

## Vizsgadolgozat

### Megoldás

**Tanszéki általános alapelvek** A pontozási útmutató célja, hogy a javítók a dolgozatokat egységesen értékeljék. Ezért az útmutató minden feladat (legalább egy lehetséges) megoldásának főbb gondolatait, és az ezekhez rendelt részpontoszámokat közli. Az útmutatónak nem célja a feladatok teljes értékű megoldásának részletes leírása; a leírt lépések egy maximális pontszámot érő megoldás vázlatának tekinthetők.

Az útmutatóban feltüntetett részpontoszámok csak akkor járnak a megoldónak, ha a kapcsolódó gondolat egy áttekinthető, világosan leírt és megindokolt megoldás egy lépéseként szerepel a dolgozatban. Így például az anyagban szereplő ismeretek, definíciók, tételek pusztán leírása azok alkalmazása nélkül nem ér pontot (még akkor sem, ha egyébként valamelyik leírt tény a megoldásban valóban szerephez jut). Annak mérlegelése, hogy az útmutatóban feltüntetett pontszám a fentiek figyelembevételével a megoldónak (részben vagy egészében) jár-e, teljes mértékben a javító hatásköre.

Részpontoszám jár minden olyan ötletért, részmegoldásért, amelyből a dolgozatban leírt gondolatmenet alkalmas kiegészítésével a feladat hibátlan megoldása volna kapható. Ha egy megoldó egy feladatra több, egymástól lényegesen különböző megoldást is elkezd, akkor legfeljebb az egyikre adható pontszám. Ha mindegyik leírt megoldás vagy megoldásrészlet helyes vagy helyessé kiegészíthető, akkor a legtöbb részpontot érő megoldáskezdeményt értékeljük. Ha azonban több megoldási kísérlet között van helyes és (lényeges) hibát tartalmazó is, továbbá a dolgozathoz nem derül ki, hogy a megoldó melyiket tartotta helyesnek, akkor a kevesebb pontot érő megoldáskezdeményt értékeljük (akkor is, ha ez a pontszám 0). Az útmutatóban szereplő részpontoszámok szükség esetén tovább is oszthatók. Az útmutatóban leírtól eltérő jó megoldás természetesen maximális pontot ér.

Aritmetikai hiba esetén elszámolásonként 1-1 pont vonandó le a feladatokból. Ez alól kivétel, ha az elszámolás lényegesen egyszerűsíti vagy módosítja a feladat felépítését. Ilyen esetekben azon feladatrészekért, amik az elszámolás okán fel sem merültek, nem jár pont.

1. Legyen  $X$  és  $Y$  két valószínűségi változó pozitív és véges szórással.

- (a) Hogyan jelöljük  $X$  és  $Y$  korrelációját és hogyan definiáljuk ezt a korrelációt  $X$  és  $Y$  kovarianciája és szórásai segítségével?
- (b) Amennyiben  $X$  és  $Y$  függetlenek, milyen összefüggés áll fenn ezen két valószínűségi változó szórásnégyzetei és az összegük szórásnégyzete között?

*Megjegyzés:* A szórásra, szórásnégyzetre és kovarianciára használhatjuk az előadásról ismert jelöléseket, ezen jelöléseket és a szórás(négyzet) és kovariancia jelentését nem kell megmagyarázni. A (b) részben a lehető legegyszerűbb helyes képletet adjuk meg.

*Megoldás.*

(a)

(3 pont) Jelölés:  $\text{corr}(X, Y)$

(7 pont)  $= \frac{\text{cov}(X, Y)}{\mathbb{D}(X)\mathbb{D}(Y)}.$

(b)

(10 pont)  $\mathbb{D}^2(X + Y) = \mathbb{D}^2(X) + \mathbb{D}^2(Y).$

*Pontozási instrukciók:*

(a)

Ha valahogy szövegesen megfogalmazza a képletet, az is jó, ilyen esetben nem kellene a cov és  $\mathbb{D}$  jelölések. De ha a szórás helyett a szórásnégyzet gyökét írja oda, akkor max. 5 pont, mert a kérdésben szórás volt. Ha nem az előadáson szereplő, de értelmes jelölést használ, például Var vagy  $\sigma$ , akkor max. 4 pont, attól függően, hogy szemantikailag és szintaktikailag mennyire értelmes, amit ír.

Ha nem jó a képlet, de a  $\text{cov}(X, Y)$  jelölése jó, arra jár max. 1 pont, és a  $\mathbb{D}(X)$  vagy  $\mathbb{D}(Y)$  jelölésre (elég az egyik) is jár max. 1 pont.

(b)

A megoldás szövegesen is elfogadható, pl.  $X$  és  $Y$  összegének szórásnégyzete  $X$  és  $Y$  szórásnégyzeteinek összege. Ha valahogy úgy fogalmazza meg, hogy  $X$  és  $Y$  közül nem szerepel benne mindkettő, akkor max. 9 pont.

Ha odaírja a  $+2\text{cov}(X, Y)$  tagot is, akkor max. 8 pont.

