Matematika „A” 12. évfolyam

6. modul
Statisztika és valószínűség

Készítette: Lövey Éva

A modul célja	A valószínűség fogalmának elmélyítése. A valószínűségszámítás eszköztárát felhasználva az élet valós helyzeteinek elemzése. Az eseményalgebra és a kombinatorika mint segédeszközök felelevenítése.
Időkeret	4 óra
Ajánlott korosztály	12. évfolyam
Modulkapcsolódási pontok	Tágabb környezetben: Hétköznapi szituációk. Humán, reál tudományterületeken az esélyek felmérése. Döntések előkészítése a valós életben. Szűkebb környezetben: Statisztikai adatsokaságok elemzése. Szöveges feladatok értelmezése. Grafikonok elemzése. Ajánlott megelőző tevékenységek: Előző években tanultak: permutáció, ismétléses permutáció, kombináció (ismétlés nélkül), variáció ismétléssel és ismétlés nélkül kis elemszámok esetén. Valószínűség fogalma, kombinatorikus valószínűség kis elemszámok esetén.
A képességfejlesztés fókuszai	Számolás, számlálás, számítás: Számolás nagyon nagy és nagyon kis abszolútértékű számokkal, normálalakkal. Szöveges feladatok, metakogníció: A valóságból merített szöveges feladatok alapján felismerni az alkalmazandó eljárást, képletet. A megkapott végeredmény értelmezése. Szövegben előforduló tartalmi összefüggések megkeresése. Rendszerezés, kombinatív gondolkodás: Adatok kiolvasása és elemzése táblázatokból, illetve valós életből merített szövegekből.

A TANANYAG JAVASOLT ÓRABEOSZTÁSA:

1. óra: Adatsokaságok
2. óra: Középértékek
3. óra: Középértékek alkalmazhatósága
4. óra Nagy elemszámú adatsokaságok jellemzése

ÉRETTSÉGI KÖVETELMÉNYEK:

	Középszint	**Emelt szint**
Valószínűség	Véges sok kimenetel esetén szimmetriamegfontolásokkal számítható valószínűségek (egyenlő esélyű elemi eseményekből) egyszerű feladatokban. Esemény, eseménytér konkrét példák esetén. A klasszikus (Laplace)-modell ismerete. Szemléletes kapcsolat a relatív gyakoriság és a valószínűség között. Valószínűségek kiszámítása visszatevéses mintavétel esetén, binomiális eloszlás.	Ismerje és alkalmazza a következő fogalmakat: események egyesítésének, metszetének és komplementerének valószínűsége, feltételes valószínűség, függetlenség, függőség. A nagy számok törvényének szemléletes tartalma (nagyobb n-ekre valószínűbb, hogy $\left\|\frac{k}{n}-p\right\|<\delta$). Geometriai valószínűség. A binomiális eloszlás (visszatevéses modell) és a hipergeometrikus eloszlás (visszatevés nélküli modell) tulajdonságai és ábrázolása. Várható érték, szórás fogalma és kiszámítása a diszkrét egyenletes és binomiális eloszlás esetén. A binomiális eloszlás alkalmazása. A minta relatív gyakoriságának becslése a sokaság paraméterének ismeretében.
Statisztika	Statisztikai adatok és ábrázolásuk (kördiagram, oszlopdiagram, gyakorisági diagram stb.), számtani közép, medián, módusz; adatok szóródásának mérése. Relatív gyakoriság. Szórás kiszámolása adott adathalmaz esetén számológéppel. Tudjon adathalmazokat összehasonlítani a tanult statisztikai mutatók segítségével.	Tudjon hisztogramot készíteni, és adott hisztogramról információt kiolvasni. Ismerje az adathalmazok egyesítése és átlaguk közötti kapcsolatot.

MODULVÁZLAT

	Lépések, tevékenységek	Kiemelt készségek, képességek	Eszköz/ Feladat/ Gyűjtemény
I. Adatsokaságok			
1.	Nagy adathalmaz elemzése	Rendszerezés, kombinatív gondolkodás, szabatos fogalmazás	1. mintapélda
2.	Gyakorisági táblázat készítése	Rendszerezés, kombinatív gondolkodás	
3.	Sávszélesség	Rendszerezés, kombinatív gondolkodás	1. és 2. mintapélda
4.	Oszlopdiagram a gyakoriságról		1. mintapélda
5.	Módusz, átlag	Szövegértés, rendszerezés, kombinatív gondolkodás, adatok képletbe rendezése	1. mintapélda
6.	Relatív gyakoriság, kördiagram, sávdiagram		2. mintapélda, 1. és 3. feladat

II. Középértékek			
1.	Középértékek és összehasonlításuk	Rendszerezés, kombinatív gondolkodás, szabatos fogalmazás	3. mintapélda (a–c)
2.	Szórás	Rendszerzés, logikus gondolkodás	3. mintapélda (d), 4. mintapélda, 4. és 5. feladat

III. Középértékek alkalmazhatósága			
1.	Középértékek összehasonlítása	Kombinatív gondolkodás, adatok képletbe rendezése	5. mintapélda, 6. feladat
2.	Számtani közép	Számlálás, logikus gondolkodás	6. mintapélda

IV. Nagy elemszámú adatsokaságok jellemzése			
1.	Adatok típusai, minősítése és méréses ismérvek	Szövegértés, kombinatív gondolkodás, adatok képletbe rendezése	7. mintapélda, 8. feladat
2.	Nagy adathalmaz feldolgozása, kvartilis, percentilis	Szövegértés, kombinatív gondolkodás, adatok képletbe rendezése, becslés	8–10. mintapélda, 9–10. feladat

I. Adatsokaságok

Ennek a fejezetnek - érettségi előtt pár hónappal - az a célja, hogy átismételjük a statisztikai alapfogalmakat és azok kapcsolatát a valószínűségszámítással. Fontos cél az is, hogy az érettségin előforduló feladatokat meg tudják oldani a tanulók, és középiskolai tanulmányaik végéhez közeledve felvértezzük őket olyan tudással, ami a mindennapi életben szükséges.

A fogalmak áttekintése a 10. évfolyam 2. moduljának kislexikonában található meg.

Mintapélda$_1$

Egy háziorvos feljegyezte, 150 betege 1 év alatt hányszor kereste fel a rendelőjében:

3	2	6	2	6	5	22	3	1	10	5	9	7	2	5	1	5	4	9	7	25	19	8	2	5
8	10	16	15	5	7	8	3	6	6	21	6	9	4	5	6	6	22	8	11	23	8	5	9	6
8	7	15	10	16	11	13	1	7	3	2	18	0	16	4	9	8	5	9	17	7	9	5	19	12
1	10	3	5	7	13	18	8	7	8	7	7	13	0	5	14	7	20	1	9	4	6	24	9	6
11	5	6	28	7	7	22	1	17	4	11	8	1	4	12	13	9	23	14	5	2	6	6	11	3
14	6	8	4	4	6	8	29	18	5	8	8	17	4	4	5	18	7	3	11	23	20	10	6	6

a) A doktort legtöbbször, illetve legkevesebbszer felkereső betege között hány alkalom eltérés volt?

b) Van-e a feljegyzett adatok között leggyakrabban előforduló esetszám, azaz mondhatja-e a doktor: „A legtöbb betegem … alkalommal keresett fel.”?

c) Átlagosan hány alkalommal keresték fel betegei az orvost?

Megoldás:

a) Az orvost legkevesebbszer felkereső beteg nem is járt ebben az évben az orvosnál, azaz 0 alkalommal kereste fel, az orvost legtöbbször felkereső beteg 29-szer járt ott.

A legkisebb és legnagyobb érték közti különbség tehát $29 - 0 = 29$. Ezt az értéket az **adatsokaság terjedelmének** nevezzük.

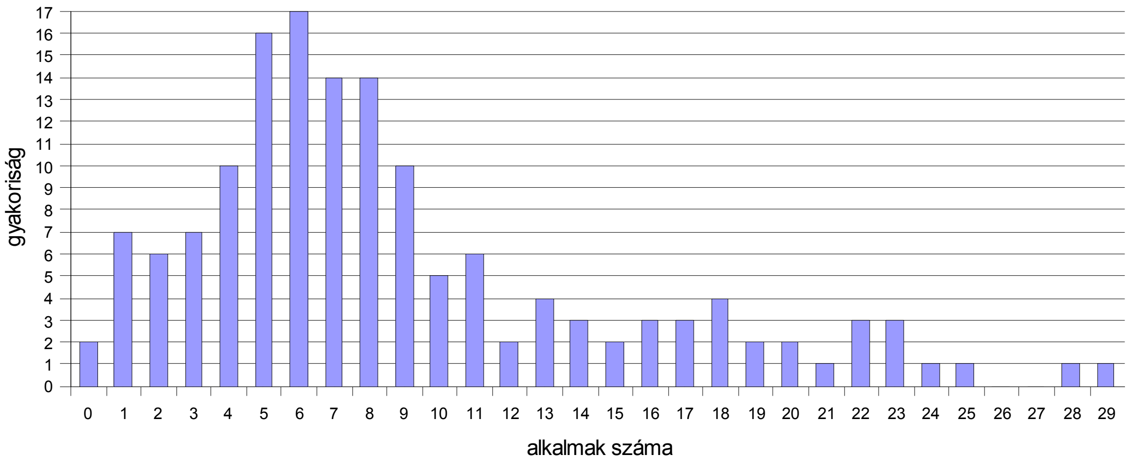
b) Azt az adatot keressük, amely a legtöbbször fordul elő az adatsokaságban, azaz az adatsokaság móduszát.

