Vizsgadolgozat Valószínűségszámítás és statisztikából (ÚJ Valszám), 2024.01.18.

1. (a) Legyen $(\Omega, \mathcal{F}, \mathbb{P})$ valószínűségi mező.
   i. Legyenek $A, B \in \mathcal{F}$ események. Milyen feltétel teljesülése esetén mondjuk azt definíció szerint, hogy A és B függetlenek?
   ii. Legyenek X és Y valószínűségi változók ezen a valószínűségi mezőn. Milyen feltétel teljesülése esetén mondjuk azt definíció szerint, hogy X és Y függetlenek? (Ezen kérdés megválaszolásához az i. részben már definiált fogalma(ka)t nem kell újra definiálni.)

   (b) Legyen $X_1, X_2, \ldots, X_n, \ldots$ valószínűségi változók egy sorozata és Z egy valószínűségi változó. Jelöljük X_i eloszlásfüggvényét F_{X_i}-vel ($i = 1, 2, \ldots$ esetén) és Z eloszlásfüggvényét F_Z-vel. Milyen feltétel teljesülése esetén mondjuk azt definíció szerint, hogy X_n eloszlásban konvergál Z-hez (amint $n \to \infty$)?

Megoldás.
(a) (11 pont)
i.
(5 pont) A és B (pontosan) akkor függetlenek, ha $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$.
Szorzatjel nem kötelező. Ha $\mathbb{P}(A|B) = \mathbb{P}(A)$-t vagy $\mathbb{P}(B|A) = \mathbb{P}(B)$-t ír, akkor 2 pont, ha hozzáírja, hogy a feltétel valószínűsége pozitív kell, hogy legyen, akkor még 1 pont.
(jegyzet 2.1.1. Definíció)
ii.
(4 pont) X és Y (pontosan) akkor függetlenek, ha az $\{X < x\}$ és az $\{Y < y\}$ események függetlenek
(az is jó, ha azt írja, hogy $\mathbb{P}(X < x, Y < y) = \mathbb{P}(X < x)\mathbb{P}(Y < y)$ vagy $\mathbb{P}(\{X < x\} \cap \{Y < y\}) = \mathbb{P}(X < x)\mathbb{P}(Y < y)$, illetve ha pl. $\{X < x\}$ helyett $\{\omega \in \Omega | X(\omega) < x\}$-et, az sem, de kisebb szintaktikai hibákért vonjunk le 1 pontot, pl. ha a metszetet halmazjelölések nélkül vagy hiányos halmazjelölésekkel írja. Ha pl. $\mathbb{P}(\{X < x\})$-et ír, az nem baj)
(2 pont) minden $x, y \in \mathbb{R}$ esetén.
(jegyzet 6.1.1. Definíció)
(b) (9 pont)
(5 pont) X_n akkor konvergál eloszlásban Z-hez (amint $n \to \infty$), ha $\lim_{n\to\infty} F_{X_n}(x) = F_Z(x)$ teljesül (vagy $F_{X_n}(z) \underset{n\to\infty}{\longrightarrow} F_Z(x)$, vagy $F_{X_n}(x) \to F_Z(x)$, amint $n \to \infty$; ha csak konvergenciajel szerepel, de nem derül ki, hogy $n \to \infty$ esetén értendő a határérték, akkor max. 3 pont)
(4 pont) minden olyan $x \in \mathbb{R}$ esetén, ahol F_Z folytonos. (Kisebb szintaktikai hibákért 1–1 pont levonás. Ha azt írja, hogy minden $x \in \mathbb{R}$-re, azért 1 pont jár a 4-ből.)
(jegyzet 9.3.1. Definíció)

2. Az X és Y valószínűségi változók együttes eloszlását az alábbi táblázat adja meg.

$Y \backslash X$	1	2
0	$\frac{1}{3}$	$\frac{1}{6}$
1	$\frac{1}{12}$	$\frac{1}{4}$
2	$\frac{1}{12}$	$\frac{1}{12}$

