KLIMATOLÓGIAI STATISZTIKA FELADATOK

1. Tihanyban 22 év során 37 napon fordult elő 30 mm-t meghaladó 24 órai csapadék. Milyen valószínűséggel várható olyan év, amikor nem fordul elő 30 mm-t meghaladó 24 órai csapadékmennyiség? *(megoldás: 18,58 %)*

2. Egy hegycsúcsra telepítendő TV-torony műszaki tervezéséhez szükséges annak az ismerete, hogy mekkora a villámcsapás valószínűsége a csúcson. A környéken végzett megfigyelések szerint 8 év során 19 napon jegyeztek fel villámcsapást hegycsúcson. Mekkora annak a valószínűsége, hogy egy évben 3 napon lesz villámcsapás a csúcson? *(megoldás: 20,7 %)*

3. Mi annak a valószínűsége, hogy egy adott helyen az év 12 hónapjából egynek sem lesz magasabb a középhőmérséklete a felső kvartilisnél? *(megoldás: 3,17 %)*

4. Winnipeg januári középhőmérséklete -17,7 °C , a középhőmérsékletek szórása 4,1 °C. Normális eloszlást feltételezve határozzuk meg annak a középre szimmetrikus intervallumnak a végpontjait, amelyek közé az összes érték 50 %-a esik! *(megoldás: [-20,49; -14,91])*

5. Egy bizonyos növényfaj létfeltételeihez szükséges, hogy a vegetációs időszakban uralkodó havi középhőmérsékletek legalább 40 %-a 12 °C fölött legyen. Az *A* és *B* észlelőhelyeken a vegetációs időszak havi közepeinek megfelelő paraméterei a következők:

A: m = 11,5 °C, σ = 1 °C

B: m = 11,0 °C, σ = 4 °C

Alkalmasak-e az észlelőhelyek az illető növényfaj termesztésére? *(megoldás: A: P = 30,85 % $\Rightarrow$ nem alkalmas; B: P = 40,13 % $\Rightarrow$ alkalmas)*

6. Szegeden az 1901-1960 közötti 60 év adatai alapján a júniusi középhőmérséklet 20,5 °C. Kizárólag a hőmérsékletek feltételezetten normális eloszlását felhasználva határozzuk meg a júniusi középhőmérsékletek szórását, ha tudjuk, hogy a vizsgált 60 június közül 11-nek volt a középhőmérséklete legalább 22 °C. *(megodás: σ = 1,67 °C)*

7. Tekintsünk egy adott helyre vonatkozó több éves hőmérsékleti sort. Ez a hőmérsékleti sor havi közepekből áll. Annak a valószínűsége, hogy egy tetszőleges hónap középhőmérséklete a több éves sor közepe fölötti 0,5. Kiválasztva egy tetszőleges évszakot, mi a valószínűsége annak, hogy abban egyetlen hónap középhőmérséklete sem lesz a több éves sor átlaga

a) felett?

b) alatt?

(megoldás: a = b = 0,125)

8. Szombathelyen 140 év adatai alapján az áprilisi középhőmérséklet 9,7 °C, szórása 2,3 °C. Határozzuk meg az áprilisi középhőmérsékletek alsó és felső kvartilisét! *(megoldás: [8,1; 11,3])*

9. Szegeden a júniusi középhőmérséklet 20,4 °C, a szórás 1,2 °C. Határozzuk meg annak a számtani középre szimmetrikus intervallumnak az alsó és felső határát, amelybe 2/3 valószínűséggel esnek az értékek! *(megoldás: [19,32; 21,48])*

10. Hódmezővásárhelyen a 20 mm-t meghaladó napi csapadék átlagos évi gyakorisága 4 nap. Mi a valószínűsége annak, hogy egyetlen napon sem fordul elő egy évben és annak, hogy 5 napon fordul elő egy évben 20 mm-t meghaladó csapadékösszeg? *(megoldás: 1,8 %; 15,6 %)*

11. Budapest belterületén és külterületén 1975 júliusának és augusztusának 21 órai hőmérsékletei az alábbi esetszámok szerint oszlottak meg:

hőmérséklet, °C	**16-18**	**18-20**	**20-22**	**22-24**	**24-26**	**26-28**
belterület	5	10	16	18	10	3
külterület	8	14	25	10	5	0

Kimutatható-e, lényeges eltérés a belterület és a külterület esti hőmérsékleteinek eloszlásában? *(megoldás: nem mutatnak szignifikáns eltérést)*

12. Egy ipari városban levegőszűrő berendezésekkel csökkentették a levegőbe jutó égéstermékek mennyiségét. A beavatkozás előtti 10 és a beavatkozás utáni 3 éven át végzett megfigyelések szerint januárban 13 órakor a jó, a közepes és rossz látástávolság az alábbi esetszámban volt megfigyelhető:

látástávolság	**jó**	**közepes**	**rossz**
beavatkozás előtt	78	124	108
beavatkozás után	35	43	15

Kimutatható-e a műszaki beavatkozás a város levegőjének tisztulásában?
(megoldás: kimutatható)

13. Bresztben 60 év során a legcsapadékosabb évszakok megoszlása a következő volt:

évszak	**tél**	**tavasz**	**nyár**	**ősz**
eset	16	11	13	20

Fenntartható-e az az állítás, hogy Bresztben a csapadékmaximum egyenlő eséllyel várható bármely évszakban? *(megoldás: az állítás fenntartható)*

14. Sopronban az 1971-1975 közötti 5 év júniusi napjai közül 82 volt csapadékos. Adjunk 95 %-os megbízhatósági becslést a csapadékos nap valószínűségére júniusban! *(megoldás: $[P_1; P_2] = [0,47; 0,63]$)*

15. Pécsett 80 év megfigyelése alapján az augusztusi középhőmérséklet: 21,7 °C, a szórás: 1,3 °C. Az értékek az alábbi módon oszlanak meg:

hőmérséklet, °C	18-19	19-20	20-21	21-22	22-23	23-24	24-25
eset	2	7	12	22	23	9	5

Igazolható-e, hogy az adatok a normál eloszlás szerint oszlanak meg?
(megoldás: igazolható)

16. Orosházán hosszú megfigyelési sor alapján júliusban a 30 °C fölötti maximum hőmérsékletek bekövetkezési valószínűsége 30 %. Két év során előfordult 21 anticiklonális időjárású júliusi nap közül 14-en emelkedett a hőmérséklet 30 °C fölé. Jelentősen befolyásolja-e az anticiklonális időjárás a nyári hőség kialakulását? *(megoldás: jelentősen befolyásolja)*

17. Siófokon 10 év összes júniusi napja közül 75-ön jegyeztek föl 10 m s^{-1}-nál nagyobb széllökést. Ugyanezen időszak 60 zivataros napja közül 33-on fordult elő ekkora szélsebesség. Kimondható-e, hogy zivataros napokon jelentősen nagyobb az erős szél bekövetkezési valószínűsége? *(megoldás: kimondható)*