A fenti táblázat nehezen áttekinthető a leggyakrabban előforduló adat megtalálásához. Készítsük **gyakorisági táblázatot**, amely megmutatja, az egyes számok milyen gyakorisággal fordulnak elő a táblázatunkban:

Ha eddig még nem tettük volna, érdemes megmutatni a tanulóknak, hogy a megfelelő adatok mellé húzott „strigulákkal” hogy lehet aránylag gyorsan elkészíteni egy ilyen gyakorisági táblázatot.

alkalom	gyakoriság	alkalom	gyakoriság	alkalom	gyakoriság
0	2	10	5	20	2
1	7	11	6	21	1
2	6	12	2	22	3
3	7	13	4	23	3
4	10	14	3	24	1
5	16	15	2	25	1
6	17	16	3	26	0
7	14	17	3	27	0
8	14	18	4	28	1
9	10	19	2	29	1

Még látványosabb, ha **oszlopdiagram**on ábrázoljuk az alkalmak számának gyakoriságát:



A grafikonon és a táblázatból is láthatjuk, hogy a legtöbb beteg 6 alkalommal kereste fel az orvost, tehát **az adatsokaság módusza 6.**

c) **Egy adatsokaság átlaga az adatok számtani közepe:**

$$\overline{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}.$$

Ilyen sok adatból úgy célszerű számtani közepet számolni, hogy a gyakorisági táblázatot használjuk: $\overline{x} = \dfrac{2\cdot 0 + 7\cdot 1 + 6\cdot 2 + 7\cdot 3 + \ldots + 1\cdot 28 + 1\cdot 29}{150} \approx 8{,}91.$

Ha valóban végig akarjuk számoltatni ezt a tanulókkal, tapasztalhatjuk, hogy a sorkijelzős számológép ennyi adattal nem bír. Ha a szorzásokat fejben el is végzik, az utolsó pár adat beírásakor a gépek többsége a gép memóriájának telítettségét jelzi. Érdemes a részösszegeket kiszámítani – mondjuk oszloponként – majd az összegüket osztani 150-nel.

Ennyi adattal a számológép statisztikus módja sem bír, az átlagszámításkor tehát nem használható.

Mintapélda$_2$

A fenntartó önkormányzat jelentést kért az előző példabeli orvostól, hogy milyen gyakorisággal látogatják a betegei.

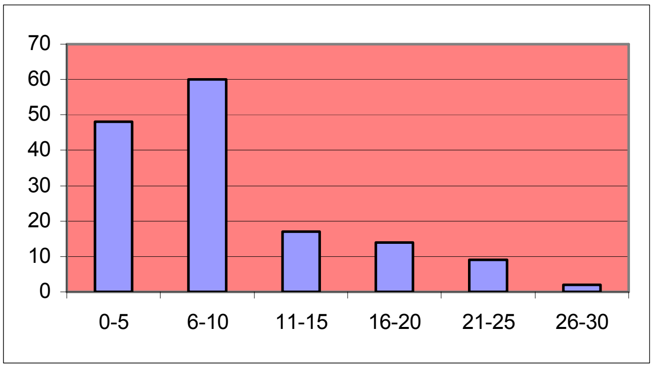
a) Soroljuk egyenlő széles osztályokba az adatokat, majd ábrázoljuk az egyes osztályok gyakoriságát!

b) Számítsuk ki annak valószínűségét, hogy Mari néni – aki az orvos régi betege – 25 alkalomnál többször kereste fel a doktort az év folyamán!

Megoldás:

a) Ha már elkészítettük a gyakorisági táblázatot az előző feladathoz, könnyű dolgunk van. Ha mégsem, írjuk fel az osztályokat, majd mellettük húzzunk strigulákat, így aránylag gyorsan el tudjuk készíteni a gyakorisági táblázatot ilyen sok adatból is.

osztály	gyakoriság
0–5	48
6–10	60
11–15	17
16–20	14
21–25	9
26–30	2



b) 150 beteg közül 2 olyan van, aki 25-nél több alkalommal látogatta. (Ez a 26–30 osztály gyakorisága.) Annak valószínűsége, hogy Mari néni közéjük tartozik:

$$P = \frac{2}{150} \approx 1{,}33\% .$$

Észrevehetjük, hogy a számított valószínűség megegyezik az osztály **relatív gyakoriságával**.

Mintapélda$_3$

Az alábbi táblázat 250 fiú testtömegét mutatja 100 gramm pontosságra kerekítve.

a) Ábrázoljuk a kumulatív (latin: összegyűjtött, felhalmozott) gyakoriságot vonaldiagramon!

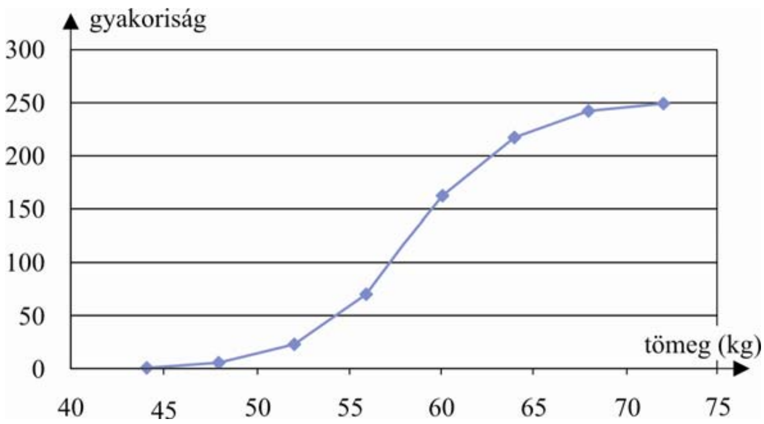
b) Számítsuk ki, mi az átlaga a 250 fiú tömegének!

Tömeg (kg)	Fiúk száma
44,0–47,9	3
48,0–51,9	17
52,0–55,9	50
56,0–57,9	45
58,0–59,9	46
60,0–63,9	57
64,0–67,9	23
68,0–71,9	9

Megoldás:

a) A gyakoriságok fokozatos összegzésével nyert gyakoriság a kumulatív gyakoriság. Készítsük el először azt a táblázatot, mely segítségével a grafikont meg tudjuk rajzolni! Az egyes sorokba most azoknak az elemeknek a gyakorisága kerül, melyek nem nagyobbak, mint az előző osztályok valódi felső határa. Például az 58,0 – 59,9 osztályba akkor tartozik egy elem, ha az kisebb, mint 59,95, hiszen e fölött kerekítéssel már a következő osztályba tartozna.

Tömeg (kg)	Fiúk száma
$m<47{,}95$	3
$m<51{,}95$	20
$m<55{,}95$	70
$m<57{,}95$	115
$m<59{,}95$	161
$m<63{,}95$	218
$m<67{,}95$	241
$m<71{,}95$	250



b) Az átlag kiszámításához szükség lenne a fiúk össztömegére. Ezt viszont pontosan nem tudjuk megadni, hiszen pontosan egyetlen fiú tömegét sem ismerjük. Ilyen esetekben az osztályhatárok számtani közepével, azaz az **osztályközéppel** számolhatunk. Hogy mely osztályhatárok számtani közepét vesszük, az teljesen mindegy, hiszen az osztályhatárok (pl. 44 és 47,9) és a valódi osztályhatárok (jelen esetben 43,95 és 47,95) számtani közepe azonos: $\frac{44+47{,}9}{2}=45{,}95=\frac{43{,}95+47{,}95}{2}$.

A fiúk össztömege tehát:

$$\frac{44+47{,}9}{2}\cdot 3+\frac{48+51{,}9}{2}\cdot 17+\frac{52+55{,}9}{2}\cdot 50+\frac{56+57{,}9}{2}\cdot 45+\frac{58+59{,}9}{2}\cdot 46+$$
$$+\frac{60+63{,}9}{2}\cdot 57+\frac{64+67{,}9}{2}\cdot 23+\frac{68+71{,}9}{2}\cdot 9=14\,636{,}5\ \text{kg}.$$

Az átlagos testtömeg pedig $\frac{14636{,}5}{250}=58{,}546\ \text{kg}.$

Feladatok

1. Egy parkolóőr megszámolta, hányan ültek azokban az autókban, melyek a parkolójába behajtottak reggel 8 és 10 között.

Emberek száma	1	2	3	4	5
Kocsik száma	41	33	18	6	2

a) Átlagosan hányan ültek egy-egy személyautóban?

