Leíró és matematikai statisztika gyakorlat
Matematikai elemző szakirány
2015/2016 tavaszi félév

Játékszabályok

- A gyakorlatokról maximum 3-szor lehet hiányozni. Aki többször hiányzik, nem kap gyakjegyet.
- 100 + x pontot lehet szerezni a félév során:
  - 30 pont: 1. ZH: III.16. (60 perces) leíró statisztikából
  - 40 pont: 2. ZH: V.11. (80 perces) matematikai statisztikából
  - 30 pont: két, egyenként 15 pontos beadandó feladat
  - x pont: szorgalmi feladatok
- Mindkét ZH-n minimálisan teljesíteni kell 30 %-ot.
- Ha egy ZH sikertelen, nem írod meg, vagy javítani szeretnél, akkor vizsgaidőszak első hetén lesz lehetőség a pótZH megírására vagy a javításra. Csak az egyik ZH anyagából javíthatsz, és a jobbik eredményt veszem figyelembe, azaz nem lehet rontani. Két sikertelen vagy meg nem írt ZH esetén gyakUV-t kell írnod, és legfeljebb 2-est kaphatsz.
- A ZH-kon használható: **számológép** és egy legfeljebb A4-es méretű lapra KÉZZEL írott "puska".
- Osztályozás:

elégtelen (1)	0	-	34,99
elégséges (2)	35	-	49,99
közepes (3)	50	-	64,99
jó (4)	65	-	79,99
jeles (5)	80	-	1000

Infók a gyakvezetőről

Név	Varga László
Tanszék	Valószínűségelméleti és Statisztika Tanszék (ELTE TTK)
Szoba	D 3-309
E-mail	`vargal4@cs.elte.hu`
Honlap	`www.cs.elte.hu/~vargal4`

Ajánlott irodalom

- Szarvas-Sugár: Példatár a Statisztika c. tankönyvhöz
- Móri-Szeidl-Zempléni: Matematikai statisztikai feladatok

1.) Döntsd el, hogy az alábbiak egy sokaságot definiálnak, a sokaság egy-egy egyedére vonatkoznak, vagy statisztikai adatok! A sokaságok esetében határozd meg a sokaság típusát is, az adatok esetében pedig azt, hogy alapadatokról, vagy leszármaztatott értékekről van-e szó!

a.) a teremben lévő lányok átlagmagassága
b.) az épület előtti parkolóban álló autók száma
c.) az épület előtti parkolóban álló autók
d.) az épület előtti parkolóban álló ASY-766 rendszámú Opel Vectra
e.) az épület előtti parkolóban álló Opelek aránya
f.) az egy hét alatt legyártott selejtes termékek
g.) a MOL nyeresége
h.) a bankszámlámon jóváírt kamatok
i.) a múlt tavaszi ELTE 5 km-en legjobb időt elérő másodéves hallgató (nem volt holtverseny)
j.) a teremben lévő hallgatók

2.) Határozd meg, hogy a következő ismérvek milyen típusúak és hogy milyen skálán mérhetők! Mennyiségi ismérvek esetén állapítsd meg, hogy az adott ismérv diszkrét vagy folytonos!

a.) szemszín	b.) testmagasság
c.) hőmérséklet	d.) munkahely
e.) születési idő	f.) egy vállalat bérköltsége

g.) a szurkolók véleménye a magyar fociválogatottról a norvégok elleni mérkőzés után

3.) Egy vállalat belső céges kimutatásaiban kényelmi okokból ezer Ft-ra kerekítve jelenítik meg az árbevétel értékét – az árbevétel rovatnál 234 654 e Ft szerepel.

a.) Határozd meg az árbevétel abszolút hibakorlátját! Határozd meg, mely intervallumban található a tényleges árbevétel!
b.) Határozd meg az árbevétel relatív hibakorlátját!

Mennyire változnak a hibakorlátok, ha a vállalat vezetői az áttekinthetőség érdekében M Ft-ra kerekítve kérik az árbevétel értéket, azaz 235 M Ft szerepel az árbevétel rovatban?

4.) Határozd meg, hogy az alábbi mondatokban milyen viszonyszámok rejtőznek, azok milyen típusúak, és add meg precíz kiszámításukat (számláló, nevező, mértékegységek)

a.) Egy 25 fős csoportban a lányok részaránya 40%.
b.) Egy 50 fős csoportban az egy lányra jutó fiúk száma 1,5.
c.) Idén 180, a tavalyihoz képest 10%-kal kevesebb hallgató vette fel az Algebra I. tantárgyat.

d.) Marika összesen 2000 km-es nyaralása alatt autója átlagfogyasztása 8 l/100 km volt.

e.) Az ELTE-n 4000 diák van, az egy tanárra jutó diákok száma 20.

5.) Egy termelő vállalatnál a fizikai munkát végzők összesen 18000 db alkatrészt állítottak elő, amiből a nők teljesítménye 8500 db volt. A vállalatnak 950 férfi fizikai dolgozója van. A nőknél az egy főre jutó termelt mennyiség 17 db/fő.
Szerkessz statisztikai táblát a megadott adatokból és töltsd ki a hiányzó adatokat! Milyen típusú a tábla és milyen típusú sorokat tartalmaz?