Kisebbszintaktikai hibák, nyilvánvalóan csak sajtóhibának számítót elírások esetén lehet pontot adni. Ha kihagyja a bal oldalt, akkor max. 6 pont. Ha kihagyja a jobb oldalt, akkor 0 pont (+ lásd a következő sort).

Ha nem jó a képlet, de a szórásnégyzet jelölése helyes (nem csak annyi, hogy  $\mathbb{D}^2$ , hanem  $\mathbb{D}^2(X), \mathbb{D}^2(Y), \mathbb{D}^2(X + Y)$  közül legalább az egyik ott van), arra jár max. 1 pont.

2. Pista a sarki péknél minden reggel egymástól függetlenül  $1/3$  valószínűséggel sós kiflit vásárol, amely 150 forintba kerül,  $1/3$  valószínűséggel pogácsát, amely 300 forintba, és  $1/3$  valószínűséggel kakaós csigát, amely 450 forintba kerül. Minden reggel pontosan egy pékárut vásárol. 40 egymást követő napon át megvizsgáljuk, hogy Pista mennyit költ az egyes reggeleken a péknél, jelölje  $X_i$  azt, hogy hány forintot költ Pista az  $i$ -edik reggelen, ahol  $i = 1, \dots, 40$ .

(a) Határozzuk meg  $X_1$  várható értékét és szórását.

(b) Jelölje  $X$  a Pista által a péknél a 40 nap alatt elköltött pénz összegét. Közelítsük alkalmas approximációval a  $\mathbb{P}(X \geq 12\,500)$  valószínűséget. (A valószínűség pontos értékét nem kell meghatározni.)

*Megoldás.*

(a) (8 pont)

(1+1 pont)  $\mathbb{E}(X_1) = \frac{1}{3}(150 + 300 + 450) = 300$  (1 pont a számolásra, ahol elég csak az általános vagy csak a konkrét képlet, de ha hiányzik az általános képlet és rossz a behelyettesítés, akkor ez az 1 pont nem jár; 1 pont a végeredményre).

(1+1 pont)  $\mathbb{D}(X_1) = \sqrt{\mathbb{D}^2(X_1)} = \sqrt{\mathbb{E}(X_1^2) - \mathbb{E}(X_1)^2}$  (a két lépés összevonható),

(1+1 pont) ahol  $\mathbb{E}(X_1^2) = \frac{1}{3}(150^2 + 300^2 + 450^2) = 105\,000$  (pontozás az első momentumhoz hasonlóan),

(1+1 pont) tehát  $\mathbb{D}(X) = \sqrt{105\,000 - 300^2} = \sqrt{15000} \approx 122,4745$  (itt már további számolás nem kell, ha az összes részeredmény jó, viszont ha hiányzik a numerikus végeredmény, akkor max. 1 pont. Minimális kerekítési hibáért itt ne vonjunk le pontot).

(b) (12 pont)

(4 pont) Sztenderdizálunk (ezt szöveggel nem kell odaírni):

$$\mathbb{P}(X \geq 12\,500) = \mathbb{P}\left(\frac{X - 40 \mathbb{E}(X_1)}{\sqrt{40} \mathbb{D}(X_1)} \geq \frac{12\,500 - 40 \mathbb{D}(X_1)}{\sqrt{40} \cdot \mathbb{D}(X_1)}\right) = \mathbb{P}\left(\frac{X - 40 \cdot 300}{\sqrt{40} \cdot 122,4745} \geq \frac{12\,500 - 40 \cdot 300}{\sqrt{40} \cdot 122,4745}\right).$$

Ennek a 4 pontnak a megoszlása: 2 pont, ha tudja, hogy hogyan kell sztenderdizálni a val. változót (általános  $\mathbb{E}(X_1)$ -re és  $\mathbb{D}^2(X_1)$ -re), 1 pont, ha jól helyettesíti be a képletbe az adatokat és 1 pont, ha ugyanezeket a transzformációkat a jobb oldalakra is megfelelően végrehajtja. 40 helyett az általános képlet felírásánál természetesen jó az  $n$  is, ha később kiderül, hogy  $n = 40$ .

A két lépés összevonható pontlevonás nélkül, ha azonosítható, hogy az előbbi képletet használja (ide értve azt is, hogy a  $\sqrt{40}$  nem csak numerikus értékkel szerepel!)

(1 pont)  $= \mathbb{P}\left(\frac{X - 12\,000}{774,5967} \geq 0,6455\right)$  (ez az előző lépéssel összevonható, de csak akkor, ha azt nem vonta össze az azelőttivel)

(1 pont) ami a centrális határeloszlás-tétel (elég: CHT; de Moivre–Laplace viszont nem jó!) miatt

(1 pont)  $\mathbb{P}(Y \geq 0,6455)$ -tel közelíthető, ahol  $Y \sim N(0; 1)$ .

(1 pont) arra, hogy valahol hivatkozik arra, hogy az  $X$  független, azonos eloszlású valószínűségi

változók összege (konkrétan az  $X_i$ -ké, de ezt nem muszáj explicite kimondani. Ha a függetlenség megvan, csak az azonos eloszlás hiányzik, és egyébként jól csinálja, akkor ne vonjuk le ezt a pontot). Emiatt helyes a sztenderdizálás (emiatt  $40 \mathbb{E}(X_1)$  az  $X$  val. változó várható értéke és  $\sqrt{40} \mathbb{D}(X_1)$  a szórása) és emiatt alkalmazható a CHT, de elég, ha egyszer hivatkozik rá.

