Geostatisztikai példatár

Geresdi István, Bugya Titusz, Pirkhoffer Ervin
TTK, Földrajzi Intézet

Pécs
2005

A kétciklusú képzés bevezetése a magyar felsőoktatásban
a természettudományi szakokon.
Alkalmazkodás a munkaerőpiac igényeihez.

HEFOP-3.3.1-P-2004-06-0016/1.0

Az Európai Szociális Alap támogatásával

Európa itt épül

Európa Terv

1. Kombinatorikai feladatok

I. ***Permutáció*: *n* egymástól különböző elem egy meghatározott sorrendben való elhelyezését az *n* elem egy permutációjának nevezzük. A permutációk lehetséges száma:**

$$n! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot (n-2) \cdot (n-1) \cdot n$$

Feladatok

1.1. Az 1, 2, 3, 4 és 5 számjegyekből hány ötjegyű számot alkothatunk, ha mindegyik számjegy csak egyszer fordulhat elő?

Megoldás

Az első helyre 5, a másodikra 4, a harmadikra 3, a másodikra 2 és az ötödikre 1 számjegyet írhatunk. Így az összes lehetséges permutációk száma $5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 5! = 120$

1.2. Az 0,2,3,6 és 9 számjegyekből hány ötjegyű számot alkothatunk, ha mindegyik számjegy csak egyszer fordulhat elő? **(96)**

1.3. Van egy piros, egy fehér, egy kék és egy zöld golyónk.

a) Hányféle sorrendben rakhatjuk egymás mellé a golyókat? **(24)**

b) Mi a megoldás akkor, ha a golyókat egy négyzet csúcsán helyezzük el, és nem tekintjük különbözőnek az egymáshoz képest elforgatott helyzeteket? **(6)**

II. ***Ismétléses permutáció*: az *n* elem között vannak egyformák is. Két permutációt akkor tekintünk különbözőnek, ha van legalább egy olyan pozíció, amelyben különböző elemek találhatóak. Ha az elemek száma *n*, és *i* olyan elem van, amely ismétlődik k_1-szer, k_2 –ször, ….. és k_i –szer, akkor a permutációk száma:**

$$\frac{n!}{k_1! \cdot k_2! \cdot \ldots \cdot k_i!}.$$

Feladatok

1.4. A 2,2,3,3,3,3 és 4 számjegyekből hány darab, különböző, hétjegyű számot lehet felírni?

Megoldás

Különböztessük meg az azonos számjegyeket valamilyen módon, pl.: $2, 2^*, 3, 3^*, 3^{**}, 3^{***}, 4$. Könnyű belátni, hogy az így megkülönböztetett számjegyek permutációja 7! . Ez után számoljuk ki, hogy a megkülönbözetett 2-eseket és a 3-asokat hányféle módon rakhatjuk sorba! A lehetséges permutációk száma 2! és 4! . Mivel ezeknek a számoknak a cseréje nem eredményez új számot, a megoldás a következő lesz:

$$\frac{7!}{2!\cdot 4!}=105$$

1.5. Van 3 db piros, 4 db kék, 2 db piros és 5 db fehér golyónk. Hányféle módon tudjuk sorba rakni a golyókat? **(25225250)**

III. *Kombináció*: Ha *n* különböző elem közül *k*-t kiválasztunk, és a kiválasztás sorrendje *nem számít*, akkor azt az *n* elem egy *k*-ad osztályú kombinációjának nevezzük. A kombinációk lehetséges száma:

$$\frac{n!}{k!(n-k)!}=\binom{n}{k}.$$

Feladatok

1.6. A LOTTÓ húzás során hány lehetséges módon lehet 90 számból 5-öt kihúzni? **(43949268)**

1.7. Mennyi a lehetséges kombinációk száma a 6-os LOTTÓ esetén (45 számból kell 6-ot húzni)? **(8145060)**

Vajon melyik LOTTÓ esetén nagyobbak a nyerési esélyek?

IV. *Variáció*: Ha *n* különböző elem közül *k*-t kiválasztunk, és a kiválasztás sorrendje *számít*, akkor azt az *n* elem egy *k*-ad osztályú variációjának nevezzük. A variációk lehetséges száma:

$$\frac{n!}{(n-k)!}.$$

Feladatok

1.8. Az 1, 2, 3, 4 és 5 számjegyekből, hány háromjegyű számot írhatunk fel, ha minden számjegy csak egyszer fordulhat elő.

Megoldás

Az 5 db számjegyet 5! féle sorrendben lehet felírni. Mivel a maradék két számjegy sorrendje nem számít, a lehetséges variációk száma:

$$\frac{5!}{2!}=60$$

1.9. Mennyivel csökkennének a nyerési esélyek az 5-ös LOTTÓ-n, ha a számok kihúzásának sorrendje is számítana? **(120)**

V. *Ismétléses variáció*: *n* különböző elem közül úgy választunk ki *k* darabot, hogy a kiválasztás után az elemet visszatesszük, és megengedjük, hogy ugyanazt újból kiválaszthassuk. A lehetséges variációk száma ebben az esetben:

$$n^k.$$

Feladatok

1.10. A 2,3, 4 és 6 számjegyekből hány háromjegyű számot írhatunk fel, ha minden számjegy többször is szerepelhet?

Megoldás

Az első helyre 4, a második helyre 4 és az utolsó helyre szintén 4 számot írhatunk. Így a megoldás 4^3 .

1.11. Hány lehetséges módon lehet kitölteni egy TOTO szelvényt? (14 helyre kell 1-est, 2-est vagy X-et írni.) **(3^{14})**

1.12. Négy kockával dobva hány lehetséges variáció állhat elő? **(6^4)**

2. Műveletek halmazokkal

A halmaz fogalma

A halmaz alapfogalom, ennek megfelelően nem definiálható, csak körülírható fogalom. Ha *h* a *H* halmaz eleme, akkor azt a

$$h \in H$$

módon jelöljük.

Ha h nem eleme a H halmaznak, akkor ezt

$$h \notin H$$

módon jelöljük.

Azt a halmazt, melynek egyetlen eleme sincsen, *üres halmaz*nak nevezzük.

Műveletek

Részhalmaz képzése, halmazok egyesítése (uniója), halmazok közös részének képzése (metszete), halmazok különbsége, halmazok szimmetrikus differenciája, Descartes-szorzat képzése.

Halmaz részhalmaza

Ha H_1 halmaz, olyan, hogy mindegyik eleme egyben H_2 halmaznak is eleme, akkor H_1 halmazt H_2 halmaz részhalmazának hívjuk és így jelöljük:

$H_1 \subset H_2$ (olvasva: H_1 részhalmaza H_2-nek)

Valódi részhalmaznak nevezzük H_1-et, ha H_2-nak van olyan eleme is, amely H_1-nem nem eleme, de H_1 mindegyik eleme, együttal eleme H_2-nek is.

Halmazok uniója

A halmazok unióját nevezzük a halmazok összeadásának is. Jelölése:

$H_1 \cup H_2$ (H_1 unió H_2-nek olvasandó)

A H_1 és H_2 uniója azt a halmazt jelenti, amelynek eleme H_1 és H_2 összes eleme. Ha

$$H_1 \cup H_2 = H_3$$

Akkor elmondható, hogy H_3-ba csak azok az elemek tartoznak, amelyek vagy H_1-ben, vagy H_2-ben, de legalább az egyikükben megtalálhatóak.

Halmazok metszete

Nevezzük halmazok közös részének, vagy halmazok szorzatának is. H_1 és H_2 halmaznak metszete az a halmaz, melynek elemei azok, melyek H_1 *és* H_2 halmaznak *is* elemei. Jelölésekkel:

$$\boldsymbol{H_1 \cap H_2} \text{ (} H_1 \text{ metszet } H_2\text{-nek olvasandó)}$$

Így ha

$$H_3 = H_1 \cap H_2$$

, akkor igaz, hogy H_3-nak csak olyan elemei vannak, melyek H_1-nek *és* H_2-nek is elemei.

Halmazok különbsége

H_1 és H_2 halmaz különbsége az a halmaz, melynek elemei H_1-nek elemei, de H_2-nek nem elemei. Jelöléssel:

$$\boldsymbol{H_1 - H_2} \text{ (} H_1 \text{ mínusz } H_2\text{-nek olvasandó)}$$

Így ha

$$H_3 = H_1 - H_2$$

, akkor H_3 olyan halmaz, melynek csak azok az elemei, amelyek H_1-nek elemei és H_2-nek nem elemei.

H_1 és H_2 halmazok különbségének jelölésére alkalmazzák a

$$H_1 \setminus H_2$$

formulát is. Ez is H_1 mínusz H_2-nek olvasandó, és nincs a kétféle jelölés között elvi különbség.

Halmazok szimmetrikus differenciája

H_1 és H_2 halmazok szimmetrikus differenciája az a halmaz, mely csak azokat az elemeket tartalmazza, amelyek csak vagy H_1 halmaznak, vagy H_2 halmaznak az elemei. Jelölve:

$$H_1 \Delta H_2$$

Ez nem egyebet jelent, mint:

$$\boldsymbol{H_1 \Delta H_2 = (H_1 - H_2) \cup (H_2 - H_1)}$$

Halmazok Descartes-szorzata

H_1 és H_2 halmaz Descartes-szorzata az a halmaz, mely tartalmazza mindazon elempárokat, melyeket úgy képzünk, hogy a pár első tagja H_1 eleme, második tagja pedig H_2 eleme. Jelölése:

$$\boldsymbol{H_1 \times H_2}$$

Így, ha H_1 halmaz h_1, h_2 és h_3 elemekből áll, H_2 pedig h_a, h_b és h_c elemkből, akkor szorzatuk:

$$\boldsymbol{h_1, h_a; h_1, h_b; \ h_1, h_c}$$

$$\boldsymbol{h_2, h_a; h_2, h_b; \ h_2, h_c}$$

$$\boldsymbol{h_3, h_a; h_3, h_b; h_3, h_c}$$

Feladatok

Legyen H_1 és H_2 halmazunk. Határozzuk meg uniójukat és metszetüket!

H_1 halmaz elemei:

1, 4, 6, 3, 12, 5, 98, 20, 22, 10

H_2 halmaz elemei:

1, 50, 22, 31, 67, 10, 9, 3, 54, 55, 16

Megoldás

$H_1 \cup H_2$ = 1, 4, 6, 3, 12 5, 98, 20, 22, 10, 50, 31, 67, 9, 54, 55, 16

$H_1 \cap H_2$ = 1, 22, 10, 3

Feladat

Határozzuk meg a fenti H_1 és H_2 halmaz különbségét és szimmetrikus differenciáját!

Megoldás

$H_1 \setminus H_2$ = **4, 6, 12, 5, 98, 20**

$H_1 \Delta H_2 = (H_1 \setminus H_2) \cup (H_2 \setminus H_1)$ = (4, 6, 12, 5, 98, 20) $\cup$ (50, 31, 67, 9, 54, 55, 16) = **4, 6, 12, 5, 98, 20, 50, 31, 67, 9, 54, 55, 16**

Feladat

Legyen H_1 és H_2 halmazunk. Határozza meg uniójukat, metszetüket, különbségüket, szimmetrikus differenciájukat és Descartes-szorzatukat. Képezzen belőlük legalább három-három részhalmazt!

H_1 halmaz elemei:

Pécs, Debrecen, Békés, Eger, Győr, Szeged, Pest

H_2 halmaz elemei:

Vas, Zala, Baranya, Pécs, Pest, Békés, Győr

3. Valószínűségszámítás

A valószínűség fogalmának bevezetése

Minden véletlen eseménynek több, esetleg végtelen sok kimenetele lehetséges.

Egy esemény relatív gyakorisága az a szám, amely megmutatja, hogy az összes megfigyelt esemény mekkora hányadában következik be a kívánt esemény. Ha a kísérletek száma *n* és a kívánt esemény *k*-szor következik be, akkor a relatív gyakoriság:

$$0 \le k/n \le 1$$

Az a számérték, amely körül valamely – azonos körülmények között – vizsgált véletlen esemény (A) relatív gyakorisága ingadozik az illető esemény valószínűsége P(A).

Valószínűségszámítás axiomatikus megalapozása

a) Minden egyes véletlen eseményhez (**A**) hozzárendelhető egy nem negatív szám (P(**A**)), amelyre teljesül:

$$0 \le P(\mathbf{A}) \le 1$$

b) Biztos esemény valószínűsége 1, a lehetetlen esemény valószínűsége 0.

c) Ha két esemény kizárja egymást ($\mathbf{A} \cdot \mathbf{B} = \emptyset$) , akkor :

$$P(\mathbf{A} + \mathbf{B}) = P(\mathbf{A}) + P(\mathbf{B})$$

d) Ha az $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{A}_3$, …..$\mathbf{A}_n$ események páronként kizárják egymást (azaz $\mathbf{A}_i \cdot \mathbf{A}_j = \emptyset$, ha $i \neq j$), akkor:

$$P(\mathbf{A}_1 + \mathbf{A}_2 + \mathbf{A}_3 \ldots. + \mathbf{A}_n) = P(\mathbf{A}_1) + P(\mathbf{A}_2) + P(\mathbf{A}_3) \ldots. + P(\mathbf{A}_2)$$

Következmények:

- Teljes eseményrendszer esetén (az $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{A}_3$, …..$\mathbf{A}_n$ események páronként kizárják egymást, de valamelyik biztosan bekövetkezik):

$$P(\mathbf{A}_1 + \mathbf{A}_2 + \mathbf{A}_3 \ldots. + \mathbf{A}_n) = P(\mathbf{A}_1) + P(\mathbf{A}_2) + P(\mathbf{A}_3) \ldots. + P(\mathbf{A}_2) = 1$$

- Egymást nem kizáró események esetén:

$$P(\mathbf{A} + \mathbf{B}) = P(\mathbf{A}) + P(\mathbf{B}) - P(\mathbf{A} \cdot \mathbf{B})$$

Feltételes valószínűség fogalma:

P(A/B) : az A esemény bekövetkeztének valószínűsége, feltéve, hogy a B esemény bekövetkezik.

$$P(A|B) = \frac{P(A \cdot B)}{P(B)}$$

Ha A és B események függetlenek akkor P (A · B) = P (A) · P(B), azaz P(A/B) = P(A).

