a felezőpont), amelynél egy ismérv szerint sorbarendezett értékei alapján a sokaság egyik fele kisebb, a másik fele nagyobb értéket vesz fel. A medián legalább sorrendi skálán értelmes mutató, de sok egyforma ismérvváltozat esetén nincs értelme, bár mechanikusan megadható. Miután az adatok a leggyakrabban egyediek (a gyakorlatban a megfigyelések egy listája/regisztere áll rendelkezésre), a medián általában egyértelműen meghatározható.

A módusz a tipikus ismérvváltozat. Értelmes minőségi ismérv esetén is. Diszkrét ismérv esetében ránézésre látszik, a leggyakrabban (vagy legnagyobb valószínűséggel) előforduló ismérvváltozat. Folytonos ismérvek/változók esetében az elméleti vagy empirikus sűrűségfüggvény maximumhelyéhez tartozó ismérvváltozat.

A szoftverek a számtani átlag mellett általában megadják a medián és módusz értékét, de utóbbinak folytonosnak tekinthető ismérv (a gyakorlatban olyan ismérv, aminek sok különböző ismérvváltozata van) esetén egyedi adatokból nincs értelme, mert véletlenszerű, hogy egy-egy ismérvváltozat többször is előfordul. Ilyenkor a móduszra érdemes becslést adni. Az osztályközös gyakorisági sorból adott becslések (interpolációk) az egyszerűbb megoldások, de akár a sűrűségfüggvényt is becsülhetjük (simított hisztogram), és ez alapján állapíthatjuk meg a móduszt.

A módusz értelmezését árnyalja, hogy léteznek többmóduszú eloszlások is, sőt egyenletes eloszlás esetén minden lehetséges érték módusz (vagy egyik sem). A gyakorlatban is előfordulnak többmóduszú eloszlások. A klasszikus statisztikában az egy mennyiségi ismérv szerinti elemzéskor egymóduszú eloszlások a kiindulópontok. A több módusz általában egy háttérváltozó szerinti heterogenitásra utal, érdemes a sokaságot olyan részekre bontani, ahol a mennyiségi ismérv szerinti eloszlás már egymóduszú.

Példaként vegyük a teljes munkaidőben alkalmazásban álló magyar munkavállalók 2023-as kereseti megoszlását. Magyarországon a KSH adatai alapján 2023-ban a havi bruttó átlagkereset 571 ezer Ft volt, míg a bruttó mediánkereset 450 ezer Ft (azaz a teljes munkaidőben alkalmazásban állók felének bruttó havi keresete 450 ezer forint alatt, felének felette volt 2023-ban). A móduszkeresetet nem számolják, mert a kereseteloszlás nem egymóduszú, a természetes sűrűsödési pont érzékelését befolyásolja, hogy a minimálbér, diplomás minimálbér körül is kialakulnak lokális sűrűsödési pontok.

A medián és módusz is (az átlagoktól eltérően) robusztus mutatók, az outlierekre nem érzékenyek.

3 Az átlag fogalma, fajtái

Az átlag a középérték egy fajtája. Az átlag fogalma, számítása szorosan kapcsolódik a valószínűségszámítás egyes területeihez is, de ezek tárgyalása meghaladja cikkünk lehetőségeit, pedig ez magyarázhatja azt is, hogy más matematikai területekhez képest az átlagok számítása történetileg nagyon későn került be a gazdasági/társadalmi gyakorlatba. L. erről Kendall (1956), Dusek (2022).

Mint az elnevezésben is benne van, a számított középértékek valamilyen számítás eredményei, ezeket a mutatókat szokás átlagoknak nevezni. Az átlagok esetén a statisztika tankönyvek sok évtizede négyfajta átlagot szoktak ismertetni, a harmonikus, mértani, számtani és négyzetes átlagot. Az *1. táblázatban* összefoglaljuk a négyféle átlagra vonatkozó legfontosabb jellemzőket, alapvetően Köves (1961) alapján.[2] A képletekben az egyszerűség kedvéért elhagyjuk az összegzésnél és szorzatnál használt futóindexek határait.

Átlagfajta	Súlyozatlan forma	Súlyozott forma	Visszavezetés a számtani átlagra
Harmonikus ($\overline{X}_h$)	$\frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\dots+\frac{1}{x_n}} = \frac{n}{\sum \frac{1}{x_i}}$	$\frac{\sum f_j}{\sum \frac{f_j}{x_j}}$	$\frac{1}{\overline{X}_h} = \frac{\sum \frac{1}{x_i}}{n}$
Mértani ($\overline{X}_g$)	$\sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n} = \sqrt[n]{\prod x_i}$	$\sqrt[\sum f_j]{\prod x_j^{f_j}}$	$\ln \overline{X}_g = \frac{\sum \ln x_i}{n}$
Számtani ($\overline{X}_a$)	$\frac{x_1+x_2+\ldots+x_n}{n} = \frac{\sum x_i}{n}$	$\frac{\sum f_j \cdot x_j}{\sum f_j}$	$\overline{X}_a = \frac{\sum x_i}{n}$
Négyzetes ($\overline{X}_q$)	$\sqrt{\frac{x_1^2+x_2^2+\ldots+x_n^2}{n}} = \sqrt{\frac{\sum x_i^2}{n}}$	$\sqrt{\frac{\sum f_j \cdot x_j^2}{\sum f_j}}$	$\overline{X}_q^2 = \frac{\sum x_i^2}{n}$

1. táblázat. Átlagok fajtái és tulajdonságaik

A leggyakrabban használt átlagfajta a számtani átlag, gyakran amikor átlagot mondunk, automatikusan a számtani átlagra utalunk. A számtani átlag definíció szerint az a szám, amit ha a számok helyébe írunk, akkor az összeg ugyanaz marad. Úgy is lehet interpretálni, ha az adatok összegét (az ún. értékösszeget) egyenletesen szétosztjuk, akkor egy megfigyelésre a számtani átlagnak megfelelő érték jut. Ez a megfogalmazás sugallja, hogy a számtani átlag akkor értelmes, ha az adatok logikailag összeadódnak, tehát értelmes az értékösszeg. Ez a legtöbb gyakorlati alkalmazás esetén igaz is (pl. átlagbérek esetén a bérek összege a kifizetett bértömeg, vagy egy termelékenységi mutatót említve az egy alkalmazott által előállított átlagos termékszám esetén az összes legyártott termék mennyisége), de nem mindig. Van, amikor ugyan összeadódnak logikailag az értékek, de a tényleges értékösszegnek csak erőltetett értelmet lehetne tulajdonítani (pl. az átlagos testmagasság vagy átlagos életkor esetén).