6.) Magyarország népességéről az alábbiakat ismerjük:

Település jellege	Népesség megoszlása 2012-ben (%)	Népesség változása 1990-ről 2012-re (%)
Budapest	17,4	-14,4
Többi város	51,9	-2,4
Községek	30,7	-0,8
Összesen	100,0	...

a.) 1990 és 2012 között évente átlagosan mennyivel változott a budapesti lakosság?

b.) Hány százalékkal változott a népesség száma 1990-ről 2012-re?

c.) Melyik településen élők részaránya csökkent?

7.) Egy vállalatnál az alkalmazottak számára vonatkozóan tartalmaz január 1-jei adatokat a következő táblázat:

Év	Alkalmazottak száma	Változás		
	január 1-jén (fő)	2012-höz képest (%)	az előző évhez képest (%)	az előző évhez képest (fő)
2010	...	...	...	...
2011	95	...	...	+5
2012	100	...	...	...
2013	...	...	+20	...
2014	...	+50	...	...

a.) Milyen típusú a tábla és milyen típusú sorokat tartalmaz?

b.) Töltsd ki a táblázatot! Értelmezzünk a táblázatban néhány értéket!

c.) Határozd meg az alkalmazottak átlagos számát a 2011-es évben, valamint 2010. január 1. és 2013. január 1. között!

d.) Jellemezd az alkalmazottak számának évi átlagos változását 2010. január 1. és 2014. január 1. között!

e.) Ábrázold az alkalmazottak számának alakulását megfelelő diagrammal!

8.) Az egyetem büféjében egy adott napon az összes vendég fogyasztását megvizsgálták, és ez alapján az elköltött összegekről az alábbi táblázatot készítették el:

Fogyasztás összege (Ft)	Vendégek száma (fő)
0 – 200	40
201 – 500	42
501 – 800	80
801 – 1200	22
1201 –	16
Összesen	200

a.) Átlagosan mennyit költöttek a büfében? Készítsünk hisztogramot!

b.) Vizsgáld az elköltött összegek koncentrációját Lorenz-görbével, koncentrációs együtthatóval és Herfindahl-indexszel!

9.) Van két piac, az elsőn 10 azonos méretű vállalat tevékenykedik, a másodikon pedig 5 azonos méretű vállalat van. Hasonlítsuk össze a két piac koncentrációját!

SZ1.) Tulajdonosa vagy egy vállalkozásnak, év végén a könyvelőd jelentést készít, amiben a következőket írja: "Remek évet zártál, a tervezetthez képest magasabb lett az árbevétel, bár a költségek csak stagnáltak. A árbevétel tervezett 10%-os növelését 10%-kal túlteljesítetted, így 20%-os növekedést hoztál össze. Az előző évhez képest 10%-os költségcsökkenést terveztél, azonban a tervezetthez képest 10%-kal magasabbak lettek a költségek, így összességében a költségek összege nem változott."
Értékeld a könyvelőd érvelését! Van-e benne valami, ami sántít; és ha igen, miért? (1p)

SZ2.) Egy vállalat alkalmazottainak száma 2010-ről 2014-re évente átlagosan 5,7371 %-kal, azaz 3 fővel nőtt. A kifizetett összes éves bér 2014-ben 54 millió Ft volt, míg a 2010-es havi átlagbér 250 ezer Ft volt.
A rendelkezésre álló adatokból készíts táblát és töltsd ki az üres rubrikákat! Átlagosan mennyivel változott az átlagbér? (2p)

SZ3.) Határozd meg a Gini-együttható lehetséges legkisebb és legnagyobb értékét! Mikor veheti fel ezeket? (2p)

SZ4.) Mutasd meg, hogy az L koncentrációs együttható valóban a koncentrációs terület 2-szerese! (3p)

SZ5.) Bizonyítsuk be, hogy $\frac{1}{k} \leq HI \leq 1$! Mikor veheti fel ezeket a szélső értékeket? (2p)

10.) Egy kereskedelmi egység három fajta paprikás chips-et árul, a következő táblázat a 2014/2015-ös értékesítésről tartalmaz adatokat:

Márka	2014.		2015.	
	Értékesített mennyiség (db)	Egységár (Ft/db)	Értékesített mennyiség (db)	Egységár (Ft/db)
Chio	25	300	30	310
Lays	20	250	30	240
Cheetos	10	200	10	220
Összesen	...	...	...	...

a.) Jellemezd az értékesítésben bekövetkezett mennyiségi, ár- és értékváltozásokat egyedi és összetett indexekkel! Értelmezd szövegesen az egyes indexeket!

b.) Számítsd ki az árváltozás miatti többletbevételt!

11.) Egy vállalat termelési értékének (árbevételének) a 35,4%-át 2014-ben az I. számú üzem, a többit pedig a II. számú üzem adta. Az I. számú üzem termékeinek egységára 2014-ről 2015-re átlagosan 5%-kal, a II. számú üzemé pedig átlagosan 3%-kal csökkent.
Számítsuk ki a vállalati termelés volumenének változását, ha ismert, hogy a vállalati termelési érték 3%-kal emelkedett! Értelmezd szövegesen a kapott volumenváltozást!