Határozzuk meg X és Y kovarianciáját és korrelációját, valamint Y-nak X-re vett regressziós egyenesét.
Megoldás.
(1 pont) $\mathrm{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$,
(1 pont) ahol $\mathbb{E}(XY) = \sum_{k \in \mathrm{Ran}(XY)} k\mathbb{P}(XY = k)$ (ha a következő lépés jó, akkor ez a pont jár az általános képlet feltüntetése nélkül is)
(1 pont) $= \frac{1}{12} \cdot 1 + (\frac{1}{4} + \frac{1}{12}) \cdot 2 + \frac{1}{12} \cdot 4 = \frac{13}{12}$,
(1 pont) valamint mivel $\mathbb{P}(X = 1) = \mathbb{P}(X = 2) = 1/2$ (ezt persze táblázatosan is fel lehet

tüntetni),
(1 pont) $\mathbb{E}(X) = \frac{1}{2} \cdot (1+2) = \frac{3}{2}$,
(1 pont) és mivel $\mathbb{P}(Y=1) = 1/3$ és $\mathbb{P}(Y=2) = 1/6$ (és így $\mathbb{P}(Y=0) = 1/2$, de ezt nem muszáj odaírni, mert a várható érték és szórás kiszámításához úgysem kell),
(1 pont) $\mathbb{E}(Y) = \frac{1}{3} \cdot 1 + \frac{1}{6} \cdot 2 = \frac{2}{3}$,
(1 pont) tehát $\mathrm{cov}(X,Y) = \frac{1}{12}$.
(1 pont) $\mathrm{corr}(X,Y) = \frac{\mathrm{cov}(X,Y)}{\mathbb{D}(X)\mathbb{D}(Y)}$,
(1 pont) ahol $\mathbb{D}(X) = \sqrt{\mathbb{E}(X^2) - \mathbb{E}(X)^2}$ és Y esetén hasonlóan (de ha csak az egyiket írja oda, azért ne vonjunk le pontot),
(1 pont) és $\mathbb{E}(X^2) = \frac{1}{2}(1^2 + 2^2) = \frac{5}{2}$,
(1 pont) tehát $\mathbb{D}(X) = \sqrt{5/2 - (3/2)^2} = \sqrt{1/4} = 1/2$,
(1 pont) valamint $\mathbb{E}(Y^2) = \frac{1}{3} \cdot 1^2 + \frac{1}{6} \cdot 2^2 = 1$,
(1 pont) így $\mathbb{D}(Y) = \sqrt{1 - (2/3)^2} = \frac{\sqrt{5}}{3}$,
(1 pont) emiatt $\mathrm{corr}(X,Y) = \frac{\frac{1}{12}}{\frac{1}{2} \cdot \frac{\sqrt{5}}{3}} = \frac{1}{2\sqrt{5}} \approx 0{,}2236$ (a tört alakú és a numerikus végeredmény közül elég, ha az egyiket írja ki).
(1 pont) A regressziós egyenes képlete $\{(x,y) \in \mathbb{R}^2 | y = \beta x + \alpha\}$,
(1+1 pont) ahol $\beta = \frac{\mathrm{cov}(X,Y)}{\mathbb{D}^2(X)} = \frac{\frac{1}{12}}{1/4} = \frac{1}{3}$
(1+1 pont) és $\alpha = \mathbb{E}(Y) - \beta\mathbb{E}(X) = \frac{2}{3} - \frac{1}{3} \cdot \frac{3}{2} = \frac{1}{6}$.

3. Egy X folytonos valószínűségi változó eloszlásfüggvénye

$$F_X \colon \mathbb{R} \to \mathbb{R}, \quad F_X(x) = \begin{cases} 0, & \text{ha } x \leq 1, \\ \ln x, & \text{ha } 1 < x < \mathrm{e}, \\ 1, & \text{ha } x \geq \mathrm{e}, \end{cases}$$

ahol $\mathrm{e} \approx 2{,}71\ldots$.

(a) Határozzuk meg X várható értékét és szórásnégyzetét.

(b) Számítógéppel szimuláltunk 100 darab független, azonos eloszlású valószínűségi változót az (a) feladatban szereplő F_X eloszlásfüggvénnyel. Becsüljük meg alkalmas approximációval annak a valószínűségét, hogy ezek összege legfeljebb 180 (a valószínűség pontos értékének kiszámításáért itt nem jár pont).

Tanács: A (b) részfeladatban érdemes az (a) részfeladat megoldásának megfelelően kerekített numerikus eredményeivel dolgozni szimbolikus értékek helyett.

