7. heti feladatsor, (statisztika. gyak.) 2012. május 3–16.

(1) Tekintsük a következő párokat. Határozzuk meg és ábrázoljuk az $ax+b$ alakú regressziós egyenest. Becsüljük meg a hiba szórásnégyzetét, és számoljuk ki az egyenes megbízhatóságát, R^2-t.

x	0	1	6	5	3
y	4	3	0	1	2

(2) Az Antarktiszon az F-12 gáz koncentrációjára az alábbi értékeket kapták.

év	1990	1992	1994	1996	1998
koncentráció (ppt)	195	216	244	260	284

a) Határozzuk meg és ábrázoljuk az $ax+b$ alakú regressziós egyenest.
b) Vizsgáljuk az $a=0$ hipotézist az $a>0$ ellenhipotézissel szemben.
c) Adjunk előrejelzést és konfidenciaintervallumot az F-12 gáz 2013. évi koncentrációjára.

(3) Ábrázoljuk egy használtautó-kereskedő alábbi, eladási árakra vonatkozó adatait. Ez alapján javasoljunk és számoljunk is ki közelítő modellt, melynek segítségével becsüljük egy tízéves modell várható árát.

kor (év)	2	3	4	6	7
ár (ezer forint)	1200	1000	850	650	550

(4) Tegyük fel, hogy egy síbolt forgalma (millió Ft-ban) az alábbiak szerint alakult.

év	2008				2009			
negyedév	I.	II.	III.	IV.	I.	II.	III.	IV.
forgalom	14	10	0	25	18	6	5	31

a) Ábrázoljuk az idősort, számítsuk ki a lineáris trendet. Teszteljük az együtthatói szignifikanciáját. b) A kapott reziduálisok idősorára számítsuk ki az első négy autokorreláció közelítését. Mire következtethetünk ebből? c) Becsüljük meg a reziduálisok szezonális komponensét a negyedéves átlagokkal. A fennmaradó idősor autokorrelációi mit mutatnak? d) A periodikus komponenstől megtisztított idősorra is ábrázoljuk és számítsuk ki a lineáris trendet. Teszteljük az együtthatóinak szignifikanciáját.

6. heti feladatsor, (statisztika. gyak.) 2012. április 25–26.

(1) Egy évfolyamon 300-an tanulnak. Megszámolták, hogy a legutóbbi vizsgaidőszakban hányszor buktak az egyes hallgatók. Az eredményeket az alábbi táblázat tartalmazza.

bukások száma	0	1	2	3	4
hallgatók száma	80	113	77	27	3

Elfogadhatjuk-e azt a nullhiptézist, hogy egy hallgató bukásainak száma binomiális eloszlású $(4; 0,25)$ paraméterrel? És azt, hogy binomiális $(4; p)$ eloszlású $(0<p<1)$?

(2) CASCO biztosítással rendelkezők éves kárszámát vizsgáltuk. 4000 vezető adatait tartalmazza az alábbi táblázat. Elfogadható-e, hogy a kárszám Poisson-eloszlású?

kárszám	0	1	2	3	4	5	>5
vezetők száma	3691	232	68	5	3	1	0

(3) Az alábbi táblázat mutatja, hogy 100 évben a csapadék mennyisége és az átlaghőmérséklet hogyan alakult.

	kevés	átlagos	sok
hűvös	15	10	5
átlagos	10	10	20
meleg	5	20	5

Tekinthető-e a csapadékmennyiség és a hőmérséklet függetlennek?

(4) Megvizsgáltak összesen 460 darab csavart, amelyek közül méretre 439 volt megfelelő, a többi selejtes. A méretre megfelelők közül szakítószilárdságra megfelelő volt 416 darab, a többi selejtes. Az összes 460 közül szakítószilárdságra selejtes volt 28 darab. Elfogadható-e, hogy a méret és a szakítószilárdság megfelelősége független tulajdonságok?