12.) Egy kisvállalkozás kétféle terméket gyárt. Termelési adatai:

Termék fajtája	Termelési érték 2015-ben (M Ft)		Volumenváltozás (2014=100%)
	folyó áron	2014-es áron	
A	50	60	110
B	60	80	120
Összesen	...	...	...

a.) Határozd meg a termelés értékindexét!

b.) Határozd meg mindkét súlyozással az ár- és volumenindexeket!

c.) Számítsd ki az volumenváltozás miatti többletbevételt 2014-es árakon!

13.) Mari néni kávézójában 3 féle kávét szolgál fel, a családi könyvelésből az alábbi adatok ismertek:

Kávéfajta	A forgalom értéke 2015-ben (e Ft)	Az árak	A forgalom értékének
		alakulása, 2015/2010 (%)	
Cappuccino	2000	130	200
Cafe Latte	1500	120	180
Espresso	1000	120	150

a.) Számíts érték-, ár- és volumenindexet a kávézó forgalmára vonatkozóan!

b.) A forgalom értékének növekedéséből hány forint volt az ár- és a volumenváltozás hatása?

SZ6.) Egy vállalat bázisidőszaki árbevétele 20 millió forint. Határozd meg a tárgyidőszaki árbevételt, ha a Fisher-féle árindex, a tárgyi súlyozású árindex és a Fisher-féle volumenindex megegyeznek! (1p)

SZ7.) Fejezd ki a Fisher-féle árindexet az egyedi árindexek súlyozott átlagaként, azaz $I_p^F = \sum_j w_j i_{p,j}$ alakban, alkalmasan választott w_j súlyokkal! (2p)

SZ8.) Egy boltban háromféle cigarettát árusítanak. A cigaretták összes forgalma (árbevétele) 2013-ról 2015-re 20%-kal emelkedett. A cigarettákra vonatkozóan az alábbi adatokat ismerjük:

Márka	Árbevétel megoszlása (%)		Árak alakulása	Árak alakulása	Árbevétel alakulása
	2013.	2014.	2014/2013 (%)	2015/2014 (%)	2015/2014 (%)
Marlboro	60	50	115	110	105
Helikon	20	30	107	102	110
Sopianae	20	20	125	95	120

Számítsd ki a Fisher-féle volumenindexet, ha a bázisidőszak 2013, a tárgyidőszak pedig 2015! (2p)

14.) Legyen az $X_1, X_2, \ldots$ val. változók közös sűrűségfüggvénye (c valós paraméter) $f(x) = \begin{cases} c, & \text{ha } 0 < x \le 1 \\ \frac{1}{x^3}, & \text{ha } x > 1 \\ 0 & \text{különben} \end{cases}$

a.) Határozd meg a c értékét!

b.) Határozd meg X_1 eloszlásfüggvényét!

c.) $P(X_1 = e^{\pi^{666}}) = ?$ $\quad P\left(\frac{1}{2} < X_1 < 2\right) = ?$

d.) Számítsd ki X_1 várható értékét!

e.) Hova és hogyan konvergál $\frac{X_1 + \ldots + X_n}{n}$ $n \to \infty$ esetén?

15.) Legyenek $X_i \sim N(10, 5^2)$ $(i = 1, \ldots, 9)$ függetlenek.

a.) Számítsuk ki a $P(\overline{X} < 9)$ mennyiséget!

b.) Közelítsük a keresett valószínűséget szimuláció segítségével!

16.) Legyen $X_1, \ldots, X_n$ független, azonos abszolút folytonos eloszlású valószínűségi változók sorozata. Adjuk meg X_1^* és X_n^* eloszlás- és sűrűségfüggvényét! A minimumnál külön is vizsgáljuk meg azt az esetet, ha az X_i változók exponenciális eloszlásúak!

17.) Legyen $X \sim \mathrm{Exp}(\lambda)$. Határozd meg X móduszát és tetszőleges kvantilisét! Hasonlítsd össze a mediánt és a várható értéket!

18.) Legyen $X \sim \text{Ind}(p)$. Határozd meg X móduszát, kvantilisfüggvényét, ferdeségét és csúcsosságát!

19.) Legyen $X \sim N(m, \sigma^2)$.

a.) Határozd meg X móduszát, mediánját, ferdeségét és csúcsosságát!

b.) Határozd meg a $P(m - k\sigma < X < m + k\sigma) = P(|X - m| < k\sigma)$ valószínűséget $k = 1, 2, 3, 4, 5, 6$ értékek esetén!

20.) Egy osztályban a diákok magassága (cm):

180 163 1500 157 165 165 174 191 172 165 1-68 186

a.) Nézzük át nagy vonalakban az adatokat, reálisak-e! Próbáljuk javítani az esetleges adathibákat!

b.) Rajzold fel a tapasztalati eloszlásfüggvényt! Mennyi a tapasztalati eloszlásfüggvény értéke a 180 helyen?

c.) Elemezd a diákok testmagasságát

- átlag;
- korrigált tapasztalati szórás;
- szórási együttható;
- kvartilisek;
- terjedelem;
- interkvartilis terjedelem;
- tapasztalati ferdeség;
- tapasztalati csúcsosság segítségével!

