Balogh Norbert

Statisztika

Jegyzet

a közép- és emelt szintű matematika érettségi előkészítő segédanyaga

Gödöllői Török Ignác Gimnázium
Gödöllő, 2022

NAT 2020

Tartalomjegyzék

1. A statisztika alapfogalmai

Definíció: A statisztika emberek vagy tárgyak csoportjait vizsgálja. A csoport tagjai az *egyedek*, a csoport a *statisztikai sokaság*. Az egyedek száma a *statisztikai sokaság mértéke*. Az egyedek vizsgált tulajdonságai az *ismérvek*. Az ismérv egy konkrét előfordulása az *adat*.

Példa:

Egy 32 fős osztályban vizsgáljuk a testsúlyt.

- A statisztikai sokaság az osztály tanulói.
- Az egyedek az osztályban lévő tanulók.
- A statisztikai sokaság mértéke 32.
- A tanulók testsúlya az ismérv.
- Az adatok az egyes testsúlyértékek.

Definíció: Az ismérvek, illetve az adatok fajtái:

- A nem számmal kifejezhető ismérvek a *minősítéses ismérvek*. (Ha az adatok rendezhetőek, akkor *rendezhető minősítéses ismérvről* beszélünk.)
- A számmal kifejezhető ismérvek a *méréses ismérvek*. (Ha az adatok csak véges sok értéket vehetnek fel, akkor *diszkrét ismérvről*, ha bármilyen értéket felvehetnek, *folytonos ismérvről* beszélünk. Valamely dolgok számlálásakor diszkrét adatokat kapunk.)

Példa:

a) Minősítési ismérv például a tanulók neme, szemszín, Budapest kerületei (nem szám jellegű mennyiség), iskolai osztályzatok (nem szám jellegű mennyiség). Ezek közül az iskolai osztályzatok rendezhetőek.

b) Méréses ismérv például a napi tévénézési idő (24-nél kisebb nemnegatív szám), testvérek száma (természetes számok). Ezek közül diszkrét ismérv a testvérek száma, folytonos ismérv a napi tévénézési idő.

Definíció: Egy adat *gyakorisága* megmutatja, hogy az adat hányszor fordul elő a többi adat között. Ezt *gyakorisági táblázatban* (gyakorisági eloszlási táblázatban) szemléltethetjük.

Definíció: Valamely ismérv *relatív gyakorisága* a gyakoriság (k) és az adatok számának (n) hányadosa, azaz $\frac{k}{n}$.

Megjegyzés:

a) A relatív gyakoriság 0 és 1 közé eső szám, azaz $0 \leq \frac{k}{n} \leq 1$.

b) A relatív gyakoriságot százalékban is kifejezhetjük.

c) Ha táblázatban adjuk meg, *relatív gyakorisági táblázatról* (relatív gyakorisági eloszlási táblázatról) beszélünk.

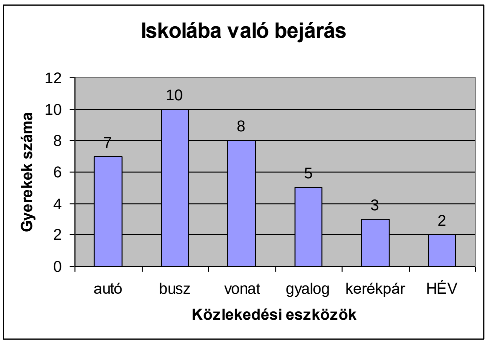
Példa:

Egy osztály iskolába járási adatainak (relatív) gyakorisági táblázata:

Közlekedési eszköz	*Gyakoriság*	*Ralatív gyakoriság*
autó	7	$\frac{7}{35} = 0{,}2$ (20%)
busz	10	$\frac{10}{35} \approx 0{,}286$ (28,6%)
vonat	8	$\frac{8}{35} \approx 0{,}229$ (22,9%)
gyalog	5	$\frac{5}{35} \approx 0{,}143$ (14,3%)
kerékpár	3	$\frac{3}{35} \approx 0{,}086$ (8,6%)
HÉV	2	$\frac{2}{35} \approx 0{,}057$ (5,7%)
Összesen:	**35**	**1 (100%)**