(5) Két dobókockával dobva a következő eredmények adódtak:

dobás	1	2	3	4	5	6
1. kocka	7	11	8	10	8	6
2. kocka	16	11	20	19	18	16

a) $\alpha=0,05$ terjedelemmel tekinthető-e a két kocka egyformának?
b) $\alpha=0,05$ terjedelemmel tekinthető-e az 1. kocka szabályosnak?
c) $\alpha=0,05$ terjedelemmel tekinthető-e a 2. kocka szabályosnak?

(6) Az alábbi táblázat CASCO biztosítással rendelkezők éves kárszámát tartalmazza 2003-ban és 2004-ben. Tekinthető-e a kárszám azonos eloszlásúnak a két esztendőben?

kárszám	0	1	2	3	4	5	>5
vezetők száma 2003-ban	3692	232	65	7	3	1	0
vezetők száma 2004-ben	3542	284	135	24	9	5	1

2012. május 9-ig beadható házi feladatok

(7) Egy cikkben, amely a szem és a kéz aszimmetrikus fejlődése közti kapcsolatot tárgyalta, a következő táblázatot találtuk. Az oszlopok arra vonatkoznak, hogy melyik szem erősebb, a sorok pedig arra, hogy melyik kéz erősebb.

	bal szem	egyforma	jobb szem
bal kéz	34	62	28
egyforma	27	28	20
jobb kéz	57	105	52

Lehet-e ezen adatok alapján arra következtetni, hogy a szem és a kéz aszimmetrikus fejlődése között van összefüggés?

(8) Egy ötkérdéses tesztnél a következő eredményeket érték el:

találatok száma	0	1	2	3	4	5
hányan érték el	3	10	15	9	2	1

$\alpha=0,05$ terjedelemmel döntsünk arról a hipotézisről, hogy a találatok száma binomiális eloszlású!

5. heti feladatsor, (statisztika. gyak.) 2012. március 28–29.

(1) Legyen $X_1,\dots,X_n$ független minta $N(m,1)$ eloszlásból. Adjunk 95%-os, illetve 98%-os megbízhatóságú kétoldali konfidenciaintervallumot m-re, ha
a) $n=10$, $\sum_{i=1}^n X_i = 349$;
b) $n=250$, $\sum_{i=1}^n X_i = 48,5$;
c) $n=10$, és a minta a következő:

1,49	3,84	3,16	2,09	1,79	1,19	2,38	2,85	3,01	1,74

(2) Valaki azt állítja, hogy a klíma változik, és ezt azzal véli bizonyítottnak, hogy az elmúlt 10 évben 2-szer is volt jégeső, pedig korábban ez egyes évekre a jégeső valószínűsége a hivatalos adatok alapján csupán $0,1$ volt. Írjuk fel a hipotéziseket, a próbát és állapítsuk meg az elsőfajú hiba valószínűségét, valamint az erőfüggvényt a $p=0,2$ pontban.

(3) Egy tejgyárban minden szállítás előtt megvizsgálják a 25 dkg-os túrók minőségét. Jelölje p annak valószínűségét, hogy egy csomag túró romlott. A $H_0: p=0.01$ nullhipotézist (ez még elfogadható a gyár számára) vizsgálják a $H_1: p>0.01$ ellenhipotézissel szemben. A gyár eljárása a következő: N csomagot bontanak fel és amennyiben legalább 2 köztük romlott, akkor újrahasznosítást rendelnek el.
a) Írjuk le a kísérletet (mintatér, kritikus tartomány, elfogadási tartomány, statisztikai próba)!
b) Milyen N-re lesz az elsőfajú hiba valószínűsége kisebb 5%-nál?
c) Írjuk fel az erőfüggvényt.

