

**7. feladatsor, statisztika, 2017. ápr. 20–26.**

- (1) Öt órán keresztül így alakult a hőmérséklet Budapesten.

időpont	6	7	8	9	10	11
°C	5	6	8	9	10	13

- (a) Illesszünk  $aX + b$  alakú lineáris modellt a mintára, és becsüljük meg az  $a$  és  $b$  együtthatókat, valamint a  $\sigma$  szórást.  
 (b) Mennyi az  $R^2$ , azaz a megmagyarázott ingadozás részaránya? Ez alapján mit mondhatunk, mennyire illeszkedik jól a lineáris modell az adatokra?  
 (c) Adjunk előrejelzést a déli hőmérsékletre. Mennyi az előrejelzésünk szórása?
- (2) Az elmúlt hónapokban így alakult a fogyasztói árindex (az infláció mérőszáma) Magyarországon.

hónap	nov.	dec.	jan.	febr.	márc.
árindex (%)	101,0	101,8	102,3	102,9	102,7

**6. feladatsor, statisztika, 2017. ápr. 5–6.**

- (1) Néhány kőzetmintában megvizsgálták a fossziliákat. Az alábbi táblázat azt tartalmazza, hogy hány olyan kőzetminta volt, amelyben adott számú élőlény nyomát találták meg.

élőlények száma	0	1	2	3	4	5	6
kőzetek száma	2	5	7	5	7	3	1

- $\alpha = 0,05$  terjedelem mellett elfogadható-e az a hipotézis, hogy az egy kőzetmintában található fossziliák száma 3 paraméterű Poisson-eloszlású?
- (2) Negyven embert kérdeztek arról, hogy milyen italokat szoktak inni. Huszonnyolcan fogyasztanak rendszeresen sört, közülük tizennyolcan rendszeresen bort is. Azok között, akik nem fogyasztanak rendszeresen sört, heten voltak, akik bort viszont rendszeresen szoktak fogyasztani.  
 (a) Állíthatjuk-e  $\alpha = 0,05$  terjedelem mellett, hogy a sör- és borfogyasztás között szignifikáns összefüggés van?  
 (b) Állíthatjuk-e  $\alpha = 0,05$  terjedelem mellett, hogy a sör- és borfogyasztás között szignifikáns pozitív korreláció van?
- (3) Hatvan téli napon át figyelték, hogy egy adott napon belül volt-e fagy, és volt-e csapadék. Az alábbi eredmények adódtak.

	fagyos	nem fagyos
volt csapadék	14	8
nem volt csapadék	25	13

$\alpha = 0,05$  terjedelem mellett állíthatjuk-e, hogy a fagy és a csapadékos időjárás között szignifikáns összefüggés van?

- (4) Két különböző helyszínről származó kőzetmintákat osztályoztak a vastartalmuk szerint. Az alábbi táblázat tartalmazza, hogy milyen vastartalmú kőzetből hányat találtak az egyes helyszíneken.

%	0 – 50	51 – 55	56 – 60	61 – 65	66 – 100
A	6	12	18	23	7
B	8	16	19	14	6

$\alpha = 0,05$  terjedelem mellett elfogadható-e az a hipotézis, hogy az A és B helyszínen a vastartalom eloszlása megegyezik?

- (5) **Házi feladat április 20-ig. (3 pont)** Egy tetszőlegesen választott szövegben számoljuk meg ötven szó hosszát (vagyis azt, hogy hány betűből állnak). A minta alapján  $\alpha = 0,05$  terjedelem mellett elfogadható-e az a hipotézis, hogy egy véletlenszerűen választott szó hossza geometriai eloszlású  $p = 1/5$  paraméterrel?

**5. feladatsor, statisztika, 2017. márc. 16–29.**

- (1) A következő vízkeménységeket mértük az A kútnál:

16.9, 11.5, 9.7, 14.2, 12.0, 10.8, 13.9, 15.6

Tudjuk, hogy a mérések szórása 2, és normális eloszlást feltételezünk.

- (a) Elfogadható-e az a hipotézis 95% szinten, hogy az átlagos vízkeménység 12?  
 (b) Elfogadható-e 98%-os szinten ugyanez a hipotézis?  
 (c) Elfogadható-e 95%-os szinten az a hipotézis, hogy az átlagos vízkeménység nem több 13-nál?  
 (d) A B kútból is kapunk néhány adatot:  
 12.0, 11.9, 9.6, 10.6, 14.5  
 Itt a mérések szórása 1.2. Állíthatjuk-e 95%-os szinten, hogy az A kút vize szignifikánsan keményebb a B kút vizénél?

- (2) Egy napon tíz budapesti helyszínen megmérték a  $\text{NO}_2$ -koncentrációt. Az átlag  $352 \mu\text{g}/\text{m}^3$ , a korrigált tapasztalati szórás 8 lett.