2. Adatok ábrázolása

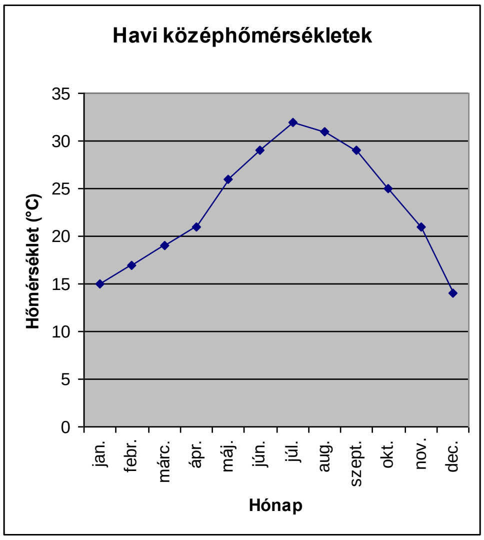
Oszlopdiagram:



Az oszlopdiagramot akkor használjuk, ha adatokat szeretnénk összehasonlítani, vagy adatok időbeli változását szeretnénk bemutatni. A számadatokat az oszlopok magassága jelzi. Az oszlopdiagram továbbfejlesztett változatai a *csoportosított oszlopdiagram* és a *halmozott oszlopdiagram*. Ezekben a kategóriákat összetevőire bontjuk és egymás mellé, illetve egymásra helyezzük az oszlopokat.

Az oszlopdiagram függőlegesen és vízszintesen is megrajzolható, ez utóbbi a *sávdiagram*. Sávdiagramot akkor célszerű használni, ha a tengelyek felirata hosszú, vagy ha a megjelenített értékek időtartamok.

Pont- és vonaldiagram:



A pont- és vonaldiagramot valamely mennyiség időbeli változásának szemléltetésére használjuk. A pontokat összekötő szakaszok a változás érzékeltetik. Ideálisan használható egyenlő intervallumok, például hónapok, negyedévek vagy gazdasági évek adataiban felfedezhető trendek ábrázolására.

Piktogram:

A piktogramok az ismérvekre emlékeztető rajzocskákból állnak. Számadatokat területtel vagy térfogattal szemléltetünk.

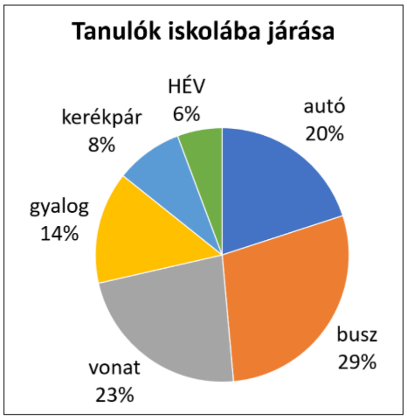
Megjegyzés: Ha adataink intervallumokba esnek, akkor speciális oszlopdiagrammal, *hisztogrammal* ábrázolhatjuk őket. Ha az intervallumok nem egyenlők, akkor az oszlopok területeit kell figyelembe venni és nem a magasságukat.

Kördiagram:

A kördiagramot akkor használjuk, ha az adatok százalékban vannak megadva. Az ábrázolt százalék-értékek a körcikk középponti szögével egyenesen arányosak.

Ne használjuk, ha túl sok adat van vagy az egyik adat nagyon nagy.

Nem szerencsés a kördiagramot térben elforgatni és koronggal ábrázolni (*tortadiagram*), mert ekkor a középponti szögek torzulnak.



3. Helyzetmértékek

Definíció: Az adatok értékösszegének és az esetszámnak (mérési adatok száma) a hányadosát *számtani középnek* (átlagnak) nevezzük. Jele: $\bar{x}$. Képlettel:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n},$$

ahol x_i az egyes adatokat, n pedig az adatok számát jelöli.

Megjegyzés: A számtani középnek az egyes értékektől való eltérésének az összege 0.

Definíció: Egy statisztikai adatsorozatban a leggyakrabban előforduló adatot *módusznak* nevezzük.

Definíció: Az adatok nagyság szerinti sorrendjében a középső adatot *mediánnak* nevezzük.

Megjegyzés: Páros elemszámnál a két középső adat számtani közepe adja a mediánt.

Példa:

Egy munkahelyen az alkalmazottak havi keresetének a megoszlását mutatja az alábbi táblázat. Állapítsa meg a bérek számtani közepét, móduszát és mediánját!