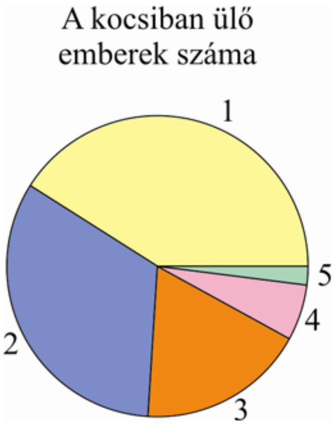
b) Szemléltesd kördiagramon az egy autóban ülők számának megoszlását!

c) Számítsd ki annak valószínűségét, hogy az elsőnek érkező kocsiban legalább négyen utaztak!

Megoldás:

a) $\bar{x}=1{,}95$

Egy autóban ülő emberek száma	1	2	3	4	5
Autók száma	41	33	18	6	2
Relatív gyakoriság	0,41	0,33	0,18	0,06	0,02
Ennek megfelelő középponti szög	147,6°	118,8°	64,8°	21,6°	7,2°



c) Mivel a 4 vagy 5 utassal érkező autók relatív gyakorisága 0,06 + 0,02, így a keresett valószínűség 0,08 körüli.

2. Az egyik, új lakásokat értékesítő vállalat úgy hirdette meg lakásait, hogy futni lehetett a kedvezményért. Ahány másodperccel korábban

odafutottak egy adott helyről az épülő házhoz, mint a meghirdetett szintidő, annyi négyzetméter burkolatot kaptak ingyen. Mielőtt a szintidőt megállapították volna, megkértek 50 embert, hogy fussa le a távot. Úgy akarták megadni a szintidőt, hogy az emberek 50%-ának sikerélménye legyen, és legalább 1 négyzetméternyi kedvezményt kaphasson. A próbaként futók a következő eredményeket érték el másodpercekben.

61	83	62	96	66	61	92	87	69	91	82	62	80	86	97	72	78
68	88	63	85	63	73	99	66	82	61	86	63	73	65	89	95	61
72	89	92	76	75	77	75	88	87	91	84	73	67	76	82	79	

a) Mekkora az adatsokaság terjedelme? (Azaz hány másodperc az időkülönbség a leggyorsabb és leglassúbb között?)

b) Milyen szintidőt javasolnál a fenti próbafutás után, ha te lennél a vállalat értékesítési vezetője?

Megoldás:

Az adatokat sorba rendezve:

61	61	61	61	62	62	63	63	63	65	66	66	67	68	69	72	72
73	73	73	75	75	76	76	77	78	79	80	82	82	82	83	84	85
86	86	87	87	88	88	89	89	91	91	92	92	95	96	97	99	

a) A terjedelem 99 – 61 = 38 (s).

b) Az adatsokaság mediánját keressük, ami $\frac{77+78}{2}=77{,}5$. Tehát 77, vagy 78 másodperces szintidőt érdemes megadni.

3. Angliában a Cornwall-félszigeten egy régi agyagbánya területén hatalmas félgömbök hívják fel magukra az arra járók figyelmét. Az Éden-terv üvegházai, a XXI. század botanikus kertje a brit kormány és az EU segítségével jött létre. A projekt 140 millió euróba került, és a tervek szerint évente 600 ezer látogatót fogad majd. A "buborékok" körül kertek borítják az öreg felszíni bányát. Bent az üvegházakban trópusi növényekben gyönyörködhet a látogató. A kert közepén nagy táblán a következő információsor olvasható:

„Ha a föld összezsugorodna egy 100 lakosú falu méretére, 57 ázsiai, 21 európai, 14 amerikai és 8 afrikai élne rajta. A százból 70 fehér és 30 nem fehér lenne. 89 lenne heteroszexuális és 11 homoszexuális. Hat tulajdonosé lenne a világ gazdaságának 59 százaléka, és mind a

6 az USA-ból származna. Nyolcvanan élnének rossz lakáskörülmények között, hetvenen nem tudnának olvasni, és ötvenen lennének alultápláltak. Egynek lenne felsőfokú képzettsége és ugyancsak egynek számítógépe."

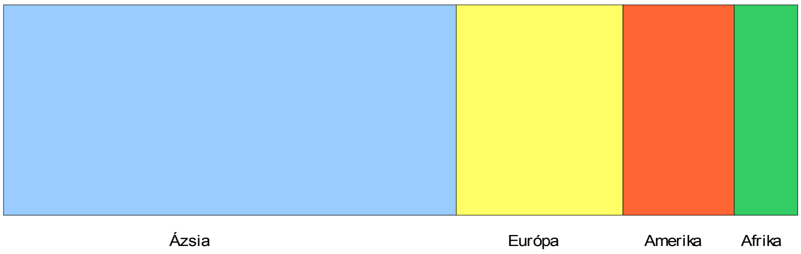
a) Ábrázold sávdiagramon a föld népességének eloszlását kontinensek szerint!

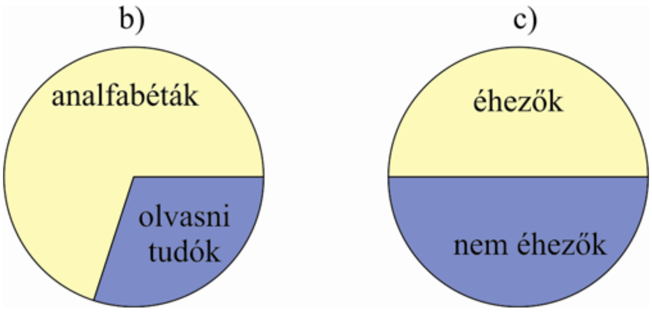
b) Ábrázold kördiagramon az olvasni tudók, illetve az analfabéták arányát!

c) Ábrázold kördiagramon az éhezők és a nem éhezők arányát!

Megoldás:

a)





II. Középértékek

Mintapélda$_4$

Az alábbi táblázat egy évfolyam fiainak testmagasságát mutatja egy középiskolában.

a) Készítsük el az egyes testmagasságokhoz tartozó gyakorisági táblázatot!

Valaki – a táblázat ismeretében – ki akarja találni, hogy a névsorban a legelső fiú milyen magas. Melyik értékre adja a voksát?

186	179	171	167	167	184	174	176	182	179	186	164	178	177	165
184	189	175	167	190	177	175	176	176	175	183	179	186	188	166
174	169	175	188	185	172	169	184	173	169	185	171	166	170	181
180	170	168	178	171	167	189	178	178	165	177	170	186	180	179
175	168	187	176	173	168	170	172	177	184	175	169	167	179	181
180	176	172	175	176	170	171	175	180	173	182	177	165	175	188
177	166	171	170	180	168	167	179	176	184	182	172			

Megoldás:

Készítsük el a különböző magasságértékek gyakorisági táblázatát. A legnagyobb gyakorisággal rendelkező magasságértékre érdemes szavazni.

164	165	166	167	168	169	170	171	172
1	3	3	9	4	4	6	5	4
173	174	175	176	177	178	179	180	181
3	2	8	7	6	4	6	5	2
182	183	184	185	186	187	188	189	190
3	1	4	2	3	1	3	2	1

A legnagyobb gyakoriságú magasságérték a 167 cm, erre érdemes tippelni. Ennek valószínűsége $\frac{9}{102} \approx 0{,}088$.

b) Ezen az évfolyamon a fiúk elhatározzák, hogy a tanév végén szerenádot adnak, és szerenádra egyforma pólót rendelnek maguknak. A pólók méretezése a következő:

S	M	L	XL	XXL
160–168	168–176	176–184	184v192	192–

Készítsünk táblázatot, amelyből kiderül, melyik méretből hányat kell rendelni.

Megoldás:

A méretezéskor a magasságokat 8 cm szélességű osztályokba sorolják. Észrevehetjük, hogy bizonyos magasságok két osztályba is illenek. Ezeknél az értékeknél döntsünk a kényelmesebb, nagyobb méret mellett.

S	M	L	XL	XXL
160–168	168–176	176–184	184–192	192–200
16	36	34	16	0

c) A boltban kiderült, hogy csak akkor kapnak kedvezményt a nagy tételű vásárlásra, ha mind a 102 pólót azonos méretben rendelik meg. Vitatkozni kezdtek azon, hogy melyik is legyen ez az egyetlen méret.

- János azt javasolta, akkora pólót vegyenek, amelyik egy **átlagos testmagasságú** fiúra illene.
- A „középutas” Marci azt javasolta, válasszanak pólót a tornasor **középső** emberének.
- Péter azt mondta, azt a méretet rendeljék, amelyből amúgy is a **legtöbbet** rendeltek volna.
- Gábor amellett kardoskodott, hogy annak a magasságnak megfelelő méretet rendeljék, amelyiknek **eltérése** az összes fiúétól **a legkisebb**, így lesz az eltérés a legkevésbé feltűnő.
- A 167 cm magas Sanyi állította, olyan pólót kell venni, ami a többségnek jó, és mivel a vele egy **magasságúak vannak a legtöbben**, ezért az S méret a nyerő.
- A colos Laci arról győzködte a többieket, hogy az **XL méretet** vegyék, hiszen míg a nagy legfeljebb egy kicsit „laza” lesz, ő bele sem fér a kisebbekbe.