Teljes valószínűség tétele:

Legyen a B_1, B_2, B_3, ...B_n teljes eseményrendszer, és A egy tetszőleges esemény, akkor :

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$

Bayes tétele:

Ha a B_1, B_2, B_3, ...B_n események teljes eseményrendszert alkotnak, és A egy tetszőleges, pozitív valószínűségű esemény, akkor

$$P(B_k|A) = \frac{P(A|B_k)P(B_k)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)}$$

Feladatok

3.1. Öt kockával dobunk, mi annak a valószínűsége, hogy a dobások összege kisebb, vagy egyenlő 8-al?

Megoldás

Az esemény bekövetkezésének valószínűségét megkapjuk, ha a kedvező esetek számát elosztjuk az összes lehetséges eset számával.

A kedvező lehetőségek a következők:

öt darab 1-es	(egyféleképpen lehetséges)
négy darab 1-es és egy darab 2 -es	(ötféleképpen lehetséges)
három darab 1-es és két darab 2-es	(tízféleképpen lehetséges)
két darab 1-es és három darab 2-es	(tízféleképpen lehetséges)
négy darab 1-es és egy darab 3-as	(ötféleképpen lehetséges)

Ez összesen 31 esetet jelent. Az összes lehetséges esetek száma: 6^5.
A keresett esemény valószínűsége:

$$P = \frac{31}{6^5}$$

3.2. Öt kockával dobunk, mi annak a valószínűsége, hogy pontosan 3 hatost dobunk? **($250/6^5$)**

3.3. Mi a valószínűsége annak, hogy a LOTTÓ húzáson 2 vagy 3 találatunk lesz? **(0,023)**

3.4. Három dobozban vannak sárga, piros és fekete golyók. Az első dobozban 3 sárga 4 fekete és 5 piros, a második dobozban 4 sárga, 2 fekete és 3 piros, a harmadikban 5 sárga 5 piros és 5 fekete golyó van.

a) Mi a valószínűsége annak, hogy véletlen szerűen kiválasztva valamelyik dobozt 1 sárga, 1 fekete és 1 piros golyót húzunk ki? A kihúzás után a golyókat nem rakjuk vissza.

Megoldás

Használjuk fel a teljes valószínűség tételét! Legyen a $\mathbf{B_i}$ esemény az, hogy valamelyik dobozt választjuk! Mivel a dobozokat egyforma valószínűséggel választjuk:

$$P(\mathbf{B_1}) = P(\mathbf{B_2}) = P(\mathbf{B_3}) = 1/3$$

Legyen az **A** esemény, hogy a feltételnek megfelelően 1 sárga, 1 fekete és 1 piros golyót húzunk!

$$P(A|B_1) = \frac{3}{12}\cdot\frac{4}{11}\cdot\frac{5}{10} = 0,0545$$

$$P(A|B_2) = \frac{4}{9}\cdot\frac{2}{8}\cdot\frac{3}{7} = 0,0476$$

$$P(A|B_3) = \frac{5}{15}\cdot\frac{5}{14}\cdot\frac{5}{13} = 0,0458$$

Ezek után $P(\mathbf{A}) = 0,333\cdot 0,0545 + 0,333\cdot 0,0476 + 0,333\cdot 0,0458 = 0,0493$

b) Mi lesz a megoldás akkor, ha az első dobozt 1/3, a második dobozt 1/2 és harmadik dobozt 1/6 valószínűséggel választjuk? **(0,0496)**

3.5. Tegyük fel, hogy véletlenszerűen választva dobozok között egy sárga golyót húzunk ki. Mi a valószínűsége annak, hogy ezt az első dobozból húztuk ki? A dobozok kiválasztásának valószínűségét megadtuk az előző feladatban.

Megoldás

A megoldáshoz használjuk fel Bayes tételét!

$$P(B_1|A) = \frac{P(A|B_1)P(B_1)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)} = \frac{1/3\cdot 0,0545}{0,0496} = 0,366$$

3.6. Egy dobozban 3 zöld és 5 kék golyó van. Két golyót húzunk ki a dobozból egymás után. Mi a valószínűsége annak, hogy a második golyó zöld, ha az első kék volt?

Megoldás

Legyen az **A** esemény az, hogy zöld golyót húzunk és a **B** esemény az, hogy kéket. A feltételes valószínűség a következő módon számítható ki:

$$P(A|B) = \frac{P(A \cdot B)}{P(B)} = \frac{\frac{3}{8} \cdot \frac{5}{7}}{\frac{5}{8}} = \frac{3}{7}$$

3.7. Egy dobozban 3 sárga 4 fehér és 5 piros golyó van. Egymás után, visszatevés nélkül két golyót húzunk ki a dobozból.

a) Mi a valószínűsége annak, hogy két egyforma színű golyót húzunk ki? **(0,29)**

b) Mi a valószínűsége annak, hogy a két kihúzott golyó színes? **(0,42)**

3.8. Két kockával dobunk. Mennyi a valószínűsége annak, hogy a két kocka valamelyike hatost mutat, feltéve, hogy a dobott számok összege kisebb, vagy egyenlő 9 -el?

Megoldás

Legyen az **A** esemény az, hogy a két kocka valamelyike hatost mutat, és legyen a **B** esemény, hogy a számok összege kisebb, mint 9. A feltételes valószínűség a következő módon számítható ki:

$$P(A|B) = \frac{P(A \cdot B)}{P(B)}$$

Először számítsuk ki, a P(**A**·**B**) valószínűséget. Két kockával dobva az összes lehetséges száma: 36. Azoknak a dobásoknak a száma, amelyek eleget tesznek mind az **A** mind a **B** feltételnek 4 (1,6; 6,1; 2,6 és 6,2). Így

$$P(A \cdot B) = \frac{4}{36}$$

A B esemény (a dobások összege kisebb, mint 9) 25-féleképpen lehetséges. Így:

$$P(B) = \frac{25}{36}$$

Tehát a keresett valószínűség:

$$P(A|B) = \frac{\frac{4}{36}}{\frac{25}{36}} = \frac{4}{25}$$

3.1 Nevezetes eloszlások

I. Binomiális eloszlás

Adva van egy *p* valószínűségű alternatív esemény (vagy az esemény, vagy az ellentettje következik be). Keressük annak valószínűségét, hogy *n*-szer megismételve a kísérletet az esemény *k*-szor következik be.

Annak valószínűsége, hogy egy *p* valószínűségű esemény *k*-szor bekövetkezik p^k. Annak valószínűsége, hogy a fennmaradó *n-k* esetben nem következik be az esemény $(1-p)^{n-k}$. Végül figyelembe véve, hogy a kedvező esetek száma annyi lehet ahányféleképpen *n*-ből kiválaszthatunk *k*-t, megkapjuk a keresett valószínűséget:

$$P(k,n) = \frac{n!}{k! \cdot (n-k)!} p^k (1-p)^{n-k}$$

.

Feladatok

3.9. Júliusban a Balatonnál a csapadékos napok előfordulásának valószínűsége 0,1. Mi a valószínűsége annak, hogy az egyhetes nyaralás során 3 nap esni fog az eső? Hány csapadékos napra számíthatunk a legnagyobb valószínűséggel, azaz milyen *k* értéknél veszi fel a maximumát a binomiális eloszlás? Hogyan változik a megoldás, ha p = 0,3 ?

Megoldás

Az adatok alapján p = 0,1, n = 7 és k = 3. Így a keresett valószínűség:

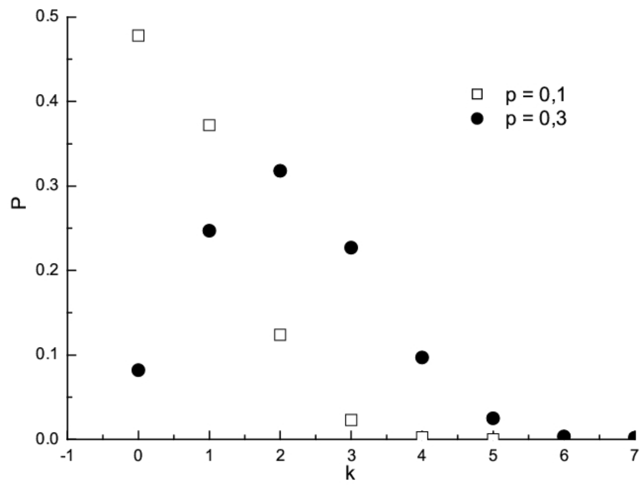
$$P(3,7) = \frac{7!}{3! \cdot 4!} 0,1^3 \cdot 0,9^4 = 0,023$$

Az eloszlás maximumának meghatározásához számítsuk ki a binomiális eloszlás értékét

k = 0,1,2,3,4,5,6 és 7 esetén. A kapott értékeket az alábbi táblázat tartalmazza.

	k=0	**k=1**	**k=2**	**k=3**	**k=4**	**k=5**	**k=6**	**k=7**
p = 0,1	0,478	0,372	0,124	0,023	0,0025	0,0002	0	0
p = 0,3	0,082	0,247	0,318	0,227	0,097	0,025	0,0035	0,0002

3.1 táblázat.
Binomiális eloszlás értékei különböző *p* valószínűségek esetén.



3.1 ábra.
Binomiális eloszlás különböző *p* valószínűségek esetén.

Az adatok alapján látható, hogy ha a csapadékos napok valószínűsége 0,1, akkor a legnagyobb valószínűsége annak van, hogy egy hét alatt egyszer sem fog esni az eső. Ha a csapadékos napok valószínűsége 0,3, akkor a legnagyobb valószínűsége annak van, hogy két csapadékos napunk lesz az egy hét folyamán.

3.10. Az elmúlt száz évben Pécsett a januári középhőmérséklet 35 esetben volt pozitív. Mi a valószínűsége annak, hogy a következő 10 évben ez 3-szor fog előfordulni?

Megoldás

Az adatok alapján az alternatív esemény bekövetkeztének valószínűsége p = 0,35, a kísérletek száma n = 10 és a kedvező esetek száma k = 3. Ezek alapján a binomiális eloszlás értéke a következő lesz:

$$P(3,10) = \frac{10!}{3! \cdot 7!} 0,35^3 \cdot 0,65^7 = 0,252$$

II. Geometriai eloszlás

Hasonlóan a binomiális eloszláshoz adva van egy *p* valószínűségű alternatív esemény. A kísérleteket egymás után végre hajtva mi a valószínűsége annak, hogy a kívánt esemény először a *k*-ik kísérlet során következik be?

Annak a valószínűsége, hogy a kísérlet ellentettje egymás után *k–1* -szer következik be:

$$(1-p)^{k-1}.$$

Ezt megszorozva az esemény bekövetkezésének *p* valószínűségével megkapjuk annak valószínűségét, hogy a kívánt esemény pont a *k*-ik kísérlet során következik be:

$$P_k = (1-p)^{k-1}\, p.$$

Feladatok

3.1. Egy folyó vízszintje átlagosan tízévente emelkedik 8 m fölé. Mi a valószínűsége annak, hogy ez pontosan 5 év múlva fog először bekövetkezni?

Megoldás

Az adatok alapján az esemény bekövetkeztének valószínűsége p = 0,1, továbbá a kísérletek száma k = 5. A keresett valószínűség:

$$P_5 = 0{,}9^4 \cdot 0{,}1 = 0{,}066.$$

III. Poisson –féle eloszlás

A Poisson-féle eloszlás a természetben előforduló véletlenszerű események egyik leggyakoribb jellemzője. Gyakran használják az ún. ritka események eloszlása elnevezést is. Néhány példa a Poisson-féle eloszlásra: az egy adott szögtartományban megszámolható csillagok száma, az egy évben született hármas ikrek száma, az augusztusi éjszakán hulló csillagok száma, vagy a radioaktív bomlások száma. A Poisson –féle eloszlás az alábbi összefüggéssel számolható ki:

$$P(k,\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$

ahol λ = n · p és *e* az ún. természetes alap, amelynek értéke 2,718. *p* a kívánt esemény bekövetkeztének valószínűsége, *n* a kísérletek száma, rendszerint nagy szám, *k* azon kísérletek száma, amikor a számunkra kedvező esemény következik be. (Be lehet bizonyítani, hogy ha *n* értéke nagy és a *p* értéke kicsi, akkor a Poisson-féle eloszlás jó közelítéssel megegyezik a binomiális eloszlással.)

Feladatok

3.12. Intenzív zivatarok esetén átlagosan 10 másodpercenként észlelhető egy villámlás. Poisson-féle eloszlást feltételezve mi annak a valószínűsége, hogy 15 másodperc alatt két villámlást észlelünk?

Megoldás

Az adatok alapján p = 0,1, n = 15, valamint λ = 1,5 és k = 2. A keresett valószínűség:

$$P = \frac{1{,}5^2 e^{-1{,}5}}{2!} = 0{,}25$$

3.13. Intenzív zivatarok esetén átlagosan 10 másodpercenként észlelhető egy villámlás. Poisson eloszlást feltételezve mi annak a valószínűsége, hogy 15 másodperc alatt kettőnél több villámlást észlelünk?

Megoldás

A kívánt esemény akkor következik be, ha az észlelt villámlások száma vagy 3, vagy 4 stb. Könnyű belátni, hogy a villámlás 0-szor, 1-szer, 2-szer stb. történő bekövetkezése teljes esemény rendszer (valamelyik biztosan bekövetkezik, és az egyik kizárja a másikat). Így felírhatjuk, hogy:

$$P(0) + P(1) + P(2) + P(3) + \ldots + P(n) \ldots = 1$$

Azaz a keresett valószínűség:

$$P(3) + P(4) + P(5) + \ldots + P(n) \ldots = 1 - P(0) + P(1) + P(2)\ .$$

A feladatot meg tudjuk oldani, ha k = 0, 1 és 2 értékhez tartozó valószínűségeket kiszámoljuk. Az előző feladat alapján P(k=0) = 0,223 (fontos tudni hogy $0! \equiv 1$), P (k=1) = 0,335 és P(k=2) = 0,25. A keresett valószínűség:

$$1 - 0{,}223 - 0{,}335 - 0{,}25 = 0{,}192.$$

3.14. Egy országban az elmúlt száz évben 500 alkalommal történt hármas ikerszülés. Mi a valószínűsége annak, hogy egy évben ez az esemény 2-szer fordul elő.