Munkabér	265 000 Ft	270 000 Ft	285 000 Ft	295 000 Ft	310 000 Ft
Fő	10	22	15	12	5

Megoldás:

a) Az adatok száma: $n = 64$, a keresett átlag:
$$\bar{x} = \frac{10\cdot 265\,000 + 22\cdot 270\,000 + 15\cdot 285\,000 + 12\cdot 295\,000 + 5\cdot 310\,000}{64} = \frac{17\,955\,000}{64} = 280\,546{,}875.$$

b) A leggyakrabban előforduló adat, azaz a módusz: 270 000.

c) A feladatban 64 elem szerepel, ezért a két középső elem (32. és 33.) számtani közepét véve a medián: 277 500.

4. Szóródási mértékek

A szóródás mérőszámai megmutatják az egyes érték-átlagtól való eltérést és azt, hogy mekkora az egymástól való eltérésük.

Definíció: A legnagyobb és a legkisebb adat mérőszámának a különbségét *terjedelemnek* nevezzük.

Definíció: Az átlagtól való eltérések abszolút értékének számtani közepét *átlagos abszolút eltérésnek* nevezzük. Jele: $\delta(\bar{x})$. Képlettel:

$$\delta(\bar{x}) = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + \cdots + |x_n - \bar{x}|}{n},$$

ahol x_i az egyes adatokat, n az adatok számát jelöli.

Definíció: Az átlagtól való eltérések négyzeteinek számtani közepét *négyzetes eltérésnek* (*szórásnégyzetnek*) nevezzük. Jele: $\sigma^2(\bar{x})$. A szórásnégyzet négyzetgyöke a *szórás*. Képlettel:

$$\sigma(\bar{x}) = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_i - \bar{x})^2}{n}}.$$

ahol x_i az egyes adatokat, n pedig az adatok számát jelöli.

Megjegyzés: Általában az átlagtól való eltérésekkel dolgozunk, de vizsgálhatjuk tetszőleges számtól is. Ilyenkor a fenti képletekbe az $\bar{x}$ helyébe az adott számot írjuk.

Példa:

100 fiatal testmagasságának az eloszlását mutatja a következő táblázat. Számítsa ki a testmagasságok terjedelmét, az átlagos eltérést, a négyzetes eltérést és a szórást! (Az egyes intervallumokon belül nem tudjuk az eloszlást, ezért az egyes intervallumok átlagmagasságával számoljunk, azaz 155,5; 165,5; 175,5; 185,5 és 195,5.)

Testmagasság	151-160	161-170	171-180	181-190	191-200
Fő	6	26	49	15	4

Megoldás:

a) A testmagasságok terjedelme: $195{,}5 - 155{,}5 = 40$.

b) Az adatok száma: $n = 100$, az átlagos testmagasság: $\bar{x} = 174$. Az átlagos abszolút eltérés:

$$\delta(174) = \frac{6\cdot|155{,}5-174|+26\cdot|165{,}5-174|+49\cdot|175{,}5-174|+15\cdot|185{,}5-174|+4\cdot|195{,}5-174|}{100} =$$
$$\frac{6\cdot 18{,}5+26\cdot 8{,}5+49\cdot 1{,}5+15\cdot 11{,}5+4\cdot 21{,}5}{100} = \frac{664}{100} = 6{,}64.$$

c) A testmagasságok szórása:

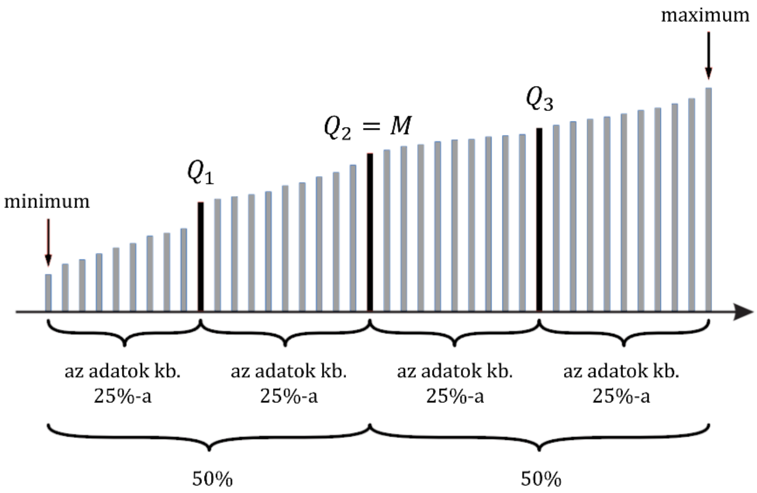
$$\sigma(174) = \sqrt{\frac{6\cdot(155{,}5-174)^2+26\cdot(165{,}5-174)^2+49\cdot(175{,}5-174)^2+15\cdot(185{,}5-174)^2+4\cdot(195{,}5-174)^2}{100}} =$$
$$\sqrt{\frac{7875}{100}} \approx 8{,}87.$$