(1 pont) Tehát a keresett valószínűség közelítőleg  $\mathbb{P}(Y \geq 0,6455) = 1 - \mathbb{P}(Y < 0,6455) = 1 - \Phi(0,6455)$  (a kettő közül elég az egyik)

(1 pont)  $\approx 1 - \Phi(0,65)$

(1 pont)  $\approx 1 - 0,7422$

(1 pont)  $= 0,2578$ .

3. Egy budapesti lakótelepi lakásban időnként előfordulnak áramszünetek. Tudjuk, hogy  $3/4$  annak a valószínűsége, hogy egy adott évben előfordul legalább egy áramszünet. Tegyük fel, hogy az egyes években egymástól függetlenül fordul elő vagy nem fordul elő áramszünet.

(a) Jelölje  $Y$  tíz egymást követő év közül azon évek számát, amelyekben előfordul legalább egy áramszünet. Mennyi  $Y$  várható értéke?

(b) Mi a valószínűsége, hogy 10 egymást követő évből legalább 8 évben lesz legalább egy áramszünet?

(c) Egy adott pillanatban elkezdjük mérni, hogy mennyi idő telik el a következő áramszünetig. Jelölje ezt a várakozási időt (évben mérve)  $X$ . Feltéhetjük, hogy  $X$  eloszlása folytonos és örökifjú, valamint a fentiek alapján  $\mathbb{P}(X \leq 1) = 3/4$ . Határozzuk meg  $X$  szórását.

(d) Feltéve, hogy 4 évnél többet kell várni az első áramszünetre, mi a valószínűsége, hogy 5 évnél is többet kell várni rá?

*Megoldás.*

(a) (3 pont)

(1 pont)  $Y$  binomiális eloszlású

(1 pont)  $n = 10$  és  $p = 3/4$  paraméterekkel (nem kell indokolni az eddigieket, csak kimondani).

(1 pont) Ezért  $\mathbb{E}(Y) = np = 10 \cdot \frac{3}{4} = 7,5$ . (Itt már elég a végeredmény, ha megvannak a paraméterek;  $np$  szerepel a táblázatban.)

(b) (5 pont)

(1 pont) A keresett valószínűség  $\mathbb{P}(Y \geq 8)$

(1 pont)  $= \mathbb{P}(Y = 8) + \mathbb{P}(Y = 9) + \mathbb{P}(Y = 10)$

(2 pont)  $= \binom{10}{8} \cdot (3/4)^8 \cdot (1/4)^2 + \binom{10}{9} \cdot (3/4)^9 \cdot (1/4) + \binom{10}{10} (3/4)^{10}$  ( $\binom{10}{k}$  helyett jó a  $\binom{10}{n-k}$  is, a  $\binom{10}{10}$  elhagyható és a  $\binom{10}{9}$  helyett elég az, hogy 10)

(1 pont)  $\approx 0,5256$ . (Nem kell részletezni, de az első tag numerikus értéke 0,2816, a másodiké 0,1877, a harmadiké pedig 0,0563.)

(c) (9 pont)

(1 pont)  $X$  exponenciális eloszlású,

(1 pont) mert folytonos és örökifjú eloszlás csak exponenciális lehet (elég annyi, hogy folytonos és örökifjú  $\Rightarrow$  exponenciális [eloszlású]).

(0 pont) A kérdések megválaszolásához az eloszlás  $\lambda$  paraméterét kell meghatározni.

(1+1 pont)  $\mathbb{P}(X \leq 1) = \mathbb{P}(X < 1) = 1 - e^{-\lambda}$ , mert  $X$  (vagy:  $X$  eloszlásfüggvénye) folytonos.

*Megjegyzés javítóknak:* Ha csak annyit ír, hogy  $\mathbb{P}(X \leq 1) = 1 - e^{-\lambda}$ , esetleg ugyanezt eloszlásfüggvénnyel kifejezve, akkor a 2 pontból 1 jár. A második pont akkor jár, ha hivatkozik a folytonosságra, viszont ebben az esetben elég  $\mathbb{P}(X \leq 1) = 1 - e^{-\lambda}$  is,  $\mathbb{P}(X \leq 1) = \mathbb{P}(X < 1)$ -et nem muszáj kiírni.

(1 pont) Tehát  $1 - e^{-\lambda} = 3/4$ , vagyis  $e^{-\lambda} = 1/4$  (elég az utóbbi),

(1+1 pont) vagyis  $-\lambda = \ln(1/4)$ , azaz  $\lambda = \ln 4 \approx 1,3863$ . (1 pont a végeredményre, 1 a számolásra. Itt még elég az  $\ln 4$  vagy  $-\ln(1/4)$ , a numerikus végeredmény csak a későbbiekben kell, viszont ha van helyes végeredmény, akkor az elég  $\ln 4$  nélkül is.)