(4) Adott egy 4 elemű független minta: X_1, X_2, X_3, X_4. Két hipotézisünk van:
$H_0: P(X_i=2)=1/4, P(X_i=3)=1/2, P(X_i=10)=1/4$
$H_1: P(X_i=2)=1/4, P(X_i=3)=1/4, P(X_i=10)=1/2$
Határozzuk meg a valószínűséghányados-próbát 5 %-os elsőfajú hibavalószínűség mellett.

(5) Az alábbi minta 5 év április 18-án Budapesten mért napi középhőmérsékleteit tartalmazza. Ellenőrizzük a $H_0: m=17$ hipotézist $\alpha=0,05$ elsőfajú hibavalószínűség mellett a $H_1: m<17$ alternatívával szemben, úgy, hogy
a) a korábbi tapasztalatok alapján tekintsük az értékek szórását 2-nek.
b) Ne használjunk a szórásra vonatkozóan előzetes információt.

középhőmérséklet (°C):	14,8	12,2	16,8	17,1	16,1

(6) Egy tejgyárban minden szállítás előtt megvizsgálják a 25 dkg-os túrók ólomtartalmát. Egy alkalommal a még megengedett szint %-ában a mérések a következők voltak: 98,5; 101,4; 99,5; 100,9; 99,7. A korábbi tapasztalatok alapján az ellenőr feltételezi, hogy 1 szórásúak. Elfogadható-e a $H_0: m=100$ nullhipotézis? Hogyan válasszuk meg a H_1 hipotézist? Mennyi lesz a p-érték (szignifikancia-szint)? Mennyi a próba erőfüggvényének értéke az $m=102$ pontban? Hány elemű mintára van szükség, ha azt szeretnénk, hogy ez legalább 0,99 legyen?

(7) Az alábbi két minta 10 - egyforma képességűnek feltételezett - sportoló súlylökésben elért eredményeit tartalmazza. A sportolók két ötfős csoportban készültek az edzőtáborban. Edzéstervük ugyanaz volt, csak az étrendjük különbözött. 2 hét felkészülés után értékelték az eredményeket. Tételezzük fel, hogy a minták normális eloszlásból származnak.

1. diéta	14,8	12,2	16,8	17,1	16,1
2. diéta	18,0	12,1	17,2	17,7	17,0

a) Melyik diéta volt jobb, ha a dobások szórását 2-nek tekintjük?
b) Melyik diéta okozott nagyobb változékonyságot az eredményekben?
c) Ha nem ismerjük a szórást, akkor tekinthetjük-e valamelyik diétát jobbnak?

2012. április 19-ig beadható házi feladatok

(8) Egy HÖK-elnök számára az elfogadható, ha a hallgatók legfeljebb 72 százaléka utálja (ez a H_0 hipotézis). Az ennél nagyobb arány esetén (H_1 hipotézis) lemond posztjáról. Minden negyedév végén 10 hallgatót kérdez meg, és akkor mond le, ha a tízből legalább 8 diák utálja. Mekkora a próba terjedelme? Várhatóan hány negyedévet fog tevékenykedni az elnök, ha minden félévben a diákok 65 százaléka utálja?

(9) 8 fiú és 8 lány testmagasságára tekintsük a következő adatokat (centiméterben mérve).

Fiúk:	185	193	185	184	188	176
Lányok:	163	169	163	170	156	160

A saját testmagasságot a megfelelő adatsorhoz hozzátéve $\alpha = 0.01$ elsőfajú hibavalószínűség mellett milyen következtetés vonható le az adatokból: bizonyítható-e, hogy a fiúk magasságának várható értéke több a lányokénál? Mennyi a p-érték?

4. heti feladatsor, (statisztika. gyak.) 2012. március 7–8.