Megoldás.
(a) (8 pont)
(1 pont) $f_X(x) = \frac{1}{x}$, ha $1 < x < \mathrm{e}$
(1 pont) és $f_X(x) = 0$ különben,
(1 pont) tehát $\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x) \mathrm{d}x = \int_1^{\mathrm{e}} 1 \mathrm{d}x$ (elég, ha a második képlet szerepel)
(1 pont) $= [x]_1^{\mathrm{e}} = \mathrm{e} - 1 (\approx 1{,}7183)$.
(1 pont) $\mathbb{D}^2(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$,
(1 pont) ahol $\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x) \mathrm{d}x = \int_1^{\mathrm{e}} x \mathrm{d}x$
(1 pont) $= [x^2/2]_1^{\mathrm{e}} = \frac{\mathrm{e}^2 - 1}{2} (\approx 3{,}1945)$.
(1 pont) Tehát $\mathbb{D}^2(X) = \frac{\mathrm{e}^2-1}{2} - (\mathrm{e} - 1)^2 \approx 0{,}2420$. (Nem baj, ha itt nem szerepel numerikus eredmény, de előbb-utóbb hiányozni fog.)
(b) (12 pont)
(1 pont) arra az ötletre, hogy a CHT-t kellene használni. Ez jár akkor is, ha ez a későbbiekben

nem sikerül, feltéve, hogy kiderül, hogy az általános CHT-ról és nem a de Moivre–Laplace-tételről van szó.
(1 pont) Jelölje $X_1, \dots, X_{100}$ a szimulált valószínűségi változókat, ekkor a keresett valószínűség $\mathbb{P}(\sum_{i=1}^{100} X_i \le 180)$.
(5 pont) Sztenderdizálunk:

$$\mathbb{P}\Big(\sum_{i=1}^{100} X_i \le 180\Big) = \mathbb{P}\Big(\frac{\sum_{i=1}^{100} X_i - 100\mathbb{E}(X_1)}{\sqrt{100}\mathbb{D}(X_1)} \le \frac{180 - 100\mathbb{E}(X_1)}{\sqrt{100}\mathbb{D}(X_1)}\Big) \approx \mathbb{P}\Big(\frac{\sum_{i=1}^{100} X_i - 171{,}8282}{4{,}9197} \le 1{,}6610\Big)$$

(Ebből 3 pont a középső lépés, 2 pont az utolsó, amiből 1 az, hogy a szórásnégyzetből gyököt vonva kiszámolja a szórást.) (Kisebb kerekítési hibákért, ha pl. már az (a) részben kiszámolta a várható értéket és a szórásnégyzetet és most a négy tizedesjegyre kerekített értékekkel számol, ne vonjunk le pontot. Ha a középső kifejezés nincs ott szimbolikusan, de ott van $\mathbb{E}(X_1)$ és $\mathbb{D}(X_1)$ már ismert értékeivel, az is jó, és a két lépés össze is vonható, ha kiderül, hogy mi micsoda. Ha a jobb oldalon csak szimbolikus értékek szerepelnek, azért ne vonjunk le pontot, de mint már említettem, a numerikus eredmények előbb-utóbb hiányozni fognak.)
(1 pont) $\approx \mathbb{P}(Y \le 1{,}6610)$, ahol $Y \sim \mathrm{N}(0;1)$,
(1 pont) ami megegyezik $\mathbb{P}(Y < 1{,}6610)$-mal, mert Y folytonos valószínűségi változó (ha nem hivatkozik a folytonosságra, azért ne vonjunk le pontot, de legalább annyi legyen kimondva, hogy a két valószínűség egyenlő. Mivel X_i és így $\sum_{i=1}^{100} X_i$ is folytonos valószínűségi változó, a „$\le$”-ról „$<$”-ra már az előző lépésben át lehet térni, de ott indokolni kell a folytonosságot.)
(1+1 pont) és $\mathbb{P}(Y < 1{,}6610)(= \Phi(1{,}6610)) \approx \Phi(1{,}66)$ (nem muszáj használni a Φ jelölést, jó a $\mathbb{P}(Y < 1{,}66)$ is)
(1 pont) $\approx 0{,}9515$.

4. Egy X és egy Y valószínűségi változó együttes sűrűségfüggvénye

$$f_{X,Y}\colon \mathbb{R}^2 \to \mathbb{R}, \quad f_{X,Y}(x,y) = \begin{cases} c, & \text{ha } 0 < y < 1 \text{ és } 0 < x < y, \\ 0, & \text{különben,} \end{cases}$$

ahol $c > 0$ egy alkalmasan választott valós szám.

(a) Határozzuk meg c értékét.
(b) Határozzuk meg az f_X és f_Y perem-sűrűségfüggvényeket.
(c) Független-e X és Y? Azonos eloszlású-e X és Y?