Értelmezd is az eredményeket!

d.) Készíts boxplot ábrát!

e.) Készíts alkalmas osztályközös gyakorisági sort, majd abból hisztogramot!

21.) A `Nyarhom.Rdata` nevű fájl a 2014. nyári napi maximum-hőmérsékleteket tartalmazza egy településen (°C).

a.) Nézzük át nagy vonalakban az adatokat, reálisak-e! Próbáljuk javítani az esetleges adathibákat!

b.) Elemezd együtt a nyári maximális hőmérséklet értékeket

- átlag;
- korrigált tapasztalati szórás;
- szórási együttható;
- kvartilisek;
- terjedelem;
- interkvartilis terjedelem;
- tapasztalati ferdeség;
- tapasztalati csúcsosság segítségével!

Értelmezd is az eredményeket!

c.) Készíts boxplot ábrát!

d.) Készíts osztályközös gyakorisági sort, majd abból hisztogramot!

22.) 2014-ben egy bútorboltban eladott konyhabútorok értékéről az alábbi adatok ismerjük:

Konyhabútor ára (e Ft)	Eladott bútorok száma (db)
– 100	10
101 – 200	18
201 – 300	25
301 – 500	30
501 –	17
Összesen	100

Jellemezd (szövegesen is) az értékesített konyhabútorok árának eloszlását alapstatisztikák (módusz, kvartilisek, átlag, szórás) segítségével! Milyen az eloszlás ferdesége? Készíts hisztogramot!

SZ9.) Legyen $X_1^* < X_2^* < \cdots < X_n^*$ az f sűrűségfüggvényű és F eloszlásfüggvényű abszolút folytonos eloszlásból vett rendezett minta.

a.) Határozd meg X_2^* eloszlás- és sűrűségfüggvényét!

b.) $\text{cov}(X_1^*, X_2^*) =?$ (2+2= 4p)

SZ10.) Vezesd le az n és p paraméterű binomiális eloszlás móduszát! (1 p)

SZ11.) Vezesd le az $\text{Exp}(\lambda)$ eloszlás ferdeségét és csúcsosságát! (2p)

23.) Február 17-én Budapesten az elmúlt 10 évben az alábbi középhőmérsékleteket mérték: 2; 2,5; 1,6; -4,5; 5,3; 7,9; 1,5; -1,6; -2,2; 1,6.

a.) Számítsuk ki és ábrázoljuk a középhőmérséklet sűrűségfüggvényének Parzen-Rosenblatt becslését, ha h=0,25 és a magfüggvényünk $k(x) = \begin{cases} \frac{1}{2} & \text{ha } |x| < 1 \\ 0 & \text{különben} \end{cases}$!

b.) Készítsük el a Parzen-Rosenblatt-féle sűrűségfüggvénybecslést Gauss-magfüggvény esetén különöző sávszélességekre (R segítségével)!

24.) Legyen $X_1, ..., X_{20}$ i.i.d. minta $N(m, 1^2)$ eloszlásból. Célunk az ismeretlen m paraméter becslése. Tekintsük az alábbi három statisztikát:

- $T_1(\mathbf{X}) = X_8$,
- $T_2(\mathbf{X}) = \frac{X_3+X_7}{2}$,
- $T_3(\mathbf{X}) = \frac{X_9+X_{19}}{8}$.

a.) A fenti statisztikák közül melyek torzítatlanok? Amelyik nem torzítatlan, hogyan tudnánk torzítatlanná tenni?

b.) Vizsgáljuk meg a fenti statisztikák közül a torzítatlanokat hatásosság szempontjából!

25.) Legyen $X_1,\ldots,X_n$ i.i.d. minta ismeretlen eloszlásból.

a.) Torzítatlan becslés-e a várható értékre nézve az átlag?

b.) Torzítatlan becslés-e a szórásnégyzetre nézve a tapasztalati szórásnégyzet? Amennyiben nem az, hogyan tudnánk torzítatlanná tenni?

c.) Mikor konzisztens becslése a várható értéknek az átlag?

d.) Adjunk torzítatlan és konzisztens becslést az eloszlásfüggvényre!

26.) $X_1,\ldots,X_n \sim$Exp(λ) i.i.d. minta esetén adjunk torzítatlan becslést $e^{-3\lambda}$-ra és $\frac{1}{\lambda}$-ra!

27.) $X_1,\ldots,X_n \sim$Poi(λ) i.i.d. minta esetén adjunk torzítatlan becslést $e^{-\lambda}$-ra és λ^2-re!

28.) Adjunk torzítatlan becslést a $[0,\vartheta]$ intervallumon egyenletes eloszlás ismeretlen ϑ paraméterére a

a.) mintaátlag

b.) maximum

segítségével. Melyik a hatásosabb? Melyik konzisztens?

29.) Mutassuk meg, hogy exponenciális eloszlású minta esetén $T(\mathbf{X}) = n \cdot X_1^*$ statisztika torzítatlan, de nem konzisztens becslése a várható értéknek!