18. San Cristobal (Galápagosz-szigetek) megfigyelő helyen az átlagosnál csapadékosabb március bekövetkezési valószínűsége 46 %. 11 olyan márciusi hónapból, amikor a környező tenger vizének hőmérséklete legalább 1 °C-kal melegebb volt a szokásosnál, 8 volt az átlagosnál csapadékosabb. Igazolható-e, hogy meleg tenger környezetében a szokásosnál több a csapadék? *(megoldás: igazolható)*

19. Budapesten az átlagosnál csapadékosabb január bekövetkezési valószínűsége 46 %. 14 átlagosnál hidegebb január közül 9 volt az átlagosnál csapadékosabb. Kimondható-e az információk alapján, hogy a hideg januárok a szokásosnál csapadékosabbak? *(megoldás: nem mondható ki)*

20. Békéscsabán 80 évi megfigyelés alapján az októberi középhőmérséklet 12,4°C, a szórás 1,4 °C. Az adatok gyakorisági eloszlása a következő:

hőm, °C	**7-8**	**8-9**	**9-10**	**10-11**	**11-12**	**12-13**	**13-14**	**14-15**	**15-16**	**16-17**
eset	2	2	8	4	13	22	15	8	4	2

Igazolható-e az adatok normális eloszlás szerinti megoszlása? *(megoldás: igazolható)*

TÍPUSFELADATOK ÉS RÉSZLETES MEGOLDÁSI MENETÜK

1. BINOMIÁLIS ELOSZLÁSSAL KAPCSOLATOS FELADAT

1.1. Feladat:
Magyarországon szeptemberben az anticiklon-centrum makroszinoptikus helyzet bekövetkezési valószínűsége 21 %. Egymást követő 4 napot kiválasztva adjuk meg annak a valószínűségét, hogy ebből a 4 napból 3 napon fordul elő anticiklon-centrum időjárási helyzet!

Megoldási menet:

1. Alapkérdés: Adott egy p alapvalószínűségű alternatív esemény. n esetből k-szori bekövetkezése milyen valószínűséggel várható? Mivel az alapkérdés alkalmazható a feladatra, ezért az vagy a binomiális-, vagy a Poisson-eloszlás segítségével oldható meg.

2. Döntsük el a feladat típusát!
   Ha $p > 0{,}03 \Rightarrow$ binomiális eloszlással számolunk
   Ha $p < 0{,}03 \Rightarrow$ Poisson-eloszlással számolunk
   Mivel $p = 21\ \% = 0{,}21 \Rightarrow$ ***az eloszlás típusa: binomiális***

3. Határozzuk meg a binomiális eloszlás paramétereit!
   $n = 4$
   $k = 3$

4. Helyettesítsünk be a binomiális eloszlás képletébe!

$$P(k;n) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \qquad (1.1.)$$

   Innen:

$$P(3;4) = \frac{4!}{3!(4-3)!} 0{,}21^3 (1-0{,}21)^{4-3} \qquad (1.2.)$$

   Azaz:

$$P(3;4) = 4 * 0{,}21^3 * 0{,}79 = 0{,}029 \qquad (1.3.)$$

5. Válaszoljunk a feladatra!
   Tehát annak a valószínűsége, hogy Magyarországon a kiindulási feltételek mellett egymást követő 4 nap közül 3 napon fordul elő anticiklon-centrum időjárási helyzet: P = 0,029.

2. POISSON ELOSZLÁSSAL KAPCSOLATOS FELADAT

2.1. Feladat:
Nyíregyházán áprilisban 50 év alatt összesen 28 napon fordult elő fagy. Mi a valószínűsége a fagymentes áprilisnak Nyíregyházán?

Megoldási menet:

1. Alapkérdés: Adott egy p alapvalószínűségű alternatív esemény. n esetből k-szori bekövetkezése milyen valószínűséggel várható? Mivel az alapkérdés alkalmazható a feladatra, ezért az vagy a binomiális, vagy a Poisson-eloszlás segítségével oldható meg.

2. Döntsük el a feladat típusát!
   Ha $p > 0{,}03 \Rightarrow$ binomiális eloszlással számolunk
   Ha $p < 0{,}03 \Rightarrow$ Poisson-eloszlással számolunk
   Mivel:
   a keresett esemény, azaz a fagyos áprilisi napok száma: k = 28 nap
   az összes esemény, azaz az összes vizsgált áprilisi napok száma:
   n = 50 év*30 nap = 1500 nap
   Ezért:

$$p = \frac{k}{n} = \frac{28}{50*30} < \frac{30}{50*30} = \frac{1}{50} = \frac{2}{100} = 0{,}02 < 0{,}03 \Rightarrow \textbf{\textit{az eloszlás típusa: Poisson}}$$

3. Határozzuk meg a Poisson eloszlás paramétereit!
   k = 0 (azaz 0 napon következzen be fagy egy tetszőleges – n = 30 nap tartamú – áprilisban)

$$n*p = 30*\frac{28}{50*30} = \frac{28}{50} = 0{,}56$$

4. Helyettesítsünk be a Poisson eloszlás képletébe!

$$P(k; n*p) = \frac{(n*p)^k * e^{-(n*p)}}{k!} \qquad (2.1.)$$

   Innen:

$$P(0; 0{,}56) = \frac{0{,}56^0 * e^{-0{,}56}}{0!} = e^{-0{,}56} = \frac{1}{e^{0{,}56}} \qquad (2.2.)$$

   Mivel az *e* természetes logaritmus alapszáma: e = 2,72, így P(0; 0,56) = 0,571.

5. Válaszoljunk a feladatra!
   Tehát annak a valószínűsége, hogy Nyíregyházán áprilisban a kiindulási feltételek mellett ne forduljon elő fagyos nap: <u>P = 0,571</u>.

3. NORMÁLIS ELOSZLÁSSAL KAPCSOLATOS FELADATOK

3.1. Feladat:
Szombathelyen az áprilisi középhőmérséklet 9,7 ºC, a középhőmérsékletek szórása 2,3 ºC. Mennyi a valószínűsége annak, hogy az áprilisi középhőmérséklet 8 ºC-nál alacsonyabb?

Megoldási menet:

1. Döntsük el a feladat típusát!
   Mivel a középhőmérsékletek eloszlása normális, ezért ***normális eloszlással kapcsolatos feladat***ról van szó.