(1 pont)  $\mathbb{D}(X) = \sqrt{\mathbb{D}^2(X)}$

(1 pont)  $= \sqrt{\frac{1}{\lambda^2}} \approx 0,7213$  (az  $1/\lambda$  vagy  $1/\ln 4$  elég, ha impliciten derül ki, viszont kell a numerikus végeredmény).

(d) (3 pont)

(1 pont) A keresett (feltételes) valószínűség  $\mathbb{P}(X > 5 | X > 4)$

(1 pont)  $= \mathbb{P}(X > 1)$  az örökifjúság miatt (ez a pont csak akkor jár, ha hivatkozik az örökifjúságra, de ha az eredmény jó, akkor a továbbiakra még lehet pontot kapni).

(1 pont)  $= 1 - \mathbb{P}(X \leq 1) = \frac{1}{4}$ .

(Ugye  $\mathbb{P}(X \leq 1)$  értékét tudjuk a (c) rész szövegéből, ezért indoklasként elég a komplementer valószínűségre hivatkozni, a végeredmény és az indoklás együtt 1 pont. Ha már a (c)-re adott megoldásában is megjelent valahol, hogy  $\mathbb{P}(X > 1) = 1/4$ , akkor OK a végeredmény újabb indoklás nélkül is.

Természetesen azt is lehet mondani, hogy  $\mathbb{P}(X \leq 1) = 1 - e^{-\lambda}$ . Itt már ne vonjunk le pontot, ha nem hivatkozik a folytonosságra, sőt akkor se, ha csak azt írja, hogy  $\mathbb{P}(X > 1) = 1 - \mathbb{P}(X < 1)$  vagy  $\mathbb{P}(X > 1) = 1 - F_X(1)$  vagy  $\mathbb{P}(X > 1) = e^{-\lambda}$ .)

4. Az  $(X, Y)$  folytonos valószínűségi vektorváltozó együttes sűrűségfüggvénye

$$f_{X,Y}(x, y) = \begin{cases} c(x+y), & \text{ha } 1 < x < 2 \text{ és } 1 < y < 2, \\ 0, & \text{egyébként,} \end{cases}$$

ahol  $c \in \mathbb{R}$  egy alkalmasan választott valós szám.

(a) Határozzuk meg  $c$  értékét.

(b) Határozzuk meg  $X$  sűrűségfüggvényét.

(c) Határozzuk meg  $Y$  várható értékét.

*Megoldás.*

(a) (7 pont)

(1 pont)  $f_{X,Y}$  akkor lesz sűrűségfüggvény, ha nemnegatív (az is jó, ha azt írja, hogy  $c$ -nek kell nemnegatívnak lennie, és nem baj, ha nem ellenőrzi a végén még egyszer, hogy ez teljesül-e)

(1 pont) és  $1 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx dy$

(1 pont)  $= \int_1^2 \int_1^2 c(x+y) dx dy$  (az előző lépéssel összevonható, ha kiderül, hogy 1-nek kell lennie az integrálnak)

(1 pont)  $= c \int_1^2 [x^2/2 + xy]_{x=1}^2 dy = c \int_1^2 \frac{3}{2} + y dy$

(1 pont)  $= c [\frac{3}{2}y + \frac{y^2}{2}]_{y=1}^2$

(1 pont)  $= c(\frac{3}{2} + \frac{3}{2}) = 3c$ ,

(1 pont) tehát  $c = \frac{1}{3}$ .

(b) (6 pont)

(1 pont)  $x \in \mathbb{R}$  esetén  $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy$ ,

(1 pont)  $= \int_1^2 \frac{1}{3}(x+y) dy$

(1 pont)  $= \frac{1}{3} [xy + y^2/2]_{y=1}^2$

(1 pont)  $= \frac{1}{3}(x + \frac{3}{2}) = \frac{1}{3}x + \frac{1}{2}$ ,

(1 pont) ha  $x \in (1, 2)$  (ha kihagyja, hogy a fenti képletek csak  $x \in (1, 2)$  esetén érvényesek, de egyébként jól számol, akkor ezt az 1 pontot vonjuk le),

(1 pont) és  $f_X(x) = 0$  egyébként.

*Megjegyzés csak javítóknak a pontozáshoz:* Ha a határokat máshogy sorolja be a két szakaszhoz, azért ne vonjunk le pontot, de ha valamelyik határponto(ka)t nem sorolja be sehova, akkor vonjunk le egyet. Bármilyen nemnegatív, minden pontban definiált függvényt elfogadunk, ami a fenti  $f_X$ -szel majdnem mindenütt megegyezik.

(c) (7 pont)

## 1. megoldás.

(1 pont)  $Y$ -nak ugyanaz a sűrűségfüggvénye, mint  $X$ -nek

(ezért az 1 pontért elég, ha ez impliciten derül ki, viszont ha  $Y$  helyett  $X$  várható értékét kezdi el kiszámolni és  $Y$ -ra nincs kellő utalás, akkor a (c)-re max. 3 pont),

(1 pont) mert  $f_{X,Y}$  szimmetrikus ( $x$ -ben és  $y$ -ban).