Több olyan általánosítása létezik a számtani átlagnak, amelyekbe a többi átlag is belefoglalható, ezek a gyakorlatban erőltetettek is lehetnek, de van,

[2]Köves Pál 63 évvel ezelőtti cikke a mai napig releváns, jó összefoglaló, de természetesen azóta is sok cikk született az átlagokról, azok fajtáiról, alkalmazásaikról; egy viszonylag friss áttekintést lehet olvasni pl. Carvalho (2016)-ban.

amikor további alkalmazásokra adnak alkalmat. Gyakori, hogy az ún. momentumot definiálják, illetve ennek további általánosításait, ezek közül a centrális momentum a legelterjedtebb. Az (1) képletek mutatják a momentumok és a centrális momentumok (súlyozatlan) számítását. A k-adik momentum (M_k) az x_i-k k-adik hatványainak számtani átlaga, ahol k természetes szám. Ha $k = 1$, akkor a számtani átlagot kapjuk, ha $k = 2$, akkor a négyzetes átlag négyzetét. A k-adik centrális momentum ($M_k(\overline{X})$) a számtani átlagtól vett eltérések k-adik hatványainak számtani átlaga. Ha $k = 1$, akkor 0-t kapunk (hiszen az átlagtól vett eltérések kiegyenlítik egymást, összegük 0), a második centrális momentum a variancia, ami a szóródás mértékének egyik mutatószáma (a négyzetgyöke a szórás, amely az átlagtól vett átlagos eltérés mértékét mutatja). A centrális momentumok esetén k magasabb értékeire számolt mutatókat is használjuk a gyakorlatban, $k = 3$ esetén az aszimmetria irányát és intenzitását mérő mutatóhoz jutunk, $k = 4$ esetén pedig a kurtózis (csúcsosság-lapultság) mértékét tudjuk számszerűsíteni. A statisztikai szoftverek által számolt aszimmetria- és kurtózis-mutatók szinte minden esetben a centrális momentumokon alapuló mutatók.

$$M_k = \frac{\sum x_i^k}{n} \qquad M_k(\overline{X}) = \frac{\sum (x_i - \overline{X})^k}{n}. \tag{1}$$

Még egy formula (általánosítás) említése lehet hasznos (vannak továbbiak is, de – főleg a gazdasági/társadalomtudományi gyakorlatban – számunkra nem tűnnek fontosnak). A (2) képlet mutatja az ún. hatványközepet (H_α), ami két területen is továbblép a momentumhoz képest.

$$H_\alpha = \left(\frac{\sum x_i^\alpha}{n}\right)^{\frac{1}{\alpha}}, \qquad x_i \in I\!R_+ \setminus \{0\}. \tag{2}$$

Egyrészt a k természetes számok mellett bármilyen valós (nem nulla) szám lehet a kitevőben, másrészt a hatványozás hatását a végén visszatranszformáljuk. Az 1. táblázatban felsorolt – a gyakorlatban használt 4 átlag felfogható a hatványközép 4 speciális eseteként, $\alpha = 1$ adja a számtani, $\alpha = 2$ a négyzetes átlagot, $\alpha = -1$ esetén kapjuk a harmonikus átlagot, és a mértani átlag formulája is megkapható, ha képezzük a $\lim_{\alpha \to 0} H_\alpha$ értéket. Ezek közül egyedül az utolsó nem látszik ránézésre, de annak bizonyítása is viszonylag egyszerű, l. pl. Sykora (2009).

Itt érdemes kiemelni, milyen gyakorlati esetben használjuk igazán a harmonikus átlagot (ezt a területet szokták a rész- és összetett viszonyszámok vagy rész- és főátlagok közötti összefüggésnek nevezni), amennyiben egy $V = A/B$ viszonyszámot részsokaságokra és a sokaság egészére is számolunk, akkor az ún. aggregát forma mellett az összetett viszonyszám a részviszonyszámok B_j-kkel súlyozott számtani, vagy A_j-kkel súlyozott harmonikus átlagaként adódik. Ezt mutatják a (3) képletei. Ennek speciális esete a rész- és főátlagok közötti összefüggés, amit a (4) képletekben lehet nyomon követni.

$$V_j = \frac{A_j}{B_j} \qquad \overline{V} = \frac{\sum A_j}{\sum B_j} = \frac{\sum B_j \cdot V_j}{\sum B_j} = \frac{\sum A_j}{\sum \frac{A_j}{V_j}} \tag{3}$$

$$A_j = N_j \cdot \overline{X_j} = S_j \qquad B_j = N_j \qquad V_j = \overline{X}_j$$

$$\overline{X} = \frac{\sum S_j}{\sum N_j} = \frac{\sum N_j \overline{X}_j}{\sum N_j} = \frac{\sum S_j}{\sum \frac{S_j}{\overline{X}_j}}, \tag{4}$$

ahol S_j a j-edik részsokasághoz tartozó részletösszeg, N_j a j-edik részsokaság mérete, $\overline{X}_j$ a j-edik részsokasághoz tartozó átlag. Egy technikai (de a gyakorlatban fontos) elem, hogy a súlyozáshoz elég a gyakoriságok vagy értékösszegek megoszlása is, ahogy az (5) képletek mutatják.

$$g_j = \frac{N_j}{\sum N_j} \qquad Z_j = \frac{S_j}{\sum S_j} \qquad \overline{X} = \sum g_j \cdot \overline{X}_j = \frac{1}{\sum \frac{Z_j}{\overline{X}_j}} \tag{5}$$

Végezetül érdemes iderakni a középértékek között nagyságviszonyt is, amennyiben a megfigyelések közül legalább 2 különbözik egymástól, akkor teljesül a (6) összefüggés.

$$x_{\min} < \overline{X}_h < \overline{X}_g < \overline{X}_a < \overline{X}_q < x_{\max} \tag{6}$$

Az átlagok közötti választás kérdésében érdemes a gyakorlati szempontokat figyelembe venni. Vegyük végig, melyik átlagfajta használata milyen körülmények esetén merül fel a gyakorlatban.