2. Adjuk meg az eloszlás paramétereit!
   m = 9,7 ºC
   σ = 2,3 ºC

3. Oldjuk meg ***grafikusan*** a feladatot!
   Fölrajzoljuk az m és σ paraméterekkel megadott normális eloszlás sűrűségfüggvényét. Bejelöljük a vízszintes tengelyen a 8 ºC-ot, s az annál kisebb értékekkel jellemzett görbe alatti terület a keresett valószínűség, azaz $F(x_a = 8\ ºC)$. Ugyanis:

   $$F(x_a = 8\ ºC) = P(x < x_a = 8\ ºC) \qquad (3.1.)$$

4. Oldjuk meg ***numerikusan*** a feladatot!
   Elvégezzük az

   $$x_a \longrightarrow \frac{x_a - m}{\sigma} = d_a \qquad (3.2.)$$

   transzformációt, amely tetszőleges N (m ≠ 0; σ ≠ 1) eloszlást N (m = 0; σ = 1) eloszlássá (standard normális eloszlássá) alakít. Erre a transzformációra azért van szükség, mert az N (m = 0; σ = 1) eloszlásnak, és csak ennek ismert az eloszlásfüggvénye. Azaz az $F(x_a) = P(x < x_a)$ minden egyes x_a-ra csak az N (m = 0; σ = 1) eloszlásra ismert, mely táblázatosan adott. Ily módon bármelyik normális eloszlással kapcsolatos feladat ennek a transzformációnak a segítségével oldható csak meg. Ez a transzformáció egy kölcsönösen egyértelmű megfeleltetés, mely az eredeti eloszlás minden egyes értékéből egy új értéket képez. Számunkra az x_a = 8 ºC, illetve az ennek megfelelő transzformált érték ($x_a \longrightarrow \frac{x_a - m}{\sigma} = d_a = -0{,}74$) az érdekes, mert a transzformációval előállított d_a független változóhoz tartozó $F(d_a)$ függvényérték a standard normális eloszlás eloszlásfüggvényének a táblázatából kikereshető (tankönyv, 297. oldal, 5.4. táblázat; Péczely, 1979). Eszerint, lineáris interpolációval $F(d_a) = 18{,}6\ \%$.
   Most alkalmazva az $F(x_a) = P(x < x_a)$ valószínűségi eloszlásfüggvény definícióját: az x_a = 8 ºC az az érték, amelynél kisebb értékek előfordulási valószínűsége a mintában 18,6 %. Vagyis képletben leírva: $F(8\ ºC) = P(x < 8\ ºC) = 18{,}6\ \%$.

5. Válaszoljunk a feladatra!
   Tehát annak a valószínűsége, hogy Szombathelyen a kiindulási feltételek mellett 8 ºC-nál alacsonyabb legyen az áprilisi középhőmérséklet: 18,6 %.

3.2. Feladat:
Szombathelyen az áprilisi középhőmérséklet 9,7 ºC, a középhőmérsékletek szórása 2,3 ºC. Határozzuk meg annak a számtani középre szimmetrikus intervallumnak az alsó és felső határát, amelybe 50 %-os valószínűséggel esnek az értékek!

Megoldási menet:

1. Döntsük el a feladat típusát!
   Mivel a középhőmérsékletek eloszlása normális, ezért ***normális eloszlással kapcsolatos feladat***ról van szó.

2. Adjuk meg az eloszlás paramétereit!
   m = 9,7 ºC
   σ = 2,3 ºC

3. Oldjuk meg ***grafikusan*** a feladatot!
   Fölrajzoljuk az m és σ paraméterekkel megadott normális eloszlás sűrűségfüggvényét. Bejelöljük a vízszintes tengely fölé a minta középértéke által megfelezett, s a teljes minta 25-25 %-át tartalmazó intervallumokat. Ennek az intervallumnak a végpontjaira vagyunk kíváncsiak.

4. Oldjuk meg ***numerikusan*** a feladatot!
   Elvégezzük a 3.1. feladatból már ismert

$$x_a \longrightarrow \frac{x_a - m}{\sigma} = d_a \qquad (3.2.)$$

   transzformációt, amely tetszőleges N (m ≠ 0; σ ≠ 1) eloszlást N (m = 0; σ = 1) eloszlássá (standard normális eloszlássá) alakít. Tudjuk továbbá, hogy minden egyes d_a-hoz tartozik egy és csak egy $F(d_a)$ a standard normális eloszlás eloszlásfüggvényének a táblázatából (tankönyv, 297. oldal, 5.4. táblázat; Péczely, 1979). Azaz: $d_a \longrightarrow F(d_a)$.
   A 3.2. feladat azonban a 3.1. feladathoz képest fordított. Ugyanis ez esetben az $F(d_a)$-t ismerjük, s a d_a-t kell meghatároznunk. Méghozzá két feladat megoldásáról van szó. A valószínűségi eloszlásfüggvény definíciójának figyelembe vételével egyrészt $F(d_{a1})=25\ \%$, amelyhez tartozó intervallumnak a d_{a1} alsó végpontját keressük; másrészt $F(d_{a2})=75\ \%$, amelyhez tartozó intervallumnak a d_{a2} felső végpontját keressük. Azaz:

$$\begin{aligned} &\text{a)}\quad F(d_{a1})=25\ \%, \longrightarrow d_{a1} = ? \\ &\text{b)}\quad F(d_{a2})=75\ \%, \longrightarrow d_{a2} = ? \end{aligned} \qquad (3.3.)$$

   Most már egyszerű a feladatunk, hiszen az ismert $F(d_{a1})$ és $F(d_{a2})$ függvényértékekhez tartozó d_{a1} és d_{a2} független változókat kikeressük a már említett táblázatból (tankönyv, 297. oldal, 5.4. táblázat; Péczely, 1979). Ezt követően, a d_{a1} (= –0,73) és d_{a2} (= +0,73) értékek ismeretében a (3.2.) egyenlet segítségével az eredeti x_{a1} (= 8,02) és x_{a2} (= +11,38) értékek könnyen kiszámíthatók.

5. Válaszoljunk a feladatra!
   Tehát Szombathelyen a kiindulási feltételek mellett áprilisban az a számtani középre szimmetrikus intervallum, amelybe 50 %-os valószínűséggel esnek az értékek, a következő: <u>(8,02 ºC; 11,38 ºC)</u>.

4.1. χ^2-PRÓBA TISZTA ILLESZKEDÉS-VIZSGÁLATTAL KAPCSOLATOS FELADAT

4.1. Feladat:
Budapesten az 1941-1960 közötti 20 év megfigyelései alapján az áprilisi maximum-hőmérséklet az alábbi gyakoriságban fordult elő a hónap első, illetve második felében:

	április	
	I. fele	II. fele
Észlelt gyakoriság (É)	5	15
Feltételezett gyakoriság (F)	10	10

Feltehetjük-e azt, hogy Budapesten az áprilisi maximum-hőmérséklet a hónap első, illetve második felében egyenlő eséllyel várható?

Megoldási menet:

1. Állítsuk föl a 0-hipotézist!
   Mivel a feladatban eldöntendő kérdésre kell válaszolnunk, egy a „priori" (előzetes, a tapasztalattól független, a számításokat megelőző) feltevést teszünk. Eszerint azt mondjuk (minden hasonló kérdésfeltevés esetén), hogy az áprilisi maximum-hőmérsékletek nem különböznek lényegesen a hónap első, illetve második felében.

2. Döntsük el a feladat típusát!
   <u>1. közelítés:</u>
   Ahhoz, hogy ezt az előzetes (a priori) feltevésünk igaz, vagy hamis voltát statisztikai alapon eldönthessük, a χ^2-próba módszerét alkalmazzuk a feladat megoldására.
   Ha nincs lényeges eltérés, akkor ez az a priori feltevésünk szerint azt jelenti [azaz azt feltételezzük (F)], hogy 10 alkalommal a hónap első felében, 10 alkalommal pedig a hónap második felében lépett föl az áprilisi maximum-hőmérséklet.
   <u>2. közelítés:</u>
   A feladatban fölvetett kérdést úgy is megfogalmazhatjuk, hogy az adott 20 éves mintában észlelt (É) kéttagú eloszlás <u>mennyire jól illeszkedik</u> az általunk a 0-hipotézis révén feltételezett (F) kéttagú egyenletes eloszláshoz? Ezen utóbbi, módosított kérdésfelvetés révén a feladattípust pontosíthatjuk: χ^2-próba, illeszkedésvizsgálat.
   <u>3. közelítés:</u>
   Mivel a feltételezett minta (F) – amelyhez az észlelt minta (É) illeszkedését vizsgáljuk – egyetlen elemével adott, ezért e feladat esetében ***χ^2-próba, tiszta illeszkedésvizsgálat***ról beszélünk.