Természetesen az is jó, ha kiszámolja  $x$  szerinti integrálással és úgy hozza ki ugyanezt az eredményt. Akkor is max. 2 pont, de ha kihagyja, hogy  $(1,2)$ -n kívül 0 a sűrűségfüggvény, azért ne vonjunk le még egy pontot abban az esetben, ha  $(1,2)$ -n jó képletet ad meg.

(1+1 pont)  $\mathbb{E}(Y) = \int_{-\infty}^{\infty} y f_Y(y) dy = \int_1^2 y \left(\frac{1}{3}y + \frac{1}{2}\right) dy$  (a két lépés összevonható)

(0 pont)  $= \int_1^2 \frac{1}{3}y^2 + \frac{1}{2}y dy$

(1 pont)  $= \left[ \frac{y^3}{9} + \frac{y^2}{4} \right]_{y=1}^2$

(1 pont)  $= \frac{8}{9} + \frac{4}{4} - \frac{1}{9} - \frac{1}{4}$

(1 pont)  $= \frac{55}{36} (\approx 1,5278)$ .

## 2. megoldás.

(1 pont)  $\mathbb{E}(Y) = \int_{-\infty}^{\infty} y f_{X,Y}(x, y) dx dy$

(1 pont)  $= \int_1^2 \int_1^2 \frac{1}{3}xy + \frac{1}{3}y^2 dx dy$

(1 pont)  $= \int_1^2 \left[ \frac{1}{6}x^2y + \frac{1}{3}y^2x \right]_{x=1}^2 dy$

(1 pont)  $= \int_1^2 \frac{4}{6}y + \frac{2}{3}y^2 - \frac{1}{6}y - \frac{1}{3}y^2 dy = \int_1^2 \frac{1}{2}y + \frac{1}{3}y^2 dy$

(3 pont) ugyanúgy, mint az 1. megoldás utolsó 3 pontjánál. (Természetesen az is OK, ha először  $y$  és aztán  $x$  szerint integrál, pontozás analóg módon.)

5. Egy rövid menetidejű repülőjáraton a menetidő jó közelítéssel normális eloszlású  $\sqrt{2}/100$  óra szórással. Hat alkalommal megmértük ezen a járaton a menetidőt és (órában mérve, tizedestört alakban) az alábbi eredményeket kaptuk:

$$x_1 = 0,97 \quad x_2 = 0,98 \quad x_3 = 0,99 \quad x_4 = 1,01 \quad x_5 = 1,00 \quad x_6 = 0,99.$$

Feltehetjük, hogy az egyes alkalmakkor mért menetidők egymástól függetlenek.

(A szórás az eredeti vizsgafeladatban ismeretlen volt, ezt módosítottam.)

- Adjunk a minta alapján 0,95 szintű konfidenciaintervallumot a mintaelemek várható értékére. Adjuk meg a minta korrigált empirikus szórását is.
- Döntsük el  $\varepsilon = 0,02$  terjedelmű próbával, hogy a mintaelemek várható értéke megegyezik-e 1 órával. Adjuk meg  $H_0$ -t,  $H_1$ -et, a próbastatisztika értékét, azt, hogy mi alapján döntünk, és magát a döntést is.
- Adjuk meg (indoklással) a minta empirikus eloszlásfüggvényének értékét az  $x = 0,995$  helyen.
- Adjuk meg a minta korrigált empirikus szórását.

## Megoldás.

(a) (7 pont)

(1 pont) a konfidenciaintervallum (ismert szórás esetére vonatkozó) helyes általános képletére; ha jó konkrét paraméterekkel írja fel, az is jó.

(1 pont)  $n = 6$  (ez a mintaelemszám),

(1 pont)  $\bar{x}_6 = \frac{1}{6} \sum_{i=1}^6 x_i = 0,99$  (csak 1 pont jár rá, ezért OK indoklás nélkül is,  $\bar{x}_n$  jelöléssel is jó),

(1 pont) a szükséges kvantilis pedig  $u_{\varepsilon/2} = u_{0,025}$

(1 pont)  $= 1,96$ .

(1+1 pont) A képletbe behelyettesítve a konfidenciaintervallum:

$$\left[ \bar{x}_n - \frac{\sigma t_{n-1, \varepsilon/2}}{\sqrt{n}}, \bar{x}_n + \frac{\sigma t_{n-1, \varepsilon/2}}{\sqrt{n}} \right] = \left[ 0,99 - \frac{1,96 \cdot 0,0141}{\sqrt{6}}, 0,99 + \frac{1,96 \cdot 0,0141}{\sqrt{6}} \right] = [0,9786, 1,0014].$$

Az általános képletre itt nem jár pont. 1 pont arra, hogy behelyettesíti a képletbe a kapott értékeket, és 1 pont a numerikus végeredményre. Ha már mindent kiszámolt és utána csak a numerikus végeredményt adja meg, akkor ne vonjunk le pontot.

Megjegyzés: ha a  $\frac{\sqrt{2}}{100}$  pontos értékkel számolunk, akkor is ez az eredmény jön ki 4 tizedesjegyre kerekítve.

(b) (7 pont)

*Először a jó megoldást közlöm, a hibás alternatíva vagy próbatípus esetén járó részpontokat lásd lentebb.*

(1 pont)  $H_0: \mu = 1$  vs.  $H_1: \mu \neq 1$ ,

(1 pont) tehát mivel a szórás ismeretlen, a kétoldali, egymintás  $t$ -próba képletét kell használni.