(1) Legyen $Z_1, Z_2, \ldots, Z_n$ független minta a következő diszkrét eloszlásból:

$$P(Z_i = 1) = c,\ P(Z_i = 2) = 3c,\ P(Z_i = 3) = 1 - 4c,\ i = 1, \ldots, n,$$

ahol c az ismeretlen paraméter.
a) Határozzuk meg c maximum likelihood becslését!
b) Határozzuk meg c momentum-módszerrel kapható becslését!
c) Írjunk fel egydimenziós elégséges statisztikát az ismeretlen paraméterre!
d) Adjunk torzítatlan becslést c-re Z_1 segítségével!
e) Az előzőben megadottnál konstruáljunk jobb becslést a Rao–Blackwell–Kolmogorov-tétel segítségével.

(2) Legyen a $Z_1, Z_2, \ldots, Z_n$ független minta $N(2m + 5, 1/d^2)$ eloszlású. Becsüljük meg az ismeretlen paramétereket momentum-módszerrel!

(3) $X_1, X_2, \ldots, X_n$ független minta $b > 0$ paraméterű Poisson-eloszlásból. Mi a b paraméter maximum likelihood becslése?

(4) Határozzuk meg az ismeretlen paraméter maximum likelihood becslését, ha a minta
a) Pascal-eloszlású;
b) binomiális eloszlású (a rend ismert);
c) $(a, 1)$ intervallumon egyenletes eloszlású.

2012. március 22-ig beadható házi feladatok

(5) Egy ezerfős közvélemény-kutatásnál 650-en válaszolták azt, hogy szeretik a kutyákat. Becsüljük meg maximumlikelihood– és momentummódszerrel a tényleges arányt! Hogyan módosul ez a becslés, ha már tudjuk, hogy a kutyákat szeretők 10%-a hazudik, a kutyákat nem szeretőknél pedig 20% ez az arány?

(6) Legyen $X_1, X_2, \ldots, X_n$ Pascal-eloszlású minta.
a) Adjunk X_1 függvényeként torzítatlan becslést $p(1-p)$-re!
b) Az a)-ban megadottnál konstruáljunk jobb becslést a Rao–Blackwell–Kolmogorov-tétel segítségével!

3. heti feladatsor, (statisztika. gyak.) 2012. február 29–március 1.

(1) Legyen $X_1, X_2, \ldots, X_n$ független minta valamely véges szórású eloszlásból, és tekintsük az $T(X_1, X_2, \ldots, X_n) = a_1X_1 + a_2X_2 + \ldots + a_nX_n$ alakú lineáris becsléseket, ahol $a_1, \ldots, a_n$ valós számok. Feltéve, hogy a becslésünk torzítatlan a várható értékre, mely $a_1, \ldots, a_n$ együtthatókra lesz minimális a $T(X_1, X_2, \ldots, X_n)$ mennyiség szórásnégyzete?

(2) Torzítatlan, illetve konzisztens becslés-e a mintaátlag reciproka geometriai eloszlás paraméterére?

(3) Adjunk meg torzítatlan becslést a $[0, \vartheta]$ intervallumon egyenletes eloszlás paraméterére a) mintaátlag; b) a maximum; c) a minimum segítségével. Számítsuk ki a becslések szórását is. Melyik hatásosabb? Melyik konzisztens?

(4) A paraméter mely függvényeit lehet torzítatlanul becsülni a) indikátor b) geometriai c) Poisson-eloszlásból származó minta esetén a mintaátlag és a tapasztalati szórásnégyzet segítségével?
Igaz-e, hogy a Poisson-eloszlás esetén $\overline{X}^2$ torzítatlan becslése a paraméter négyzetének?

(5) Legyen $X_1, \ldots, X_n$ független, b paraméterű Poisson-eloszlású minta ($b > 0$). Számítsuk ki a Fisher-információt! Adjunk elégséges statisztikát! Mi történik $a > 0$ paraméterű exponenciális minta esetén?

2012. március 14-ig beadható házi feladatok

(6) Legyen $X_1, \ldots, X_n$ független minta $Bin(4, p)$-ből, továbbá $Y_1, \ldots, Y_n$ független minta $Bin(6, p)$-ből, és tegyük fel, hogy a két minta egymástól is független. Milyen a és b számokra lesz $a\overline{X} + b\overline{Y}$ a p paraméter torzítatlan becslése? Ezen számpárok közül melyikre lesz a becslés szórása minimális?