5. Kvartilisek

A medián által megadott információt kiegészíti az ún. *alsó és felső kvartilis* értékének megadása. Ezek a mediánhoz hasonló középértékek, de nem az adathalmaz felező értéke, hanem az alsó kvartilis a negyedelő, a felső kvartilis pedig a háromnegyedelő érték (vagy alsó negyedelő, felső negyedelő).

<u>Definíció</u>: Az *alsó kvartilis* az a szám, amelynél az adatoknak (közelítőleg) a negyede kisebb vagy egyenlő. Jele: Q_1. A *felső kvartilis* az a szám, amelynél az adatoknak (közelítőleg) a negyede nagyobb vagy egyenlő. Jele: Q_3.

Megjegyzés: A negyedelés határait jelentő 5 értéket szokás Q_0, Q_1, Q_2, Q_3, Q_4-gyel jelölni. Ezzel a jelöléssel Q_0 az adathalmaz minimuma, Q_1 az alsó kvartilis, Q_2 a középső kvartilis, azaz a medián, Q_3 a felső kvartilis, Q_4 az adathalmaz maximuma.



Az alsó és felső kvartilisek értékének meghatározása: Rendezzük nagyság szerint növekvő sorrendbe az adatokat, majd a mediánnal osszuk ketté az adatsokaságot (páratlan elemszám esetén a mediánt elhagyjuk). Majd külön-külön vesszük a kapott két adatsokaság mediánját.

<u>*Példa*</u>:

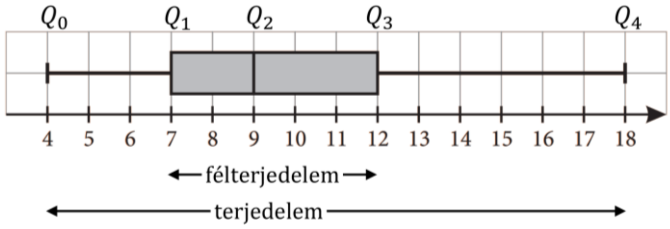
	adatok	alsó kvartilis	medián	felső kvartilis
8 adat	1 2 ∣ 3 3 ∣ 4 5 ∣ 6 6	2,5	3,5	5,5
9 adat	1 2 ∣ 3 3 ~~4~~ 5 5 ∣ 6 6	2,5	4	5,5
10 adat	1 2 ~~3~~ 3 4 ∣ 5 5 ~~6~~ 6 7	3	4,5	6
11 adat	1 2 ~~3~~ 3 4 ~~5~~ 5 6 ~~6~~ 7 8	3	5	6

<u>Definíció</u>: A felső kvartilis és az alsó kvartilis különbségét *félterjedelemnek* (*interkvartilis terjedelemnek*) nevezzük.

Megjegyzés: Az elemek (közelítőleg) 50%-a a félterjedelem sávba esik, így a félterjedelem megmutatja, hogy az adatok középső fele milyen széles skálán helyezkedik el. Ez általában nem a terjedelem fele.

6. Dobozdiagram

A terjedelem, a félterjedelem, a medián, a legkisebb és a legnagyobb érték ábrázolására szolgáló grafikus eszköz a *dobozdiagram* (más néven *boksz-plot diagram* vagy *sodrófadiagram*). A félterjedelmet egy dobozzal (téglalappal) szemlélteti, ebben van behúzva a medián, míg a maximum és minimum értékeket egy-egy talppal ábrázoljuk. A doboz elhelyezkedése a teljes talphoz viszonyítva, illetve a medián helyzete a dobozon belül információt ad az eloszlásról.