Állapítsuk meg, melyik méretet kell venni János, Péter, Gábor, illetve Marci javaslatára!

Megoldás:

János: Számítsuk ki az adatsokaság átlagát, azaz számtani közepét!

$$\overline{x} = \frac{1\cdot 164 + 3\cdot 165 + 3\cdot 166 + 9\cdot 167 + ... + 2\cdot 189 + 1\cdot 190}{102} \approx 175{,}41\,.$$

Átlag: 175,41. Ennek megfelelő méret az M, de ha ezt az értéket egészekre kerekítjük, megállapodásunk szerint már az L méretet kellene választani.

Péter: Az M osztály gyakorisága a legnagyobb (36).

Marci: ugyanezt javasolta, hiszen a „tornasorban” a **középső a magasságadatok mediánja**. Ebben az esetben nincs középső ember a tornasorban, ugyanis páros számú adatunk van. Ilyenkor **a medián a két középső adat számtani közepe**. Nézzük, milyen adat áll az 51. és az 52. helyen: Mindkét adat a 175 cm, így az adatsokaság mediánja a 175, ennek megfelelő pólóméret az M.

Gábor: Keressük azt a számot, amelyre az eltérések abszolútértékének összege a lehető legkisebb. Tudjuk, hogy **egy adatsokaság mediánja az szám, melynek átlagos**

abszolút eltérése az adatoktól a lehető legkisebb, így Gábor is azt javasolja, hogy a mediánnak megfelelő méretet vásárolják.

A vita eredményeként végül az M méretet választották.

d) Számítsuk ki a fenti adatsokaság esetén az átlagtól vett átlagos abszolút eltérést, a módusztól vett átlagos abszolút eltérést, valamint a mediántól való átlagos abszolút eltérést!

Megoldás:

164	165	166	167	168	169	170	171	172
1	3	3	9	4	4	6	5	4
173	174	175	176	177	178	179	180	181
3	2	8	7	6	4	6	5	2
182	183	184	185	186	187	188	189	190
3	1	4	2	3	1	3	2	1

Az adatsokaság átlaga $\bar{x} = 175{,}41$. Az átlagtól való átlagos abszolút eltérés:

$$\Delta\bar{x} = \frac{1\cdot|164-\bar{x}|+3\cdot|165-\bar{x}|+3\cdot|166-\bar{x}|+\ldots+2\cdot|189-\bar{x}|+1\cdot|190-\bar{x}|}{102} \approx 5{,}56.$$

Az adatsokaság módusza $Mo = 167$. A módusztól vett átlagos abszolút eltérés:

$$\Delta\bar{x} = \frac{1\cdot|164-Mo|+3\cdot|165-Mo|+3\cdot|166-Mo|+\ldots+2\cdot|189-Mo|+1\cdot|190-Mo|}{102} \approx 8{,}65$$

Az adatsokaság mediánja $Me = 175$. A mediántól vett átlagos abszolút eltérés:

$$\Delta x_{Me} = \frac{1\cdot|164-Me|+3\cdot|165-Me|+3\cdot|166-Me|+\ldots+2\cdot|189-Me|+1\cdot|190-Me|}{102} \approx 5{,}55$$

Az említett közepek közül a **medián** olyan tulajdonságú, hogy **a tőle való átlagos abszolút eltérés** a lehető legkisebb. Tehát előbbi példánkra utalva, ezt a méretet választva lesz a méretbeli különbségek átlaga a legkisebb.

A **módusznak** megvan az a jó tulajdonsága, hogy erre tippelve találjuk el a legnagyobb valószínűséggel a leggyakoribb értéket, hiszen **ennek gyakorisága**, így relatív gyakorisága **a legnagyobb**.

Az **átlag** alkalmas arra, hogy megadjon egy olyan értéket, mellyel **helyettesítve az adatsokaság értékeit az összeg változatlan lesz.**

Van a középértékektől való átlagos abszolút eltérésen kívül egy másik mutató is, ami jelzi, mennyire térnek el az adatok a középértéktől. Ha az eltérések abszolút értéke helyett azok négyzetének számítjuk az átlagát, **átlagos négyzetes eltérésről** beszélünk. Ha az átlagtól vett eltérések négyzetének vesszük az átlagát, az a leggyakrabban használt **szórásnégyzet**.

Ez a legtöbb zsebszámológépen egy billentyűvel kiszámolható.

$$\boxed{\sigma^2 = \frac{(x_1-\overline{x})^2+(x_2-\overline{x})^2+\ldots+(x_n-\overline{x})^2}{n}}$$

A szórást úgy kapjuk meg, hogy négyzetgyököt vonunk a szórásnégyzetből:

$$\boxed{\sigma = \sqrt{\frac{(x_1-\overline{x})^2+(x_2-\overline{x})^2+\ldots+(x_n-\overline{x})^2}{n}}}$$

Mintapélda$_5$

Számítsuk ki az előző feladatban szereplő adatok átlagos négyzetes eltérését a

a) módusztól; b) mediántól; c) átlagtól.

Számítsuk ki a szórást is!

Megoldás:

$$\frac{(x_1-Mo)^2+(x_2-Mo)^2+\ldots+(x_n-Mo)^2}{n} =$$
$$= \frac{1\cdot(164-167)^2+3.(165-167)^2+\ldots+2\cdot(189-167)^2+1\cdot(190-167)^2}{102} \approx 115{,}55.$$

$$\frac{(x_1-Me)^2+(x_2-Me)^2+\ldots+(x_n-Me)^2}{n} =$$
$$= \frac{1\cdot(164-175)^2+3.(165-175)^2+\ldots+2\cdot(189-175)^2+1\cdot(190-175)^2}{102} \approx 44{,}96.$$

$$\sigma^2 = \frac{(x_1-\overline{x})^2+(x_2-\overline{x})^2+\ldots+(x_n-\overline{x})^2}{n} =$$
$$= \frac{1\cdot(164-175{,}5)^2+3.(165-175{,}5)^2+\ldots+2\cdot(189-175{,}5)^2+1\cdot(190-175{,}5)^2}{102} \approx 44{,}80.$$

Látható, hogy az átlagtól számított szórásnégyzet a legkisebb.

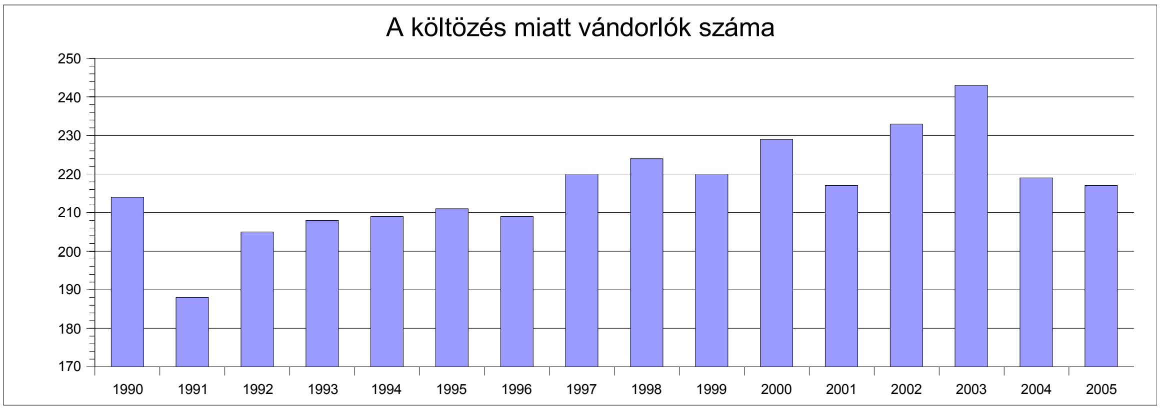
A szórás: $\sigma = \sqrt{\sigma^2} \approx 6{,}69$.

