Gyakorló feladatok Elemi Statisztikából

1. Egy TV-ben folyó valóság show egy közvélemény kutató cég adatai szerint 25%-os részesedést ért el. Ez azt jelenti, hogy országosan a bekapcsolt TV készülékek 25%-án ezt nézték az adott időben. Tegyük fel, hogy egy hirdető cég meg akarja vizsgálni, hogy igaz-e ez a felmérési adat és felmérést végez 10 véletlenül kiválasztott háztartás felhasználásával.
   a) Mi a valószínűsége, hogy egyikben sem nézik az adott show-t?
   b) Mi a valószínűsége, hogy legalább egy háztartásban nézik a show-t?
   c) Mi a valószínűsége, hogy a háztartásoknak legalább a felében nézik a show-t?
   d) Ha csak két olyan háztartást találnak, amiben ezt a műsort nézik, következik-e ebből, hogy a 25%-os adat rossz?

2. A terhességek hossza normális eloszlást követ 268 napos átlaggal és 15 napos szórással. A normális eloszlás egy klasszikus alkalmazását inspirálta az a levél, amelyben egy tengerész felesége 309 nappal a férje egy rövid hazalátogatása után adott életet gyermekének. Mindezek ismeretében, határozzuk meg, mi annak a valószínűsége, hogy egy terhesség 309 vagy annál több napig tart?

3. Vizsgáljuk meg, hogy az alábbi számok egy normális eloszlásból származhatnak-e!
   a) -0.5, 1.9, 4.2, 7.4 és 13.1
   b) -1.2, 2.1, 4.8, 15.7 és 62.2
   Készíts mindkét esetben normál kvantilis plotot! Szerinted egy 5 elemű minta alapján biztosan ki lehet-e jelenteni, hogy a populáció milyen eloszlású?

4. Egy kórházban egy évben született 144 gyermek. Súlyuk átlaga 3420 g, szórása pedig 495 g. Adjunk pontbecslést és 99%-os konfidenciaintervallum becslést az országban született gyermekek átlagos súlyára!

5. Felmérést kívánunk készíteni a Linux használók arányáról az egyetemisták között. Az arányt legfeljebb 5 %-os hibával akarjuk meghatározni. Feltéve, hogy mindenki igazat mond, legalább mennyi embert kell megkérdezni ehhez, ha 90% konfidenciaszintet használunk?

6. Vizsgáljuk meg, hogy ismeretlen szórás esetén a mintaelemszám növelése hogyan befolyásolja a becslés pontosságát! Hasonlítsuk össze a 10 és 20 elemű minta esetét!

7. A világ legkisebb emlőse egy denevérféle, ami kb. akkora, mint egy dongó. Néhány ilyen állat mért tömegét mutatjuk be grammokban:
1.51, 1.68, 1.34, 2.00, 1.60, 2.15, 1.66, 1.80, 2.18, 1.33
A 95%-os konfidencia szintet használva adjunk pontbecslést és intervallumbecslést az állatok tömegének szórására!

8. Az Öreg Hűséges (Old Faithful) gejzír kitöréseit vizsgálva a következő adatokat kapták a kitörés hosszára és időtartamára:

időtartam (s)	240	120	178	234	235	269
magasság (m)	43	33	38	37	43	37

   (a) Van-e lineáris kapcsolat a két mennyiség között 0.05-ös illetve 0.01-es szignifikanciaszinten?

   (b) A kitörések magasságának varianciáját hány százalékban határozza meg a kitörés időtartama?

9. Lásd be a lineáris korrelációs együtthatóra vonatkozó $-1 \leq r \leq +1$ egyenlőtlenséget!

10. Megmértük néhány kivágott fa kerületét és meghatároztuk életkorát. Vizsgáld meg a korrelációjukat, illetve illessz regressziós egyenest az adatokra!

kerület (cm)	280	131	139	245	207	182
életkor (év)	189	41	52	161	93	68

11. Az alábbi adatokra vizsgáld meg x és y kapcsolatát! Számítsd ki a lineáris korrelációs együtthatót és vizsgáld meg, hogy látszik-e szignifikáns lineáris kapcsolat a változók között? Készíts szórásdiagramot. Milyen kapcsolatot látsz a két változó között? Készítsd el a reziduumok diagramját is. Van-e benne bármilyen szabályosság? Milyen kapcsolatra utal a reziduumok diagramja a két változó között?

x	1	2	18	38	72	91	145	176	195
y	1	4	335	1463	5235	8304	21078	31051	38207

12. Ha egy null hipotézist elutasítunk az $\alpha = 0.01$-es szignifikancia-szinten, akkor el kell-e utasítanunk a 0.05-ös szignifikancia-szinten is? Miért igen, miért nem? És ha nem tudjuk elutasítani 0.01-es szignifikancia szinten, akkor mit tudunk mondani 0.05-ös szignifikancia szinten?

13. 31 véletlenszerűen kiválasztott férfi karmester átlagéletkora 74.3 év, ellentétben a férfi lakosságra várható 70.5 évvel. Az említett karmesterek élethosszának szórása 8.2 év. Egy $\alpha = 0.01$ szignifikancia-szintű vizsgálatot végezve teszteljük azt a feltételezést, hogy a karmesterek átlagosan tényleg hosszabb ideig élnek mint 70.5 év!

14. Egy feladatsorra az abc szerinti első 25 hallgatóból 15-nek sikerült elérnie a kapható pontok felét. Ez alapján vizsgáljuk meg azt a feltevést, hogy az összes hallgatónak legalább 80%-a érte el ezt a ponthatárt! Mi lesz a null és az alternatív hipotézis, és el tudjuk-e vetni a null hipotézist az $\alpha = 0.05$-ös szignifikancia szinten? Mi a konklúzió?

15. Próbáld meg megbecsülni magadban 10 másodperc hosszát egymás után tízszer! Minden mérésnél stopperrel ellenőrizd, mennyire voltál pontos! Adj 99%-os konfidencia-szinten intervallum-becslést az átlagra és a szórásra! Ezután ismételd meg újra a tíz mérést, és ezúttal is adj 99%-os konfidencia-szinten intervallum-becslést az átlagra és a szórásra is! (Tételezzük fel, hogy az adatok normális eloszlást mutatnak!) A második 10 mérés adatait felhasználva végezz kétoldalú tesztet a P-érték módszerrel arra a null hipotézisre, miszerint belső órád 10 másodperces méréseinek populáció-átlaga valóban 10 másodperc! Az első és a második 10 mérés adatait összevetve pontosodott az időérzéked?

16. Az alábbi táblázat almák átmérőjét és tömegét tartalmazza:

átmérő(mm)	91	74	80	103	94	62	85
tömeg (g)	235	142	170	339	282	80	200

Milyen összefüggést várunk a két mennyiség között? Miért?

Van-e lineáris kapcsolat a két mennyiség között?

Milyen transzformációval tudjuk ellenőrizni a várt összefüggést? Mit tapasztalunk, ha ennek a kapcsolatnak az erősségét megvizsgáljuk?

17. Különböző autótípusok tömegét és fogyasztását tartalmazza az alábbi táblázat:

tömeg (kg)	1443	1568	1465	1811	1109	1136	1040
fogyasztás (l/100 km)	8.7	8.1	8.7	9.8	6.4	6.9	6.4

(a) Vizsgáljuk meg, hogy van-e lineáris korreláció az autó tömege és fogyasztása között?

(b) Adjuk meg egy 1700 kg-os autó fogyasztásának 95 %-os becslési intervallumát.