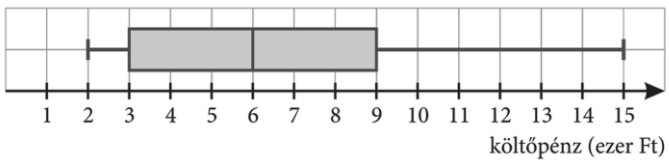


Megjegyzések:

- A dobozdiagram lehet akár vízszintes, akár függőleges.
- A dobozdiagramon néha feltüntetjük az adatok átlagát a dobozban egy ponttal.
- Az adatokból egyértelműen elkészíthető a dobozdiagram, visszafelé azonban ez nem teljesül. A dobozdiagram alapján nem lehet felsorolni az adatokat, csak az adatok „összességéről” nyerünk információt.
- Szokás még *kiugró adatokról* beszélni. Kiugró adat az, amely az alsó (felső) kvatilisnél legalább a félterjedelem másfélszeresével kisebb (nagyobb). A kiugró adatot a dobozdiagram esetén sokszor külön ponttal jelölik. Ha van kiugró adat, akkor a dobozból kiinduló vonalak nem a minimumig, illetve a maximumig tartanak, hanem a legkisebb, illetve legnagyobb olyan adatig, ami még nem számít kiugró adatnak.

Példa:

A diákok felmérést készítettek, hogy ki hány forint költőpénzt vitt magával egy kirándulásra. A kapott eredményeket az alábbi box-plot diagram szemlélteti.



a) A diagram alapján:

A minimum: $Q_0 = 2000$, alsó kvartilis: $Q_1 = 3000$, medián: $Q_2 = 6000$, felső kvartilis: $Q_3 = 9000$, maximum: $Q_4 = 15\,000$. Terjedelem: 13 000, félterjedelem: 6000.

b) Eldönthető-e a diagram alapján a következő állításokról, hogy igazak vagy hamisak?

- Átlagosan 6000 Ft volt a gyerekeknél. (*nem eldönthető*)
- A gyerekek kb. 75%-ánál volt legalább 9000 és legfeljebb 15 000 Ft. (*hamis*)
- Volt olyan gyerek a megkérdezettek között, akinél 9000 Ft volt. (*nem eldönthető*)
- A megkérdezett gyerekek kb. felénél volt legalább 3000 és legfeljebb 9000 Ft közötti összeg. (*igaz*)

Felhasznált irodalom, források

- HAJNAL IMRE, SZÁMADÓ LÁSZLÓ, BÉKÉSSY SZILVIA (2004): *Matematika a gimnáziumok 9. évfolyama számára*
  Nemzeti Tankönyvkiadó, Budapest

- INGER JÁNOS (2013): *Tehetséggondozás a matematikában: Statisztika*
  https://users.itk.ppke.hu/itk_dekani/files/matematika/pdfs/22.pdf
  Pázmány Péter Katolikus Egyetem Információs Technológiai és Bionikai Kar, Budapest (Utolsó letöltés ideje: 2020. július. 25.)

- JUHÁSZ ISTVÁN, OROSZ GYULA (2020): *Matematika 9. tankönyv*
  https://www.tankonyvkatalogus.hu/pdf/OH-MAT09TB__teljes.pdf
  Oktatási Hivatal, Budapest (Utolsó letöltés ideje: 2020. augusztus 24.)

- JUHÁSZ ISTVÁN, OROSZ GYULA (2021): *Matematika 10. tankönyv*
  https://www.tankonyvkatalogus.hu/pdf/OH-MAT10TB__teljes.pdf
  Oktatási Hivatal, Budapest (Utolsó letöltés ideje: 2021. július 12.)

- JUHÁSZ ISTVÁN, OROSZ GYULA (2022): *Matematika 11. tankönyv*
  https://www.tankonyvkatalogus.hu/pdf/OH-MAT11TB__teljes.pdf
  Oktatási Hivatal, Budapest (Utolsó letöltés ideje: 2022. július 4.)

- KOSZTOLÁNYI JÓZSEF, KOVÁCS ISTVÁN, PINTÉR KLÁRA, URBÁN JÁNOS, VINCZE ISTVÁN (2011): *Matematika tankönyv 9.*
  Mozaik Kiadó, Szeged

- MAHLER ATTILA, OROSZ GYULA (2022): *Gyűjtemény a Matematika emelt szintű oktatásához 11–12.*
  https://www.tankonyvkatalogus.hu/pdf/OH-MAT1112BE__teljes.pdf
  Oktatási Hivatal, Budapest (Utolsó letöltés ideje: 2022. július 4.)

- OKTATÁSI HIVATAL (2022): *Központi írásbeli feladatsorok*
  https://www.oktatas.hu/kozneveles/erettsegi/feladatsorok
  (Utolsó letöltés ideje: 2022. július 4.)