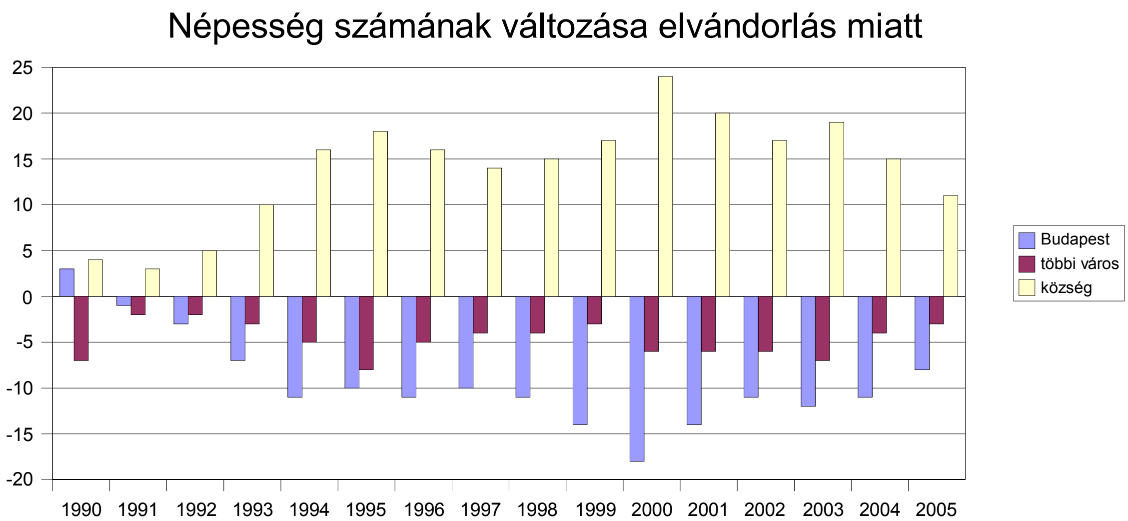
A szórás a terjedelemhez hasonló jellegű mutató. Kis szórása annak az adatsokaságnak van, ahol az értékek kevéssé térnek el az átlagtól.

Feladatok

4. Az alábbi grafikonok azt mutatják, hogy egy 16 éves időszakban (1990–2005)

- hányan változtattak lakhelyet az egyes években (az számít lakhelyváltoztatásnak, ha másik településre költözik);
- hogyan változott (költözések miatt) az egyes településtípusok lélekszáma.





a) Számítsd ki, mekkora volt a tizenhat éves időszakban az átlagos népességmozgás!

b) Igaz-e, hogy a magyarok nem szívesen változtatnak lakhelyet? Számítsd ki a vándorlók átlagának relatív gyakoriságát. (Magyarország népességét 10 millióra kerekítheted.)

c) Milyen következtetéseket tudsz levonni a második grafikonról? Mivel magyarázod, hogy míg például 2005-ben a más településre költözők száma 218, addig Budapest, más városok és a faluk népességének változása összesen is csak 22?

d) Számítsd ki mindhárom településtípus esetén a változás átlagát a tizenhat évre, valamint a szórást!

Megoldás:

Először célszerű mindhárom grafikonról táblázatot készíteni:

év	**1990**	**1991**	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**
elköltözöttek száma	214	188	205	208	209	211	209	220	224	220	229	217	233	243	219	217

év	**1990**	**1991**	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**
Budapest	3	-1	-3	-7	-11	-10	-11	-10	-11	-14	-18	-14	-11	-12	-11	-8
többi város	-7	-2	-2	-3	-5	-8	-5	-4	-4	-3	-6	-6	-6	-7	-4	-3
község	4	3	5	10	16	18	16	14	15	17	24	20	17	19	15	11

a) $\bar{x} = \dfrac{3466}{16} = 216{,}625$

b) A relatív gyakoriság$= \dfrac{216{,}625}{10\,000\,000} = 0{,}0000217$. Ez azt jelenti, hogy minden ötvenezredik ember szánja rá magát a településváltásra, ami valóban nagyon kevés.

c) Pl.: Többen hagyják el Budapestet, mint ahányan odaköltöznek. Ha valaki egyik kisvárosból másik kisvárosba, vagy egyik faluból másik faluba költözik, az nem jelenik meg a grafikonon. Másrészt, ez az adatsor előállhat úgy is, hogy például 116 ember elköltözik Budapestről közülük 51 községbe, 65 kisvárosba. Ugyanakkor 40 ember valamelyik községből, 68 pedig valamelyik kisvárosból Budapestre költözik. Az a tanulság, hogy a grafikonról leolvasott adatokból nem lehet biztos következtetéseket levonni!

d) $\bar{x}_{Bp} = -9{,}3125$ $\quad \bar{x}_{város} = -4{,}6875$ $\quad \bar{x}_{község} = 14$.

$\sigma_{Bp} \approx 5{,}059$ $\quad \sigma_{város} \approx 1{,}793$ $\quad \sigma_{község} = 5{,}766$.

5. Egy kúszónövény indáinak hosszúságát mérték, és az adatokat a következő táblázatba sorolták be (az indák hosszát mm-ben mérték):

Inda hossza	1–10	11–20	21–30	31–40	41–50	51–60	61–70	71–80	81–90
Gyakoriság	7	25	63	52	36	27	9	12	1

a) Add meg az osztályok szélességét!

b) Számítsd ki egy inda átlagos hosszát!

c) Készítsd el azt az oszlopdiagramot, mely az egyes osztályok gyakoriságát ábrázolja!

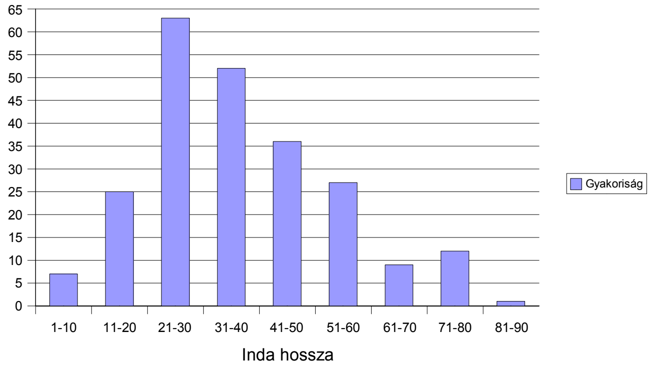
d) Mi a valószínűsége, hogy egy inda hossza 31 és 70 mm közé essen?

Megoldás:

a) Az osztályok szélessége 10 (mm).

b) Az osztályok középértékeivel kell számolni: $\overline{x} = 37{,}05$.

c)



d) $\mathbf{P} = \dfrac{52+36+27+9}{232} \approx 0{,}5345$.

III. Középértékek alkalmazhatósága

Ebben a fejezetben azzal foglalkozunk, hogy mikor, melyik középérték nyújtja számunkra a legtöbb információt.

Mintapélda$_6$

Valaki a frissen szerzett szakirányú képesítésével állást keres. Több ajánlat közül is választhat, és igyekszik a vállalatoknál uralkodó bérviszonyokról információt szerezni. Értékeljük együtt, milyen adatokból milyen következtetéseket lehet levonni!

- Az A vállalatnál fizetett havi bérek terjedelme 700 000 Ft.
- A B vállalatnál a havi fizetések átlaga 270 000 Ft.
- A C vállalatnál a havi fizetések módusza 240 000 Ft.
- A D vállalatnál a havi fizetések mediánja 280 000 Ft.

Feltételezzük, hogy mind a négy vállalat profilja hasonló, és körülbelül 20 embert alkalmaznak. Legyen ez például egy tervezővállalat, ahol valószínűleg van egy vezető, sok mérnök és 1-2 kisegítő alkalmazott.

Megoldás:

A havi bérek terjedelme egy álláskeresőnek nem sokat mond. Ez valószínűleg a kisegítő kézbesítő minimálbére és az igazgató csúcsfizetése közti különbség. Mivel egy szakirányú pályakezdő valószínűleg nem az igazgatói vagy a kézbesítői állást pályázza meg, az információ semmitmondó. Lehet, hogy a másik 18 alkalmazott fizetése egységesen havi 80 000 Ft, de az is lehet, hogy a többi 18 alkalmazott mind 400 000 Ft körül keres.

A B vállalatnál az átlagból megtudhatjuk az összes kiosztott fizetést. Ha nem tudjuk, hogy pontosan hány kis fizetésű (kisegítő) alkalmazott van, és nehezen becsüljük a vezető fizetését, akkor igen nagy csalódások érhetnek minket. Nézzük meg az alábbi két becslést:

Legyen 2 fő fizetése 70 000 Ft, és a vezető fizetése 400 000 Ft, ekkor a 17 másik (egymáshoz hasonló fizetésű alkalmazottra fejenként

$$\frac{20\cdot 270\,000-\left(2\cdot 70\,000+400\,000\right)}{17}\approx 285\,882 \text{ Ft jut.}$$

Ha azonban csak 1 alacsony fizetésű alkalmazott van, havi 80 000 Ft-os fizetéssel, és a vezető fizetése havi 800 000 Ft, akkor a másik 18 alkalmazottra fejenként

$$\frac{20\cdot 270\,000-\left(80\,000+800\,000\right)}{18}\approx 251\,111 \text{ Ft jut, ami lényeges eltérés.}$$

A C vállalatnál a leggyakoribb fizetés a 240 000 Ft. Valószínűleg ez nem azt jelenti, hogy például ketten forintra ennyit kapnak, hanem azt, hogy a leggyakoribb a 240 000 Ft körüli fizetés. Ha feltételezzük, hogy a dolgozók zömének fizetése hasonló, akkor pályázónk várható fizetése is e körül az érték körül mozog.

A D vállalatnál a medián 280 000 Ft. Talán ebben az esetben ez a legtöbbet mondó adat. Egy adatsokaság mediánja nem változik meg attól, ha a legkisebb és legnagyobb értéket másik kicsire, illetve nagyra változtatjuk, vagy akár el is hagyjuk. Úgy mondjuk, a medián nem érzékeny a szélsőséges adatokra. Valószínűleg pályázónk is e körül az érték körül várhatja fizetését.