- A leggyakoribb számtani átlag használatáról már szó esett, minden olyan esetben, amikor az adatok tartalmilag összeadódnak, ennek az átlagnak a használata a logikus. Egy szempontra érdemes felhívni a figyelmet, a számtani átlag használata esetén legalább különbségi szintű mérést feltételezünk, ugyanakkor számos esetben számolnak számtani átlagot alacsonyabb szintű mérési szintű adatok esetén is, ami lehet a gyakorlatban célravezető, hasznos számítás, de érdemes a sajátosságaival/korlátaival tisztában lenni.

- A négyzetes átlag gyakorlati helye viszonylag egyértelmű. Ezt az átlagfajtát általában akkor használjuk, ha el akarunk tekinteni az adatok előjelétől, mert pl. azok összege nulla. Ez a helyzet a szóródás számszerűsítése esetén, amikor az átlagtól vett eltérések átlagát számoljuk (ezek számtani átlaga nulla), vagy az olyan modelleknél, ahol a hiba átlagos nagysága a kérdés, de a hibák kiegyenlítik egymást, az összegük szintén nulla. Az előjeltől eltekintés két módja az abszolút érték és a négyzetre emelés, mindkét logikára épülnek mutatószámok, de főleg következtető statisztikai megfontolások miatt az abszolút értékre alapozott mutatók kiszorultak a gyakorlatból. A szóródás mértékének leggyakrabban használt mutatója a szórás, ami az átlagtól vett eltérések négyzetes átlaga.

- A mértani átlag számolási módja mutatja, hogy akkor használatos, ha az adatok logikailag nem összeadódnak, hanem összeszorzódnak, az adatok szorzatának van tárgyi (gazdasági vagy egyéb) értelme. Éppen ezért a leggyakoribb felhasználása idősorok esetén a láncviszonyszámok átlagos értékének számolása (átlagos növekedési ütem). Több tankönyv kizárólag ezt a példát hozza az alkalmazásra. Azért használata máshol is megjelenik, főleg az indexszámítás területén (egyedi indexeknek esetleg nem számtani, hanem mértani átlagát veszik, erről l. részletesebben Köves (1957), vagy keresztezett indexformulák esetén használják (l. pl. Törnquist-index, vagy a legismertebb Fisher-index).

Az indexszámítási gyakorlatban több esetben alternatíva, hogy számtani vagy mértani átlagot számoljunk. Példaként a fogyasztóiár-index számítási gyakorlatából mutatunk egy illusztrációt. Egy termékcsoport indexe számolható pl. az abban a csoportban található reprezentánsok árindexeinek súlyozatlan átlagaként. A gyakorlat kétfajta indexet is használ, az egyik a reprezentáns árindexek súlyozatlan mértani, a másik a számtani átlaga. Ezek képlete (és szokásos elnevezése):

- Carli-féle termékcsoport-index (számtani átlag): $I^C_{P(1/0)} = \frac{1}{n}\sum_j \frac{p_{1j}}{p_{0j}}$
- Jevons-féle termékcsoport-index (mértani átlag): $I^J_{P(1/0)} = \sqrt[n]{\prod_j \frac{p_{1j}}{p_{0j}}}$

A Jevons-féle index mellett szól nemcsak az, hogy indexszámokat átlagolunk (bár itt ezek az indexek nem szorzódnak össze ténylegesen, nem láncindexekről van szó), hanem az is, hogy az ezzel számolt formula több indexpróbát is teljesít[3] (l. Hüttl-Vita [2005]). Ennek ellenére a gyakorlat inkább a számtani átlag formát használja, amiatt, hogy a mértani átlag hajlamos közelebb húzódni a szélsőségesen alacsony egyedi indexhez, ami a gyakorlatban is előfordulhat.

Az alábbi számsor egy termékcsoportra vonatkozó megfigyeléseket tartalmazza 2023 decemberében (előző év azonos időszakához képest).

1,11	1,18	1,25	1,04	1,06	1,19	1,01	1,05	1,18	*0,1*	1,13	1,08
1,00	1,25	1,1	1,3	1,04	1,21	1,08	1,14	1,05	1,17	1,1	

Látható, hogy 23 reprezentáns árváltozása áll rendelkezésre, a termékcsoport árváltozása ezen indexek súlyozatlan átlaga. Lehet számtani vagy mértani átlagot használni. A számtani átlag értéke 1,079, a mértani átlagé 1,009, azaz a termékcsoport átlagos árváltozásának mértéke 7,9 vagy 0,9%, elég különböző átlagos tendenciát mutatnak. A példa azt az esetet illusztrálja, amikor a termék ára a legtöbb reprezentáns árváltozása alapján nőtt, 0 és 30% között szóródnak az áremelkedés mértékét jelző értékek, illetve van egy „kilógó” reprezentáns árváltozás, 0,1-es értékkel, azaz 90%-os árcsökkenéssel.

[3] A Jevons-féle mértani átlag index teljesíti az összemérhetőségi (az index értéke független a mennyiségi adatok mértékegységétől), tranzitivitási ($I_{p(2/0)} = I_{p(2/1)} \cdot I_{p(1/0)}$), arányossági (az index az egyedi indexek átlaga) és időpróbát (az időszakok felcserélése reciprok indexet eredményez), míg a Carli-féle index nem tesz eleget a tranzitivitási és időpróbának.