Megoldás

Az eloszlás λ paraméterének értéke 5, a k = 2. Így a megoldás:

$$P = \frac{5^2 e^{-5}}{2!} = 0{,}08$$

4. Észlelési sorok matematikai statisztikai jellemszámai

Hajtsunk végre egy kísérletet n-szer, és legyen a kísérlet kimenetele rendre: $x_1, x_2, x_3, \ldots x_n$. Az adatokat az alábbi módon meghatározott átlaggal és szórással jellemezhetjük. A szórás azt fejezi ki, hogy a megfigyelési adataink milyen mértékben térnek el az átlagtól.

I. Átlag

$$M = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n} = \frac{1}{n}\sum_{k=1}^{n} x_k$$

II. Szórás (korrigált)

$$\sigma = \sqrt{\frac{(x_1 - M)^2 + (x_2 - M)^2 + \ldots + (x_n - M)^2}{n-1}}$$

Feladatok

4.1. Dobjon egy kockával 10-szer! Jegyezze fel a dobások eredményét! Határozza meg az átlagot és a szórást! Mekkora a 6-os dobások relatív gyakorisága? Ismételje meg a kísérletet úgy, hogy most 20-szor dob! Válaszoljon ismét a fentiekben feltett kérdésekre!

4.2. Az alábbi táblázat egy hónap napi átlaghőmérsékleteit tartalmazza. Az adatok alapján határozza meg a havi átlaghőmérsékletet és a szórást!

nap	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
t(°C)	10,2	11,1	13,2	12,4	13,0	14,5	15,8	15,4	15,8	16,0	13,9	14,2	14,5	14,8	14,1
nap	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
t(°C)	13,8	13,9	13,6	12,7	13,0	14,0	13,8	14,4	15,0	15,1	13,8	14,2	14,1	14,6	14,5

Megoldás

A kísérlet kimenetelei az egyes napokhoz tartozó átlaghőmérsékletek. A kísérletek száma 30. Először határozzuk meg az átlagértéket! Az átlag: 14,0 °C. Az átlag ismeretében kiszámolható szórás is: 1,3 °C.

III. Súlyozott átlag

Az átlag számítása során gyakran előfordul, hogy a rendelkezésre álló adatokat nem azonos mértékben kell figyelembe venni. Ennek több oka is lehet. Előfordulhat, hogy vannak olyan adatok, amelyeket megbízhatóbbnak tartunk, mint más adatokat. Az is lehetséges, hogy bizonyos adatok nagyobb halmazt reprezentálnak, mint más adatok. Ezekben az esetekben hibás eredményt

kapnánk, ha minden adatot egyformán vennénk figyelembe. A súlyozott átlagot az alábbi összefüggéssel számoljuk ki:

$$M_S = \frac{a_1 x_1 + a_2 x_2 + \ldots + a_n x_n}{A} = \frac{1}{A}\sum_{k=1}^{n} a_k x_k \,,$$

ahol:

$$A = a_1 + a_2 + \ldots + a_n = \sum_{k=1}^{n} a_k$$

és a_1, a_2 , … a_n az ún. súlytényezők.

Feladatok

4.3. Egy négyzet alakú ország területe 3000 km^2. Az ország területének fele sík vidék, területének 1/3 dombos és 1/6 hegyvidéki. Mindegyik tájkörzetben található egy csapadékmérő állomás. A síkvidéki állomáson mért éves csapadékösszeg 300 mm, a dombvidéken elhelyezett állomáson egy év alatt 400 mm-t, a hegyvidéken lévő állomáson egy év alatt 600 mm csapadékot mértek. Mennyi csapadék hullik az országra átlagosan egy év alatt?

Megoldás

Az természetesen nem vezet a helyes eredményre, ha a három mért értéknek vesszük az átlagát. Ugyanis az egyes állomások más-más nagyságú terület reprezentálnak. Legyenek a súlytényezők az egyes állomáshoz tartozó területek, azaz a_1 = 1500, a_2 = 1000 és a_3 = 500. A területek alapján súlyozott átlagot a következő módon számíthatjuk ki:

$$M_S = \frac{1500 \cdot 300 + 1000 \cdot 400 + 500 \cdot 600}{3000} = 383$$

Tehát az ország 1 m^2 –nyi felületére átlagosan 383 mm csapadék hullik egy év alatt.

VI. Középérték meghatározása osztályokba sorolt adatokból

Sokszor találkozunk azzal, hogy a rendelkezésre álló adatok már eleve csak csoportokba soroltan érhetők el. Ebben az esetben az alapvető statisztikai mutatók kissé körülményesebben ugyan, de kielégítő pontossággal meghatározhatóak. Erre különösen akkor lehet szükség, ha azcsoportosított adatokat más, nem csoportba sorolt adatokkal kell összevetnünk.

Feladat

Határozzuk meg az alábbi táblázat adataiból, hogy az 1896-tól 1995-ig terjedő időszakot vizsgálva mennyi volt Budapesten a januári középhőmérséklet!

Csoportok	1896-1995
-8,9 – -8,0	1
-7,9 – -7,0	0
-6,9 – -6,0	1
-5,9 – -5,0	4
-4,9 – -4,0	2
-3,9 – -3,0	9
-2,9 – -2,0	12
-1,9 – -1,0	14
-0,9 – 0,0	8
0,1 – 1,0	18
1,1 – 2,0	13
2,1 – 3,0	7
3,1 – 4,0	5
4,1 – 5,0	5
5,1 – 6,0	0
6,1 – 7,0	1

Megoldás

Használjuk fel az alábbi egyenletet:

$$\overline{x} = \frac{\frac{x_a+x_f}{2} + \frac{y_a+y_f}{2} + \cdots + \frac{n_a+n_f}{2}}{N}$$

x_a, y_a, n_a – az osztályköz alsó határa

x_f, y_f, n_f – az osztályköz felső határa

N – a sokaság elemszáma

Vagyis meg kell határoznunk az elemszámot, amit az egyes osztályok gyakoriságainak összegeként kapunk meg. Jelen esetben ez 100 (mint az várható is volt, hiszen 100 év adatairól van szó).

Most képezzük egyenként az osztályok alsó- és felső széleinek középértékét, ahogyan az az egyenlet számlálójában szerepel: n_a+n_f/2. Így az alábbit írhatjuk fel:

((-8,9+$^{-}$8)/2)+((-7,9+$^{-}$7)/2)+((-6,9+$^{-}$6)/2)+ . . . +((6,1+7)/2)=$^{-}$**15,2**

Mivel -15,2/100=**-0,152**

ezért azt mondhatjuk, hogy Budapesten az 1896 és 1995 közötti 100 évben adatai alapján a januári középhőmérséklet -0,152 °C.

Érdemes megjegyezni, hogy a csoportosítatlan adatokból -0,36 °C adódik, mint középérték, vagyis az elövetett hiba nem számottevő, mindössze 0,21 °C.

Feladat

Az alábbi táblázat a Budapesten 1896 és 1995 között mért éves középhőmérsékletek megoszlását tartalmazza. Határozza meg az átlagukat!

Hőmérsékleti osztály, °C	Gyakoriság
8 – 9	0
9,1 – 10	5
10,1 – 11	35
11,1 – 12	53
12,1 – 13	7

Megoldás

Az átlag **11,18 °C.**

(Az összes, nem osztályokba sorolt adatból számolva 11,15°C)

V. Empirikus eloszlásfüggvény

Eloszlásfüggvény szemléletes jelentése:

Egy adott x_a értékhez tartozó $F(x_a)$ függvényérték azt fejezi ki, hogy milyen valószínűséggel fordulnak elő az x_a-nál kisebb, vagy azzal egyenlő kísérleti eredmények, azaz:

$$F(x_a) = P\ (x \leq x_a)$$

Az eloszlásfüggvény fontosabb tulajdonságai:

a) A függvény monoton növekvő, azaz $f(x_1) \leq f(x_2)$ ha $x_1 < x_2$

b) A függvény minimuma 0, maximuma 1.

Empirikus eloszlásfüggvény készítése:

i. Az adatokat nagyság szerint sorba rendezzük

ii. A fentiekben megadott valószínűséget a relatív gyakorisággal közelítjük, azaz az eloszlásfüggvény értéke az x_a-nál:

$$F(x_a) = \frac{k}{n}$$

, ahol *n* az összes adat, k azon adatok száma, amelyekre teljesül, hogy $x < x_a$.

IV. Empirikus sűrűségfüggvény (hisztogram)

Sűrűségfüggvény szemléletes jelentése:

A görbe alatti terület az x_1 és az x_2 pontok között annak valószínűségét fejezi ki, hogy a kísérlet eredménye az x_1 és az x_2 értékek közé esik, azaz:

$$\int_{x_1}^{x_2} f(x)dx = P(x_1 \le x < x_2)$$

Sűrűségfüggvény fontosabb tulajdonságai:

a) A függvény minden pontjára igaz, hogy $f(x) \ge 0$.

b) A görbe alatti terület egységnyi.

Empirikus sűrűségfüggvény (hisztogram) készítése:

i. Az adatok nagyság szerinti sorba rendezése

ii. Gyakorisági intervallumok meghatározása (*Fontos, hogy egy-egy intervallumba megfelelő számú eset kerüljön!*)

iii. Az intervallumokba eső adatok számának meghatározása

iv. A sűrűségfüggvény értékének meghatározása (a sűrűségfüggvény alatti területet téglalapokkal adjuk meg):

$$f(x_i) = \frac{k_i}{n \cdot \Delta_i}$$

ahol *n* az adatok száma, k_i az i-ik intervallumba eső esetek száma, Δ_i az i-ik intervallum szélessége, $f(x_i)$ az i-ik intervallumhoz tartozó téglalap magassága.

V. Medián

Medián alatt azt az értéket (x_m) értjük, amelynél kisebb és nagyobb adatok száma megegyezik, azaz ahol az eloszlásfüggvény értéke pontosan 0,5- el egyenlő:

$$F(x_m) = \frac{1}{2}$$

Nagyság szerint sorba rendezett minta esetén a medián pontosan a középső adattal lesz egyenlő, ha az adatok száma páratlan. Ha az adatok száma páros (összes esetszám *n*) a medián a két középső adat számtani közepével lesz egyenlő:

$$x_m = x_{\frac{n+1}{2}} \text{ , ha n páratlan és } x_m = 0,5\left(x_{\frac{n}{2}+1} + x_{\frac{n}{2}}\right) \text{ ha } n \text{ páros}$$

VI. Módusz

A módusz azon értékek, ahol a sűrűségfüggvény lokálisan felveszi a maximumát (akár több maximum is lehetséges). Tehát a módusz a nagyobb gyakori-

sággal előforduló adatokkal egyezik meg. Ha csak egy maximum van, akkor a módusz a legnagyobb gyakorisággal előforduló adattal egyenlő. ***Fontos: a módusz nem feltétlen egyezik meg az átlaggal, azaz nem mindig a leggyakoribb érték az átlagos.***

Feladatok

4.4. Az alábbi táblázatban olvasható hőmérséklet értékekből készítsen empirikus eloszlásfüggvényt és empirikus sűrűségfüggvényt! Határozza meg az eloszlás mediánját és móduszát is!

Maximum hőmérséklet Budapesten
július 1-én

év	t_{max}	év	t_{max}	év	t_{max}
1901	27,2	1921	28,3	1941	24,7
1902	32	1922	22,5	1942	28,3
1903	25,5	1923	25,5	1943	17,8
1904	27,3	1924	30,2	1944	25
1905	32,1	1925	22,6	1945	25,4
1906	17,8	1926	18,2	1946	32,8
1907	32,5	1927	33,7	1947	34,1
1908	26,5	1928	31	1948	17
1909	19,2	1929	25,9	1949	21,8
1910	29,9	1930	32,8	1950	38,3
1911	24,1	1931	30,8	1951	23,1
1912	28,9	1932	31,8	1952	32,4
1913	14,4	1933	22,9	1953	28,6
1914	21,3	1934	25,1	1954	31,4
1915	20,8	1935	30,3	1955	25,3
1916	28	1936	30,2	1956	28,6
1917	31,3	1937	21,8	1957	33
1918	21,2	1938	32,9	1958	26,7
1919	21,8	1939	32,8	1959	21,2
1920	31,1	1940	27,8	1960	17,3

Megoldás

a) Empirikus eloszlásfüggvény meghatározása

Írjuk le növekvő sorrendbe a hőmérsékleti értékeket:

14,4; 17,0; 17,3; 17,8; 17,8; 18,2; 19,2; 20,8; 21,2; 21,2; 21,3; 21,8; 21,8; 21,8; 22,5; 22,6; 22,9; 23,1; 24,1; 24,7; 25,0; 25,1; 25,3; 25,4; 25,5; 25,5; 25,9; 26,5; 26,7; 27,2; 27,3; 27,8;

28,0; 28,3; 28,3; 28,6; 28,6; 28,9; 29,9; 30,2; 30,2; 30,3; 30,8; 31,0; 31,1; 31,3; 31,4; 31,8;

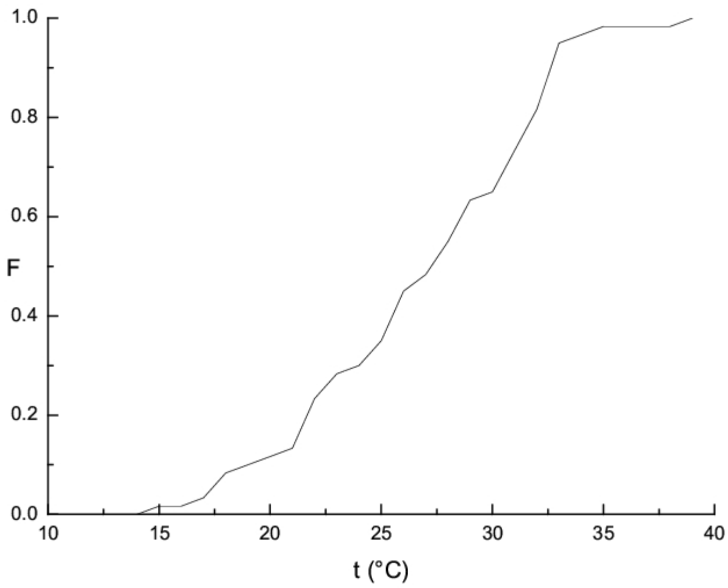
32,0; 32,1; 32,4; 32,5; 32,8; 32,8; 32,8; 32,9; 33,0; 33,7; 34,1; 38,3.