3. Helyettesítsünk be a χ^2-próba képletébe!
   A próba képlete a következő:

$$\chi^2 = \sum_{i=1}^{n} \frac{(É_i - F_i)^2}{F_i} = \frac{(É_1 - F_1)^2}{F_1} + \frac{(É_2 - F_2)^2}{F_2} + \ldots + \frac{(É_n - F_n)^2}{F_n} \qquad (4.1.)$$

   Behelyettesítve:

$$\chi^2 = \sum_{i=1}^{n} \frac{(É_i - F_i)^2}{F_i} = \frac{(5-10)^2}{10} + \frac{(15-10)^2}{10} = 2,5 + 2,5 = 5 \qquad (4.2.)$$

Ahhoz, hogy az általunk imént kapott χ^2-érték ismeretében dönthessünk a 0-hipotézis teljesedését, vagy nem teljesedését illetően, meg kell határoznunk a szabadsági fokok számát. A szabadsági fok (sz.f.) az általunk fölállított kontingencia-táblázatban heurisztikusan az oszloponként szabadon megválasztható paraméterek számát jelenti. Azaz:

$$\text{sz.f.} = \text{o} - 1 \qquad (4.3.)$$

ahol o = az oszlopok száma a kontingencia-táblázatban.

A feladatunkban – mivel a kontingencia-táblázat oszlopainak a száma = 2, ezért: sz.f. = (2 – 1) = 1.

Ily módon, a kapott χ^2-érték (χ^2 = 5) és szabadsági fok (sz.f. = 1) mellett adott valószínűségi szinten meghatározhatjuk a 0-hipotézis teljesedését, vagy nem teljesedését a tankönyv, 315. oldalán található 5.11. táblázat alapján (Péczely, 1979).

A táblázatban található valószínűségi szintek (v.sz. = 10 %, 5 %, 1 %, 0,1 %) jelentése a következő. Ezek a valószínűségek az igaz 0-hipotézis elvetésének a valószínűségét jelentik. Azaz, ha pl. az 5 %-os valószínűségi szinten döntünk (általában ezt választjuk), akkor olyan küszöbérték ($\chi^{2;sz.f.}_{v.sz.} = \chi^{2;1}_{0,05}$ = 3,86) tartozik az sz.f. = 1 értékhez, hogy 100 esetből átlagosan 5 esetben ennél a küszöbértéknél nagyobb számított χ^2-értéket kapunk, s elvetjük a 0-hipotézist, holott az valójában igaz. Csupán az adatok véletlenszerű elrendeződése folytán kapunk olyan nagy χ^2-értéket, hogy az túllépi a küszöböt (tankönyv, 315. oldal, 5.11. táblázat; Péczely, 1979).

Itt rögtön meg kell jegyezzük, hogy amikor e táblázat alapján döntünk, kétféle hibát követhetünk el: a) igaz 0-hipotézist elvetünk (elsőfajú hiba), illetve b) hamis 0-hipotézist elfogadunk (másodfajú hiba).

Ha – bármely tetszőleges szabadsági fok mellett – csökken a valószínűségi szint, akkor nő a χ^2-küszöbérték. Ez azt jelenti, hogy ha csökken a valószínűségi szint, akkor csökken ugyan az igaz 0-hipotézis elvetésének a valószínűsége, vagyis az elsőfajú hiba (hiszen nő a χ^2-küszöbérték!), másrészről viszont – ezzel párhuzamosan – nő a hamis 0-hipotézis elfogadásának a valószínűsége, vagyis a másodfajú hiba (pontosan amiatt, hogy nő a χ^2-küszöbérték!). Tudnunk kell, hogy a mindenkori döntésünk a 0-hipotézis teljesedését, vagy nem teljesedését illetően csak statisztikai értelemben igaz. Ugyanis minden döntésünk tartalmaz bizonyos százalékban mind elsőfajú hibát, mind másodfajú hibát. Célunk az, hogy mindkét hibát minimalizáljuk.

Ha a χ^2-próba képletébe történő behelyettesítést követően olyan nagy χ^2-értéket kapunk, hogy az a 0,1 %-os valószínűségi szinten is szignifikáns, az azt jelenti, hogy ehhez a valószínűségi szinthez olyan magas χ^2-küszöbérték ($\chi^{2,sz.f.}_{v.sz.}$) tartozik, hogy 1000 esetből átlagosan mindössze egy esetben vetünk el igaz 0-hipotézist. Ugyanakkor látnunk kell, hogy ha csökken a valószínűségi szint, akkor – ezzel párhuzamosan – nő a másodfajú hiba, amelynek értékéről közelebbit nem tudunk mondani.

Ha a χ^2-próba képletébe történő behelyettesítést követően nagyon kicsi χ^2-értéket kapunk, akkor elmondhatjuk, hogy (ha $\chi^2_{számított} < \chi^2_{küszöbérték}$) teljesül az adott valószínűségi szinten a 0-hipotézis. Sőt, további számítással a 0-hipotézis teljesedésének a konkrét %-os valószínűsége is meghatározható.

A konkrét feladatunk esetében tehát $\chi^2 = 5$; sz.f. = 1; a választott valószínűségi szint: 5 %. A tankönyv 315. oldal, 5.11. táblázat alapján (Péczely, 1979) azt kapjuk, hogy

$$\chi^2_{számított} = 5{,}00 > \chi^2_{küszöbérték} = \chi^{2;sz.f.}_{v.sz.} = \chi^{2;1}_{0,05} = 3{,}86\ . \qquad (4.4.)$$

A döntési kritériumunk a következő:

ha $\chi^2_{számított} < \chi^2_{küszöbérték}$ (= $\chi^{2,sz.f.}_{v.sz.}$), akkor megtartjuk a 0-hipotézist,
ha $\chi^2_{számított} > \chi^2_{küszöbérték}$ (= $\chi^{2,sz.f.}_{v.sz.}$), akkor elvetjük a 0-hipotézist a választott valószínűségi szinten.

4. Válaszoljunk a feladatra!
   Számításaink alapján a 0-hipotézis nem teljesül. Tehát nem igaz az, hogy Budapesten áprilisban a maximum-hőmérséklet a hónap első, illetve második felében egyenlő eséllyel várható.