(7) Mutassuk meg, hogy exponenciális eloszlású minta esetén
a) $n \cdot \min(X_1, \ldots, X_n)$ torzítatlan, de nem konzisztens a várható értékre.
b) $1/\overline{X}$ nem torzítatlan, de konzisztens a paraméterre.

2. heti feladatsor, (statisztika. gyak.) 2012. február 22-23.

(1) Tegyük fel, hogy február 17-e középhőmérséklete Budapesten az elmúlt 10 évben így alakult (Celsius-fokban):

2,0	8,5	1,6	-4,5	3,3	7,9	0,2	-1,6	-2,2	5,6

Számítsuk ki és ábrázoljuk a középhőmérséklet sűrűségfüggvényének Parzen–Rosenblatt-becslését, ha $h = 0,25$ és a magfüggvény:

$$k(x) = \begin{cases} 1/2, & \text{ha } |x| < 1 \\ 0 & \text{különben.} \end{cases}$$

(2) Mi a mintatér az alábbi esetekben:
a) Három kockadobás.
b) Öten addig lottóznak hetente egy-egy szelvénnyel, míg legalább két találatuk nem lesz. A megfigyelés, hogy kinek hányadik héten sikerül ez először.
c) Húsz napon át feljegyezzük egy diák felkelésének időpontját.

(3) Adjunk torzítatlan becslést a valószínűségszámítás vizsga bukási arányára, ha 300-ból 100-an buktak meg. Adjunk a becslés szórására felső becslést.

(4) n elemű $a > 0$ paraméterű exponenciális minta esetén adjunk torzítatlan becslést $1/a$-ra és $\exp(-3a)$-ra.

(5) n elemű $b > 0$ paraméterű Poisson-eloszlású minta esetén adjunk torzítatlan becslést b-re, b^2-re és $\exp(-b)$-re.

1. heti feladatsor, (statisztika. gyak.) 2012. február 15.

(1) Legyen X standard normális eloszlású valószínűségi változó, mennyi annak valószínűsége, hogy
$2X + 3 < 2$?

(2) n elemű független mintára határozzuk meg a tapasztalati közép és a tapasztalati szórásnégyzet várható értékét!

(3) 16 darab független, 10 várható értékű és 25 szórásnégyzetű normális eloszlás átlaga milyen valószínűséggel nagyobb 9-nél?

(4) Tekintsünk n elemű exponenciális mintát, azaz legyenek $X_1, X_2, \ldots, X_n$ független, azonos λ paraméterű exponenciális eloszlású valószínűségi változók. Határozzuk meg a legkisebb mintaelem eloszlását!

(5) Határozzuk meg a mintaközepet, tapasztalati szórásnégyzetet és szórást, a harmadik tapasztalati momentumot, a szórási együtthatót és a tapasztalati eloszlásfüggvényt az alábbi mintákon:

a) 1, 3, 0, 1;

b) négy kockadobás;

c) fogyasztói árindex:

1993	1994	1995	1996	1997	1998	1999	2000
122,5	118,8	128,2	123,6	118,3	114,3	110,0	109,8

2001	2002	2003	2004	2005	2006	2007	2008
109,2	105,3	104,7	106,8	103,6	103,9	108,0	106,1

(6) Dobjunk fel háromszor egy dobókockát.

a) Határozzuk meg az alapstatisztikákat: mintaközép, korrigált tapasztalati szórásnégyzet és szórás, kvartilisek, rendezett minta.

b) Határozzuk meg az F_n tapasztalati eloszlásfüggvényt, és a $\sup_x |F(x) - F_n(x)|$ statisztika értékét, ahol F az $\{1, 2, 3, 4, 5, 6\}$ pontokon egyenletes eloszlás eloszlásfüggvénye.