Mintapélda$_7$

A cukrászoknak szóló receptkönyvben a receptek nem úgy kezdődnek, hogy vegyél 5 tojást, hanem úgy, hogy hozzávalók: tojássárgája: 0,04 kg. Ha valaki az otthoni sütéshez mégis ezt akarja használni, meg kell tudnia, hány tojást is üssön föl anélkül, hogy minden alkalommal patikamérleggel mérné a tojássárgája tömegét. Egy különösen nagy adag rántotta elkészítése alkalmával valaki rászánta az időt, és mielőtt fölverte volna a tojásokat, megmérte 10 tojássárgájának tömegét külön-külön. A következő adatsokaságot kapta:

gramm:	21	21	19	20	21	18	18	22	19	21

Értékeljük az adatsokaságot! Melyik közép jellemzi legjobban ezt a sokaságot?

Megoldás:

Ennek az adatsokaságnak is van módusza és mediánja $Mo = 21$, $Me = 20{,}5$, de senki sem gondolja, hogy van jelentősége annak, hogy a megmért tojásokat sorba rendezve melyik tömege áll középen, vagy annak, hogy a 10 méréseredmény között 4-szer fordult elő a 21. Sokkal inkább jellemzi az átlag. Kiszámíthatjuk, hogy a mérés átlaga 20 g = = 0,02 kg, tehát ha egy süteménybe 0,04 kg tojássárgája kell, 2 tojás sárgáját teszünk bele.

Feladatok

6. Egy éjjel-nappal nyitva tartó áruház hűtőpultjának hőmérsékletét háromóránként ellenőrizni kell, és az adatokat föl kell jegyezni egy ott függő táblára. Az elmúlt három nap alatt a következő adatokat jegyezték fel:

Leolvasás	**0 h**	**3 h**	**6 h**	**9 h**	**12 h**	**15 h**	**18 h**	**21 h**
Hőmérséklet (ºC)	–22	–22	–21	–18	–19	–19	–17	–21
	–23	–22	–22	–19	–19	–18	–16	–19
	–22	–23	–21	–20	–20	–21	–19	–20

a) Ábrázold közös koordináta-rendszerben a három nap mérési eredményeit! Az egyes mérési pontokat kösd össze, feltételezve, hogy két mérés között a hőmérséklet egyenletesen változik.

b) Számítsd ki az első napon mért hőmérsékletek átlagát!

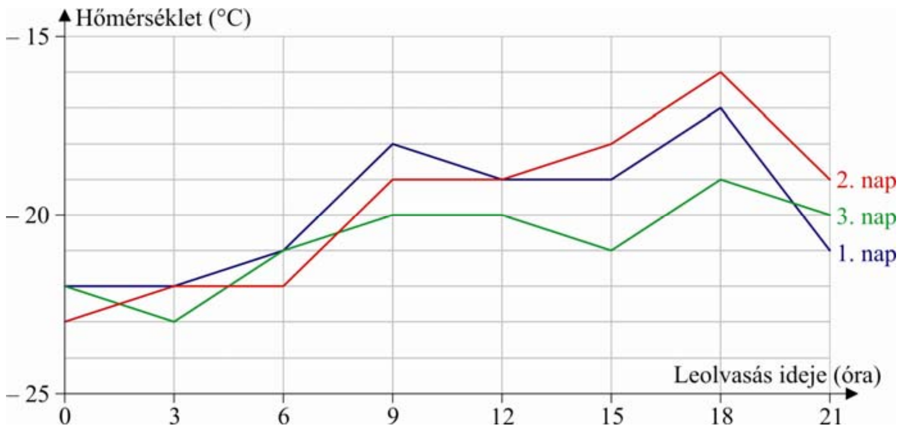
c) Számítsd ki a három nap csúcsforgalomban mért (18 h) hőmérsékleteinek átlagát!

d) A hűtő hőmérséklete nem mehet tartósan a –18 ºC-os érték fölé. Ha tartósan (6 órán át) ennél magasabb értéket mérnek, az árut meg kell semmisíteni. Szükséges-e most ez az intézkedés?

e) A grafikon alapján olvasd le, mi a valószínűsége annak, hogy a 3. napon érkező ellenőr –20 ºC-nál magasabb értéket talál?

Megoldás:

a)



b) $\overline{x}_{1.nap} = -19{,}88\,°\mathrm{C}$ c) $\overline{x}_{csúcs} = -17{,}33\,°\mathrm{C}$.

d) Nem szükséges.

e) Az időtartam, amíg a hőmérséklet – 20ºC fölött volt (a grafikon alapján) 4,5 óra, tehát a valószínűség $\mathbf{P} = \frac{4,5}{24} = 0,1875$, azaz 18,75%.

7. Egy kieséses tenisztornán minden játékról följegyezték, hogy hány „game”-ből állt:

„game”-ek száma	Gyakoriság	Osztály szélesség	Relatív gyakoriság
15–24	6		
25–34	12		
35–44			0,3
45–54	15		
55–64	3		
65–84	6		

a) Töltsd ki a táblázat hiányzó részeit!

b) Készíts kördiagramot a relatív gyakoriságokból!

c) Melyik középértéket számítanád ki az adatokból ahhoz, hogy megállapítsd, mennyi ideig tart egy 20 mérkőzésből álló bajnokság?

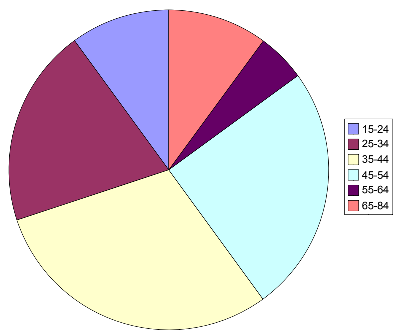
d) Mekkora a valószínűsége annak, hogy az elsőnek kiesett játékos több mint 65 játékot (game-et) játszott?

Megoldás:

a)

„game”-ek száma	Gyakoriság	Osztály szélesség	Relatív gyakoriság
15–24	**6**	10	0,1
25–34	**12**	10	0,2
35–44	18	10	**0,3**
45–54	**15**	10	0,25
55–64	**3**	10	0,05
65–84	**6**	20	0,1

b)



c) Ha tudom, általában mennyi ideig tart egy játék (game), a mérkőzések átlagos játékszámát érdemes kiszámítani.

d) $\mathbf{P} = \frac{6}{60} = 0,1$, ami megegyezik a relatív gyakorisággal.

8. Az Egészségbiztosítási Felügyelet 2007-es honlapján található táblázatokból kiderül, hogy az egyes kórházak osztályain milyen az ágykihasználtság, illetve az átlagos ápolási idő.

Osztálynév	Kórházi esetszám	Ágykihasználtság (%)	Átlagos ápolási idő (nap)
I. Belgyógyászat	406	74,9	10,8
II. Belgyógyászat	778	60,2	11,3
III. Belgyógyászat	585	82,4	10,3
Sebészet	2142	65,1	6,0
Traumatológia	2162	48,5	3,3
Szülészet-nőgyógyászat	3373	98,0	4,2
Fül-orr-gégészet	1204	67,1	8,1
Szemészet	2133	68,0	2,9
Ideggyógyászat	1097	77,9	7,8
Ortopédia	807	21,2	3,8
Urológia	1540	62,9	4,8
Intenzív	255	49,4	9,2
Tüdőgyógyászat	837	68,7	12,0
Sugárterápia, onkoradiológia	1919	47,4	9,6
Anyagcsere, endokrinológia	640	77,7	11,1
Gasztroenterológia	629	72,1	10,5
Érsebészet	274	49,9	6,6
Stroke	374	57,5	8,4
Immuno / nephrológia	521	55,1	10,8
Kardiológia	968	63,5	9,6

a) Add meg az osztályok ágykihasználtságának mediánját, és értelmezd is ezt az adatot!

b) Add meg ennek az értéknek a szórását!

c) Számítsd ki, hány beteg fordult meg ebben az évben a kórházban!

d) Számítsd ki, átlagosan hány napot töltöttek a kórházban!

Megoldás:

a) 64,3; b) 15,81 (%); c) 22 644 beteg;

d) Az egyes osztályokon eltöltött napszámot súlyozni kell az osztályok betegszámával:

$$\frac{152\,698{,}7}{22\,644} \approx 6{,}74\,.$$

IV. Nagy elemszámú adatsokaságok jellemzése

A gyakorlatban a statisztika sok adattal, hatalmas adatsokasággal dolgozik. Amikor ezeket a statisztikusok feldolgozzák, mi csak az értékelés eredményével találkozunk. Próbáljuk meg a lehető legtöbb információt kinyerni egy-egy táblázatból, grafikonból!

Mintapélda$_8$

Az OKI (Országos Közoktatási Intézet) felmérést készített a tanulók terheléséről. Ennek keretében azt kérték, írjanak a négyzetbe egy számot, attól függően, milyen osztályba járnak:

1-et, ha normál tantervűbe; **2-t**, ha tagozatosba vagy emelt szintűbe; **3-at** egyéb esetben.