Ezt a terméket kivezették a kínálatból, és drasztikus leárazás volt. Outlierként ezt az árváltozást egyáltalán nem indokolt kihagyni, hiszen az adott termékcsoport árváltozásának egy fontos jelzője, ugyanakkor az elvileg jobb tulajdonságú mértani átlag helyett a gyakorlat a számtani átlagot preferálja a mértani átlag említett alacsony értékhez húzása miatt.

- A legellentmondásosabb átlagforma a harmonikus átlag. Definíciószerűen a harmonikus átlag akkor értelmes, ha az adatok reciprokösszegének van tárgyi értelme. A harmonikus átlag reciprokát írva az adatok helyébe, így lesz változatlan az eredeti adatok reciprok összege. Egyes esetekben említik, hogy fordított intenzitási viszonyszámok esetén értelmes. Egyedi adatok esetén nehézkes jó példát adni ennek használatára, több tankönyv meg is jegyzi ezt. Ugyanakkor láttuk, hogy a harmonikus átlag használata mennyire gyakran és logikusan adódik, amikor részviszonyszámok (vagy részátlagok) számlálóval súlyozott harmonikus átlagformáját kell használnunk az összetett viszonyszám (főátlag) számításához. Ez a harmonikus átlag igazi terepe. Érdemes megnézni két példát, amit statisztika tankönyvek előszeretettel mutatnak a harmonikus átlag súlyozatlan számítására, hogy lássuk ennek a kérdéskörnek a problémáit.

Tisztán látszik a fenti probléma azon a gyakori példán, amikor átlagsebességet számolunk. Amennyiben súlyozatlan harmonikus átlagformára szeretnénk példát mutatni, egyenlő hosszúságú utakat tételezünk fel. Legyen egy egyszerű példa, egy kerékpáros felteker egy hegyre 4 km/h átlagsebességgel, majd lejön 16 km/h átlagsebességgel. Ebben az esetben az átlagsebesség a két szakaszon mért átlagsebesség súlyozatlan harmonikus átlagaként is számolható, de csak azért lehet súlyozatlan formát használni, mert a két megtett út (fel a hegyre, illetve le a hegyről) nyilvánvalóan azonos (S) hosszúságú.

$$\overline{X}_h = \frac{S+S}{\frac{S}{4}+\frac{S}{16}} = \frac{2}{\frac{1}{4}+\frac{1}{16}} = 6,4\,\mathrm{km/h}.$$

Ugyanez az eredmény adódik számtani átlag formában is, ebben az esetben azonban nem egyforma a súly, hiszen lefelé 4-szer olyan gyorsan megy a kerékpáros, mint felfelé, azaz negyedannyi idő alatt ér le, mint fel, tehát az összes kerékpározásra fordított idő 20-80%-os arányban oszlik meg.

$$\overline{X}_a = 0{,}2 \cdot 16 + 0{,}8 \cdot 4 = 6{,}4\,\mathrm{km/h}.$$

A számtani/harmonikus átlag és súlyozás nehézségeinek illusztrálására nézzünk egy másik típuspéldát, egységnyi pénz vásárlóértékét, vagy ennek reciprokaként adódó egységárak átlagolását. Tegyük fel, hogy egy piacon 3 standon lehet tojást venni, 100 Ft/db, 50 Ft/db és 33,33 Ft/db egységáron. Mennyi a tojás átlagára és 100 Ft átlagos vásárlóereje? Ebben az esetben kétféle gondolatmenet lehetséges.

Az egyik szerint az átlagár (100+50+33,33)/3 azaz 61,1 Ft/db, 100 Ft vásárlóereje az átlagár reciprokának 100-szorosa, azaz 1,64 db/100 Ft. Amennyiben ezt a számítási módot követjük, feltételezzük, hogy minden standon 1 (vagy azonos mennyiségű) tojást vásárlunk, ezért írható fel az átlagár súlyozatlan számtani átlag formában. Ebben az esetben azonban az egyes standokon nem ugyanakkora összeget költünk, hiszen rendre 100, 50 és 33,33 Ft-ba kerül 1 tojás. Azaz ha a pénz átlagos vásárlóerejét számolnánk, ami standonként 1 db/100 Ft, 2 db/100 Ft, 3 db/100 Ft, akkor súlyozott formát kell használni:

$$\frac{100 \cdot 1 + 50 \cdot 2 + 33,33 \cdot 3}{100 + 50 + 33,33} = 1,64\,\mathrm{db}/100\,\mathrm{Ft}$$

A másik szerint a vásárlóerő (1+2+3)/3=2 db/100 Ft, de ez feltételezi, hogy mind a 3 standon ugyanakkora összeget költünk. Ebben az esetben az átlagár 50 Ft/db. Az átlagár itt is az egyedi árak átlaga, de súlyozott átlaga, hiszen más a vásárolt mennyiség:

$$\frac{100 \cdot 1 + 50 \cdot 2 + 33,33 \cdot 3}{6} = 50\,\mathrm{Ft}/\mathrm{db}.$$

Mind a négy felsorolt esetben felírható a harmonikus átlagforma is, ebben az esetben nem a nevező, hanem a számláló megoszlással súlyozunk. Ahol a számtani átlag súlyozatlan volt, ott a harmonikus átlag súlyozott lesz, és fordítva. A négy harmonikus átlagforma rendre:

$$\frac{100 + 50 + 33{,}33}{\frac{100}{100} + \frac{50}{50} + \frac{33{,}33}{33{,}33}} = 61{,}1\,\mathrm{Ft}/\mathrm{db},$$

$$\frac{1 + 1 + 1}{\frac{1}{1} + \frac{1}{2} + \frac{1}{3}} = 1{,}64\,\mathrm{db}/100\,\mathrm{Ft},$$

$$\frac{6}{\frac{1}{1} + \frac{2}{2} + \frac{3}{3}} = 2\,\mathrm{db}/100\,\mathrm{Ft},$$

$$\frac{1 + 1 + 1}{\frac{1}{100} + \frac{1}{50} + \frac{1}{33{,}33}} = 50\,\mathrm{Ft}/\mathrm{db}.$$

Összefoglalva: a harmonikus átlag a gyakorlatban általánosan súlyozott formában jelenik meg, és tipikusan akkor használjuk, ha részviszonyszámok (vagy részátlagok) számláló megoszlással súlyozott harmonikus átlagformáját számoljuk. Olyan speciális esetben, amikor a számláló egyenletesen oszlik meg a részek között, ez felfogható súlyozatlan átlagnak.