4.2. χ^2-PRÓBA BECSLÉSES ILLESZKEDÉS-VIZSGÁLATTAL KAPCSOLATOS FELADAT

4.2. Feladat:
Békéscsabán 80 évi megfigyelés alapján az október középhőmérséklete 12,4 ºC, s az októberi középhőmérsékletek szórása 1,4 ºC. Az értékek osztályközös gyakorisági eloszlása az alábbi:

hőmérséklet, ºC	7 – 8	8 – 9	9 – 10	10 – 11	11 – 12	12 – 13	13 – 14	14 – 15	15 – 16	16 – 17
gyak., db (É)	2	2	8	4	13	22	15	8	4	2
gyak., db (F)										

Igazolható-e, hogy az adatok a normális eloszlás szerint oszlanak meg?

Megoldási menet:

1. Állítsuk föl a 0-hipotézist!
   Mivel a feladatban eldöntendő kérdésre kell válaszolnunk, egy a „priori" (előzetes, a tapasztalattól független, a számításokat megelőző) feltevést teszünk. Eszerint azt mondjuk, hogy nincs lényeges eltérés (mivel 0-hipotézist állítunk föl, ezért 0 az eltérés) a fönti táblázatban az észlelt osztályközös gyakorisági értékek (É), illetve a normális eloszlás feltételezésével számított osztályközös gyakorisági értékek (F) eloszlása között.

2. Döntsük el a feladat típusát!
   <u>1. közelítés:</u>
   Ahhoz, hogy ezt az előzetes (a priori) feltevésünk igaz, vagy hamis voltát statisztikai alapon eldönthessük, a χ^2-próba módszerét alkalmazzuk a feladat megoldására.
   <u>2. közelítés:</u>
   Arra vagyunk kíváncsiak, hogy az eredeti minta osztályközös gyakorisági eloszlása mennyire jól illeszkedik a normalitás feltételezésével számított minta osztályközös gyakorisági eloszlásával. Tehát a feladat típusa illeszkedésvizsgálat.
   <u>3. közelítés:</u>
   Mivel a feltételezett minta elemeit (F) – amelyhez az észlelt minta (É) illeszkedését vizsgáljuk – az észlelt minta paramétereiből (m =12,4 ºC és σ = 1,4 ºC) becsüljük, ezért e feladat esetében ***χ^2-próba, becsléses illeszkedésvizsgálat***ról beszélünk.

3. Határozzuk meg a normalitás feltételezésével számított minta osztályközös feltételezett gyakorisági értékeit (F)!
   Fontos megjegyezni, hogy a χ^2-próba képletében az összeadandó tagok szélsőségesen nagy értékeit (ezáltal a potenciálisan torz számított χ^2-értéket) elkerülendő, az észlelt (É) és a feltételezett (F) cellaelemek értéke nem lehet kisebb 5-nél. Hogy ezt szükség esetén elérjük, a szomszédos oszlopok cellaelemeit soronként összevonjuk. A későbbi kevesebb számolás miatt ezt az összevonást az 5-nél kisebb észlelt (É) értékek esetében már akkor elvégezzük, amikor a hozzájuk tartozó feltételezett (F) értékek még nem ismeretesek. Az összevonásokat mindaddig folytatjuk, amíg a legkisebb cellaelem értéke eléri, vagy meghaladja az 5-öt. Mivel a fentiek alapján a feladatban elvégezzük a szükséges összevonásokat, az alábbi kontingencia-táblázathoz jutunk:

hőmérséklet, ºC	7 – 10	10 – 12	12 – 13	13 – 14	14 – 15	15 – 17
gyak., db (É)	12	17	22	15	8	6
gyak., db (F)						

Ily módon az összevonásokkal kapott – kevesebb oszlopból álló – kevesebb osztályközös gyakorisági érték, ennélfogva kevesebb számolás révén oldjuk meg a feladatot.
Minden egyes (a normalitás feltételezésével) számított cellaelem (F-érték) meghatározása egy-egy normális eloszlással kapcsolatos feladat megoldását jelenti. Azaz, rendre:

3.1. Határozzuk meg az észlelt minta adatai (cellaelemei) (É) ismeretében – azok feltételezetten normális eloszlása alapján – annak a valószínűségét, hogy az értékek a [7 ºC – 10 ºC] intervallumba esnek!

Nyilván a 7 ºC-nál kisebb értékek bekövetkezésének is lesz egy csekély valószínűsége, azonban ezt a töredékvalószínűséget hozzáadjuk a [7 ºC – 10 ºC] intervallumba esés valószínűségéhez. Ily módon a [7 ºC – 10 ºC] intervallumba esés valószínűsége megegyezik a 10 ºC-nál kisebb értékek bekövetkezési valószínűségével a mintában. Ha ezt a $p_{10\,ºC} = F(x_a = 10\,ºC) = P(x < x_a = 10\,ºC)$ valószínűséget meghatározzuk, akkor a $p = \frac{k}{n}$ formula felhasználásával a $k = n*p$ osztályközös gyakoriság már kiszámítható (itt n a minta összes elemszáma, azaz n = 80).
Mivel a végtelen sok normális eloszlás közül csak az N (m = 0; σ = 1) eloszlásnak ismert az eloszlásfüggvénye, ezért – hogy megoldhassuk a feladatot – elvégezzük a már ismert

$$x_a \longrightarrow \frac{x_a - m}{\sigma} = d_a \qquad (3.2.)$$

transzformációt, amely tetszőleges N (m ≠ 0; σ ≠ 1) eloszlást N (m = 0; σ = 1) eloszlássá (standard normális eloszlássá) alakít. Tudjuk továbbá, hogy minden egyes d_a-hoz tartozik egy és csak egy $F(d_a)$ a standard normális eloszlás eloszlásfüggvényének a táblázatából (tankönyv, 297. oldal, 5.4. táblázat; Péczely, 1979). Azaz: $d_a \longrightarrow F(d_a)$. Mivel az eredeti és a transzformált normális eloszlású sűrűségfüggvényekben az eredeti és a transzformált intervallumvégpontok közé eső görbe alatti területek (azaz valószínűségek) nem változnak, ezért $F(d_a) = F(x_a)$. Következésképpen az osztályközös valószínűségek ily módon meghatározhatók, amelyekből az osztályközös gyakorisági értékek a $k = n*p$ formulával kiszámíthatók.
A mi esetünkben x_a = 10 ºC lesz, azaz a vizsgált intervallum felső végpontja. Mivel m= 12,4 ºC, és σ = 1,4 ºC, ekkor:

transzformáció	d_a	$F(d_a)$	$P = F(x_b) - F(x_a) = P(x < x_b) - P(x < x_a)$	Feltételezett gyakoriság (F), db, k = n*p	Felt. gyak. (F), db (kerekítve)
$\frac{x_a - m}{\sigma} = d_a = \frac{10 - 12,4}{1,4} = -\frac{2,4}{1,4} = -1,71 = d_a \rightarrow F(d_a) = 0,04$				80*0,04 = 3,2	3

3.2. Határozzuk meg az észlelt minta adatai (cellaelemei) (É) ismeretében – azok feltételezetten normális eloszlása alapján – annak a valószínűségét, hogy az értékek a [10 ºC – 12 ºC] intervallumba esnek!

transzformáció	d_a	$F(d_a)$	$P = F(x_b) - F(x_a) = P(x < x_b) - P(x < x_a)$	Feltételezett gyakoriság (F), db, k = n*p	Felt. gyak. (F), db (kerekítve)
$\frac{x_a - m}{\sigma} = d_a = \frac{12 - 12,4}{1,4} = -\frac{0,4}{1,4} = -0,28 = d_a \rightarrow F(d_a) = 0,39$			0,39 – 0,04 = 0,35	80*0,35 = 28,0	28