Megoldás.
(a) (5 pont)
(1 pont) $f_{X,Y}$ akkor lesz együttes sűrűségfüggvény, ha (c nemnegatív – ez meg van adva, ezért nem jár rá pont – és) $1 = \iint_{-\infty}^{\infty} f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y$
(1 pont) $= \int_0^1 \int_0^y c\mathrm{d}x\mathrm{d}y$ (ha ez jó, akkor az előző pontot az előző képlet hiányában is megadjuk)
(1 pont) $= c\int_0^1 [x]_{x=0}^y \mathrm{d}y = c\int_0^1 y\mathrm{d}y$
(1 pont) $= c[y^2/2]_{y=0}^1 = c/2$,
(1 pont) amiből $c = 2$.
Alternatív megoldás: a geometriai valószínűségi mezővel érvelünk. A sűrűségfüggvény konstans azon a tartományon, ahol nem 0. Ez a tartomány pedig az $\{(x,y) \in \mathbb{R}^2 \colon 0 < x < y < 1\}$ derékszögű egyenlőszárú háromszög, vagyis az egységnégyzet $y = x$ szögfelező-egyenes feletti fele, amelynek amelynek területe $1/2$. Az (X,Y) valószínűségi vektorváltozó 1 valószínűséggel ebbe a tartományba esik, és itt a valószínűség a területtel arányos (mert a sűrűségfüggvény konstans), ami azt jelenti, hogy c-szer a háromszög területe éppen 1, vagyis $c = 2$. (5 pont, indoklás szükséges, de ha lényegében helyes, csak nem elég precíz, akkor ne vonjunk le túl sok pontot.)
(b) (8 pont)

(1 pont) Ha $0 < x < 1$, akkor $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\mathrm{d}y$
(1 pont) $= \int_x^1 2\mathrm{d}y$ (ha ez jó, akkor az előző pontot az előző képlet hiányában is megadjuk)
(1 pont) $= \Big[2y\Big]_{y=x}^{1} = 2 - 2x$,
(1 pont) és ha $x \leq 0$ vagy $x \geq 1$, akkor $f_X(x) = 0$,
(1 pont) valamint ha $0 < y < 1$, akkor $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\mathrm{d}x$
(1 pont) $= \int_0^y 2\mathrm{d}x$ (ha ez jó, akkor az előző pontot az előző képlet hiányában is megadjuk)
(1 pont) $= \Big[2x\Big]_{x=0}^{y} = 2y$,
(1 pont) és ha $y \leq 0$ vagy $y \geq 1$, akkor $f_Y(y) = 0$.
(c) (7 pont)
Önmagában arra a kijelentésre, hogy nem függetlenek, nem jár pont.
(1 pont) Nem függetlenek, mert... (valahogy megindokolja, akkor is, ha rosszul).
(2 pont) X és Y pontosan akkor függetlenek, ha $f_X(x)f_Y(y) = f_{X,Y}(x,y)$ minden $x, y \in \mathbb{R}$ esetén teljesül (ha kihagyja, hogy minden $x, y \in \mathbb{R}$ esetén, akkor max. 1 pont).
(2 pont) arra, hogy választ egy olyan x és egy olyan y értéket, amelyre $f_X(x)f_Y(y) \neq f_{X,Y}(x,y)$, pl. $x = y = 1/4$.
(1 pont) arra az ötletre, hogy pontosan akkor lennének azonos eloszlásúak, ha ugyanaz lenne a sűrűségfüggvényük (mivel mindkét peremeloszlás folytonos, de ezt nem kell odaírni. Szabad az eloszlásfüggvényeket is kiszámolni és az alapján következtetni, de fölösleges),
(1 pont) arra, hogy ez itt nem teljesül, ezért nem azonos eloszlásúak. (Ezt nem kell részletezni, de $t \in (0,1) \setminus \{1/2\}$ esetén $f_X(t) \neq f_Y(t)$. Az utolsó két pont természetesen összevonható.)

5. Egy trópusi barlangban a hőmérséklet egész évben közel állandó. A hőmérsékletet 6 időpontban megmértük és (Celsius-fokban mérve) az

$$x_1 = 18{,}6 \quad x_2 = 18{,}0 \quad x_3 = 19{,}2 \quad x_3 = 17{,}4 \quad x_5 = 16{,}8 \quad x_6 = 18{,}0$$

adatpontokat kaptuk. Az $\mathbf{x} = (x_1, \ldots, x_6)$ vektort egy $\mathbf{X} = (X_1, \ldots, X_6)$ fae. $\mathrm{N}(\mu; 4)$ eloszlású minta realizációjának tekintjük, ahol a μ várható érték ismeretlen.