30.) Legyen $X_1,...,X_n$ i.i.d. minta valamely véges szórású eloszlásból, és tekintsük a T($\mathbf{X}$)= $a_1X_1 + ... + a_nX_n$ alakú lineáris becsléseket, ahol $a_1,...,a_n \in \mathbb{R}$. Feltéve, hogy T($\mathbf{X}$) a várható érték torzítatlan becslése, mely $a_1,...,a_n$ számokra lesz minimális a $D^2(T(\mathbf{X}))$?

SZ12.) Öt véletlen számot jegyeztünk fel: 100,32,76,52,17. Ha tudjuk, hogy ezek az $\{1,2,...,N\}$ halmazból vett véletlen minta elemei, akkor hogyan becsülnénk az N paramétert? (1p)

SZ13.) Adjunk torzítatlan becslést a $[0,\vartheta]$ intervallumon egyenletes eloszlás paraméterére a minimum segítségével. Hatásosabb a becslés, mint a 28. feladat a.) részében kapott torzítatlan becslés? Konzisztens a becslés? (2p)

SZ14.) Legyen $X_1,\ldots,X_n$ i.i.d. minta az $E(a,b)$ eloszlásból, a és b paraméterek. Mutassuk meg, hogy b ML-becslése nem torzítatlan, de aszimptotikusan torzítatlan becslése b-nek! (2p)

SZ15.) Legyen $X_1,...,X_n$ i.i.d. minta Bin(k,p)-ből, $Y_1,...,Y_n$ i.i.d. minta Bin(l,p)-ből, és tegyük fel, hogy a két minta egymástól is független. Milyen (a,b) számpárokra lesz $a\overline{X} + b\overline{Y}$ a p paraméter torzítatlan becslése? Ezen számpárok közül melyikre lesz a becslés szórása minimális? (2p)

31.) Határozzuk meg az ismeretlen paraméter(ek) ML becslését, ha a minta

a.) Exp(λ) eloszlású;

b.) Poi(λ) eloszlású;

c.) E(a,b) eloszlású, ahol $a<b$, mindkettő paraméter.

Torzítatlan a becslés? Ha nem az, próbáljuk torzítatlanná tenni! Konzisztens a becslés?

32.) Legyen $X_1,...,X_n$ Pascal-eloszlású (geometriai eloszlású) minta p paraméterrel.

a.) Adjunk meg X_3 függvényeként torzítatlan becslést $p(1-p)^4$-re!

b.) Adjunk maximum likelihood becslést $p(1-p)$-re!

33.) Legyenek X_1,...,X_n és Y_1,...,Y_m egymástól független λ illetve $\frac{1}{\lambda}$ paraméterű exponenciális eloszlású minták. Határozzuk meg az ismeretlen paraméter (együttes) ML becslését!

34.) Tegyük fel, hogy a minta kétparaméteres eloszláscsaládból származik, a paraméterek a és b.

Ekkor mutassuk meg, hogy az $\left\{\begin{array}{lcl} E_{a,b}X & = & m_1 \\ E_{a,b}X^2 & = & m_2 \end{array}\right.$ egyenletrendszer megoldása megegyezik az $\left\{\begin{array}{lcl} E_{a,b}X & = & m_1 \\ D^2_{a,b}X & = & s_n^2 \end{array}\right.$ egyenletrendszer megoldásával.

35.) Becsüld a paramétert momentum-módszerrel, ha a minta eloszlása:

a.) Exp(λ);

b.) Poi(λ);

c.) E(a,b);

d.) E($-a,a$);

e.) N($2m+5,(\frac{1}{d})^2$) .

36.) Legyen a $Z_1,\ldots,Z_5$ minta **I.)** N($m,2^2$) **II.)** N($2m+5,2^2$) eloszlású. A megfigyelt értékek a következők: 6; 4,5; 2,5; 2; 1.

a.) Határozzunk meg 95%-os (99%-os) megbízhatóságú konfidenciaintervallumot m-re!

b.) Hány elemű mintára van szükségünk, ha azt szeretnénk, hogy a konfidenciaintervallum legfeljebb 0,01 hosszúságú legyen?

c.) Mi változik az a.) esetben, ha a szórást nem ismerjük?

d.) Adjunk a szórásra 98%-os megbízhatóságú konfidenciaintervallumot.

$\sum_{i=1}^{5} z_i = 16 \quad \sum_{i=1}^{5}(z_i-\overline{z})^2 = 16,3$

$\chi^2_{4;0,01} = 0,3 \;\; \chi^2_{4;0,99} = 13,28$

37.) Az előző évben figyelemmel kísértük a sárkányföldi tőzsdeindex, a SÜSÜX változását. Az alapstatisztikák: átlag: 1113,18; szórás: 95,31. A tőzsde 200 napon keresztül volt nyitva. Adjunk ezek alapján 95%-os megbízhatóságú konfidenciaintervallumot az index adott évre vonatkozó várható értékére!