Befejezésül megemlítjük, hogy ugyan definíciószerűen mind a négy átlagfajtát használhatjuk alapadatok átlagolására, a gyakorlatban ez a számtani és négyzetes átlagra jellemző, a mértani és harmonikus átlag esetében inkább származtatott mutatók átlagolásáról van szó. (Mértani átlag esetében leggyakrabban a láncviszonyszámok átlaga, harmonikus átlag esetében megoszlási/intenzitási viszonyszámok vagy részátlagok átlaga.) A harmonikus átlagnál éppen az okozza a megértésbeli nehézséget, ha egy viszonyszámot alapadatként fogunk fel.

4 A számtani átlag és gyakorlati variációi

A leggyakrabban tehát a számtani átlagot használjuk. Érdemes a számtani átlag tulajdonságait felsorolni:

- A számtani átlag (mint minden átlag) minden megfigyelési értéktől függ, egyértelműen meghatározható.
- Az átlag érzékeny a kilógó (outlier) értékekre, a mutató nem robusztus (nem úgy, mint a helyzeti középértékek).
- Amennyiben a megfigyelt értékeken ugyanazon lineáris transzformációt hajtjuk végre, a transzformált értékek átlaga is örökli ezt, azaz

$$\overline{A + B \cdot x_i} = A + B \cdot \overline{X}$$

- A számtani átlag 3 szempontból is közepes helyzetű, a) a legkisebb és legnagyobb megfigyelt érték közé esik, b) a tőle vett eltérések kiegyenlítik egymást, c) minimálja az elemek négyzetes eltérésösszegét. Formálisan a (7) képletek mutatják az összefüggéseket.

$$\begin{gathered} a)\ \ x_{\min} < \overline{X} < x_{\max} \qquad b)\ \ \sum \left(x_i - \overline{X}\right) = 0 \\ c)\ \ (x_i - A)^2 \to \min \quad \text{esetén} \quad A = X \end{gathered} \tag{7}$$

Többek között egy-két fenti tulajdonságra is reagálnak a számtani átlag egyes variációi, ezek közül négyet sorolunk itt fel, és közülük a 4. variációt újszerűsége miatt alaposabban is megnézzük.

1. Nyesett átlag ($\overline{X}_\beta$): a nyesett átlag éppen arra reagál, hogy az outlierek (vagy akár mérési/adatfelvételi hibák) befolyásolhatják az átlag értékét, elhúzhatják azt lefelé vagy felfelé. Ezért egy variáció, hogy elhagyjuk a legkisebb és legnagyobb β%-át a megfigyeléseknek, és ezek nélkül számítunk számtani átlagot.

   Ezen a ponton érdemes kis kitérőt tenni a robusztusság témakörében, mert az a középértékek gyakorlati használatát jelentősen befolyásolja. Egy mutatószám vagy módszer robusztusságán azt értjük, hogy a sokasági adatok nem jelentős változása nem, vagy csak kismértékben hat a mutatószámra vagy a módszer eredményére. A középértékeknél maradva, milyen mértékben változik a középérték, ha a sokasági adatok kevéssé változnak. Ennek leggyakoribb esete a középérték outlierre való reakciója. Minden hivatkozott tananyag említi, hogy a helyzeti középértékek robusztusak, az átlagok kevésbé, mert azok értékét az outlier(ek) jelentősebben befolyásolják. Az outlier jelenléte alapvetően az átlag közepes (centrális) helyzetét kérdőjelezi meg. A helyzeti középértékek esetében pl. a medián egyáltalán nem érzékeny arra, hogy a szélsőséges értékek mennyire szélsőségesek.

Gyakorlati tapasztalataim szerint nem az az igazi probléma, ha egy sokasági adattömegben vannak outlierek és azok hatnak az átlagra, mert az outlier megkeresésével (jellemzésével, hogy miért outlier) és elhagyása utáni átlagszámítással az eredmények hitelesek, az outlier jellemzése gazdagítja az elemzést. A nehezebb probléma mintavételi becsléseknél léphet fel. Ilyenkor (ha a minta nem arányosan rétegzett, vagy különböző okok miatt sérül a reprezentativitás, amit súlyozással hozunk helyre), ha egy kialakított rétegben van egy outlier, és azt súlyozzuk, az indokolatlanul felnagyíthatja az ismérv hatását az outliert tartalmazó rétegben. Éppen ezért a gyakorlatban a nemreprezentatív minták súlyozása esetében egy outlierkereső eljárás után az outliereknek tekintett megfigyelésekhez a súlyvektorban 1-es értéket rendelhetnek, és nem a rétegre számított súlyszámot.

2. Kronologikus átlag ($\overline{X}_k$): Ha egy időközönként megfigyelt állapotidősor átlagos szintje a kérdés, pl. készletek nagysága, valamilyen népességszám stb., akkor az átlagos szint jellemzése kronologikus átlaggal történik
$$\overline{X}_k = \frac{\frac{1}{2} \cdot x_0 + x_1 + \ldots + x_{n-1} + \frac{1}{2} \cdot x_n}{n}.$$
A kronologikus átlag tulajdonképpen az átmenet az álló jelenségről az időbelire (mozgóra).

3. Mozgóátlag ($\hat{X}_t$): használatára akkor kerül sor, ha egy idősorban van tendencia, és ezt a tendenciát (trendet) átlagolással mutatjuk be. Ilyenkor az átlag értékét az idősor közepéhez igazítjuk. Ha páros a tagszám, akkor formailag egy kronologikus átlagot számolunk, de középre (konkrét időszakhoz) igazítva. A mozgóátlag egyszerűbb (súlyozatlan) esetei mellett számos súlyozott formát is használnak.