3.3. Határozzuk meg az észlelt minta adatai (cellaelemei) (É) ismeretében – azok feltételezetten normális eloszlása alapján – annak a valószínűségét, hogy az értékek a [12 ºC – 13 ºC] intervallumba esnek!

transzformáció	d_a	$F(d_a)$	$P = F(x_b) - F(x_a) = P(x < x_b) - P(x < x_a)$	Feltételezett gyakoriság (F), db, k = n*p	Felt. gyak. (F), db (kerekítve)
$\frac{x_a - m}{\sigma} = d_a = \frac{13-12,4}{1,4} = \frac{0,6}{1,4} = 0,42 = d_a \rightarrow F(d_a) = 0,66$			0,66 – 0,39 = 0,27	80*0,27 = 21,6	22

3.4. Határozzuk meg az észlelt minta adatai (cellaelemei) (É) ismeretében – azok feltételezetten normális eloszlása alapján – annak a valószínűségét, hogy az értékek a [13 ºC – 14 ºC] intervallumba esnek!

transzformáció	d_a	$F(d_a)$	$P = F(x_b) - F(x_a) = P(x < x_b) - P(x < x_a)$	Feltételezett gyakoriság (F), db, k = n*p	Felt. gyak. (F), db (kerekítve)
$\frac{x_a - m}{\sigma} = d_a = \frac{14-12,4}{1,4} = \frac{1,6}{1,4} = 1,14 = d_a \rightarrow F(d_a) = 0,87$			0,87 – 0,66 = 0,21	80*0,27 = 16,8	17

3.5. Határozzuk meg az észlelt minta adatai (cellaelemei) (É) ismeretében – azok feltételezetten normális eloszlása alapján – annak a valószínűségét, hogy az értékek a [14 ºC – 15 ºC] intervallumba esnek!

transzformáció	d_a	$F(d_a)$	$P = F(x_b) - F(x_a) = P(x < x_b) - P(x < x_a)$	Feltételezett gyakoriság (F), db, k = n*p	Felt. gyak. (F), db (kerekítve)
$\frac{x_a - m}{\sigma} = d_a = \frac{15-12,4}{1,4} = \frac{2,6}{1,4} = 1,85 = d_a \rightarrow F(d_a) = 0,97$			0,97 – 0,87 = 0,10	80*0,10 = 8,0	8

3.5. Határozzuk meg az észlelt minta adatai (cellaelemei) (É) ismeretében – azok feltételezetten normális eloszlása alapján – annak a valószínűségét, hogy az értékek a [14 ºC – 15 ºC] intervallumba esnek!

Ez a feladatot viszont már egyszerűen meg tudjuk oldani. Ugyanis, ha az összes mintaelemszámból (n = 80) levonjuk a normalitás feltételezésével eddig kiszámított osztályközös gyakoriságokat (F-értékeket), akkor az utolsó osztályköz feltételezett gyakorisági értéke: k = 80 – (3 + 28+ 22 + 17 + 8) = 80 – 78 = 2 lesz. Nyilván az előző eljárás alkalmazásával ugyanezt az eredményt kapnánk.
Most írjuk be a feltételezett értékeket az összevonásokkal már leegyszerűsített táblázatunkba!

hőmérséklet, ºC	7 – 10	10 – 12	12 – 13	13 – 14	14 – 15	15 – 17
gyak., db (É)	12	17	22	15	8	6
gyak., db (F)	3	28	22	17	8	2

Látjuk, hogy 5-nél kisebb feltételezett értékeket kaptunk a [7 ºC – 10 ºC], illetve a [15 ºC – 17 ºC] intervallumokban. Ezért újabb összevonásokra kerül sor. Fontos megjegyeznünk, hogy ha valamely oszlopban csak az egyik cellaelem kisebb 5-nél, akkor is annak mindkét cellaelemét összevonjuk a szomszédos oszlop megfelelő cellaelemeivel. Ennek figyelembe vételével a következő kontingencia-táblázathoz jutunk:

hőmérséklet, ºC	7 – 12	12 – 13	13 – 14	14 – 17
gyak., db (É)	29	22	15	14
gyak., db (F)	31	22	17	10

4. Helyettesítsünk be a χ^2-próba képletébe!

$$\chi^2 = \sum_{i=1}^{n} \frac{(É_i - F_i)^2}{F_i} = \frac{(29-31)^2}{31} + \frac{(22-22)^2}{22} + \frac{(15-17)^2}{17} + \frac{(14-10)^2}{10} = \frac{4}{31} + 0 + \frac{4}{17} + \frac{16}{10} = 1{,}96$$

Abban az esetben, ha a feladat annak eldöntése, hogy egy adott minta tekinthető-e normális eloszlásúnak (jelen feladat), akkor az észlelt minta osztályközös gyakorisági értékeihez ($É_i$) a feltételezett minta megfelelő osztályközös gyakorisági értékeit úgy kapjuk meg, hogy utóbbiakat az észlelt minta paramétereiből (m; σ) becsüljük, miután feltettük, hogy az észlelt minta eloszlása normális. Ebben az esetben a szabadsági fokok meghatározás a következő:

$$\text{sz.f.} = o - b - 1,$$

ahol b = a becsült paraméterek száma.
Mivel a feladat alapján 2 becsült paraméterünk van (m és σ), a szabadsági fokok száma:

$$\text{sz.f.} = o - b - 1 = 4 - 2 - 1 = 1\ .$$

Fontos megjegyezni, hogy a szabadsági fokokat mindig az összevonások után kapott oszlopszámok figyelembe vételével határozzuk meg!
Ezután a tankönyv, 315. oldal, 5.11. táblázata (Péczely, 1979) alapján döntünk a 0-hipotézisünk teljesedését, illetve nem teljesedését illetően:

$$\chi^2_{számított} = 1{,}96 < \chi^{2;sz.f.}_{v.sz.} = \chi^{2;1}_{0,05} = 3{,}84\,.$$

ahol v.sz. a választott valószínűségi szint, azaz a mi esetünkben: v.sz. = 0,05.

5. Válaszoljunk a feladatra!
Mivel az általunk számított χ^2-érték kisebb, mint a választott valószínűségi szinthez és a számított szabadsági fokhoz tartozó küszöbérték, ezért a 0-hiptézist elfogadjuk. Azaz nincs lényeges különbség az észlelt osztályközös gyakorisági értékek (É), illetve a normális eloszlás feltételezésével számított osztályközös gyakorisági értékek (F) eloszlása között. Azaz, az észlelt minta normális eloszlásúnak tekinthető.