(Ha ezt nem írja oda, de jól alkalmazza ezt a próbát, akkor ezt az 1 pontot ne vonjuk le.)

(1+1 pont) A próbastatisztika értéke  $u(=u(x_1, \dots, x_6)) = \frac{\bar{x}_n - \mu_0}{\sigma} \sqrt{n} = \frac{0,99 - 1}{0,0141} \sqrt{6} \approx -1,7372$ . (1 pont arra, hogy jól behelyettesíti a táblázatban megadott képletbe, ideértve azt is, hogy  $\mu_0 = 1$ . Természetesen jó  $n$  helyett 6-ot írni mindenhova. 1 pont pedig arra, hogy kiszámolja a helyes numerikus eredményt. Ha az utóbbi megvan és a szükséges mennyiségek értékei  $\mu_0$ -t kivéve már az (a)-ban kiderültek, akkor adjuk meg az első pontot is, akkor is, ha nem írja oda, hogy  $\mu_0 = 1$ ).

(1 pont)  $H_0$ -t akkor fogadjuk el/nem utasítjuk el, ha  $|u| \leq u_{\varepsilon/2} = u_{0,01}$  (bármilyen, ezzel ekvivalens alak jó, az is, ha azt írja, hogy mikor utasítjuk el. Ne vonjunk le pontot, ha rossz helyre teszi az egyenlőséget.)

(1 pont)  $= 2,33$ .

(1 pont) Tehát  $H_0$ -t elfogadjuk (azaz azt mondjuk, hogy a mintaelemek várható értéke meg egyezik 1 órával, illetve nincs elég bizonyítékunk ahhoz, hogy azt mondjuk hogy ez nem igaz).

*Pontozás a (b)-ben hibás próbatípus, illetve alternatíva esetén:*

Ha egyoldali  $u$ -próbát hajt végre  $H_0: \mu \geq 1$  vs.  $H_1: \mu < 1$  alternatívával, akkor max. 6 pont.

Ha egyoldali  $u$ -próbát hajt végre  $H_0: \mu \leq 1$  vs.  $H_1: \mu > 1$  alternatívával, az még rosszabb, mert az alacsony átlag alapján nem gyanakodhatunk arra, hogy a várható érték 1-nél nagyobb, így ekkor max. 5 pont.

(c) (3 pont)

(2 pont) Megadja az empirikus eloszlásfüggvény adott pontbeli értékének definícióját:  $F_n^*(x)$  az  $x$ -nél kisebb mintaelemek száma, osztva  $n$ -nel. Jó szövegesen is, a  $\frac{1}{n} |\{i \in \{1, \dots, n\} : x_i < x\}|$  definícióval és a rendezett mintaelemekkel kifejezett ( $F_n^*(x) = \frac{i}{n}$ , ha  $x_i^* < x \leq x_{i+1}^*$ ) definícióval is. Utóbbi esetben ne vonjunk le pontot, ha a határpontokban nem jó a definíció (hanem pl. jobbról folytonos) vagy ha hiányzik az  $x < x_1^*$  és az  $x \geq x_n^*$  eset, mert ezek itt nem relevánsak. Kisebb szintaktikai hibák esetén lehet 1 pontot adni a 2-ből.

Az is jó, ha eleve a konkrét  $n$ -re és  $x$ -re mondja el ugyanezt, de ha nincs általános képlet és a konkrét esetben rosszul írja fel, akkor 0 pont.

(1 pont) Esetünkben ez  $F_6^*(0,995) = \frac{1}{6} \cdot 4 = \frac{2}{3}$ . (Ezért az egy pontért elég a végeredmény is, ha van előtte valami indoklás vagy legalább minimális számolás. Tizedestört alakban és nem egyszerűsített  $\frac{4}{6}$  alakban is jó.)

d) (3 pont)

(0 pont)  $s_n^* = \sqrt{(s_n^*)^2}$

(1 pont)  $= \sqrt{\frac{1}{5} \sum_{i=1}^6 (x_i - 0,99)^2}$  (ez akkor is OK, ha esetleg csak  $(s_n^*)^2$ -et számolja ki és  $s_n^*$ -ot nem, viszont az általános képlet pusztán lemásolására nem jár pont, csak ha  $n - 1$ ,  $n$  és  $\bar{x}_n$  képletét is behelyettesíti)

(2 pont)  $= \sqrt{\frac{1}{5} (0,02^2 + 0,01^2 + 0 + 0,02^2 + 0,01^2 + 0)} = \sqrt{\frac{1}{5} \cdot \frac{4+1+4+1}{10\,000}} = \sqrt{1/5000} = \frac{\sqrt{2}}{100} = 0,0141$  (ebből 1 pont arra, hogy van valami részletszámolás, ahol a különbségek négyzetei valamilyen értelmes formában megjelennek, nem kell ennyire részletesen és a 0-kat el lehet hagyni,

1 pont pedig a numerikus végeredményre jár),

6. \* (A többdimenziós normális eloszlás fogalmának hiányában átdolgozott feladat.)