4. Végül egy olyan súlyozott számtani átlagfajtát mutatunk, ami nem magától értetődő, de éppen gyakorlati szempontból figyelemreméltó. Induljunk ki a részviszonyszámok (vagy részátlagok) súlyozott számtani átlag formájából. Mint ismeretes (és szó volt róla itt az előbbiekben is), a $\overline{V} = \frac{\sum B_j \cdot V_j}{\sum B_j}$ összetett viszonyszám a részviszonyszámok nevezőben levő értékekkel súlyozott számtani átlaga. Felmerül, hogy számoljuk a részviszonyszámok számlálóval súlyozott számtani átlagát, azaz marad az átlagforma típusa, csak más a súly. A számolás módja tehát:
$$\overline{V}_{\text{é}} = \frac{\sum A_j \cdot V_j}{\sum A_j}.$$
A $\overline{V}_{\text{é}}$ jelölés az érzékelt átlag elnevezésre utal, az érzékelt jelző majd világos lesz az alkalmazási példa ismertetése után.

Ez első ránézésre szokatlan, annyira elterjedt, hogy vagy a nevezővel súlyozott számtani, vagy a számlálóval súlyozott harmonikus átlagot számoljuk.

Utánanéztünk, van-e előzménye ennek az átlagfajtának, de egyetlen cikket találtunk, ahol ez szerepel, Ertl (1983) leírását. Az alapötlet és a példa, amivel illusztráljuk a problémát, a cikkből származik, a szokásos számtani átlaggal való összehasonlítás és az eltérés okainak elemzése saját hozzájárulás a témához. Az érzékelt átlag (számlálóval való súlyozás) akkor merülhet fel, ha más az, ami egy jelenségből ténylegesen megvalósul, és amit a felhasználó ebből észlel/érzékel. (Az érzékelhető jelenség szerepelhet a számlálóban.) A legplasztikusabb példa erre a járatok utaskihasználtsága. De sok egyéb példa is hozható (népsűrűség, egy oktatóra jutó tanuló stb.). Ha vonatok, tömegközlekedési eszközök esetén számszerűsítjük a járatok kihasználtságát (a viszonyszám alapvetően a tényleges utasok száma/kapacitás), akkor „belefuthatunk" abba a jelenségbe, hogy átlagosan nem túl magas átlagos kihasználtság mellett egyes utasok egész mást érzékelhetnek. Emögött a kihasználtság szóródása áll, minél nagyobb a szóródás mértéke, annál inkább lesznek tömött járatok (ahol sokan érzékelnek zsúfoltságot), és kevesen laza üzemmenetet, hiszen olyankor kevesen ülnek a járművön. Ilyen esetekben pl. az érzékelt átlagnak lehet létjogosultsága. Szélsőséges esetben legyen egy oda-vissza járat 100 fős kapacitással, ahol odafelé tele van a járat, visszafelé senki nem utazik. Az átlagos kihasználtság a hagyományos súlyozással (100*1+100*0)/200=0,5, azaz 50%. Ugyanakkor ebben a konkrét szélsőséges esetben az érzékelt zsúfoltság 100%-os lesz, hiszen ezt odafelé mindenki így érzi, visszafelé pedig nincs érzékelő személy, mert senki nem utazik a járaton. Az érzékelt átlag ezt jelzi is, (100*1+0*0)/100=1, azaz 100%.

Vezessük le, mi az összefüggés a hagyományos átlag és az érzékelt átlag között. Az egyszerűség kedvéért feltételezzük, hogy a nevezőben levő értékek megegyeznek egymással (példánkban minden járat azonos kapacitású, legyen ez B). Az eredeti átlag a (8) alatt látható módon írható át (n a megfigyelések száma)

$$\overline{V} = \frac{\sum B \cdot V_j}{\sum B} = \frac{\sum A_j}{n \cdot B} = \frac{\overline{A}}{B}\,. \tag{8}$$

Azt kaptuk tehát logikus módon, hogy az eredeti átlag (példánkban az átlagos kapacitáskihasználtság) a megfigyelésenkénti átlagos (utas)szám és a konstans kapacitás hányadosa. Az érzékelt átlagot alakítsuk át olyan módon, hogy lássuk, hogyan függ az értéke a tényleges átlagtól. (S az A_j-k összege, azaz az értékösszeg, a relatív értékek négyzetösszege a Herfindahl-index (HI), ami a koncentráció fokát méri, V az A_j-k relatív szórása).

$$\overline{V}_{\text{é}} = \frac{\sum A_j^2/B}{\sum A_j} = \frac{S}{B}\sum Z_j^2 = \frac{n \cdot \overline{A}}{B} \cdot HI = \overline{V} \cdot n \cdot HI = \overline{V}(V^2+1) \tag{9}$$

A képlet mutatja, hogy egyrészt az érzékelt átlag nagyobb vagy egyenlő, mint a tényleges, és egyenlőség akkor és csak akkor áll fenn, ha az A_j adatok szórása 0, azaz minden adat megegyezik egymással (azaz minden járaton ugyanannyian utaznak), másrészt az is látszik, annál inkább meghaladja az érzékelt átlag a ténylegest, minél erősebb a relatív koncentráció mértéke.

Az érzékelt átlagra szerepeljen itt egy tényleges gyakorlati példa is. A magyar vasút egyik rövidebb szárnyvonalán véletlenszerűen kiválasztott októberi

2 nap (az egyik hétköznap, a másik ugyanazon a héten vasárnap) utasszámait, a járatok helykihasználását jellemezzük. Mind hétköznap, mind hétvégén 6 járat van egy zsáktelepülés és a 70 km-re levő megyeközpont között. Mint az alapadatok is mutatják, és ahogy ez várható is, hétköznap nagyobb az utasszám, és erősebb a napközbeni szezonalitás mértéke, mint hétvégén. Az alapadatokat és a számítások eredményeit a *2. táblázatban* mutatjuk.