4.3. χ^2-PRÓBA HOMOGENITÁS-VIZSGÁLATTAL KAPCSOLATOS FELADAT

4.3. Feladat:
Egy dombos területen két nyári hónapon át naponta mértük a hajnali minimum-hőmérsékletet egy kiválasztott domb tetején és a nála 30 m-rel mélyebben fekvő völgyfenéken. Az észlelt (É) értékek az alábbi osztályközös gyakoriságokban fordultak elő:

hőmérséklet, ºC	6 – 8	8 – 10	10 – 12	12 – 14	14 – 16	16 – 18	18 – 20	*sorösszeg*$(\sum_{.j})$
dombtető (É)	1	5	13	12	18	7	4	60
völgyfenék (É)	7	10	20	12	6	5	–	60
oszlopösszeg$(\sum_{i.})$	8	15	33	24	24	12	4	120

Igazolható-e, hogy a dombtető és a völgyfenék hajnali minimum-hőmérsékleteinek eloszlásában nincs lényeges eltérés?

Megoldási menet:

1. Állítsuk föl a 0-hipotézist!
   Mivel a feladatban eldöntendő kérdésre kell válaszolnunk, egy a „priori” (előzetes, a tapasztalattól független, a számításokat megelőző) feltevést teszünk. Eszerint azt mondjuk, hogy a dombtető és a völgyfenék hajnali minimum-hőmérsékleteinek eloszlása nem tér el lényegesen egymástól.

2. Döntsük el a feladat típusát!
   <u>1. közelítés:</u>
   Ahhoz, hogy ezt az előzetes (a priori) feltevésünk igaz, vagy hamis voltát statisztikai alapon eldönthessük, a χ^2-próba módszerét alkalmazzuk a feladat megoldására.
   <u>2. közelítés:</u>
   Itt más a helyzet, mint a tiszta, illetve a becsléses illeszkedés-vizsgálat esetében. Ugyanis az utóbbiaknál rendelkezésünkre állt egy észlelt minta, s annak egy feltételezett mintához való illeszkedésének a jóságát vizsgáltuk. Viszont ennél a feladatnál mindkét minta észlelt adatokat tartalmaz. Emiatt a feladat más megközelítést igényel. Itt arra vagyunk kíváncsiak, hogy a két észlelt minta azonos alapsokaságból származik-e, azaz tekinthetők-e azonos eloszlásúaknak, függetlenül az eloszlás típusától. Az ilyen típusú feladatokat a ***χ^2-próba, homogenitás-vizsgálat*** segítségével oldjuk meg.

3. Határozzuk meg a χ^2-próba alkalmazásához a hajnali minimum-hőmérsékletek feltételezett gyakorisági értékeit (F) mind a dombtetőn, mind a völgyfenéken!
   Mielőtt hozzákezdenénk a feltételezett értékek meghatározásához, rögtön látjuk, hogy összevonásokra van szükség, mivel előfordulnak 5-nél kisebb értékek mindkét észlelt mintában. Az összevonások után az alábbi kontingencia-táblázatot kapjuk:

hőmérséklet, ºC	6 – 10	10 – 12	12 – 14	14 – 16	16 – 20	*sorösszeg*$(\sum_{.j})$
dombtető (É)	6	13	12	18	11	60
völgyfenék (É)	17	20	12	6	5	60
oszlopösszeg$(\sum_{i.})$	23	33	24	24	16	120

Mivel a feladatban mindkét minta észlelt adatokat tartalmaz, ezért a χ^2-próba, homogenitás-vizsgálat esetében a feltételezett értékek meghatározása mindkét észlelt mintára külön-külön, cellánként történik.
A feltételezett érték meghatározása egy tetszőleges cellára (F_{ij}) az imént kapott kontingencia-táblázat alapján a következő formulával történik:

$$F_{ij} = \frac{\sum_{.j} * \sum_{i.}}{\sum_{ij}} \tag{4.5.}$$

azaz egy adott cellához tartozó feltételezett értéket úgy kapunk meg, hogy az adott cella sorösszegét megszorozzuk az adott cella oszlopösszegével, s e szorzatot elosztjuk a minta összes elemszámával (ez a feladat szerint n = 120).
Határozzuk meg most minden egyes cellára a feltételezett értékeket!

$$F_{11} = \frac{60*23}{120} = 11{,}5\,;\; F_{12} = \frac{60*33}{120} = 16{,}5\,;\; F_{13} = \frac{60*24}{120} = 12\,;\; F_{14} = \frac{60*24}{120} = 12\,;\; F_{15} = \frac{60*16}{120} = 8$$

$$F_{21} = \frac{60*23}{120} = 11{,}5\,;\; F_{22} = \frac{60*33}{120} = 16{,}5\,;\; F_{23} = \frac{60*24}{120} = 12\,;\; F_{24} = \frac{60*24}{120} = 12\,;\; F_{25} = \frac{60*16}{120} = 8$$

Jól látható, hogy az azonos oszlophoz tartozó feltételezett cellaelem értékek a két mintában (a két sorban) azonosak. Azonban ez csak akkor van így, ha a két észlelt mintához tartozó sorösszegek megegyeznek. Ekkor tehát nem is kell külön kiszámolni a második sorra a feltételezett értékeket, hanem az azonos oszlopban, az első sorban kapott cellaértékeket írjuk be a második sorba is. Ha a két minta (két sor) sorösszege nem egyezik, akkor viszont a második sorban is ki kell számolni cellánként a feltételezett értékeket.
Írjuk be most a kapott feltételezett értékeket a megfelelő cellákba. A könnyebb megkülönböztetés érdekében cellánként az észlelt értékek normál betűvel a bal alsó sarokban, míg a feltételezett értékek dőlt betűvel a jobb felső sarokban találhatók. Ily módon a következő kontingencia-táblázathoz jutunk:

hőmérséklet, ºC	6 – 10	10 – 12	12 – 14	14 – 16	16 – 20	*sorösszeg*$(\sum_{.j})$
dombtető (É)	*11,5* 6	*16,5* 13	*12* 12	*12* 18	*8* 11	60
völgyfenék (É)	*11,5* 17	*16,5* 20	*12* 12	*12* 6	*8* 5	60
oszlopösszeg$(\sum_{i.})$	23	33	24	24	16	120

4. Helyettesítsünk be a χ^2-próba képletébe!
   Mivel az észlelt (É) és a feltételezett (F) értékek minden egyes cellához adottak, ezért elvégezhetjük a cellánkénti behelyettesítést a χ^2-próba képletébe. Eszerint:

$$\chi^2 = 2*\sum_{i=1}^{n}\frac{(É_i - F_i)^2}{F_i} = 2*\left[\frac{(6-11,5)^2}{11,5}+\frac{(13-16,5)^2}{16,5}+\frac{(12-12)^2}{12}+\frac{(18-12)^2}{12}+\frac{(11-8)^2}{8}\right]=$$

$$=2*\left[\frac{30,25}{11,5}+\frac{12,25}{16,5}+0+3+\frac{9}{8}\right]=15,0$$

   A szabadsági fokok (sz.f.) száma a (4.3.) egyenlet alapján:

$$\text{sz.f.} = (o - 1) = (5 - 1) = 4$$

   Fontos megjegyezni, hogy a szabadsági fokokat mindig az összevonások után kapott oszlopszámok figyelembe vételével határozzuk meg!
   Ezután a tankönyv, 315. oldal, 5.11. táblázata (Péczely, 1979) alapján döntünk a 0-hipotézisünk teljesedését, illetve nem teljesedését illetően:

$$\chi^2_{számított} = 15,0 > \chi^{2;sz.f.}_{v.sz.} = \chi^{2;4}_{0,05} = 9,49$$

   ahol v.sz. a választott valószínűségi szint, azaz a mi esetünkben: v.sz. = 0,05.