A sorba rendezés után határozzuk meg az eloszlásfüggvény értékét 14 és 39 °C között, 1 °C-onként. (Ennél lehet finomabb és durvább beosztást is választani!).

Mivel a 14°C-nál kisebb, vagy az azzal egyenlő értékek száma nulla, így az x = 14–nél a függvény értéke is 0 lesz. 15°C–nál kisebb, vagy az azzal egyenlő értékek száma 1, így a függvény értéke x = 15–nél 1/60 lesz. A számításokat hasonló módon folytathatjuk egészen 39°C-ig. Az ennél az értéknél kisebb, vagy egyenlő esetek száma 60, így az eloszlásfüggvény értéke ebben a pontban 60/60 = 1 lesz. A számítási eredményeket az alábbi táblázatban foglaljuk össze.

x	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**
F(x)	0	1/60	1/60	2/60	5/60	6/60	7/60	8/60	14/60	17/60	18/60	21/60	27/60

x	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**	**37**	**38**	**39**
F(x)	29/60	33/60	38/60	39/60	44/60	49/60	57/60	58/60	59/60	59/60	59/60	59/60	60/60



4.1 ábra.
Az empirikus eloszlásfüggvény grafikus ábrázolása.

b) Empirikus sűrűségfüggvény meghatározása

A számítások elvégzése során használjuk fel az előzőekben már nagyság szerint sorba rendezett adathalmazt.

Határozzuk meg az ábrázoláshoz szükséges intervallum szélességeket:

Annak bemutatására, hogy az eredmény mennyire függ az intervallumok szélességétől, végezzük el a számításokat a Δ = 1°C és a Δ = 5 °C-os beosztásra is. (Most az egyszerűség kedvéért feltételeztük, hogy minden intervallum szélessége egyforma.)

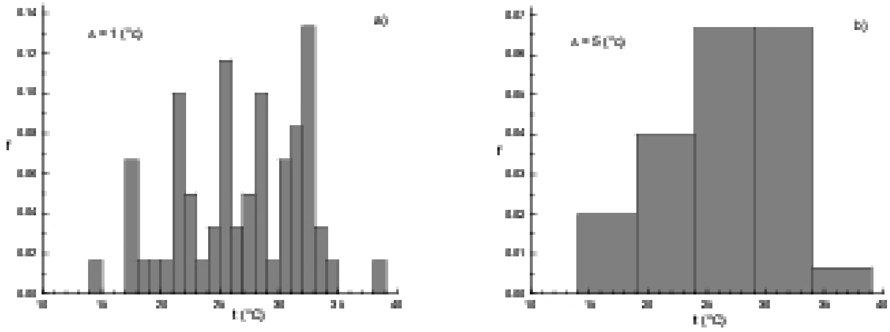
Az első intervallum a $\Delta = 1°C$ –os beosztás esetén 14 –től 14,9-ig terjed. Ebben az intervallumban egy adat található, így az intervallumhoz tartozó téglalap magassága 1/60/1 lesz. A második, 15-től 15,9-ig terjedő intervallumban nincsen adat, így a téglalap magassága 0 lesz. Hasonló lesz a magassága a 16-től 16,9 terjedő intervallumhoz tartozó téglalapnak is. A 17-től 17,9-ig terjedő intervallumban 4 adat található, így a téglalap magassága 4/60/1 lesz. A számításokat hasonló módon folyathatjuk egészen a 38-tól 38,9-ig terjedő intervallumig, amelyben egy adat található. A számítási eredményeket az alábbi táblázatban foglaljuk össze.

intervallum	k_i	f_i	**intervallum**	k_i	f_i
14 - 14,9	1	1/60	27 - 27,9	3	1/20
15 - 15,9	0	0	28 - 28,9	6	1/10
16 - 16,9	0	0	29 - 29,9	1	1/60
17 - 17,9	4	1/15	30 - 30,9	4	1/15
18 - 18,9	1	1/60	31 - 31,9	5	1/12
19 - 19,9	1	1/60	32 - 32,9	8	2/15
20 - 20,9	1	1/60	33 - 33,9	2	1/30
21 - 21,9	6	1/10	34 - 34,9	1	1/60
22 - 22,9	3	1/20	35 - 35,9	0	0
23 - 23,9	1	1/60	36 - 36,9	0	0
24 - 24,9	2	1/30	37 - 37,9	0	0
25 - 25,9	7	7/60	38 - 38,9	1	1/60
26 - 26,9	2	1/30	39 - 39,9	0	0

Mint ahogy a fenti táblázatban is látszik, az egy-egy intervallumba jutó adatok száma erősen ingadozik, és sok intervallumba csak 1 vagy 0 adat jut. Ez azzal magyarázható, hogy túlságosan keskenyre választottuk az intervallumok szélességét. Ismételjük meg a számításokat a $\Delta = 5$ °C-os beosztásra is. Most az első intervallum 14-től 18,9 –ig terjed. Ebbe az intervallumba 6 adat jut, így az intervallumhoz tartozó téglalap magassága 6/60/5. A számítási eredmények összefoglalása az alábbi táblázatban olvasható:

intervallum	k_i	f_i	**intervallum**	k_i	f_i
14 - 18,9	6	1/50	29 - 33,9	20	1/15
19 - 23,9	12	1/25	34 - 38,9	2	1/150
24 - 28,9	20	1/15	39 - 43,9	0	0

Látható, hogy a $\Delta = 5$ °C-os beosztás esetén az egy-egy intervallumba jutó esetek száma már megfelelő, és a hisztogram is sokkal simább lefutású.



4.2 ábra.
Hisztogram Δ = 1 °C-os (a) és Δ = 5 °C (b) beosztás esetén.

c) Medián meghatározása

Mivel az esetek száma páros a medián a sorba rendezett adatsor 30. és 31. elemének számtani közepével lesz egyenlő, azaz:

$$x_m = 0{,}5 \cdot (27{,}2 + 27{,}3) = 27{,}25 .$$

d) Módusz meghatározása

A módusz értéke függ az intervallum beosztástól:

Δ = 1°C–os beosztás esetén több helyen van lokális maximuma van a hisztogramnak. Ezek a helyek rendre a következők: 17,5; 21,5; 25,5; 28,5; 32,5 és 38,5. (A maximum helyeként az intervallumok közepét adhatjuk meg.)

Δ = 5°C–os beosztás esetén két intervallumhoz rendelhető lokális maximum, a t = 26,5 és a t = 31,5 helyeken.

4.5. Bizonyítsa be, hogy a 4.4. feladatban meghatározott hisztogramok esetében a téglalapok összterülete egységnyi!

4.6. A 4.4. feladatban meghatározott eloszlásfüggvény segítségével adja meg annak valószínűségét, hogy a maximum hőmérséklet 25°C alatt marad! **(1/3)**

4.7. A 4.4. feladatban meghatározott hisztogram segítségével határozza meg annak valószínűségét, hogy a hőmérséklet 24 és 30°C között lesz! A számításokat mind a két beosztással kapott hisztogram felhasználásával végezze el!

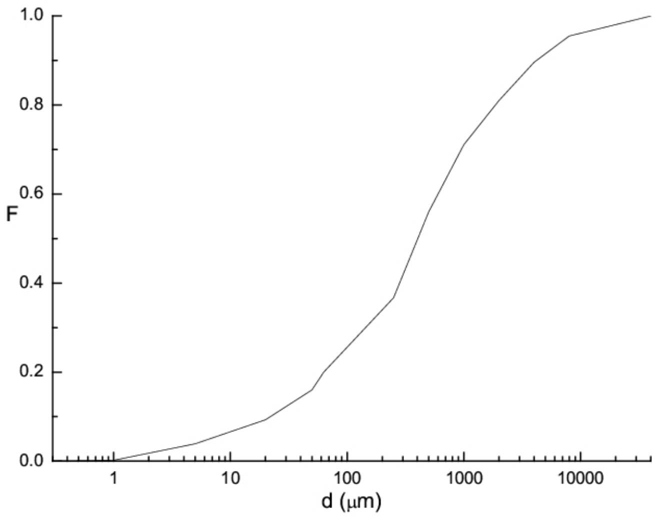
4.8. Talajfizikai laboratóriumban végzett mérés során 15 különböző lyukátmérőjű szitát használtunk egy talaj típus vizsgálatára. A mérés során (lásd alábbi táblázat) meghatároztuk, hogy mekkora tömegű anyag marad fent egy-egy szitán. A táblázatban közölt adatok alapján készítsen empirikus eloszlásfüggvényt! Határozza meg az eloszlás mediánját!

Szitalyuk átmérő (μm)	Szitán fennmaradt tömeg (g)
40000	0
8000	36
4000	47
2000	68
1000	79
500	120
250	154
100	88
63	45
50	32
20	53
5	43
2	17
1	12
0.3	2
0	0

Megoldás

a) Empirikus eloszlásfüggvény készítése. Fontos észrevenni, hogy ebben a feladatban darab szám helyett tömegeloszlást kell vizsgálni. Ehhez először számoljuk ki a teljes tömeget. Összeadva a második oszlopban lévő számokat a teljes tömeg 796 g-mal egyenlő. Ezek után határozzuk meg, hogy a teljes tömeg hányadrésze hullott át az adott lyukátmérőjű szitán. Az alábbi táblázat tartalmazza azt a tömeget, ami az adott lyukátmérőjű szitán áthullott, valamint a számított tömegarányt. A tömegarány megfelel az általunk keresett eloszlásfüggvénynek. Az eloszlás mediánja kb. 500 μm, mivel az eloszlásfüggvény ennél az értéknél egyenlő 0,5-del.

Szitalyuk átmérő (μm)	Szitán áthullott tömeg (g)	tömegarány
40000	796	1,000
8000	760	0,955.
4000	713	0,896
2000	645	0,810
1000	566	0,711
500	446	0,560
250	292	0,367
100	204	0,256
63	159	0,200
50	127	0,160
20	74	0,093
5	31	0,039
2	14	0,018
1	2	0,002
0.3	0	0,000
0	0	0,000



4.3 ábra.
Eloszlásfüggvény. A vízszintes tengelyen a *d* átmérőt logaritmikus skálán adtuk meg, mivel több, mint három nagyságrendnyi tartományt kell ábrázolni.

VI. Osztályokba sorolt adatok módusza

Feladat

Az alábbi táblázatban egy észak-alföldi meteorológiai állomáson, az 1955-től 2003-ig mért január, február és március havi csapadékértékek eloszlása olvasható.

Becsüljük meg az adatsor leggyakoribb értékét!

Havi csapadékösszeg, mm	Gyakoriság
0–15	41
15,1–30	45
30,1–45	23
45,1–60	17
60,1–75	6
75,1–90	7
90,1–105	2

Megoldás

Használjuk az alábbi összefüggést:

$$M_o = m_o + \frac{f_1}{f_1 + f_2} h$$

m_o a modális osztály alsó határának az értéke

f_1 a modális osztály gyakoriságának és a vele alulról szomszédos osztály gyakoriságának a különbsége

f_2 a modális osztály gyakoriságának és a vele felülről szomszédos osztály gyakoriságának a különbsége

h a modális osztály szélessége

Jelen esetben a modális osztály (vagyis a legnagyobb gyakoriságú osztály a 2., vagyis a 15,1–30-ig terjedő értékű. Ebből kiindulva a felírható egyenlet:

$$M_o=15{,}1+(4/(4+22))\times 14{,}9$$

öszevonások után:

$$M_o=15{,}1+(4/(26))\times 14{,}9$$

vagyis:

$$M_o=15{,}1+0{,}15\times 14{,}9 =\mathbf{17{,}39}$$

Mivel a csapadékmérések legfeljebb 0,1 mm pontossággal történnek, ezért a kapott értéket kerekíthetjük, ekkor az eredmény 17,4 mm.

A valós, osztályba sorolás nélküli, eredeti adatok alapján a mondott időszakban és hónapokban a leggyakoribb mért érték 17,1 mm volt. Ennek alapján elmondható, hogy igen pontosan sikerült meghatároznunk a keresett értéket.

Meg kell azonban jegyeznünk, hogy a módszer használatának alapfeltétele a normáleloszlású, vagy ahhoz közel álló sokaság.

Feladat

Az alábbi táblázatban a Budapesten 1896-1995 közötti júliusi középhőmérsékletek szerepelnek, 2°C-os intervallumokba sorolva. Becsülje meg a sokaság móduszát!

Középhőmérséklet, °C	**Gyakoriság**
18–20	8
20,1–22	48
22,1–24	38
24,1–26	5
26,1––28	0
28,1–30	1

Megoldás

21,6 °C.

Az eredeti adatokból meghatározott módusz értéke egyébként 21,2 °C.

VII. Folytonos eloszlások sűrűségfüggvényei

Sokszor előfordul, hogy a valószínűségi változó a valóságban folytonosan változik, a diszkrét mérései adatsor csak a mérőeszközünk pontatlanságának és/vagy a véges számú mérés következménye. Erre példa a fenti 4.4 feladatban ismertetett hőmérsékleti adatsor. A mérési adatok 0,1°C -es pontossággal vannak megadva, mivel a hőmérők ilyen pontosak. Elvileg azonban a hőmérséklet ennél jóval több tizedesnyi pontossággal is mérhető lenne. (Az persze kérdéses, hogy van-e értelme annak, hogy a levegő hőmérsékletét akár csak 2 tizedes pontossággal adjuk meg.) A folytonosan változó adatok jellemzését megkönnyíti a sűrűségfüggvény használata. A diszkrét (empirikus) adatok elemzését is elősegíti, ha a hisztogramot valamilyen sűrűségfüggvénnyel tudjuk közelíteni.