(a) A minta alapján adjunk 97% szintű konfidenciaintervallumot a μ paraméterre.

(b) Egy régi földrajztankönyv azt állítja, hogy $\mu = 17{,}5$ °C. Felmerült, hogy a klímaváltozás miatt ez az állítás már sajnos nem tartható. Vizsgáljuk $\varepsilon = 0{,}05$ terjedelmű próbával, hogy a minta alapján igaz-e az állítás. Adjuk meg H_0-t, H_1-et, a próbastatisztika értékét, a döntést és azt, hogy mi alapján döntöttünk.

Megoldás.
(a) (9 pont)
(1 pont) arra, hogy mivel a szórás ismert, az ismert szóráshoz tartozó (a képletgyűjteményben az u-próbánál szereplő) képletet használjuk a konfidenciaintervallumra. Elég, ha ez impliciten derül ki.
(1 pont) $n = 6$ és $\varepsilon = 0{,}03$ (elég, ha ezek is csak impliciten derülnek ki a későbbiekben), (1 pont) $\overline{x_6} = \frac{1}{6}\sum_{i=1}^{6} x_i$
(1 pont) $= 18$,
(1 pont) és $\sigma = \sqrt{4} = 2$.
(1 pont) $u_{\varepsilon/2} = \Phi^{-1}(1 - \varepsilon/2) = \Phi^{-1}(0{,}985)$
(1 pont) $\approx 2{,}17$ (a normális eloszlás táblázata alapján).
(2 pont) Így a konfidenciaintervallum $(\bar{x}_n - \frac{\sigma u_{\varepsilon/2}}{\sqrt{n}}, \bar{x}_n + \frac{\sigma u_{\varepsilon/2}}{\sqrt{n}}) = (18 - \frac{2 \cdot 2{,}17}{\sqrt{6}}, 18 + \frac{2 \cdot 2{,}17}{\sqrt{6}}) \approx (16{,}2282, 19{,}7718)$ (itt csak a számolásra, illetve a végeredményekre jár pont, mert arra már adtunk pontot, hogy melyik képletet kell használni).
(b) (11 pont)

(1 pont) ismét jár arra, hogy mivel a szórás ismert, u-próbát kell használni, és a próba egymintás, (elég, ha ez később impliciten derül ki).
(2 pont) arra, hogy kétoldali próbára van szükség (ez a földrajztankönyvben szereplő állításból következik). Ezt nem kell expliciten kimondani, ha jó a nullhipotézis.
Ha ezt elrontja a hallgató, akkor ezt a két pontot elveszti, illetve az alternatíva lentebbi értékelésénél a $H_0\colon \mu \leq 17{,}5$ vs. $H_1\colon \mu > 17{,}5$ alternatívára maximum 1 pont jár, a $H_0\colon \mu \geq 17{,}5$ vs. $H_1\colon \mu < 17{,}5$ alternatívára pedig nem jár pont. Ezután viszont a továbbiakban úgy értékeljük a megoldást, mintha az alternatíva helyes lenne (feltéve, hogy szintaktikailag értelmezhető), tehát az adott egyoldali alternatívának megfelelően az $1-\varepsilon$-kvantilis egyszerese, illetve -1-szerese a helyes és azzal kell összehasonlítani az u-statisztikát és az alapján dönteni.
(2 pont) Tehát az alternatíva $H_0\colon \mu = 17{,}5$ vs. $H_1\colon \mu \neq 17{,}5$ (itt μ jelöli a mintaelemek várható értékét, de ezt nem muszáj odaírni).
(1 pont) (Mivel egymintás u-próbáról van szó), a próbastatisztika értéke $u(x_1,\ldots,x_6) = \frac{\bar{x}_n-\mu_0}{\sigma}\sqrt{n}$ (elég, ha ez impliciten derül ki a következő lépésből. Ha csak u-t ír $u(x_1,\ldots,x_6)$ helyett, azért ne vonjunk le pontot, mert a táblázatban is az van)
(1 pont) $= \frac{18-17{,}5}{2}\sqrt{6} \approx 0{,}6124$.
(1 pont) H_0-t pontosan akkor fogadjuk el, ha $|u(\mathbf{x}_1,\ldots,x_n)| \leq u_{\varepsilon/2} = \Phi^{-1}(1-\varepsilon/2)$,
(1 pont) ahol ε a terjedelem, vagyis 0,05, tehát $\Phi^{-1}(1-\varepsilon/2) = \Phi^{-1}(0{,}975)$
(1 pont) $\approx 1{,}96$.
(1 pont) Ezért H_0-t elfogadjuk (vagyis arra jutunk, hogy a várható érték [még] 17,5 °C-kal egyenlőnek tekinthető).