5. Válaszoljunk a feladatra!
   Mivel az általunk számított χ^2-érték nagyobb, mint a választott valószínűségi szinthez és a számított szabadsági fokhoz tartozó küszöbérték, ezért a 0-hiptézist elutasítjuk. Azaz a dombtető és a völgyfenék nyári hajnali minimum-hőmérsékleteinek eloszlása lényegesen különbözik egymástól.

5.1. VALÓSZÍNŰSÉGEK KONFIDENCIA INTERVALLUMÁVAL KAPCSOLATOS FELADAT

5.1. Feladat:
Egy tóparti megfigyelőhelyen a szabad vízfelület az észlelőponttól számított északkelet-délkeleti irányok által meghatározott 90º-os szektorba esik. A tapasztalat szerint nyáron a nappali órákban elég gyakori a tavi szél az állomáson. Hosszú megfigyelések alapján júliusban az óránkénti szélmegfigyelések szerint 30 % annak a valószínűsége, hogy az állomáson a fenti szektorból fúj a szél. 5 vizsgált év júliusi napjain déli 12 órakor az esetek 47 %-ában volt víz felőli szélirány.
Eltér-e ez a kis mintára kapott relatív gyakoriság oly mértékben a hosszú megfigyelési időszakra megállapított alapvalószínűségtől, hogy kimondhassuk a nappali tavi szél törvényszerű fellépését?

Megoldási menet:

1. Állítsuk föl a 0-hipotézist!
   Mivel a feladatban eldöntendő kérdésre kell válaszolnunk, egy a „priori” (előzetes, a tapasztalattól független, a számításokat megelőző) feltevést teszünk. Eszerint azt mondjuk, hogy a kis mintára kapott relatív gyakoriság nem tér el lényegesen a hosszú megfigyelési időszakra vonatkozó alapvalószínűségtől. Azaz, azt mondjuk, hogy a déli órákban nem növekszik meg lényegesen a tó felőli szélirány relatív gyakorisága.

2. Döntsük el a feladat típusát!
   E feladat elméleti háttere a következő. Ismert valamely esemény bekövetkezésének P valószínűsége egy teljes adatsokaságra. Majd bizonyos szempont szerint kiválasztunk részsokaságokat, s ezekre is meghatározzuk az esemény bekövetkezésének p valószínűségét (illetve – mivel általában kis elemszámú mintákról van szó – relatív gyakoriságát). Majd azt vizsgáljuk, hogy az esemény bekövetkezését előidéző ok ugyanúgy hatott-e a kiválasztott részsokaságokban, mint a teljes adatsokaságban.
   Az ilyen típusú feladatok megoldása a ***valószínűségek konfidencia intervallumá***nak meghatározásával lehetséges. Az eljárás kidolgozása Neumann János nevéhez fűződik.

3. Mutassuk be a módszer lényegét!
   Legyen adott a teljes adatsokaságra valamely esemény P alapvalószínűsége. Válasszunk a teljes adatsokaságból véletlenszerűen n elemű részsokaságokat, majd ezen minták mindegyikére határozzuk meg az esemény valószínűségét (p) (azaz, kis mintákról lévén szó: relatív gyakoriságát). A kapott p értékek véletlenszerűen szóródnak P körül, s eloszlásukra a normális eloszlás tételezhető fel. Az n elemű részmintákra meghatározott p valószínűségek szórása (σ_p) az alábbi képlettel adható meg.

$$\sigma_p = \sqrt{\frac{p(1-p)}{n}} \qquad (5.1.)$$

   A konfidencia-intervallum P_1 alsó és P_2 felső határát a következő összefüggések szolgáltatják:

$$P_1 = p - d^*\sigma_p$$

$$P_2 = p + d^*\sigma_p \qquad (5.2.)$$

A konfidencia-határokat adott valószínűségi szintre vonatkoztatjuk. Általában az 5 %-os, az 1 %-os, illetve a 0,1 %-os valószínűségi szinteket alkalmazzuk, amelyekhez tartozó d-értékek rendre 1,96, 2,58, illetve 3,29.
[P_1; P_2] $\Rightarrow$ a 0-hipotézist a választott valószínűségi szinten megtartjuk;[P_1; P_2] $\Rightarrow$ a 0-hipotézist a választott valószínűségi szinten elvetjük.
A választott valószínűségi szint – a χ^2-próbával kapcsolatos feladatokhoz hasonlóan – az igaz 0-hipotézis elvetésének a valószínűségét jelenti. Tehát, ha pl. az 5 %-os valószínűségi szintet választjuk, akkor az ahhoz tartozó d-érték alapján számított [P_1; P_2] konfidencia-intervallum olyan széles lesz, hogy 100 esetből átlagosan 5 esetben vetünk el igaz 0-hipotézist azáltal, hogy a véletlenszerű adatok alapján számított [P_1; P_2] konfidencia-intervallumon kívülre esik P. Az is látható, hogy ha a valószínűségi szint (vagyis az igaz 0-hipotézis elvetésének a valószínűsége, ami azonos az elsőfajú hibával) csökken, akkor nő a hozzá tartozó d-érték. Ennélfogva a [P_1; P_2] konfidencia-intervallum szélesebbé válik. Ez azt jelenti, hogy ha csökken az elsőfajú hiba, azzal párhuzamosan növekszik a másodfajú hiba (azaz a hamis 0-hipotézis elfogadásának az esélye). Ennek épp az az oka, hogy a konfidencia-intervallum ez esetben szélesebb lesz. Itt is fontos megjegyezni, hogy az elsőfajú hiba csökkenése a másodfajú hiba növekedésével jár és viszont. A velük kapcsolatos statisztikai eljárások döntési kritériumainak egyik fontos célja, hogy egyik se kerüljön túlsúlyba a másik rovására.

4. Alkalmazzuk a módszert a feladatra!
   A feladat megoldásához szükséges paraméterek értékei a következők:
   P = 30 % = 0,3
   p =47 % = 0,47
   n = 5*31 = 155 (azaz 5 év júliusi napjainak a száma)
   E kiindulási paraméterek ismeretében helyettesítsünk be σ_p képletébe!

$$\sigma_p = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0,47(1-0,47)}{155}} = 0,04$$

   Ezután válasszuk az 5 %-os valószínűségi szintet, s határozzuk meg a konfidencia-intervallum végpontjait!

$$P_1 = p - d^*\sigma_p = 0,47 - 1,96^*0,04 = 0,39$$

$$P_2 = p + d^*\sigma_p = 0,47 + 1,96^*0,04 = 0,55$$

   Innen azt kapjuk, hogy P = 0,3 [P_1; P_2] = [0,39; 0,55].

5. Válaszoljunk a feladatra!
   Tehát a megfigyelőhelyen a déli órában jelentősen megnő a tó felőli szélirány.

Hivatkozás

Péczely, Gy., 1979: Éghajlattan. Budapest: Tankönyvkiadó, 336 pp