Az alábbiakban megadunk néhány gyakran használt sűrűségfüggvényt.

a) Egyenletes eloszlás

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{ha } a < x < b \\ 0 & \text{egyébként} \end{cases}.$$

b) Normális (Gauss) eloszlás

$$f(x) = \frac{1}{\sqrt{2\pi}\cdot\sigma} e^{-\frac{(x-m)^2}{2\cdot\sigma^2}}$$

,ahol *m* (az eloszlás várhatóértéke), σ (az eloszlás szórása) sűrűségfüggvény paraméterei.

c) Exponenciális eloszlás

$$f(x) = \lambda \cdot e^{-\lambda x}$$

, ahol λ az eloszlás paramétere.

Feladatok

4.9. Ábrázolja a normális eloszlást a –6,0 és 7,0 közötti intervallumban, a következő feltételek mellett:

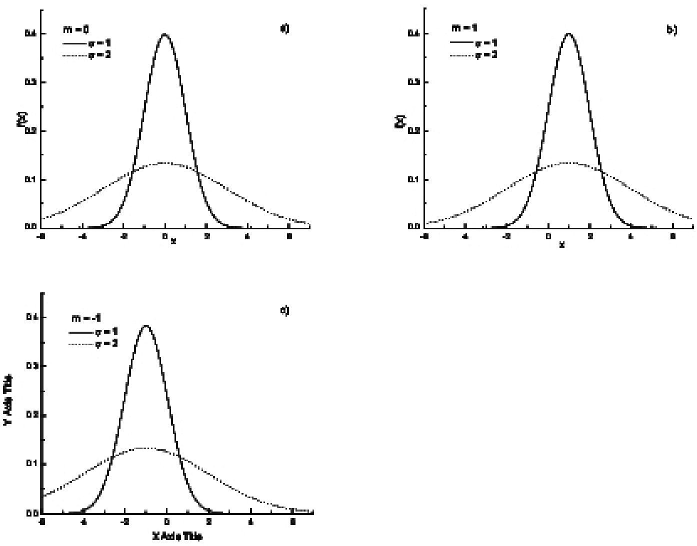
a) m = 0, σ = 1, és m = 0, σ = 3

b) m = 1, σ = 1, és m = 1, σ = 3

c) m = -1, σ = 1, és m = -1, σ = 3

Az m és a σ paraméterek módosítása következtében milyen változások tapasztalhatóak a függvény alakjában?

d) Határozza meg a görbe alatti területet az $m = 0$, $\sigma = 2$ esetben!



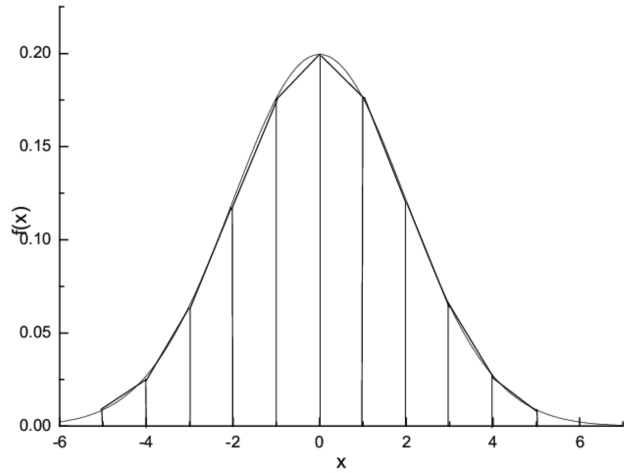
4.4 ábra.
Normális eloszlás különböző várhatóérték és szórás esetén.

Megoldás

Határozzuk meg a függvény értékét –6 és +7 között minden egész értéknél! A kapott értékeket az alábbi táblázat tartalmazza:

x	**-6,0**	**-5,0**	**-4,0**	**-3,0**	**-2,0**	**-1,0**	**0,0**
f(x)	$2{,}21\times10^{-3}$	$8{,}77\times10^{-3}$	0,027	0,0648	0,121	0,176	0,2
x	**1,0**	**2,0**	**3,0**	**4,0**	**5,0**	**6,0**	**7,0**
f(x)	0,176	0,121	0,0648	0,027	$8{,}77\times10^{-3}$	$2{,}21\times10^{-3}$	$4{,}36\times10^{-4}$

Közelítsük a görbe alatti területet trapézokkal (lásd 4.5. ábra), és számítsuk ki a fenti adatok alapján a 12 trapéz területét.



4.5. ábra.
Görbe alatti terület közelítése trapézokkal.

A *k*-ik trapéz területét a következő összefüggéssel számíthatjuk ki:

$$T_k = \frac{f(x_k) + f(x_{k+1})}{2}(x_{k+1} - x_k)$$

T_1	T_2	T_3	T_4	T_5	T_6
0,0055	0,0179	0,0459	0,0929	0,1485	0,1877
T_7	**T_8**	**T_9**	**T_{10}**	**T_{11}**	**T_{12}**
0,1877	0,1485	0,0929	0,0459	0,0179	0,0055

Összegezve a területeket a következő közelítő értéket kapjuk a görbe alatti területre: 0,9968.

4.10. Határozza meg a normális eloszlás által meghatározott görbe alatti területet a fent ismertetett eljárással, ha m = 1 és σ = 2 !

4.11. Ábrázolja az exponenciális eloszlást leíró görbét a [0,6] intervallumban, ha:

a) λ = 1,

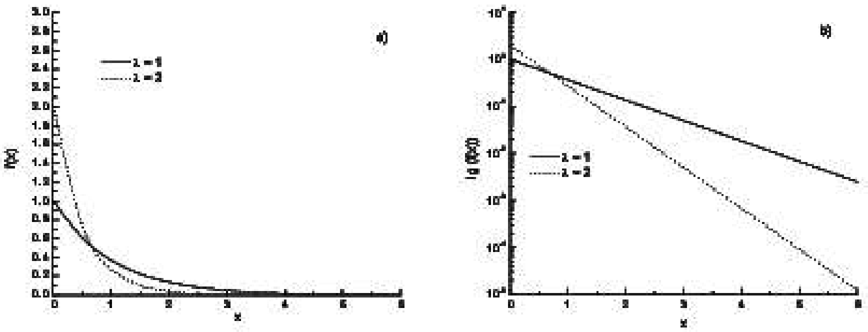
b) λ = 2.

Az ábrázolás során a függőleges tengelyen használjon 10-es alapú logaritmikus skálát!

Megoldás

A függvény értékeket a következő táblázatban adjuk meg.

x	0	1	2	3	4	5	6
f(x) (= 1)	1	3,6810 $^{-1}$	1,35 10^{-1}	4,98 10^{-2}	1,83 10^{-2}	6,74 10^{-3}	2,48 10^{-3}
f(x) (= 2)	2	2,7110 $^{-1}$	3,66 10^{-1}	4,96 10^{-3}	6,71 10^{-4}	9,09 10^{-5}	1,23 10^{-5}



4.6 ábra.
Exponenciális lineáris (a) és logaritmikus (b) függőleges skálabeosztás esetén.

A sűrűségfüggvények által meghatározott görbe alatti területek pontos meghatározása nagyon fontos probléma. Ugyanis mint a fentiekben láttuk, a görbe alatti területtel fejezhető ki annak valószínűsége, hogy adataink egy adott intervallumba esnek. A fentiekben bemutattunk egy lehetséges közelítő eljárást. Az exponenciális eloszlás esetében létezik egy jóval egyszerűbb módszer is, ugyanis az exponenciális eloszlást leíró függvény integrálható, így a keresett valószínűség a következő módon fejezhető ki:

$$P(x_1 \leq x < x_2) = \int_{x_1}^{x_2} \lambda e^{-\lambda x} dx = e^{-\lambda x_1} - e^{-\lambda x_2}$$

Nincs ilyen egyszerű módszer a normális eloszlás esetében. Az alábbiakban megmutatjuk, hogy hogyan számolhatjuk ki a normális eloszlást leíró görbe alatti területet tetszőleges $[x_1,x_2]$ intervallumban. A számítások elvégzéshez az A1 táblázatot kell felhasználni. A táblázat egy olyan normális eloszlásra vonatkozik, amely esetében m = 0,0 és σ =1,0. A második oszlopban látható számok azt fejezik ki, hogy mekkora a görbe alatti terület az adott x értéktől jobbra eső tartományban (lásd ábra). Ha például arra vagyunk kíváncsiak, hogy mekkora a görbe alatti terület az x_1 = -2 és az x_2 = 1 intervallumban, akkor ki kell vonni az x = 1-hez tartozó 0,8413-ból az x = -2-höz tartozó 0,0228-et. Tehát a görbe alatti terület a [-2,1] intervallumban 0,8185 lesz. Sajnos azonban a normális eloszláshoz tartozó várható érték és szórás csak nagyon ritkán egyezik meg a fentiekben megadott értékekkel. Azért teljesen általános esetben az intervallumok határait módosítanunk kell mielőtt a táblázatot használnánk. A módosítás lépései a következők:

i. Az eredeti intervallum határokat a várhatóérték nagyságával kell eltolni, azaz az új határok a következők lesznek:

$$x_1' = x_1 - m \text{ és } x_2' = x_2 - m$$

ii. Figyelembe kell venni azt is, hogy a szórás sem egységnyi. A fentiekben megadott módosított határokat osztva a szórással megkapjuk azokat az új intervallum határokat, amelyek segítségével már az A1 táblázatból meg tudjuk határozni a görbe alatti területet:

$$x_1'' = \frac{x_1 - m}{\sigma} \quad \text{és } x_2'' = \frac{x_2 - m}{\sigma}$$

4.12. Tegyük fel, hogy a hőmérséklet eloszlását egy normális eloszlással közelíthetjük, amelynek paraméterei m = -2,3 és σ =1,8. Mekkora annak a valószínűsége, hogy a hőmérséklet -3,0 és -2,0 °C között lesz?

Megoldás

A valószínűség meghatározásához a fentiekben megadott módon ki kell számolni az új intervallum határokat.

$$-3 \rightarrow \frac{-3,0-(-2,3)}{1,8} = -0,39 \quad \text{és} \quad -2 \rightarrow \frac{-2,0\ -\ (-2,3)}{1,8} = 0,17$$

Az A1 táblázatból a –0,39-hez tartozó terület ≈ 0,3446 és 0,17 –hez tartozó érték ≈0,5596.

A –3,0 és a –2,0 közötti pontok területet a következő különbség adja 0,5596 - 0,3446 = 0,215 . Tehát a keresett valószínűség 0,215.

4.13. Illesszen a 4.4 feladathoz mellékelt táblázatban megadott adathalmazra normális eloszlást! Vizsgálja meg, hogy a hisztogramok (Δ = 1°C és a Δ = 5 °C-os beosztások) mennyire térnek el a normális eloszlástól!

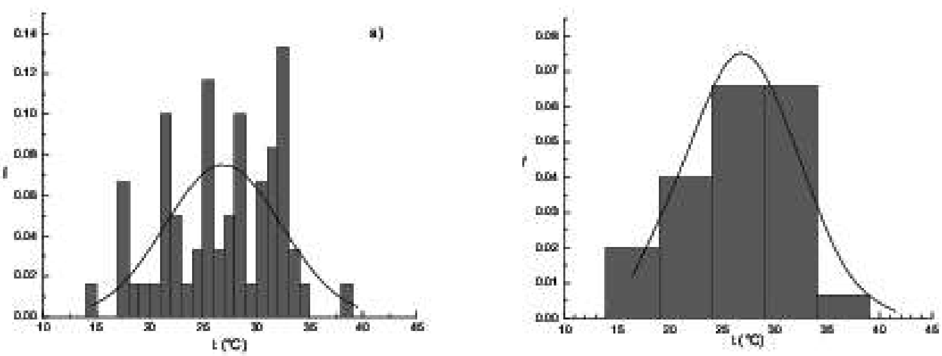
Megoldás

A normális eloszlás megrajzolásához ismernünk kell az *m* és a σ paramétereket. Az illesztés során ezek a mennyiségek megfelelnek az adathalmazból számított átlagnak és empirikus szórásnak, azaz *m = 26,9,* σ = *5,3.* Behelyettesítve a fenti értékeket a normális eloszlást leíró függvénybe, a következő összefüggést kapjuk:

$$f(x) = \frac{1}{\sqrt{2\pi} \cdot 5,3} e^{-\frac{(x-26,9)^2}{2 \cdot 5,3^2}}$$

Ezek után tetszőleges x értékre ki tudjuk számolni a függvény értékét. A hisztogramokkal való könnyebb összehasonlíthatóság kedvéért a függvény értékét számoljuk ki rendre a 14,5; 15,5; …stb pontokban.

x	f(x)	x	f(x)
14,5	0,0049	27,5	0,0748
15,5	0,0074	28,5	0,0719
16,5	0,0110	29,5	0,0667
17,5	0,0156	30,5	0,0598
18,5	0,0214	31,5	0,0517
19,5	0,0284	32,5	0,0431
20,5	0,0363	33,5	0,0346
21,5	0,0448	34,5	0,0269
22,5	0,0533	35,5	0,0201
23,5	0,0613	36,5	0,0146
24,5	0,0679	37,5	0,0102
25,5	0,0727	38,5	0,0069
26,5	0,0751	39,5	0,0045



4.7 ábra.
Normális eloszlás illesztése empirikus sűrűségfüggvényre.

VIII. Koncentráció

A koncentráció grafikus megjelenítése célszerűen Lorenz-görbe szerkesztésével valósítható meg. Számszerű értéke képezhető a Lorenz-görbe alatti terület és a görbét befoglaló négyzet fél területének hányadosaként, vagy a

$$K = G/2x'$$

ahol G= a sokaság átlagos különbsége, x'= a sokaság számtani középértéke.

Mivel a Lorenz-görbe elkészítésének előfeltétele az adatok kumulált relatív értékösszegsorba rendezése, ezért a feldoldogást ezzel érdemes kezdeni.

A Lorenz-görbe úgy ábrázolja a koncentrációt, hogy megmutatja: az egyik sokaság 10-, 20- … 100%-nyi részére a másik sokaság hány százaléka esik.