SZ16.) Határozzuk meg az ismeretlen paraméter ML becslését, ha a minta $\mathrm{E}(-a, a)$ eloszlású! Torzítatlan a becslés? Ha nem az, próbáljuk torzítatlanná tenni! Konzisztens a becslés? (3p)

SZ17.) Egy CASCO biztosítás kárai 2013-ban 200, 1200, 1800, 125, 485 ezer Ft voltak. A káreloszlásról feltételezzük, hogy (α, β) paraméterű Pareto-eloszlású, azaz az eloszlásfüggvény

$$F(x) = \begin{cases} 1 - \left(\frac{\beta}{\beta+x}\right)^{\alpha} & \text{ha } x > 0 \\ 0 & \text{különben} \end{cases}$$

Számítsd ki a Pareto-eloszlás várható értékét és szórásnégyzetét, majd határozd meg az ismeretlen paraméterek momentum módszeres becslését a minta alapján! (2p)

SZ18.) Tegyük fel, hogy az n elemű mintánk lognormális eloszlású, azaz a mintaelemek logaritmusa $\mathrm{N}(m, \sigma^2)$ eloszlású. Határozd meg az ismeretlen paraméterek maximum likelihood és momentum becslését! Segítség: használhatod a Wikipédiáról a sűrűségfüggvény képletét és a kiszámított várható értéket/szórásnégyzetet. (2p)

SZ19.) Mutasd meg, hogy az n szabadságfokú Student-féle t-eloszlás eloszlásban a standard normális eloszláshoz tart, ha $n \to \infty$. (1p)

38.) Valaki azt állítja, hogy a klíma változik, és ezt azzal véli bizonyítottnak, hogy az elmúlt 10 évben 2-szer is volt jégeső, pedig korábban az egyes évekre a jégeső valószínűsége a hivatalos adatok alapján csupán p=0.1 volt. Írjuk fel a hipotéziseket, a próbát és állapítsuk meg az elsőfajú hiba valószínűségét, valamint az erőfüggvényt a p=0.2 pontban!

39.) Az X valószínűségi változó egyenletes eloszlású a (-b; 1+2b) intervallumon. A H_0: b=0 hipotézist szeretnénk ellenőrizni a H_1: b>0 hipotézis ellenében, e célból a következő próbát alkalmazzuk: egy megfigyelést végzünk és ha ez a (0,1; 0,85) intervallumba esik, elfogadjuk H_0-t, különben elvetjük.
Írjuk fel a próba erőfüggvényét! Mekkora a próba terjedelme?

40.) Legyen $X_1, ..., X_n$ minta az $f(x) = \begin{cases} \frac{2x}{a^2} & \text{ha } 0 \le x \le a \\ 0 & \text{különben} \end{cases}$ sűrűségfüggvényű eloszlásból.

a.) Tekintsük a következő hipotéziseket: H_0: a=1 H_1: 0<a<1
Adjunk meg X_n^* függvényében 5 %-os terjedelmű próbát, keressük a kritikus tartományt $\mathcal{X}_k = \{\underline{x} : T(\underline{x}) < c\}$ alakban! Mi lesz az erőfüggvény?

b.) Tekintsük a következő hipotéziseket: H_0: a=1 H_1: $a = \frac{1}{2}$
Adjunk meg α terjedelemhez egyenletesen legerősebb próbát!

41.) Legyen két megfigyelésünk a $(3; p)$ paraméterű binomiális eloszlásból. Adjuk meg a legjobb olyan próbát az alábbi hipotézisekre, melynek elsőfajú hiba valószínűsége $0,04$: $H_0: \ p = \frac{1}{2}$ $H_1: \ p = \frac{1}{4}$

42.) Az alábbi minta 4 év október 18-án Budapesten mért napi középhőmérséklet adatait tartalmazza. Ellenőrizzük a H_0: m =15 hipotézist α =0.05 elsőfajú hibavalószínűség mellett *értelmes* alternatív hipotézissel szemben.
Középhőm. (C fok) adatok:

14,8	12,2	16,8	11,1

a.) A korábbi tapasztalatok alapján tekintsük az értékek szórását 2-nek. Adjuk meg a p-értéket is.

b.) Ne használjunk a szórásra vonatkozóan előzetes információt.

43.) Tegyük fel, hogy az emberi magasság normális eloszlású.

a.) Végezzünk statisztikai próbát arra vonatkozóan, hogy a gyakorlaton lévő lányok átlagmagassága 170 cm!

b.) Végezzünk statisztikai próbát arra vonatkozóan, hogy a gyakorlaton lévő fiúk átlagmagassága 180 cm!

44.) A Természettudományi Kar II. évfolyamán az egyik gyakorlati csoportban 10-en írtak statisztika zárthelyit. Két feladatsor volt, mindkettőben 30 pontot lehetett elérni. Tegyük fel, hogy az elért pontszámok normális eloszlásúak. A pontszámokat tartalmazza az alábbi táblázat:

1. feladatsor	12	11	8	14	10
2. feladatsor	15	14	9	16	11

a.) Vajon az első feladatsor nehezebb volt?

b.) Mennyiben változik a helyzet, ha nem 10 diákról, hanem csak 5-ről van szó, és a 2. feladatsor a pótZH eredménye?