Legyenek  $Z$  és  $W$  független, standard normális eloszlású változók és legyen  $X := \sqrt{5}Z + 2W - 1$ ,  $Y := \sqrt{5}Z - 2W + 2$ .

- Határozzuk meg az  $(X, Y)^T$  valószínűségi vektorváltozó várható érték vektorát és kovarianciamátrixát. *Ez az eredeti vizsgafeladatnak nem volt része, de az előadás jelenlegi anyaga alapján is megoldható és nem 6-os nehézségű, valamint kell a (b) és (c) kitűzéséhez.*
- Legyen  $U = X + Y$  és  $V = X - Y$ . Határozzuk meg az  $\begin{pmatrix} U \\ V \end{pmatrix}$  valószínűségi vektorváltozó várható érték vektorát és kovarianciamátrixát.
- Ez alapján adjuk meg az  $\mathbb{E}(V|U)$  regressziót.

*Megoldás.*

(a) (pontozás nélkül)

$\mathbb{E}(X) = \mathbb{E}(\sqrt{5}Z + 2W - 1) = \sqrt{5}\mathbb{E}(Z) + 2\mathbb{E}(W) - 1 = -1$  a várható érték linearitása miatt, hasonlóan  $\mathbb{E}(Y) = \mathbb{E}(\sqrt{5}Z - 2W + 2) = 2$ , tehát a várható érték vektor  $(-1, 2)^T$ .

$\mathbb{D}^2(X) = \mathbb{D}^2(\sqrt{5}Z + 2W - 1) = \mathbb{D}^2(\sqrt{5}Z + 2W) = \mathbb{D}^2(\sqrt{5}Z) + \mathbb{D}^2(2W) = 5\mathbb{D}^2(Z) + 4\mathbb{D}^2(W) = 9$ , ahol az összegre bontásnál  $W$  és  $Z$  függetlenségét használtuk.

Hasonló indoklással:  $\mathbb{D}^2(Y) = \mathbb{D}^2(\sqrt{5}Z - 2W + 2) = \mathbb{D}^2(\sqrt{5}Z - 2W) = \mathbb{D}^2(\sqrt{5}Z) + \mathbb{D}^2(-2W) = 5\mathbb{D}^2(Z) + 4\mathbb{D}^2(W) = 9$ .

Továbbá  $\text{cov}(X, Y) = \text{cov}(\sqrt{5}Z + 2W - 1, \sqrt{5}Z - 2W + 2) = \text{cov}(\sqrt{5}Z + 2W, \sqrt{5}Z - 2W) = \sqrt{5} \cdot \sqrt{5} \text{cov}(Z, Z) + \sqrt{5} \cdot (-2) \cdot \text{cov}(Z, W) + 2 \cdot \sqrt{5} \cdot \text{cov}(W, Z) + 2 \cdot (-2) \cdot \text{cov}(W, W) = 5 + 0 + 0 - 4 = 1$ , ahol a kovariancia eltolás-invarianciáját („additív konstans nem számít”), bilinearitását és az utolsó lépésben  $Z$  és  $W$  függetlenségét használtuk.

Így a kovarianciamátrix  $\begin{pmatrix} 9 & 1 \\ 1 & 9 \end{pmatrix}$ .

(b) (12 pont)

(1 pont)  $\mathbb{E}(U) = \mathbb{E}(X) + \mathbb{E}(Y)$  (ha kimondja, hogy a várható érték linearitását használja, akkor egyből mehet a  $-1 + 2$  anélkül, hogy odaírná, hogy ez  $\mathbb{E}(X) + \mathbb{E}(Y)$ -nal egyenlő)

(1 pont)  $= -1 + 2 = 1$ .

(1 pont)  $\mathbb{E}(V) = \mathbb{E}(X) - \mathbb{E}(Y) = -3$  (itt már nem jár még egyszer pont az indoklásra, elég a végeredmény).

(1 pont)  $\mathbb{D}^2(U) = \mathbb{D}^2(X) + \mathbb{D}^2(Y) + 2\text{cov}(X, Y)$

(1 pont)  $= 9 + 9 + 2 = 20$ .

(2 pont)  $\mathbb{D}^2(V) = \mathbb{D}^2(X) + \mathbb{D}^2(Y) - 2\text{cov}(X, Y)$  (ha ez a képlet jól szerepel, akkor nem kell jobban részletezni. Ha  $U$  szórásnégyzetének kiszámolásánál már szerepelt az összeg szórásnégyzetének képlete és most jól csinálja, akkor elég konkrét számokkal, de ha előjelhiba van és hiányzik az általános képlet, akkor ez a 2 pont sem jár. Ha azt írja, hogy  $\mathbb{D}^2(V) = \mathbb{D}^2(X) + \mathbb{D}^2(-Y) + 2\text{cov}(X, -Y)$  és később rontja el, akkor itt 1 pont jár, a másik nem, és később lehet részpontokat adni).

(1 pont)  $= 9 + 9 - 2 = 16$ .