Menetrend szerinti idő	Férőhely	Utasok száma		Kihasználtság (%)	
		hétköznap	hétvége	hétköznap	hétvége
Hajnali	155	25	8	16,1	5,2
Reggeli	155	144	12	92,9	7,7
Délelőtti	155	52	20	33,5	12,9
Délutáni	155	42	19	27,1	12,3
Kora esti	155	98	11	63,2	7,1
Késő esti	155	138	8	89,0	5,2
Összesen	930	499	78	53,7	8,4
Érzékelt átlag				70,4	9,5

2. táblázat. Utasszám és kihasználtság hétköznap és hétvégén

A hagyományos átlagos kihasználtság mutató (részviszonyszámok nevezővel súlyozott átlaga) mutatja, hogy a vasúti helykihasználás hétköznap átlagosan 53,7%-os volt, hétvégén sokkal alacsonyabb, 8,4%-os. Az érzékelt átlag (részviszonyszámok számlálóval súlyozott harmonikus átlaga) az utasok által érzékelt átlagos helykihasználtságot mutatja. Ez hétköznap 70,4%-os, hétvégén 9,5%-os volt. A hétköznap számolt érzékelt átlag azért haladja meg jelentősebben a hagyományos átlagmutatót, mert a szezonalitás miatt hétköznap nagyobb az adatok szóródása, relatív koncentrációja.

5 Példa a középértékek használatára

A Budapesten praktizáló főállású szakorvosokat (n=2114) jellemezzük havi bruttó keresetük (ezer forint) alapján[4]. A fő jellemzőket (alapstatisztikák) a *3. táblázatban* foglaltuk össze. Az összes orvos keresete eredetileg 561 ezer és 34146 ezer forint között vette fel az értékét, de van két kiugró (outlier) érték, egy 17 és egy 34 MFt körüli fizetés, az összes számolást a két kilógó érték nélkül is elvégeztük. Látható, hogy az outlierek a helyzeti középértékekre nem hatnak, se a medián, se a módusz[5] nem változik az elhagyásukkal. Az átlag némileg kisebb lesz, és minél magasabb átlagtól vett eltérés hatványon alapuló mutatóról van szó, annál érzékenyebben reagál a mutató az outlierre. A szoftverek által előszeretettel számolt alfa3 aszimmetria mutató outlierek esetén értelmetlen értéket ad. Az *1. ábrán* látható a hisztogram (outlierek nélkül), amely mutatja az empirikus kereslet eloszlást.

A három középérték egymáshoz képesti viszonya (és az egyéb mutatók, valamint az 1. ábra is) baloldali aszimmetriát mutat, ami a kereseteloszlások

[4] Az adatok forrása a KSH Egyéni bérek és keresetek felmérése, 2022 decemberi állapotot tükröznek.

[5] A móduszt az osztályközös gyakorisági sorból becsültük, ezt az 1. ábra adatai alapján akár rekonstruálni is lehet.

esetén általában jellemző, $Mo < Me < \overline{X}$. Az 1. ábra alapján az orvosok keresetek szerinti eloszlása lognormális eloszlást követ. Ezt alátámasztja a *2. ábra,* ami az adatok logaritmusa alapján készült hisztogramot, és az arra illesztett ugyanilyen átlagú és szórású normális eloszlásnak megfelelő gyakoriságokat mutatja.

Konkrét esetben (a két kilógó érték nélkül) az átlagkereset 1936 EFt, a szórás 651 EFt, a mediánkereset 1823 EFt. Érdemes a lognormális eloszlás néhány elméleti tulajdonságát összevetni az empíriával. A lognormális eloszlás sűrűségfüggvénye (m, σ) paraméterekkel:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - m)^2}{2\sigma^2}}, \qquad x > 0\,. \tag{10}$$

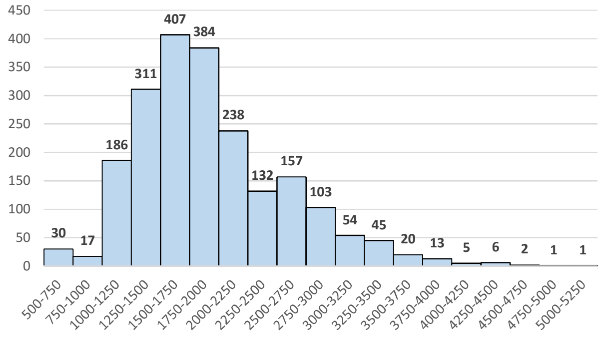
Az eloszlás különböző középértékei, szórása és a paraméterek között összefüggések írhatók fel, amelyek közül itt négyet emelünk ki. Jelöljük az eloszlás átlagát A-val, szórását B-vel. Ezeket a mintából becsült értékekkel adjuk meg a továbbiakban. Ismert összefüggések:

$$\begin{gathered} 1.\ m = \ln A - \frac{1}{2}\ln\Big(1 + \frac{B^2}{A^2}\Big) \qquad 2.\ \sigma^2 = \ln\Big(1 + \frac{B^2}{A^2}\Big) \\ 3.\ Mo = e^{m-\sigma^2} \qquad 4.\ Me = e^m\,. \end{gathered} \tag{11}$$

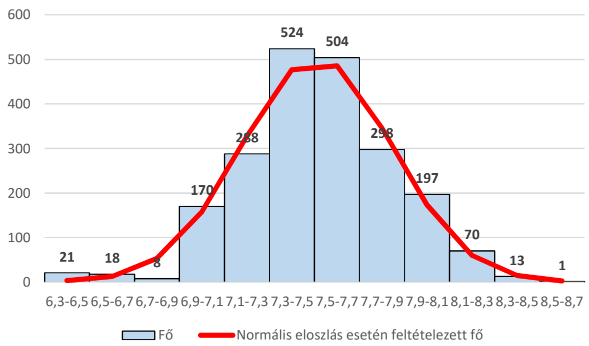
Ezek alapján az eloszlás két paraméterének becslése a momentumok alapján $m = 7{,}515$, $\sigma = 0{,}327$. A becsült paramétereket behelyettesítve 3. és 4. alatti összefüggésekbe $Mo = 1649$, $Me = 1835$, amely értékek egybecsengenek az empirikus módusszal és mediánnal, illetve azok alternatív becslései lehetnek. Még egy érdekesség említhető, bár itt a mértani közép empirikusan nem értelmes, de a lognormális eloszlás esetén a mértani közép megegyezik a mediánnal. (Ez abból, hogy a lognormális eloszlás esetén az adatok logaritmusa szimmetrikus eloszlást követ, azaz a számtani átlag és a medián ugyanaz, valamint az 1. táblázatban szereplő azon összefüggésből következik, hogy a mértani átlag logaritmusa megegyezik az adatok logaritmusainak számtani átlagával.)