45.) Az alábbi két minta 10 egyforma képességűnek feltételezett sportoló súlylökésben elért eredményeit tartalmazza. A sportolók két ötfős csoportban készültek az edzőtáborban. Edzéstervük ugyanaz volt, de az első csoportban készülők minden reggel fejenként 10 tojást és 25 túró rudit ettek meg. A második csoportban készülőknek reggel és este 1-1 kg szalonnát és 1-1 kg madártejet kellett megenni. 2 hét felkészülés után értékelték az eredményeket. Tételezzük fel, hogy normális eloszlásból származnak a minták és a terjedelem 5%.

1. csoport	15,8	15,2	16,3	17,1	16,1
2. csoport	19,0	12,1	17,2	14,7	21,0

a.) Melyik diéta volt jobb, ha a dobások szórását 2-nek tekintjük?

b.) Állíthatjuk-e, hogy a második csoportban nagyobb változékonyságot mutat a sportolók teljesítménye?

c.) Ha nem ismerjük a szórást, akkor tekinthetjük-e valamelyik diétát jobbnak?

$F_{4,4}^{0,95} = 4,4$ $F_{5,5}^{0,95} = 5,05$ $F_{4,4}^{0,975} = 9,6$ $F_{5,5}^{0,975} = 7,15$

46.) Tegyük fel, hogy az emberi magasság normális eloszlású. Végezzünk alkalmas statisztikai próbát arra vonatkozóan, hogy a gyakorlaton lévő lányok átlagmagassága megegyezik a fiúk átlagmagasságával!

SZ20.) Oldd meg a 40.) feladatot abban az esetben, ha baloldali kritikus tartományt választunk: $\mathcal{X}_k = \{\underline{x} : T(\underline{x}) > c\}$. (1p)

SZ21.) Egy érme szabályosságát (a H_1: $p>0,5$ ellenhipotézissel szemben; p a fejdobás valószínűsége) az alábbi módszerrel teszteljük: n-szer feldobjuk az érmét, és ha legalább 2 írást dobtunk, akkor elfogadjuk H_0-t.

a.) Mekkora legyen n, hogy az elsőfajú hiba kisebb legyen, mint 0,05?

b.) Adjuk meg a próba erőfüggvényét! (1+1= 2p)

SZ22.) A Politikatudományi Kar HÖK elnöke nagyon fontosnak tartja népszerűségét. Amennyiben a hallgatók legfeljebb 70%-a utálja, az számára elfogadható (H_0 hipotézis). Az ennél nagyobb arány esetén (H_1 hipotézis) lemond. Minden negyedév végén 10 hallgatót kérdez meg (közvéleménykutatást tart). Az elnök akkor mond le, ha a tízből legalább 8 diák utálja.

a.) Mekkora a próba terjedelme?

b.) Várhatóan hány negyedévet fog tevékenykedni az elnök, ha stabilan a diákok 65%-a utálja? (2+2= 4p)

SZ23.) A Hurka húsgyárban minden szállítás előtt megvizsgálják a kolbászok szulfáttartalmát. 2014. január 8-án a még megengedett szint %-ban a mérések a következők voltak: 98,5; 101,4; 99,5; 100.9 és 100,7. A korábbi tapasztalatok alapján az ellenőr az eredményekről feltételezi, hogy 1 szórásúak.

a.) Elfogadható-e a H_0: m=100 nullhipotézis α=0,05 elsőfajú hibavalószínűség mellett? Megfelelően válasszuk meg a H_1 hipotézist!

b.) Mennyi lesz a p-érték?

c.) Mennyi a próba erőfüggvényének az értéke az m=102 pontban?

d.) Hány elemű mintára van szükség, ha azt szeretnénk, hogy ez az érték legalább 0,99 legyen? (1+0,5+1,5+1= 4p)

SZ24.) Legyen X_1 minta az $f(x)$ sűrűségfüggvényű eloszlásból. Tekintsük a következő hipotéziseket:

H_0: $f(x) = f_0(x) = 2(1-x) \cdot I(0 < x < 1)$

H_1: $f(x) = f_1(x) = 2x \cdot I(0 < x < 1)$

Adjunk meg α terjedelemhez egyenletesen legerősebb próbát! (1p)

47.) Az alábbi táblázatban adatok találhatók azon személyek számáról, akik lórúgás következtében haltak meg 10 porosz hadtestben 20 év alatt (1875–1894) (összesen $10 \cdot 20 = 200$ adat):

halálesetek száma	0	1	2	3	4
gyakoriság	109	65	22	3	1

Ellenőrizzük azt a hipotézist, hogy a halálesetek száma egy hadtestben egy év alatt Poisson-eloszlású!

48.) Rendelkezésünkre áll a következő minta: 0,55; 0,59; 0,34; 0,69; 0,95; 0,34; 0,53; 0,54; 0,03; 0,11; 0,15; 0,67; 0,48; 0,09; 0,55; 0,02; 0,37; 0,76; 0,83; 0,92. A megoldás során alkalmazzunk diszkretizálást, azaz képezzünk alkalmas gyakorisági sort az adatokból.

a.) Elfogadhatjuk-e azt a hipotézist, hogy a minta (0,2) intervallumon egyenletes eloszlású? Vizsgáljuk meg Q-Q plot-tal is!

b.) Elfogadhatjuk-e azt a hipotézist, hogy a minta egyenletes eloszlású? Vizsgáljuk meg Q-Q plot-tal is!

c.) Elfogadhatjuk-e azt a hipotézist, hogy a minta exponenciális eloszlású? Vizsgáljuk meg Q-Q plot-tal is!