Feladat

Városi lakosság koncentráció-változásának meghatározása. Az alábbi táblázat 1949-re és 2001-re vonatkozóan tartalmazza a 2001-ben Magyarországon városi jogállású települések lakosságának a településnagyság szerinti megoszlását.

a) Először 1949-es lakosságuk nagysága szerint növekvő sorba állítottuk a településeket.

b) Ezután megnéztük, hogy a települések első 10%-ában, vagyis 25 településen összesen, a teljes, 2001-ben városi jogállású településállomány lakosságának hány százaléka lakott.

c) Folytattuk azzal, hogy megnéztük, a települések első 20%-ában, vagyis 50 településen összesen, a teljes, 2001-ben városi jogállású településállomány lakosságának hány százaléka lakott.

És így tovább 100%-ig.

d) Ugyanezt elvégeztük a 2001-es adatsorra vonatkozóan is.

e) Az adatokat táblázatba foglaltuk:

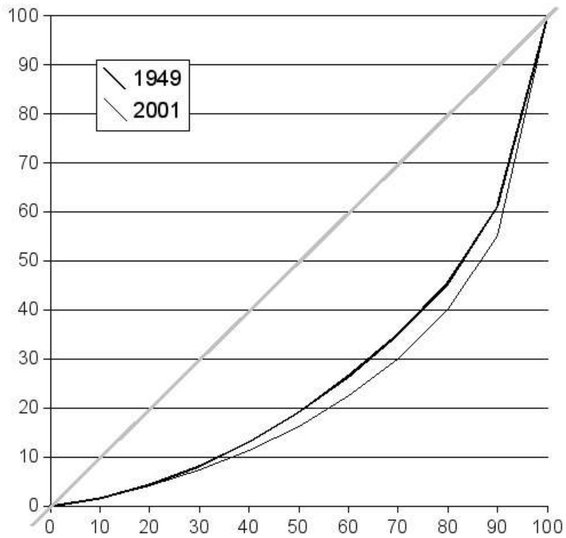
	1949	**2001**
10,00%	1,55%	1,63%
20,00%	4,32%	4,13%
30,00%	8,11%	7,43%
40,00%	13,03%	11,39%
50,00%	19,13%	16,31%
60,00%	26,47%	22,39%
70,00%	35,08%	30,04%
80,00%	45,54%	40,20%
90,00%	61,30%	55,02%
100,00%	100,00%	100,00%

Ábrázolja az adatsorokat Lorenz-görbével és adja meg számszerűen is a koncentrációkat! Melyik időszakban volt nagyobb a lakosság koncentrációja?

Megoldás

A Lorenz-görbe voltaképpen nem is görbe, hanem töröttvonal. A „görbe” vonala ugyanis az 10%-onként vett értékekhez tartozó pontokat köti össze egyenes szakaszokkal. Olyan négyzetbe rajzoljuk, melynek két, egymásra merőleges oldalát 0–100%-ig terjedően egyenletesen és maradék nélkül felosztjuk, mint koordinátarendszer tengelyeit. Az alsó tengelyre vesszük fel a független változót (vagyis a táblázat első oszlopát), a bal oldalira pedig a függő változó aktuális értékeit (vagyis jelen esetben a többi oszlop értékeit). Ezután a megfelelő pontokat összekötjük, mint az ábrán is látható. Az ábrán ugyan feltüntettük vastag szürke vonallal a négyzet átlóját is, de ez egyébként nem szükséges.

Látható, hogy a 2001-es adatok nagyobb koncentrációt mutatnak, mint az 1949-esek, mert távolabb futnak a négyzet átlójától. Vagyis elmondható, hogy 2001-re, 1949-hez képest a magyarországi városállomány úgy változott, hogy a legnagyobb városok népessége gyorsabban nőtt, mint a kisebbeké: a népesség koncentrációja emelkedett.



Számszerűen vagy az átlagos különbség segítségével, vagy a görbe alatti területek arányaival fejezhető ki a változás.

Mivel az átlagos különbség meghatározása roppantul számolásigényes feladat, célszerűbb a második megoldást választani.

Nyilvánvaló, hogy a z 1949-es töröttvonal alatti terület kiszámolható a háromszög és trapézok területének összegeként. Az első területegység a vízszintes tengelyen mérve 0–10%-ig terjedő rész. Ez nem más, mint egy derékszögű háromszög, melynek egyik oldala 0,1 (vagyis 10%), másik oldala pedig az ott mért magassága, vagyis 0,0155 (1,55%). Így a területe (T_1) a (0,1*0,0155)/2 összefüggéssel fejezhető ki.

A második egység már derékszögű trapéz. Ennek egyik párhuzamos oldala az iménti 0,0155 (1,55%), másik párhuzamos oldala pedig a 20%-nál mért adattal egyezik meg, vagyis 0,0432 (4,32%). Magassága, mivel derékszögű, a 10%-os és a 20%-os osztás közötti hosszal egyezik meg, vagyis 0,1. Így területe: ((0,0155+0,0432)/2)*0,1.

T_1-től T_{10}-ig a részterületek az alábbiak:

T_1=(0,1*0,0155)/2

T_2=((0,0155+0,0432)/2)*0,1

T_3=((0,0432+0,0811)/2)*0,1

T_4=((0,0811+0,1303)/2)*0,1

T_5=((0,1303+0,1913)/2)*0,1

T_6=((0,1913+0,2647)/2)*0,1

T_7=((0,2647+0,3508)/2)*0,1

T_8=((0,3508+0,4554)/2)*0,1

T_9=((0,4554+0,6130)/2)*0,1

T_{10}=((0,6130*1)/2)*0,1

Ezeket összegezve a teljes görbe alatti terület=**0,1839.**

A megfelelő értékek behelyettesítésével ugyanilyen módon kiszámolható a 2001-es adatokhoz tartozó terület is, mely =**0,1610**.

Feladat

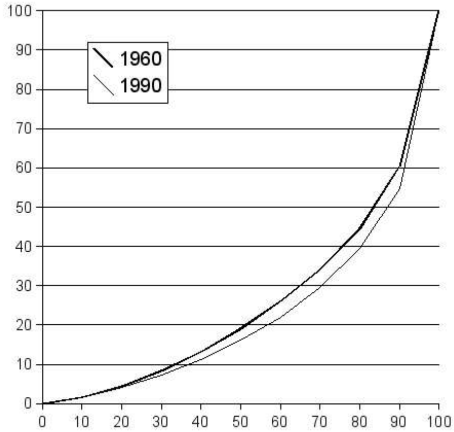
Az előzőekhez hasonlóan állapítsa meg a táblázat adatai alapján, hogyan változott a 2001-ben Magyarországon városi jogállású települések lakosságának a településnagyság szerinti megoszlása 1960 és 1990 között

	1960	**1990**
0,00%	0,00%	0,00%
10,00%	1,65%	1,60%
20,00%	4,44%	4,12%
30,00%	8,25%	7,39%
40,00%	13,14%	11,24%
50,00%	19,07%	16,10%
60,00%	26,00%	22,04%
70,00%	34,19%	29,61%
80,00%	44,49%	39,56%
90,00%	60,23%	54,61%
100,00%	100,00%	100,00%

.

Megoldás

A Lorenz görbe:



Az 1960-hoz tartozó terület: **0,1813.**

Az 1990-hez tartozó terület: **0,1589.**

IX. Aszimmetria meghatározása

A torzult normál eloszlású sokaságok jellemzésére alkalmas az aszimmetria meghatározása.

Feladat

Az alábbi táblázat egy észak-alföldi megfigyelőhelyen mért januári csapadékösszegeket tartalmazza. Határozza meg a Pearson, valamint a Köppen-féle aszimmetria-mutató értékét!

Év	Csapadék (mm)	Év	Csapadék (mm)
1955	86,4	1971	51,5
1956	3,1	1972	11,8
1957	30,4	1973	3
1958	37	1974	29,2
1959	59,8	1975	17,1
1960	28,2	1976	22,7
1961	25,3	1977	86,6
1962	25,3	1978	17,1
1963	69,4	1979	94,9
1964	0,6	1980	26,3
1965	43,3	1981	16,8
1966	52,7	1982	23,1
1967	30,5	1983	21
1968	21,5	1984	32,7
1969	8,8	1985	24,6
1970	53,4	1986	45,4
		1987	59,8

Megoldás

A Pearson -féle aszimmetria meghatározása az alábbi:

$$A_p = \frac{\overline{x} - Mo}{\sigma}$$

x: a minta középértéke

Mo: a minta módusza

σ : a minta szórása

A minta módusza könnyen meghatározható, értéke **17,1.**

A minta középértékének meghatározása ugyancsak nem okozhat gondot: **35,13**.

A szórás meghatározására a példatár más részeiben már voltak gyakorlatok. Az ott leírtak szerint meghatározva a szórást: **24,37**.

Behelyettesítve most a fenti összefüggésbe:

A_p=(35,13-17,1)/24,37

A_p=0,74

A Köppen-féle aszimmetria meghatározása pedig:

$$A_k = 1 - \frac{2n_a}{n}$$

n_a: a minta azon elemeinek száma, melyek a minta számtani középértékénél kisebbek,

n: a minta elemszáma.

Jelen esetben n_a=**21**,

n=**33**.

Vagyis A_k= 1-(42/33)

A_k**= –0,27**.

Feladat

Az alábbi táblázat az előzőekben is használt észak-alföldi adatsorból származik, de a júniusi csapadékösszegeket tartalmazza. Határozza meg, hogy milyen az eloszlás Pearson-, illetve Köppen-féle aszimmetriája, mennyivel tér el a januári adatsorétól!

Év	Csapadék (mm)	Év	Csapadék (mm)
1955	80,8	1972	69,6
1956	129,2	1973	139,8
1957	51,5	1974	121,8
1958	196,7	1975	135,6
1959	111	1976	29,9
1960	108,4	1977	123,4
1961	47,6	1978	92,3
1962	24,5	1979	71,1
1963	32,7	1980	132,2
1964	41,3	1981	40,9
1965	171,4	1982	72,2
1966	80,7	1983	52,5
1967	51	1984	73,2
1968	7,7	1985	52,3
1969	118,8	1986	39,9
1970	131,2	1987	27,5
1971	70,7		

Megoldás

A_p= Mivel a mintának nincs módusza, nem kell meghatároznunk.

A_k= **–0,21.**

Vagyis, bár egyik minta sem mutat jelentős aszimmetriát, a júniusi csapadék-összegek eloszlásának aszimmetriája kisebb, mint a januáriaké.

5. Összefüggés vizsgálat

Gyakori feladat, hogy olyan mennyiségek között kell összefüggést keresnünk, amelyek közötti függvénykapcsolat nem ismert. Ennek két alapvető oka lehet. Vagy a kapcsolat bonyolult jellege nem teszi lehetővé az általános érvényű függvénykapcsolat meghatározását, vagy pedig a létező összefüggés felfedezésének első lépésinél tartva még nagyon sok, nem kontrollált, zavaró tényező akadályoz bennünket az összefüggés felismerésében.

I. A korrelációs együttható

Az elemzést a tapasztalati függvénykapcsolat ábrázolásával kezdjük. Az ábra segít a kvalitatív következtetések levonásában. Ennél azonban többet tehetünk, számszerűsíthetjük az összefüggés erősségét. Ennek kifejezésére a korrelációs együtthatót használjuk:

$$r = \frac{\sum_{i=1}^{n}(x_i - M_x)(y_i - M_y)}{\sqrt{\sum_{i=1}^{n}(x_i - M_x)^2 (y_i - M_y)^2}},$$

ahol x_i és y_i a két adatsor érték párjai, M_x és M_y a két adatsor átlagértékei. Az r korrelációs együttható –1 és +1 közötti értéket vehet fel. Ha az r értéke 0-hoz közeli akkor csak gyenge kapcsolat van a két mennyiség között. Ha az r abszolút értéke 1-hez közeli, akkor a kapcsolat erős.

A két mennyiség közötti kapcsolatot valamilyen függvény segítségével is kifejezhetjük. Első közelítésként többnyire lineáris függvénnyel leírható összefüggést tételezünk fel a két mennyiség között:

$$y = a \cdot x + b,$$

$$\text{ahol } a = r\frac{\sigma_y}{\sigma_x} \quad \text{és } b = M_y - a \cdot M_x$$

ahol σ_x és σ_y a két adatsorra jellemző szórás.

Feladatok

5.1. Az alábbi táblázat Budapest január havi átlaghőmérsékletét és csapadékösszegét tartalmazza 1901 és 1960 között. Vizsgáljuk meg a csapadék és a hőmérséklet közötti kapcsolatot!