- (a) 95%-os szinten elfogadható-e az a hipotézis, hogy a koncentráció a  $350 \mu\text{g}/\text{m}^3$  tájékoztatási küszöbérték alatt van?  
 (b) Elfogadható-e ugyanez a hipotézis 99%-os szinten?  
 (c) Londonban 20 mérésből az átlagos koncentráció 376, a korrigált tapasztalati szórás 16 lett.  $\alpha = 0,05$  terjedelem mellett állíthatjuk-e, hogy Londonban szignifikánsan nagyobb a  $\text{NO}_2$  koncentrációja?

- (3) Két különböző helyszínről származó minták porozitását mértük meg. Az első helyszínen  $n_1 = 15$  mintából számolva a mérések átlaga  $\bar{X}_1 = 0,32$  lett, a korrigált tapasztalati szórás pedig  $s_{n,1}^* = 0,0016$ . A második helyszínen  $n_2 = 30$  mintából számolva az átlag  $\bar{X}_2 = 0,46$ , a korrigált tapasztalati szórásnégyzet pedig  $s_{n,2}^* = 0,0009$ . A porozitás eloszlását normális eloszlásnak feltételezzük, és azt is feltesszük, hogy a két helyszínen az eloszlások szórása megegyezik.

- (a) Elfogadható-e 95%-os szinten, hogy az első helyszínen a porozitás 0,3?  
 (b) Elfogadható-e 99%-os szinten, hogy a második helyszínen a porozitás legfeljebb 0,3?  
 (c) 0,05 terjedelem mellett igaz-e, hogy a két helyszínen szignifikánsan különböző a talaj porozitása?

- (4) **Házi feladat április 6-ig (3 pont).** Hat férfi és hat nő testmagasságára az alábbi adatok adódtak:

férfiak	178	183	189	194	169	172
nők	182	162	174	168	179	171

Az adatsort egészítse ki a saját testmagasságával. Az így kapott minta alapján  $\alpha = 0,05$  terjedelem mellett állíthatjuk-e, hogy a férfiak testmagassága szignifikánsan nagyobb a nőkénel? (Tegyük fel, hogy a férfiak és nők magasságának szórása azonos, az eloszlások pedig normális eloszlásúak.)

4. feladatsor, statisztika, 2017. márc. 8–9 .

- (1) Két különböző helyszínről származó minták porozitását mértük meg. Az első helyszínen  $n_1 = 15$  mintából számolva a mérések átlaga  $\bar{X}_1 = 0,32$  lett, a korrigált tapasztalati szórás pedig  $s_{n,1}^* = 0,04$ . A második helyszínen  $n_2 = 30$  mintából számolva az átlag  $\bar{X}_2 = 0,46$ , a korrigált tapasztalati szórás pedig  $s_{n,2}^* = 0,03$ . A porozitás eloszlását normális eloszlásnak feltételezzük.

- (a) Adjunk 95%-os megbízhatósági szintű kétoldali konfidenciaintervallumot az első helyszínen lévő talaj porozitásának várható értékére.  
 (b) Adjunk 98%-os megbízhatósági szintű kétoldali konfidenciaintervallumot az első helyszínen lévő talaj porozitásának várható értékére.  
 (c) Adjunk 95%-os megbízhatósági szintű kétoldali konfidenciaintervallumot a második helyszínen lévő talaj porozitásának várható értékére.

- (d) Tegyük fel, hogy az első helyszínen a talaj porozitásának szórása 0,042. Ezt felhasználva adjunk újra 95%-os megbízhatósági szintű kétoldali konfidenciaintervallumot az első helyszínen lévő talaj porozitásának várható értékére.  
 (e) A (b), (c), (d) kérdésekre kapott intervallumok közül melyik (és miért) rövidebb az (a) feladatban kapott intervallumnál?  
 (2) Egy helyszínről származó nyolc kőzetminta vastartalmát megmérve az alábbi eredményeket kaptuk (százalékban). Feltételezhetjük, hogy a vastartalom normális eloszlású.

52,8 64,2 63,3 66,4 70,6 69,6 58,2 58,3

- Adjunk 95%-os megbízhatósági szintű kétoldali konfidenciaintervallumot a vastartalom várható értékére.  
 (3) **Házi feladat március 23-ig (3 pont).** Néhány (legalább nyolc) véletlenszerűen kiválasztott ismerős testmagassága alapján adjunk 95%-os, illetve 98%-os megbízhatósági szintű kétoldali konfidenciaintervallumot a testmagasság várható értékére.

3. feladatsor, statisztika, 2017. márc. 1–2 .

- (1) Egy sportoló három éven át feljegyezte, hogy az egyes heteken hány napon volt edzése. Az alábbi eredmények adódtak.

gyakoriság	0	1	2	3	4	5	6	7
2014	3	7	9	11	9	5	5	3
2015	0	2	6	5	8	13	15	3
2016	1	1	9	16	15	7	2	1

- (a) Készítsünk hisztogramot az egyes évek adataiból külön-külön.  
 (b) A hisztogram alapján tippeljünk meg, hogy a három minta közül melyiknek a legnagyobb és melyiknek a legkisebb a ferdesége.  
 (c) A hisztogram alapján tippeljünk meg, hogy a három minta közül melyiknek a legnagyobb és melyiknek a legkisebb a lapultsága.