49.) Az Informatikai Kar III. évfolyamán 300-an tanulnak. Megszámolták, hogy a legutóbbi vizsgaidőszakban hányszor buktak az egyes hallgatók. Az eredményeket tartalmazza az alábbi táblázat.

Bukások száma	0	1	2	3	4
Hallgatók száma	80	113	77	27	3

a.) Elfogadhatjuk-e azt a hipotézist, hogy egy hallgató bukásszáma Bin(4; 0,25) eloszlású?

b.) és azt, hogy Bin(4;p) eloszlású?

50.) Az alábbi táblázat CASCO biztosítással rendelkezők éves kárszámát tartalmazza 2012-ben és 2013-ban:

Kárszám	0	1	2	3	4	5	>5
Vezetők száma	3692	232	65	7	3	1	0
Vezetők száma	3542	284	135	24	9	5	1

a.) Vajon tekinthető-e a 2012-es kárszám Poisson-eloszlásúnak?
b.) Vajon tekinthető-e a kárszám azonos eloszlásúnak a két évben?

51.) Az alábbi kontingencia-táblázat mutatja, hogy 100 évben a csapadék mennyisége és az átlaghőmérséklet hogyan alakult.

Hőmérséklet \ Csapadék	Kevés	Átlagos	Sok
Hűvös	15	10	5
Átlagos	10	10	20
Meleg	5	20	5

(A cellákban az egyes esetek gyakoriságai találhatóak.) Tekinthető-e a csapadékmennyiség és a hőmérséklet függetlennek?

SZ25.) 100 napon keresztül feljegyezték egy város energiafogyasztását. Az alábbi táblázat azt tartalmazza, hogy az egyes intervallumokba hány megfigyelés esett, valamint azt is, hogy az adott intervallumba eső értékeknek mennyi az átlaga. Az energiafogyasztást normális eloszlásúnak tekinthetjük?

Intervallumok	< 5000	$5000 - 6000$	$6000 - 7000$	> 7000
Gyakoriságok	20	31	28	21
Átlagok	3875	5700	6500	7800

(2p)

52.) Legyenek adottak a következő (x,y) párok:

x_i	0	1	6	5	3
y_i	4	3	0	1	2

a.) Határozzuk meg és ábrázoljuk is az $aX + b$ alakú regessziós egyenest.
b.) Számoljuk ki a reziduálisokat és becsüljük meg a hiba-szórásnégyzetet.
c.) Adjunk előrejelzést x=10-re a regressziós egyenes alapján.

53.) A január 8-án tartott Statisztika II. vizsgát 8 hallgató írta meg, akiktől megkérdeztük, mennyi órát készültek a vizsgára, hány pontot szereztek a tantárgy előfeltételének számító Statisztika I. tantárgyból a vizsgán és milyen magasok:

Statisztika II. pontszám	49	55	56	62	65	70	78	92
Hány órát készült (ó)	15	16	14	13	12	19	21	24
Statisztika I. pontszám	60	50	66	53	67	76	88	87
Testmagasság (cm)	160	174	178	182	173	168	191	167

a.) Vizsgáljuk meg lineáris regresszióval a tanulási idő hatását a Statisztika II. pontszámra! Ábrázoljuk a regressziós egyenest!
b.) Illesszünk négyzetes regressziós függvényt a Statisztika II. pontszámra, ha a magyarázó változó a tanulási idő! Ábrázoljuk a regressziós egyenest!
c.) Illesszünk lineáris regressziót a Statisztika II. pontszámára, ha a magyarázó változók a tanulási idő és a Statisztika I. pontszám!
d.) Illesszünk lineáris regressziót a Statisztika II. pontszámára, ha a magyarázó változók a tanulási idő, a Statisztika I. pontszám és a testmagasság!
e.) Vessük össze a modelleket!
f.) A Statisztika II. vizsga sikeres, ha a hallgató legalább 50 pontot elér. Juli 175 cm magas, a Statisztika I.-ből 60 pontot szerzett és a Statisztika II.-re 10 órát tervez tanulni. Várhatóan át fog menni a Statisztika II. vizsgán?

54.) Keressük meg "kézzel" és **R** segítségével a legjobb (legkisebb négyzetes) becslést (a, b és c paraméterek)!

Adatok / Modell

a.)

x	-1	1	2
y	1	2	-1

$y = ax + b$

b.)

x	-1	0	1	2
y	8	8	4	16

$y = ax^2 + bx + c$

c.)

x	0	$\frac{1}{2}$	1
y	1	3	7

$y = a \cdot \cos(\pi x) + b \cdot \sin(\pi x)$

d.)

x	0	-1	1	1
y	1	0	-1	1

$a \cdot (x^2 + y^2) + b \cdot (x + y) = 1$

SZ26.) Keressük meg a

x	-1	1	2	3
y	1	2	1	3

pontokat a legkisebb négyzetek módszerével legjobban közelítő, a $(2, 2)$ ponton áthaladó egyenest (elsőfokú polinomfüggvényt)! (2p)