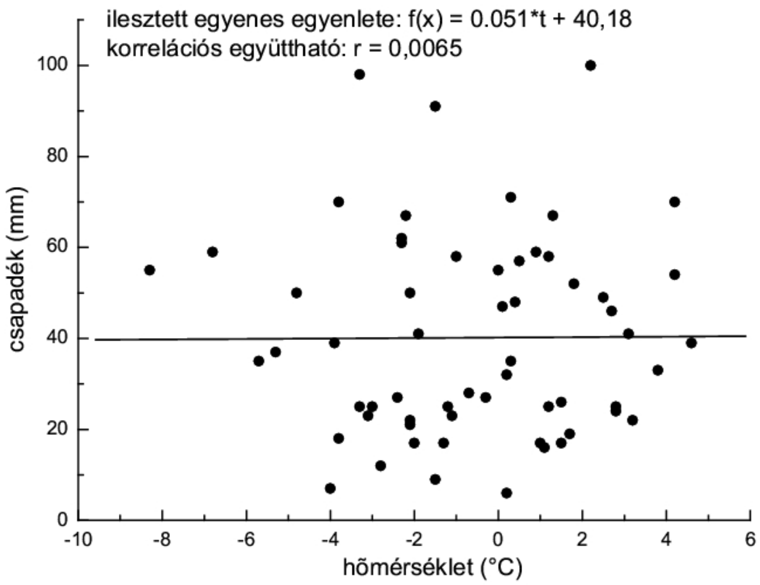
A levegő január havi átlaghőmérséklete és január havi csapadékösszege Budapesten

év	t (°C)	csap (mm)	év	t (°C)	csap (mm)	év	t (°C)	csap (mm)
1901	-5,3	37	1921	4,6	39	1941	-2,1	50
1902	2,8	25	1922	-1,5	91	1942	-8,3	55
1903	-1,1	23	1923	1,8	52	1943	-3,9	39
1904	-2,1	21	1924	-3,1	23	1944	3,2	22
1905	-3,8	18	1925	0,2	6	1945	-3,3	98
1906	-0,7	28	1926	0,2	32	1946	-2,8	12
1907	-2,3	61	1927	2,7	46	1947	-5,7	35
1908	-2,4	27	1928	-0,3	27	1948	4,2	70
1909	-3,0	25	1929	-3,8	70	1949	1,5	17
1910	0,9	59	1930	0,3	35	1950	-2,2	67
1911	1,0	17	1931	0,4	48	1951	2,8	24
1912	-3,3	25	1932	-1,2	25	1952	1,2	58
1913	-1,9	41	1933	-2,1	22	1953	1,3	67
1914	-4,0	7	1934	-1,5	9	1954	-4,8	50
1915	2,2	100	1935	-2,0	17	1955	-1,0	58
1916	3,8	33	1936	4,2	54	1956	1,5	26
1917	0,5	57	1937	-2,3	62	1957	-1,3	17
1918	1,1	16	1938	0,0	55	1958	0,3	71
1919	2,5	49	1939	1,7	19	1959	0,1	47
1920	3,1	41	1940	-6,8	59	1960	1,2	25

Megoldás

A probléma megértése egyszerűbb, ha először grafikusan ábrázoljuk a két mennyiség közötti összefüggést. A vízszintes tengelyen megadott mennyiséget nevezzük független változónak, a függőleges tengelyen megadott mennyiségét pedig függő változónak. A pontok mutatják, hogy adott hőmérsékletértékhez mekkora csapadék érték tartozik. Az ábra azt sejteti, hogy nem lehet szoros kapcsolat a két mennyiség között.

Ennek bizonyításához meg kell határoznunk a korrelációs együtthatót. A részletes számításokat később, egy egyszerűbb példán fogjuk bemutatni. Az ábrán az EXCEL táblázat kezelő segítségével meghatározott értéket láthatjuk. Az adatok bevitele után a táblázat kezelő segítségével ábrázolja a két mennyiség közötti kapcsolatot leíró görbét, határozza meg a korrelációs együtthatót, valamint a pontokra illeszthető egyenes egyenletét! (Az illesztett egyenest a vastag egyenes vonal jelöli, az egyenes egyenlete a bal felső sarokban látható.)



5.1. ábra.
Csapadék a havi átlaghőmérséklet függvényében (pontok). A vastag vonal az adatokra illesztett egyenest jelöli.

5.2. Az alábbi táblázat az Északi-középhegységben található helységek tengerszint feletti magasságát és a hótakarós napok számát tartalmazza. Milyen szoros kapcsolat van a tengerszint feletti magasság és a hótakarós napok száma között? Írja fel az illesztett egyenes egyenletét! Várhatóan mennyi a hótakarós napok száma abban a helyiségben, amelynek tengerszint feletti magassága 500 m?

helység	magasság (m)	hótakarós napok száma
Gyöngyös	160	31
Eger	170	36
Terény	200	43
Rudabánya	280	55
Mátrafüred	340	56
Mátraháza	670	88
Kékestető	990	113

Megoldás

Válasszuk független változónak (x) a tengerszint feletti magasságot és függő változónak (y) a hótakarós napok számát. A korrelációs együttható kiszámításához készítsük el a következő táblázatot:

x_i	$x_i - M_x$	y_i	$y_i - M_y$	$(x_i - M_x)(y_i - M_y)$	$(x_i - M_x)^2$	$(y_i - M_y)^2$
160	-241,4	31	-29,3	7073,0	58274,0	858,5
170	-231,4	36	-24,3	5623,0	53546,0	590,5
200	-201,4	43	-17,3	3484,2	40562,0	299,3
280	-121,4	55	-5,3	643,4	14738,0	28,1
340	-61,4	56	-4,3	264,0	3370,0	18,5
670	268,6	88	27,7	7440,2	72146,0	767,3
990	588,6	113	52,7	31019,2	346450,0	2777,3
				= 55547,1	= 589485,0	= 5339,4

Az M_x és az M_y a tengerszint feletti magasságok és a hótakarós napok számának átlaga. A fenti táblázatban meghatározott adatok segítségével kiszámolhatjuk a korrelációs együtthatót (*r*) és az illesztett egyenes két paraméterét (*a*) és (*b*):

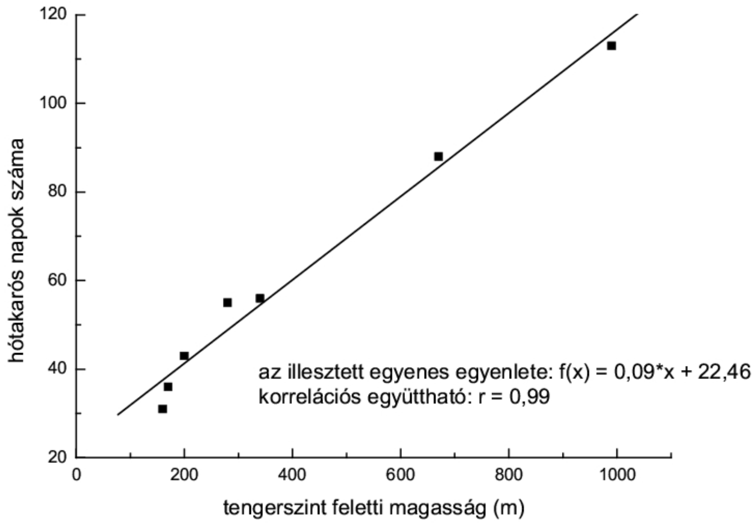
$$r = \frac{55574,1}{\sqrt{589485,0 \cdot 5339,4}} = 0,99$$

és

$$a = r\frac{\sigma_y}{\sigma_x} = 0,99\frac{\sqrt{5339,4}}{\sqrt{589485,0}} = 0,094 \text{ és } b = M_y - a \cdot M_x = 60,3 - 0,094 \cdot 401,4 = 22,5$$

Az *a* paraméter kiszámítása során vegyük észre, hogy a fenti táblázat utolsó két oszlopában szereplő tagok összegéből vont négyzetgyök csak egy konstansban (adatok száma –1) különbözik a szórástól.

A regressziós egyenes egyenlete: y = 0,094 x + 22,5. Ez alapján a tengerszintje felett 500 m-rel a hótakarós napok száma: y = 0,094 · 500 + 22,5 = 69,5 nap.



5.2 ábra.
A hótakarós napok száma a tengerszint feletti magasság függvényében. A vastag folytonos vonal az adatokra illesztett egyenest jelöli.

5.3. Az 5.1 feladatban láttuk, hogy a január hónapi középhőmérséklet és a január havi csapadék között nem mutatható ki összefüggés. Vizsgálja meg ezen két mennyiség közötti kapcsolatot egy nyári hónapban! Az alábbi táblázat egy húsz éves adatsort tartalmaz. Határozza meg a korrelációs együtthatót és a regressziós egyenes egyenletét! A számítások eredményét ellenőrizze az EXCEL táblázatkezelő segítségével! (**r = -0,593; a = - 0,025; b = 23,5**).

Július csapadékösszeg (mm-ben) és középhőmérséklet (C°-ban) kapcsolata Budapesten

év	**csapadék (x)**	**hőmérséklet (y)**
1941	61	21,2
1942	75	21,2
1943	26	22,0
1944	44	21,2
1945	25	22,8
1946	25	23,9
1947	18	23,8
1948	98	20,3
1949	61	21,8
1950	26	24,4
1951	63	22,4
1952	8	24,2
1953	49	23,0
1954	67	19,9
1955	76	21,2
1956	34	21,8
1957	95	22,2
1958	60	22,1
1959	124	22,7
1960	88	19,9

5.4. Az alábbi táblázat adatai alapján határozza meg a január havi átlaghőmérséklet és az átlagos hótakaró-vastagság közötti összefüggést!

a) Ábrázolja grafikusan a hótakaró vastagságát az átlaghőmérséklet függvényében!

b) Határozza meg a korrelációs együtthatót! **(r = -0,89)**

c) Határozza meg a pontokra illeszthető egyenes egyenletét! **(f(x) = -1,064·x + 1,942)**

d) Várhatóan mennyi lenne a hótakaró vastagsága –1°C-os átlaghőmérséklet esetén? **(3 cm)**

	átlaghőmérséklet januárban	hótakaró vastagsága (cm)
1940	-9.6	10.0
1941	-3.7	5.0
1942	-10.2	10.0
1943	-6.2	8.0
1944	-0.4	1.0
1945	-3.2	1.0
1946	-4.8	10.0
1947	-8.5	15.0
1948	3.5	0.0
1949	0.0	2.0
1950	-5.0	10.0
1951	1.9	0.8
1952	0.2	0.1
1953	0.0	0.5
1954	-8.2	12.5
1955	-2.2	4.5
1956	0.8	1.5
1957	-3.2	5.9
1958	-2.2	5.4
1959	-0.6	2.2
1960	-2.5	4.2

II. A Rang-korreláció

Ha valamely jelenségről csak annyit tudunk, hogy elemeinek valamely szempont szerint mi a sorrendjük (rangjuk), és egy másik jelenségről szintén csak annak sorrendiségét ismerjük, továbbá ha a közöttük levő korrelációs kapcsolat erősségét meg kell határoznunk, a rangkorrelációs együttható megállapítása vezet eredményre.

A ragkorrelációs együttható értelmezése mindenben megegyezik a korrelációs együtthatóéval.

Feladat

A 2001-es népszámlálás adataiból ismert a megyék népesség szerinti sorrendje, valamint a megyeszékhelyek népesség szerinti sorrendje is. Mondjuk meg, hogy van-e összefüggés, ha igen milyen irányú és erősségű, a megyék népessége és a megyeszékhelyek népessége között. Másképpen fogalmazva: igaz-e, hogy a nagyobb megyéknek nagyobb a székhely-népessége is? Az adatok az alábbi táblázatból olvashatók ki. (Pest megyét és Budapestet elhagytuk, mert a klasszikus értelemben vett megye-megyeszékhely reláció nem értelmezhető.)

	megyei sorrend	megyeszékhelyi sorrend
Bács-Kiskun	4	7
Baranya	9	4
Békés	10	13
B-A-Z	1	2
Csongrád	7	3
Fejér	6	8
Gy-M-S	5	5
Hajdú-Bihar	3	1
Heves	13	16
J-N-Sz	8	10
K-E	14	11
Nógrád	18	17
Somogy	12	12
Sz-Sz-B	2	6
Tolna	17	18
Vas	16	9
Veszprém	11	14
Zala	15	15

Megoldás

A rang-korrelációs együttható meghatározása:

$$r_s = 1 - \frac{6\sum_{i=1}^{n} {d_i}^2}{n^3 - n}$$

n – a minta elemszáma

d_i – a rangértékek különbsége

A minta elemszáma egyszerűen meghatározható: 18. A rangértékek különbsége azt jelenti, hogy minden egyes csoportra (jelen esetben megyér) egyenként meghatározzuk a kétféle sorrend (a megye lakossága és a megyeszékhely lakossága) alapján vett helyezések különbségét. Ha ez megvan, akkor ezen értékeket egyenként négyzetre kell emelnünk, majd az így keletkezett mennyiségek összegét vennünk. Ezt legegyszerűbben egy újabb táblázatban valósíthatjuk meg.

	Megyei sorrend	megye-székhelyi sorrend	A rangértékek különbsége	A rangértékek különbségének négyzete
Bács-Kiskun	4	7	-3	9
Baranya	9	4	5	25
Békés	10	13	-3	9
B-A-Z	1	2	-1	1
Csongrád	7	3	4	16
Fejér	6	8	-2	4
Gy-M-S	5	5	0	0
Hajdú-Bihar	3	1	2	4
Heves	13	16	-3	9

J-N-Sz	8	10	-2	4
K-E	14	11	3	9
Nógrád	18	17	1	1
Somogy	12	12	0	0
Sz-Sz-B	2	6	-4	16
Tolna	17	18	-1	1
Vas	16	9	7	49
Veszprém	11	14	-3	9
Zala	15	15	0	0

Most már csak az utolsó oszlopban foglalt értékek összegét kell képeznünk, vagyis a Σd_i^2 -et.

Ez jelen esetben 166. Ezután felírhatjuk az egyenletet, behelyettesítve az aktuális értékeket:

$$1- (6\times166)/(18^3-18).$$

Egyszerűbben:

$$1-996/5814$$

Vagyis a végeredmény:

$$1-0{,}17 = \mathbf{0{,}83}$$

Tehát elmondhatjuk, a megyék lakossága és a megyeszékhelyeik lakosságának nagysága között pozitív irányú, erős összefüggés áll fenn. Magyarán: a nagyobb lakosságú megyéknek általában a megyeszékhelye is népesebb.

Feladat

Ismert a megyeszékhely városok népesség szerinti és terület szerinti sorrendje. Határozza meg, hogy van-e milyen erősségű és irányú a kétféle sorrend közötti összefüggés!

Város	Terület szerinti sorrend	Népesség szerinti sorrend
Békéscsaba	6	13
Debrecen	1	1
Eger	17	16
Győr	8	5
Kaposvár	12	12
Kecskemét	2	7
Miskolc	5	2
Nyíregyháza	4	6
Pécs	10	4
Salgótarján	13	17
Szeged	3	3
Székesfehérvár	9	8
Szekszárd	16	18

Szolnok	7	10
Szombathely	15	9
Tatabánya	18	11
Veszprém	11	15
Zalaegerszeg	14	14

Megoldás

r_s=**0,73**

6. Hipotézisvizsgálat

A hipotézisvizsgálat egy statisztikai döntési feladat. El kell dönteni, hogy az adott, a feladatra jellemző minta alapján mely állítást fogadjuk el helyesnek.