- (d) Számítsuk ki a ferdeséget és a lapultságot mindhárom mintára.  
 (e) Melyik mintában a legnagyobb az átlag és a medián különbsége?  
 (f) Számítsuk ki a minta harmadik és negyedik tapasztalati momentumait.

- (2) Nyolc kőzetminta vastartalmát megmérve az alábbi eredményeket kaptuk (százalékban).

52,8 64,2 63,3 66,4 70,6 69,6 58,2 58,3

- (a) Számítsuk ki a minta harmadik tapasztalati momentumát.  
 (b) Számítsuk ki a minta negyedik tapasztalati momentumát.  
 (c) Számítsuk ki a minta ferdeségét.  
 (d) Számítsuk ki a minta lapultságát.

2. feladatsor, statisztika, 2017. febr. 22–23.

- (1) Tekintsük az alábbi NO<sub>2</sub>-koncentráció adatokat ( $\mu\text{g}/\text{m}^3$ , Budapest, 2017. február 15.)

108 123 167 163 139 148 109 97 90 95

- (a) Mi a rendezett minta?  
 (b) Hogyan tudnánk becsülni annak valószínűségét a minta alapján, hogy a koncentráció a  $100 \mu\text{g}/\text{m}^3$ -os egészségügyi határérték alatt marad? Mennyi  $\hat{F}_n(100)$ ?  
 (c) Mennyi  $\hat{F}_n(130)$ ?  
 (d) Ábrázoljuk a minta tapasztalati eloszlásfüggvényét.  
 (e) Mennyi a minta első kvartilise?  
 (f) Mennyi a minta mediánja?  
 (g) Mennyi a minta harmadik kvartilise?  
 (h) Készítsünk boxplotot a mintából.  
 (i) Mennyi a minta 90%-os kvantilise?

- (2) Nyolc kőzetminta vastartalmát megmérve az alábbi eredményeket kaptuk (százalékban).

52,8 64,2 63,3 66,4 70,6 69,6 58,2 58,3

- (a) Mi a rendezett minta?  
 (b) Mennyi  $\hat{F}_n(60)$ , azaz a tapasztalati eloszlásfüggvény értéke a 60 helyen?  
 (c) Ábrázoljuk a minta tapasztalati eloszlásfüggvényét.  
 (d) Mi a minta első kvartilise?  
 (e) Mi a minta mediánja?  
 (f) Mi a minta harmadik kvartilise?  
 (g) Készítsünk boxplotot a mintából.  
 (h) Mennyi a minta 87,5%-os kvantilise?

**Házi feladat március 1-ig (3 pont)** Nyolc véletlenszerűen választott ismerős testmagasságából készítsünk adatsort, majd erre a mintára válaszoljunk meg a (2) feladat (a)-(g) kérdéseit (a (b) kérdést így módosítva: mennyi  $\hat{F}_n(172)$ ).

1. feladatsor, statisztika, 2017. febr. 15–16.

- (1) 20-szor dobtunk két dobókockával. A dobott számok összegére az alábbi eredmények adódtak:

10	6	7	8	12	12	11	3	3	5
6	9	7	6	8	9	9	6	3	4

- (a) Mi a rendezett minta?  
 (b) Készítsünk hisztogramot a mintából.  
 (c) Mennyi a minimum, maximum, terjedelem, medián, módusz?  
 (d) Mennyi a mintaátlag?  
 (e) Mennyi az elemek összege, illetve négyzetösszege?  
 (f) Mennyi a tapasztalati szórás, illetve a korrigált tapasztalati szórás?  
 (g) Mennyi lenne egy dobásnál az összeg, illetve a szórás?  
 (h) Mennyi a valószínűsége, a 20 dobásból a maximum pontosan 12?

**Házi feladat február 22-ig.** Ismételjünk meg a feladatban szereplő kísérletet, és adjuk meg a választ az (a)-(f) kérdésekre az így kapott mintára vonatkozóan.

- (2) Nyolc kőzetminta vastartalmát megmérve az alábbi eredményeket kaptuk (százalékban).

52,8 64,2 63,3 66,4 70,6 69,6 58,2 58,3

- (a) Mi a rendezett minta?  
 (b) Készítsünk hisztogramot a mintából.  
 (c) Mennyi a minimum, maximum, terjedelem, medián?  
 (d) Mennyi a mintaátlag?  
 (e) Mennyi a tapasztalati szórás, illetve a korrigált tapasztalati szórás?  
 (3) Megmértük kilenc férfi és hét nő testmagasságát, az alábbi értékeket kaptuk (centiméterben számolva, kerekítve).

férfiak	168	183	175	172	187	177	175	179	167
nők	171	169	168	162	165	166	160		

- (a) Hasonlítsuk össze a kapott adatokat az átlag, a medián és a korrigált tapasztalati szórás szempontjából is. Mire következtethetünk ebből?  
 (b) Számítsuk ki az átlagot, a mediánt és a korrigált tapasztalati szórást az összesített 16 elemű mintából is.  
 (c) Készítsünk hisztogramot a két adatsorból külön-külön, illetve az összesített 16 elemű mintából is.