A következő válaszok születtek:

1	2	3	nem válaszolt
1197	644	14	847

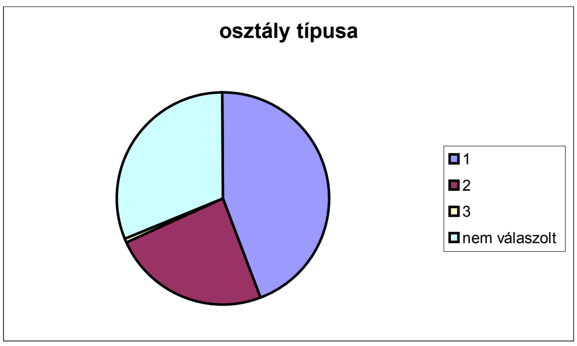
Milyen típusú ismérvre kérdez rá ez a statisztika? Milyen jellegű középnek van értelme az ilyen típusú adatok esetében? Számítsuk ki, mi a valószínűsége annak, hogy olyan gyerek adatlapja kerül a kezünkbe, aki erre a kérdésre nem válaszolt! Ábrázoljuk az adatokat kördiagramon!

Megoldás:

Beszélhetünk méréses és minősítéses ismérvekről. A méréses ismérvek mindig számok, de olyanok, amelyeket érdemes sorba rendezni, nagyságuk jellemzi az ismérvet. Ez az adatsokaság méréses ismérv alapján osztályozható. Az adatok hiába számok, más jellegű információt hordoznak. Ilyenkor nincs értelme átlagot számolni, a medián sem hordoz információt, de a módusz itt például elárulja, hogy a válaszadók közül a legtöbben normál tantervű osztályba járnak.

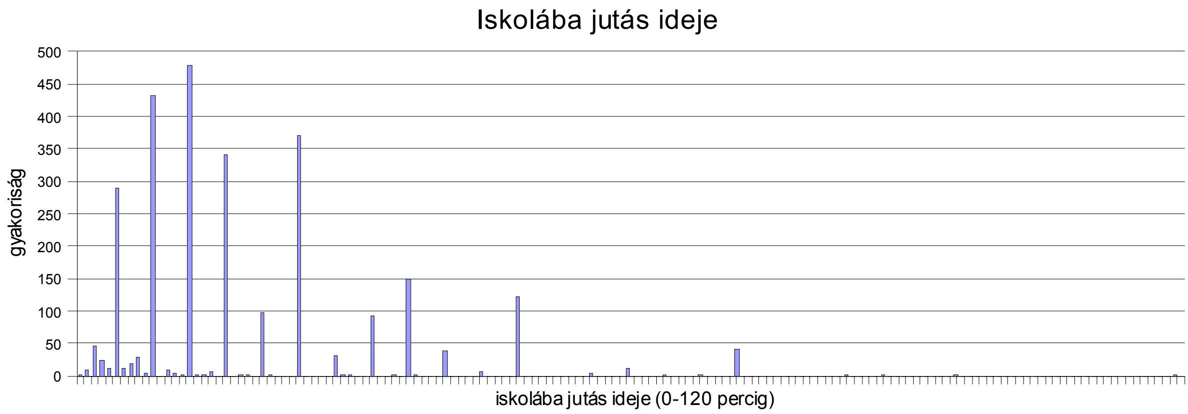
A beérkezett adatlapok száma $1197 + 644 + 14 + 847 = 2702$. Ezek közül 847-ben nem szerepelt válasz erre a kérdésre, így $\frac{847}{2702} \approx 0{,}3135$ a valószínűsége annak, hogy ilyen lap jut a kezünkbe.

A kördiagramon látszik, hogy azok száma, akik osztályuk típusát az egyéb kategóriába sorolták, elenyésző azokhoz képest, akik mivel nem tudtak dönteni, inkább üresen hagyták a mezőt.



Mintapélda9

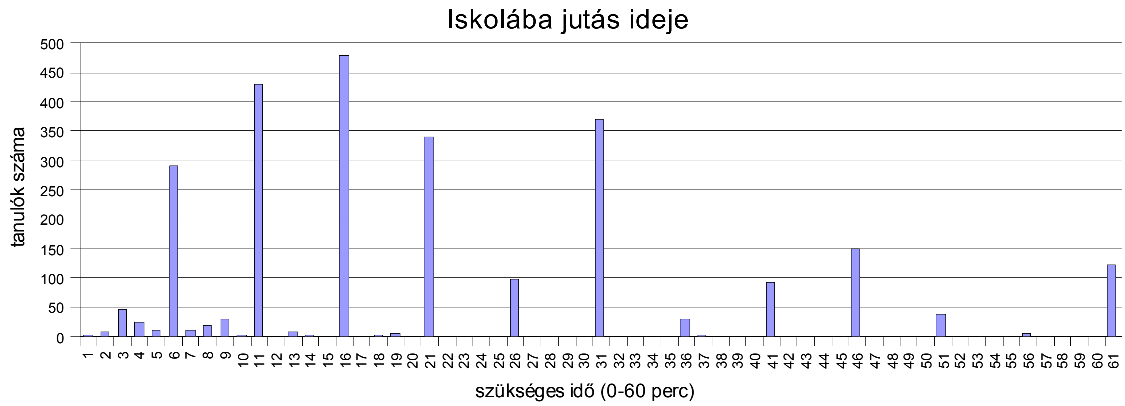
Megkérdezték a tanulókat, mennyi idejükbe telik reggelente eljutni az iskolába. A válaszok a 0–120 perc között voltak. Amikor ábrázolni akarták az egyes percek gyakoriságát, az alábbi grafikont kapták.



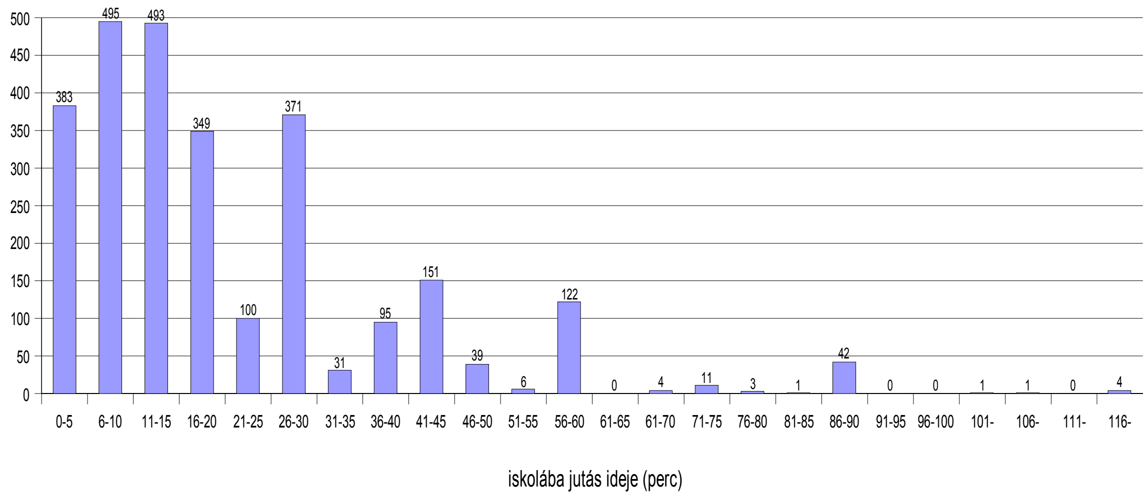
a) Próbáljuk megfogalmazni, mi is a baj ezzel a grafikonnal, és adjunk javaslatot a grafikon kijavítására!

Megoldás:

Erről az ábráról még azt sem tudjuk leolvasni, hogy melyik volt az a válasz, amit a tanulók a leggyakrabban adtak. A vízszintes tengelyen azért nem lehet beosztás, mert a terjedelem túl nagy. Látható azonban, hogy 60 percnél többet alig néhányan töltenek utazással, tehát már az is sokat segítene, ha az ábrázolt időtartamot a felére csökkentenénk.



Így már többet le lehet olvasni az ábrából, de az információk egy részét elvesztettük. Ilyenkor szokás az adatokat osztályokba sorolni, és azok gyakorisági táblázata alapján ábrázolni.



Ennek a grafikonnak a megrajzolásához az alábbi osztályközös gyakorisági táblázatot kellett elkészíteni:

0–5	6–10	11–15	16–20	21–25	26–30	31–35	36–40
383	495	493	349	100	371	31	95
41–45	46–50	51–55	56–60	61–65	61–70	71–75	76–80
151	39	6	122	0	4	11	3
81–85	86–90	91–95	96–100	101–105	106–110	111–115	116–120
1	42	0	0	1	1	0	4

Nagy adatsokaság esetén az osztályba sorolás áttekinthetőbbé teszi az adatokat.

b) Számítsuk ki, mi a valószínűsége annak, hogy egy diák 1 óránál hosszabb időt tölt utazással!