6. * Az $\begin{pmatrix} X \\ Y \end{pmatrix}$ valószínűségi vektorváltozóról tudjuk, hogy (nemelfajuló) kétdimenziós normális eloszlású, várható érték vektora $\begin{pmatrix} 42 \\ -1 \end{pmatrix}$, valamint $\mathbb{E}(Y|X) = X + d$, ahol $d \in \mathbb{R}$ egy alkalmasan választott konstans.

   (a) Határozzuk meg d értékét.

   (b) Határozzuk meg Y-nak X-re vett regressziós egyenesét.

   (c) Független-e X és Y?

   (d) Diagonális-e $\begin{pmatrix} X \\ Y \end{pmatrix}$ kovarianciamátrixa?

Megjegyzés: Ha az (a) *részben nem sikerült d-t kiszámítani, a többi részfeladatot paraméteresen is megoldhatjuk.*
Megoldás. (a) (7 pont)
(1 pont) A teljes várható érték tétele szerint
(2 pont) $\mathbb{E}(Y) = \mathbb{E}(\mathbb{E}(Y|X))$
(1 pont) $= \mathbb{E}(X + d)$.
(1 pont) $= \mathbb{E}(X) + d$ (az előző egyenlettel összevonható)
(1 pont) $= 42 + d$, aminek -1-gyel kell megegyeznie (a megadott várható érték vektor alapján),
(1 pont) tehát $d = -43$.
(b) (5 pont)
(1 pont) arra, hogy többdimenziós normális eloszlás esetén a regresszió megegyezik a lineáris regresszióval (ez a pont akkor is jár, ha a konkrét szituációra nem tudja megfelelően alkalmazni. A többdimenziós normalitásra viszont mindenképpen hivatkozni kell).
(2 pont) Ez azt jelenti, hogy $\mathbb{E}(Y|X) = \beta X + \alpha$, ahol β és α a lineáris regresszió képletében szereplő kifejezések (nem kell kiírni, hogy mi β és α képlete, de az valahogy derüljön ki, hogy a lineáris regressziónak felelnek meg).

(1 pont) Ezért $\beta = 1$ és $\alpha(= d) = -43$,
(1 pont) vagyis a regressziós egyenes $\{(x, y) \in \mathbb{R}^2 | y = x - 43\}$.
(c) (3 pont)
(1 pont) arra, hogy ha X és Y függetlenek lennének, akkor $\mathbb{E}(Y|X)$ megegyezne $\mathbb{E}(Y)$-nal.
(2 pont) Ez viszont itt nem teljesülhet, hiszen $\mathbb{E}(Y|X) = (X + d =)X - 43$, és X (vagy: $X - 43$) nem konstans valószínűségi változó (vagy: nem 1 valószínűséggel konstans) (mivel az eloszlás nemelfajuló, de nem várjuk el az erre való hivatkozást), míg $\mathbb{E}(Y)(= d = -43)$ egy konstans.
(d) (5 pont)
(1 pont) A kovarianciamátrix pontosan akkor diagonális, ha $\text{cov}(X, Y) = 0$.
(1 pont) ennek megindoklására (pl. kimondja a kovarianciamátrix definícióját, vagy azt mondja, hogy a kovarianciamátrix egy 2×2-es szimmetrikus mátrix, amelyben a főátlón kívül mindkét helyen $\text{cov}(X, Y)$ áll).
(2 pont) Ha $\text{cov}(X, Y) = 0$ teljesülne (vagyis ha X és Y korrelálatlanok lennének; ha ezt írja, nem kell indokolni, hogy a korreláció létezik, de ez is abból következik, hogy az eloszlás nemelfajuló), akkor viszont függetlenek is lennének, mert $(X, Y)^T$ kétdimenziós (vagy: többdimenziós) normális eloszlású.
(1 pont) És mivel láttuk, hogy nem függetlenek, ezért korrelálatlanok sem lehetnek, tehát a kovarianciamátrix nem diagonális.