Megnevezés	Jellemzők	
	összes orvos	2 outlier nélkül
n (elemszám, fő)	2114	2112
minimum, EFt	561	561
maximum, EFt	34146	5108
átlag, EFt	1957	1936
medián, EFt	1823	1823
módusz, EFt	1702	1702
alsó kvartilis, EFt	1490	1490
felső kvartilis, EFt	2272	2270
szórás, EFt	993	651
relatív szórás	0,51	0,34
aszimmetria F	0,15	0,15
aszimmetria alfa3	17,28	0,92

3. táblázat. Szakorvosok havi bruttó keresetének jellemzői Budapesten, 2022 decemberében



1. ábra. Budapesti szakorvosok havi kereset szerinti eloszlása (hisztogram)



2. ábra. Budapesti szakorvosok havi keresetének logaritmusa szerinti eloszlása

6 Összefoglalás

Cikkünkben a tankönyvekben használt középérték-fajtákat tekintettük át. Először röviden jellemeztük a helyzeti középértékeket, használatuk korlátait, előnyeit, a gyakorlatban felmerülő problémákat. Ezek után tértünk át az átlagok (számított középértékek) jellemzésére. Bemutattuk, hogy elméletileg többféle módon is lehet az átlagokat általánosítani, de a gyakorlatban a számtani átlag mellett a többi átlagot csak speciális esetekben használjuk. Különösen a harmonikus átlag használata esetén jelenik meg a gyakorlati alkalmazás szempontjának elsődlegessége az elméleti rendszerezéssel szemben. A leggyakrabban használt számtani átlag tulajdonságait részletesebben is szemügyre vettük. Bemutattunk egy ritkán használt, de gyakorlati szempontból hasznos átlagfajtát, az ún. érzékelt átlagot, valamint viszonyát a hagyományos számtani átlaghoz. A cikk utolsó részében egy példán illusztráltuk a középértékek számításának lehetőségeit, a középértékek egymáshoz való viszonyát, az esetlegesen felmerülő problémákat.

Irodalom

1. Carvalho, M. (2016). Mean, what do you Mean?, *The American Statistician,* 70(3), 764–776. DOI: 10.1080/00031305.2016.1148632

2. Dusek, T. (2022). A valószínűségszámítás ókori és középkori hiányának okai, *Statisztikai Szemle,* 100(4), 408–420. DOI: 10.20311/stat2022.4.hu0408
3. Ertl, I. (1983). Az intenzitási viszonyszámokkal mért átlagokról, *Statisztikai Szemle,* 61(3), 280–293.
4. Hunyadi, L. – Vita, L. (2003). *Statisztika közgazdászoknak,* KSH, Budapest.
5. Kendall, M. G. (1956). Studies in the history of probability and statistics: II. The beginnings of a probability calculus. *Biometrika,* 43(1-2), 1–14. DOI: 10.1093/biomet/43.1-2.1
6. Kerékgyártó, Gy. – L. Balogh, I. – Sugár, A. – Szarvas, B. (2009). *Statisztikai módszerek a gazdasági és társadalmi elemzésekben,* Aula kiadó, Budapest
7. Kovács, E. (2011). *Pénzügyi adatok statisztikai elemzése.*
8. Köves, P. (1957). A mértani átlag statisztikai alkalmazásai, *Statisztikai Szemle,* 35(4-5), 303–332.
9. Köves, P. (1961). Az átlagszámítás helye a statisztika elméletében, *Statisztikai Szemle,* 39(1), 3–30.
10. Köves, P. – Párniczky G. (1981). *Általános statisztika I.,* KJK Budapest.
11. Hüttl, A. – Vita, L. (2005). *Gazdaságstatisztika,* BCE Statisztika tanszék.
12. Sýkora, S. S. (2009). *Mathematical means and averages: basic properties,* Vol. 3. Stan's Library: Castano Primo.
13. Sándorné Kriszt É. – Csesznák A. – Ország G. (2013). *Statisztika I.,* Tankönyvkiadó, Budapest.

MEAN VALUES IN PRACTICE

Mean values are among the most frequently used statistical indicators/methods. The textbooks also discuss them in a fairly similar way and structure. In this article, we undertake to briefly summarize what we mean by mean value, within which we mainly deal with averages. We will review what types there are, and what are the properties of mean values/averages based on the literature (mainly economic Hungarian higher education curricula and Statistical Review articles). On the other hand, we expand the discourse on averages with some novel aspects, mainly in connection with the types of averages, their practical use, and the generalizations and variations of the most commonly used arithmetic average. First, we briefly describe the situational mean values, the limitations and advantages of their use, and the problems that arise in practice. After that, we move on to the characterization of the averages (calculated mean values). We show that, in theory, averages can be generalized in several ways, but in practice, in addition to the arithmetic average, other averages are only used in special cases. Especially in the case of using the harmonic mean, the priority of the aspect of practical application over theoretical systematization appears, and we give several examples of this. We take a closer look at the properties of the most commonly used arithmetic mean. We have presented a rarely used, but useful from a practical point of view, the so-called perceived average, as well as its relation to the traditional arithmetic average. In the last part of the article, we use an example to illustrate the possibilities of calculating the mean values, the relationship between the mean values, and the problems that may arise.