Megoldás:

Az érintett osztályokba $4+11+3+1+42+1+1+4=67$ tanuló tartozik, így a keresett valószínűség $\frac{67}{2702}\approx 0{,}0248$, azaz körülbelül 2,5%.

Az ilyen nagy adatsokaságok jellemzésére nemcsak a mediánt, azaz az adatsokaságot két egyenlő részre bontó adatot szokták megadni, hanem a kvartilis (latin: negyed) és percentilis (latin: percent=századrész) értékeket is.

A **kvartilis** az adatsokaság sorba rendezett elemeit négy egyenlő részre bontja. A negyedelő pontoknál szereplő értékeket szokás rendre első, második, harmadik kvartilisnek nevezni. A második kvartilis éppen felezi az adatsokaságot, így az megegyezik a mediánnal.

Ha az adatsokaság elég sok adatot tartalmaz, beszélhetünk a **percentilisek**ről is. Ilyenkor a sokaság sorba rendezett tagjait 100 egyforma részre bontjuk, és a 30. percentilis (30th-ként szokták írni az angol sorszámnév képző miatt) a 30. osztópontnál szereplő adat. Könnyű belátni, hogy ilyenkor az adatok 30%-a kisebb ennél az adatnál (vagy legalábbis nem nagyobb), 70%-a pedig nagyobb (vagy legalábbis nem kisebb).

Mintapélda$_{10}$

A fenti adatsokaságnál állapítsuk meg a harmadik kvartilis és a 20. percentilis értékét!

Megoldás:

2702 adatunk van. Tehát a negyedében 675,5 adat van, a háromnegyedében pedig 2026,5.

A harmadik kvartilis tehát a 2026. és a 2027. adat számtani közepe. Mindkét érték a 26-30 perces osztályba esik.

A 20. percentilis a 20. osztópont akkor, amikor az adatsokaság növekvő sorrendű tagjait 100 egyenlő részre osztottuk. Tehát az adatok 20 %-ának kell ez alatt lennie, vagyis a sorba rendezett adatokból az 541. adatot keressük, és ez a 6-10 perces osztályba esik.

Feladatok

9. Tudsz-e magyarázatot adni arra, hogy amikor a modul második mintapéldájában az iskolába jutás percre pontos idejének gyakoriságát ábrázoltuk, miért szerepeltek kiugró gyakorisági értékek periodikusan?

Megoldás:

A kiugró értékek az 5, 10, 15 stb. percértékeknél szerepeltek. Oka az lehet, hogy – hacsak nem stopperrel mérjük az időt – sokkal szívesebben adunk ilyen válaszokat, mint pl. azt, hogy 12 perc. Ez indokolja azt is, miért érdemes már a kérdést is inkább időintervallumokkal feltenni.

10. Egy osztály tanulói a következő kérdőívet töltik ki:

A: Fiú vagy lány vagy?

B: Hány éves vagy?

C: Hány centiméter magas vagy?

D: Hány kg vagy?

E: Mi a postai irányítószámod?

Állapítsd meg, a fenti kérdések alapján összegyűjtött adatsokaságokból melyek a minősítéses és melyek a méréses adatsokaságok?

Megoldás: Minősítéses: A, E; Méréses: B, C, D.

11. Az egyik honlapon található az alábbi kalkulátor. A táblázat felső felét kitöltve (megadva a gyerek nemét, korát, testtömegét és magasságát) adja ki a táblázat alsó részében szereplő statisztikai mutatókat. Értelmezd az itt kapott eredményeket!

Gyermek percentilis kalkulátor	
Adja meg a következő adatokat	
A gyermek neme	fiú
A gyermek kora	12 év és 3 hónap
Testtömeg	35 kg
Testmagasság	150 cm
Eredmények	
A korhoz tartozó testmagasság percentilis	25th – 50th
A korhoz tartozó testtömeg percentilis	10th – 25th
A testmagassághoz tartozó testtömeg percentilis	25th – 50th

Megoldás:

A testmagasság a 25. és az 50. percentilis között van, tehát a fiú magassága nagyobb, mint a vele egykorú fiúk 25%-ának, de a fiúk „tornasorának” alsó felében van.

A testtömeg a 10. és a 25. percentilis közé esik, tehát a vele egykorú fiúknak csak 10%-a könnyebb nála, de 75%-a valószínűleg súlyosabb.

A testmagassághoz viszonyított testtömege a 25. és az 50. percentilis közé esik, ez azt jelenti, hogy a vele azonos magasságú fiúk körülbelül negyedrésze nála könnyebb, de a fele nála súlyosabb, tehát valószínűleg egy vékony fiúról van szó.

Kislexikon

Egy **adatsokaság terjedelmének** az adatsokaságban a legkisebb és legnagyobb érték közti különbséget nevezzük.

Módusznak az adatsokaságban legtöbbször előforduló adatot nevezzük. (A módusz gyakorisága és relatív gyakorisága a legnagyobb.)

Medián: Ha egy adatsokaság értékeit növekvő sorrendbe rendezzük, akkor a sorban középső értéket az adatsokaság mediánjának nevezzük. Ha az adatsokaságban két középső érték van (páros sok adatból áll), akkor ennek a két értéknek a számtani közepe lesz a medián.
(Egy adatsokaság mediánja az szám, melynek átlagos abszolút eltérése az adatoktól a lehető legkisebb.)

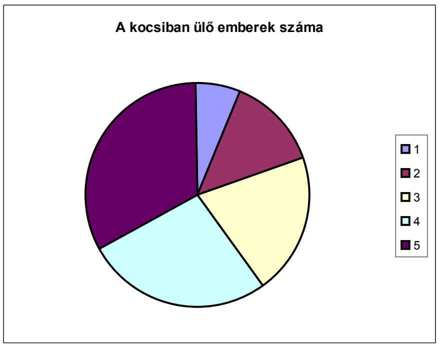
A **kvartilisek** az adatsokaság sorba rendezett elemeit négy egyenlő részre bontják. A negyedelő pontoknál szereplő értékeket szokás rendre első, második, harmadik kvartilisnek nevezni. A második kvartilis éppen felezi az adatsokaságot, így az megegyezik a mediánnal.
Ha az adatsokaság elég sok adatot tartalmaz, beszélhetünk a **percentilisekről** is. Ilyenkor a sokaság sorba rendezett tagjait 100 egyforma részre bontjuk, és például a 30. percentilis (30th-ként szokták írni az angol sorszámnév képző miatt) a 30. osztópontnál szereplő adat.

Átlag: Egy adatsokaság átlaga a benne szereplő adatok számtani közepe.

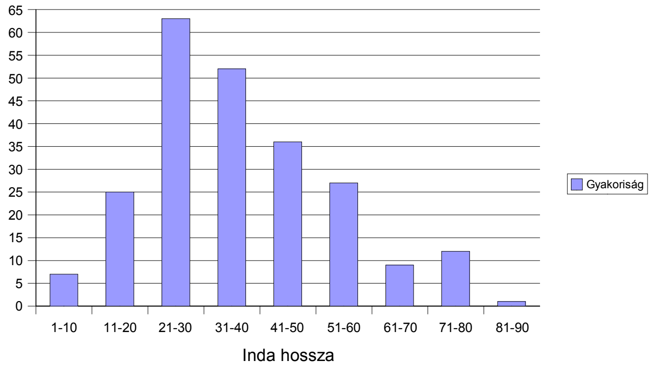
Gyakoriság: Ha megszámoljuk, hogy egy adathalmazban egy bizonyos adat hányszor szerepel, az adat gyakoriságát kapjuk meg.

Relatív gyakoriság: Ha egy adatsokaságban n adat van, és egy bizonyos adat gyakorisága k, akkor ennek az adatnak a relatív gyakorisága $\frac{k}{n}$.

DIAGRAMOK

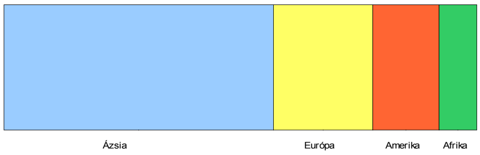


KÖRDIAGRAM



OSZLOPDIAGRAM

SÁVDIAGRAM



Az oszlopdiagram függőleges tengelyén a gyakoriságot és a relatív gyakoriságot is ábrázolhatjuk.

Egy x_i adat eltérése az $\bar{x}$ átlagtól különbségük abszolút értéke: $|x_i - \bar{x}|$. Ha az egyes adatok eltéréseinek számtani közepét vesszük egy n elemű adatsokaságban, az **átlagtól vett abszolút eltérést** kapjuk meg:

$$\Delta\bar{x} = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + \ldots + \cdot|x_n - \bar{x}|}{n}$$

Szórásnégyzet: az átlagtól vett eltérések négyzetének vesszük az átlagát.

$$\boxed{\sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{n}}$$

A **szórást** úgy kapjuk meg, hogy négyzetgyököt vonunk a szórásnégyzetből:

$$\boxed{\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{n}}}$$