(2 pont)  $\text{cov}(U, V) = \text{cov}(X + Y, X - Y) = \mathbb{D}^2(X) + \text{cov}(X, Y) - \text{cov}(X, Y) - \mathbb{D}^2(X)$  (itt 1 pont jár arra, hogy a kovariancia bilinearitását használja – nem kell megnevezni – és 1 összesen arra, hogy valaminek az önmagával vett kovarianciája a szórásnégyzet és a kovariancia szimmetrikus. A két lépés összevonható, itt is összevontuk)

(0 pont)  $= 0$ .

*Megjegyzés:* A 8. feladatsoron szerepelt, hogy ha  $X$  és  $Y$  azonos szórásuak, akkor  $X + Y$  és  $X - Y$  kovarianciája 0. Ez itt alkalmazható, hiszen  $\mathbb{D}(X) = \mathbb{D}(Y) = 3$ , de mivel ez az állítás csak gyakorlaton szerepelt (előadáson nem és Mészáros Szabolcs jegyzetében sincs benne), ezért

ha valaki erre hivatkozik, akkor a fenti 2 pontból csak 1 jár.

(1+1 pont) Tehát a várható érték vektor  $\begin{pmatrix} 1 \\ -3 \end{pmatrix}$  és a kovarianciamátrix  $\begin{pmatrix} 20 & 0 \\ 0 & 16 \end{pmatrix}$ .

Itt nem kell az általános képlet, ha jól helyettesít be a kapott eredmények alapján. Ha nem tudja kiszámolni az értékeket, de felírja, hogy milyen mennyiségeknek kellene szerepelnie a várható érték vektor és a kovarianciamátrix adott helyein a *konkrét*  $U$  és  $V$  esetén, akkor ezt az 1–1 pontot meg lehet adni.

(c) (8 pont)

(2 pont)  $U$  és  $V$  is független normális eloszlású val. változók lineáris kombinációjaként áll elő (nem kell tovább indokolni; ha azt írja, hogy független normálisok összegeként, azt is elfogadjuk). *Innen 1. megoldás:*

(2 pont) Adott független normális eloszlású val. változók lineáris kombinációi esetén [vagy: valahogy ugyanerre képlettel helyesen hivatkozik] a korrelálatlanságból következik a függetlenség. (Azt nem kell még jobban a javító szájába rágni, hogy  $U$  és  $V$  korrelálatlanok, ha ez már kijött az (a)-ban. A korrelálatlanság helyett természetesen arra is elég hivatkozni, hogy a kovariancia 0.)

Ha csak annyit ír, hogy a korrelálatlanság miatt függetlenek, de már korábban megjelent valahogy, hogy  $U$  és  $V$  is független normálisok lineáris kombinációja, azt fogadjuk el. Ha ez nem szerepelt, akkor max. 1 pont a kettőből.

(2 pont)  $\mathbb{E}(V|U) = \mathbb{E}(V)$

(1 pont)  $= -3$ ,

(1 pont) mert függetlenek. (Ha már írta, hogy függetlenek, csak most nem indokolja az előző lépést, akkor ezt az 1 pontot ne vonjuk le. De ha rosszul indokolja, akkor vonjuk le).

*2. megoldás:*

(1 pont) Adott független normálisok különböző lineáris kombinációi esetén [vagy: valahogy ugyanerre képlettel helyesen hivatkozik] (ezért az 1 pontért ennyi elég, de természetesen az is jó, ha konkrétan azt írja, hogy  $\mathbb{E}(V|U)$  megegyezik  $V$ -nek  $U$ -ra vett lineáris regressziójával, vagy ha nem szövegesen írja ezt, hanem egyértelműen a lineáris regressziónak megfelelő, helyes képletet akar használni).

Ha csak annyit ír, hogy a regresszió lineáris (vagy ennek megfelelőjét a konkrét esetre nézve), de már korábban megjelent valahogy, hogy  $U$  és  $V$  is független normálisok lineáris kombinációja, azt fogadjuk el. Ha ez nem szerepelt, akkor ez a pont nem jár.

(1 pont) Tehát  $\mathbb{E}(V|U) = \beta U + \alpha$ , (ha kiszámolja  $\beta$ -t és  $\alpha$ -t helyesen, de nem használja ezt a jelölést, az is OK)

(1+1 pont) ahol  $\beta = \frac{\text{cov}(U,V)}{\mathbb{D}^2(U)} = 0$  (ha csak  $\beta = 0$  szerepel, akkor max. 1 pont)

(1+1 pont) és  $\alpha = \mathbb{E}(V) - \beta \mathbb{E}(U) = -3$  (itt az első pont a képletre jár, ami jó konkrét számokkal is, de csak ha azonosítható, hogy ezt a képletet használja, a második pont pedig  $\mathbb{E}(V)$  azonosítására és a végeredményre,  $\mathbb{E}(U)$  értéke természetesen nem kell).

(0 pont) Tehát  $\mathbb{E}(V|U) = -3$ . (Ha  $\mathbb{E}(V|U) = \beta U + \alpha$  már szerepelt és csak a végeredmény hiányzik, akkor ne vonjunk le pontot.)

Ha regressziós egyenest ad meg, nem pedig magát a lineáris regressziót, akkor 1 pont levonás. De ha megadja a regressziót is, nem baj, ha odaírja az egyenest is.