Leggyakrabban azt vizsgáljuk, hogy:

- valamilyen statisztikai adatsorból kiválasztott rész-adatsor jellemzői (pl. várható érték, szórás, sűrűségfüggvény) eltérnek-e egymástól,
- valamilyen módon mért adatsor eltér-e egy ideális eloszlástól,
- adatsorok empirikus sűrűségfüggvényei, számtani közepei, szórásai eltérnek-e egymástól.

Null-hipotézis: különböző csoportba tartozó események, mérések matematikai jellemzői között *nincs eltérés.*

Ha a null-hipotézis teljesedési valószínűsége nagyon kicsi, akkor a null-hipotézist elvetjük, és az eltérést szignifikánsnak nevezzük.

Azt a valószínűséget, amelynél kisebb valószínűséget kapva a hipotézist elvetjük szignifikancia szintnek nevezzük. Értéke lehet tetszőleges, de általában a 0,1, 0,05 és 0,01-et szokás választani.

A hipotézisek teljesülésének valószínűségét statisztikai próbák segítségével határozzuk.

Néhány fontosabb statisztikai próba

Két mintás t-próba

Két adathalmaz számtani közepeinek összehasonlítására szolgál. A két adatsor adataiból meghatározott számtani közepek kisebb – nagyobb mértékben eltérnek egymástól. Az eltérés két okra vezethető vissza:

i. Az adatok véletlenszerű változása eredményezi a különbséget.

ii. Az eltérés valóságos, két lényegesen különböző adatsorról van szó.

Null-hipotézis: a két adathalmaz számtani közepe egyenlő. A t-próba elvégzésekor először az alábbi mennyiséget kell meghatározni:

$$t = \frac{|m_1 - m_2|}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad \text{ahol } S = \sqrt{\frac{(n_1 - 1)\sigma_1^2 + (n_2 - 1)\sigma_1^2}{n_1 + n_2 - 1}}$$

ahol m_1, σ_1 és n_1 az adatokból képzett számtani közép, az adatok szórása és az adatok száma az első halmazban, m_2, σ_2 és n_2 ugyan ezen mennyiségek a második halmazra vonatkozólag. A t-próba során feltételezzük, hogy a két adathalmazra jellemző szórás megegyezik. (Ez egy újabb hipotézis, amelyet F-próbával ellenőrizhetünk. Tehát az adatokból kiszámolt szórásoknak nem feltétlen kell megegyeznie.)

Feladatok

6.1. A táblázat egy tóparti meteorológiai állomáson mért átlagos nappali és éjszakai szélirányokat mutatja. Hasonlítsuk össze a nappal és az éjszaka megfigyelt értékeket! Annak eldöntésére, hogy a parti-tavi cirkuláció iránya a napszaknak megfelelően változik-e alkalmazzon t-próbát! Ellenőrizze, hogy szignifikánsan különbözik-e a nappali és az éjszaki szélirányok átlaga!

Nap	szélirány nappal	szélirány éjszaka	Nap	szélirány nappal	szélirány éjszaka
1	120	170	16	140	155
2	130	235	17	120	180
3	190	190	18	150	200
4	150	170	19	160	220
5	130	195	20	170	100
6	140	130	21	180	210
7	120	230	22	110	150
8	250	190	23	210	220
9	180	230	24	160	220
10	155	210	25	180	240
11	180	215	26	190	180
12	230	250	27	180	210
13	185	190	28	170	240
14	170	200	29	160	210
15	160	290	30	210	190
			31	150	180

Megoldás

Az adatok alapján a nappali időszakra vonatkoztatható várható érték és szórás a következő

$$m_1 = 165{,}5 \text{ és } \sigma_1 = 32{,}7.$$

Ugyan ezek a mennyiségek az éjszakai időszakra vonatkoztatva:

$$m_2 = 200 \text{ és } \sigma_2 = 37{,}2.$$

A fenti összefüggésbe behelyettesítve ezeket az értékeket kapjuk, hogy $t = 3{,}83$.

A rendszer szabadsági foka, $szf = n_1 + n_2 - 2 = 60$

Illesszük be ezt az értéket a t-próbához javasolt táblázatba!

szabadsági fok	0,90	0,95	0,99	
40	1,684	2,021	2,704	
60	*1,671*	*2,000*	*2,660*	***3,83***
120	1,658	1,980	2,617	

Azaz annak valószínűsége, hogy a két várható érték megegyezik kisebb, mint 0,01. Tehát a mérési adatok alapján szignifikánsan különbözik a nappali és az éjszakai időszakra jellemző szélirány.

6.2. Az 5.1 feladathoz tartozó táblázat adatait felhasználva állapítsa meg, hogy az1901-30 időszak és az 1931-60 időszak átlaghőmérsékletei szignifikánsan különböznek-e! (t = 0,76, szabadsági fok = 58, a két időszakra vonatkozó átlaghőmérsékletek 0,1 –nél nagyobb valószínűséggel megegyeznek, azaz a null-hipotézist miszerint a két időszakra vonatkozó átlaghőmérsékletek megegyeznek elfogadhatjuk)

χ^2 –próba

Több különböző típusú hipotézisvizsgálat tartozik ebbe a csoportba. Ezek a következők:

a) Illeszkedésvizsgálat. Az illeszkedésvizsgálat során azt vizsgáljuk, hogy a mérési adatokból készített empirikus sűrűségfüggvény egy adott elméleti sűrűségfüggvénnyel írható-e le.

b) Függetlenségvizsgálat. A függetlenségvizsgálat során azt vizsgáljuk, hogy két mérési eljárásból származó adatsorok függetlennek tekinthetők- e vagy sem.

c) Homogenitásvizsgálat. A homogenitásvizsgálat során azt vizsgáljuk, hogy a két mérési eljárásból származó adatsorból készített hisztogarmok azonosnak tekinthetők vagy sem.

Az alábbiakban részletesen az illeszkedésvizsgálattal foglalkozunk. Számoljuk ki az alábbi mennyiséget:

$$\chi^2 = \sum_{i=1}^{r} \frac{(k_i - n \cdot p_i)^2}{n \cdot p_i}$$

ahol az empirikus sűrűségfüggvény (hisztogram intervallumainak a száma), k_i az i-ik intervallumba eső adatok száma, *n* az adatok száma és p_i az elméleti sűrűségfüggvény segítségével meghatározható valószínűség az *i*-ik

6.3. Legyen a null-hipotézis az, hogy a 4.4 feladathoz kapcsolódó táblázatban közölt adatok normális eloszlást követnek. Milyen valószínűséggel teljesül ez a feltételezés? (A számításokhoz a Δ = 5 °C-os intervallum szélességgel készített hisztogram adatait használja!)

Megoldás

i. Az adatok alapján meghatározzuk a sűrűségfüggvény két paraméterét, szórást és a várhatóértéket. m = 26,71, σ = 5,28

ii. Végezzük el az intervallumok határainak transzformációját!

Például:

$$14 \rightarrow \frac{14-26,71}{5,28} = -2,40$$

a többi határ esetében:

19	$\rightarrow$	-1,46
24	$\rightarrow$	-0,51
29	$\rightarrow$	0,43
34	$\rightarrow$	1,38
39	$\rightarrow$	2,33
44	$\rightarrow$	3,27

iii. Az A1 táblázat segítségével határozzuk meg a p_i valószínűségeket!

Felhasználva az A1 táblázatot az alábbi táblázatban megadtuk a (ii) pontban meghatározott *x* értékekhez tartozó *F(x)* függvény értékeket.

A fenti táblázat adatai alapján

p_1 = 0,0668 – 0,0080 = 0,0586

p_2 = 0,3085 – 0,0668 = 0,2417

p_3 = 0,6554 – 0,3085 = 0,3469

p_4 = 0,9192 – 0,6554 = 0,2638

p_5 = 0,9863 – 0,9192 = 0,0671

p_6 = 0,9994 – 0,9863 = 0,0131

Számítások ellenőrzése végett adjuk össze a fentiekben kiszámolt valószínűségeket, ha pontosan számoltunk, akkor jó közelítéssel 1-et kell kapnunk.

$$p_1 + p_2 + p_3 + p_4 + p_5 + p_6 = 0,9912$$

iv. Határozzuk meg az egy – egy intervallumba eső adatok számát.

intervallum	k_i	intervallum	k_i
14 - 18,9	6	29 - 33,9	20
19 - 23,9	12	34 - 38,9	2
24 - 28,9	20	39 - 43,9	0

v. Határozzuk meg a χ^2 értékét!

$$\chi^2 = \sum_{i=1}^{r} \frac{(k_i - n \cdot p_i)^2}{n \cdot p_i} = \frac{(6-60 \cdot 0,0586)^2}{60 \cdot 0,0586} + \frac{(12-60 \cdot 0,2417)^2}{60 \cdot 0,2417} + \dots \frac{(0-60 \cdot 0,0131)^2}{60 \cdot 0,0131} = 5,12$$

vi. Határozzuk meg a szabadsági fokok számát!

Szabadsági fokok száma = intervallumok száma – származtatott paraméterek száma –1

intervallumok száma = 6 ,

származtatott paraméterek száma = 2 (várhatóérték és szórás),

Így a szabadsági fokok száma = 6 – 2 – 1 = 3 .

vii. Az A3 táblázat segítségével határozzuk meg a null-hipotézis teljesülésének valószínűségét!

szabadsági fok	**1 – p**		
	0,90	**0,95**	**0,99**
2	4,61	5,99	9,21
3	*6,25*	*7,81*	*11,34*
4	7,78	9,49	13,28

Az általunk kiszámított értéket beillesztve a megfelelő sorba megállapíthatjuk, hogy a null-hipotézis teljesülésének valószínűsége nagyobb, mint 10 % . Így a feltételezést, miszerint a hőmérsékleti adatok normális eloszlást követnek elfogadhatjuk.

x	F(x)	x	F(x)
-3.2	0.0007	**0.0**	0.5000
-3.1	0.0010	**0.1**	0.5398
-3.0	0.0013	**0.2**	0.5793
-2.9	0.0019	**0.3**	0.6179
-2.8	0.0026	**0.4**	0.6554
-2.7	0.0035	**0.5**	0.6915
-2.6	0.0047	**0.6**	0.7257
-2.5	0.0062	**0.7**	0.7580
-2.4	0.0082	**0.8**	0.7881
-2.3	0.0107	**0.9**	0.8159
-2.2	0.0139	**1.0**	0.8413
-2.1	0.0179	**1.1**	0.8643
-2.0	0.0228	**1.2**	0.8849
-1.9	0.0287	**1.3**	0.9032
-1.8	0.0359	**1.4**	0.9192
-1.7	0.0446	**1.5**	0.9332
-1.6	0.0548	**1.6**	0.9452
-1.5	0.0668	**1.7**	0.9554
-1.4	0.0808	**1.8**	0.9641
-1.3	0.0968	**1.9**	0.9713
-1.2	0.1151	**2.0**	0.9772
-1.1	0.1357	**2.1**	0.9821
-1.0	0.1587	**2.2**	0.9861
-0.9	0.1841	**2.3**	0.9893
-0.8	0.2119	**2.4**	0.9918
-0.7	0.242	**2.5**	0.9938
-0.6	0.2743	**2.6**	0.9953
-0.5	0.3085	**2.7**	0.9965
-0.4	0.3446	**2.8**	0.9974
-0.3	0.3821	**2.9**	0.9981
-0.2	0.4207	**3.0**	0.9987
-0.1	0.4602	**3.1**	0.999
0.0	0.5000	**3.2**	0.9993

A1. táblázat.
A normális eloszlást leíró görbe alatti terület kiszámolásához szükséges adatok.

szabadsági fok	1 – p		
	0,90	**0,95**	**0,99**
1	6,314	12,706	63,657
2	2,920	4,303	9,925
3	2,353	3,182	5,841
4	2,132	2,776	4,604
5	2,015	2,571	4,032
6	1,943	2,447	3,707
7	1,895	2,365	3,499
8	1,860	2,306	3,355
9	1,833	2,262	3,250
10	1,812	2,228	3,169
11	1,796	2,201	3,106
12	1,782	2,179	3,055
13	1,771	2,160	3,012
14	1,671	2,145	2,977
15	1,753	2,131	2,947
16	1,746	2,120	2,921
17	1,740	2,110	2,898
18	1,734	2,101	2,878
19	1,729	2,093	2,861
20	1,725	2,089	2,845
21	1,721	2,080	2,831
22	1,717	2,074	2,819
23	1,714	2,069	2,807
24	1,711	2,064	2,797
25	1,708	2,060	2,787
26	1,706	2,056	2,779
27	1,703	2,052	2,771
28	1,701	2,048	2,763
29	1,699	2,045	2,756
30	1,697	2,042	2,750
40	1,684	2,021	2,704
60	1,671	2,000	2,660
120	1,658	1,980	2,617

A2. táblázat
Student – (t) eloszlás

szabadsági fok	1 – p		
	0,90	**0,95**	**0,99**
1	2,71	3,84	6,63
2	4,61	5,99	9,21
3	6,25	7,81	11,34
4	7,78	9,49	13,28
5	9,24	11,07	15,09
6	10,64	12,59	16,81
7	12,02	14,07	18,48
8	13,36	15,51	20,09
9	14,68	16,92	21,67
10	15,99	18,31	23,21
11	17,28	19,68	24,73
12	18,55	21,03	26,22
13	19,81	22,36	27,69
14	21,06	23,68	29,14
15	22,31	25,00	30,58
16	23,54	26,30	32,00
17	24,77	27,59	33,41
18	25,99	28,87	34,81
19	27,20	30,14	36,19
20	28,41	31,41	37,57
21	29,62	32,67	38,93
22	30,81	33,92	40,29
23	32,01	35,17	41,64
24	33,20	36,42	42,98
25	34,38	37,65	44,31
26	35,56	38,88	45,64
27	36,74	40,11	46,96
28	37,92	41,34	48,28
29	39,09	42,56	49,59
30	40,26	43,77	50,89
40	51,81	55,76	63,69
60	74,40	79,08	88,38
120	140,23	146,57	158,95

A3. táblázat
χ^2 -eloszlás