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 részpontszámokat 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 részpontszámok 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 puszta 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.

Részpontszám 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 dolgozatból 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 részpontszámok szükség esetén tovább is oszthatók. Az útmutatóban leírttól 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. Válaszoljuk meg az alábbi kérdéseket az előadáson elhangzott definíciók alapján.

   (a) Legyenek A, B események. Mikor mondjuk azt definíció szerint, hogy A és B függetlenek?

   (b) Mikor nevezünk egy valószínűségi változót egyszerűnek?

   (c) Legyen (X, Y) folytonos valószínűségi vektorváltozó. Hogyan nevezzük és hogyan jelöljük azt a függvényt, amelynek értékét az (x, y) pontban $f_X(x) \neq 0$ esetén az

   $$\frac{f_{X,Y}(x,y)}{f_X(x)}$$

   képlettel definiáljuk, $f_X(x) = 0$ esetén pedig definíció szerint 0 az értéke?

   (d) Legyen $x \in \mathbb{R}$. A (c) részfeladatban szereplő szituációban hogyan számíthatjuk ki $f_X(x)$ értékét, ha $f_{X,Y}(x, y)$ értékét ismerjük minden $y \in \mathbb{R}$ esetén?

Megoldás.
(a)
(4 pont) Ha $\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$.
(b)
(3 pont) Ha értékkészlete véges/ha csak véges sok értéket vesz fel.
(c)
(3 pont) $f_{Y|X}(y|x)$-szel jelöljük (nem kell odaírni, hogy $(x, y) \mapsto f_{Y|X}(y|x)$)
(5 pont) és az Y-nak X-re vett (vagy X-re vonatkozó) feltételes sűrűségfüggvényének nevezzük.
(d)
(5 pont) $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\mathrm{d}y$.

Megjegyzések javítóknak, részben hibás megoldások értékelése.
(a)
Szövegesen is elfogadható, ha egyértelmű, hogy erre gondol.
Ha az $A \cap B$ helyett más jelölést használ, akkor abban az esetben lehet részpontot adni, ha egyértelmű, hogy a metszetet akarja körülírni, max. 3 pont. Ha uniót vagy arra utaló kifejezést ír, akkor 0 pont!
Ha feltételes valószínűségekkel akarja megadni a definíciót, akkor max. 2 pont, és abból is 1 pont arra jár, hogy odaírja, hogy a feltétel valószínűségének pozitívnak kell lennie.
(b)
A $\mathrm{Ran}(X)$ jelölés itt nem kötelező az értékkészletre, és ha az egyetlen hiba az, hogy valahogy rosszul jelöli, pl. Ran_X, de egyértelműen erre utal, azért ne vonjunk le pontot.
Ha azt mondja, hogy maga a val. változó véges, akkor max. 1 pont.
(c)
Ha csak annyit ír, hogy Y feltételes sűrűségfüggvénye, akkor az 5-ből max. 4 pont; ha csak annyit ír, hogy feltételes sűrűségfüggvény, akkor max. 3 pont.
A következő megjegyzés kizárólag a javítóknak szól, elolvasását még nem vizsgázott hallgatóknak kifejezetten nem ajánlom: Ha más forrásokban szereplő, de az előadásról/jegyzetből nem ismert jelölést használ, – ilyenek bőven vannak, pl. $f_{Y|X}(x,y)$, $f_{Y|X=x}(y|x)$, $f_{Y|X=x}(x,y)$ – akkor a 3-ból max. 2 pont, és 2 csak akkor, ha úgy tűnik, hogy ez a jelölés valamilyen, egyetemi oktató által írott forrásban tényleg szerepel és nincs benne egyéb szintaktikai hiba.
Ha összekeveri X és Y szerepét, akkor a 3-ból max. 1 pont, ami csak a tökéletes $f_{X|Y}(x|y)$ jelölés esetén jár.
(d)
Nem baj, ha nem írja oda, hogy „$f_X(x) =$”.
Kisebb szintaktikai hibák esetén adható részpont.
Ha az integrálási határok nem jók vagy hiányoznak, akkor max. 3 pont.
Ha x szerint integrál vagy nem világos, hogy mi szerint integrál, akkor egyből 0 pont!

2. Legyenek $X_1, \ldots, X_{100}$ független, azonos eloszlású valószínűségi változók, amelyek sűrűségfüggvénye

$$f_{X_1}(x) = \begin{cases} \frac{3}{7}x^2, & \text{ha } 1 < x < 2, \\ 0, & \text{egyébként.} \end{cases}$$

   (a) Bizonyítsuk be, hogy f_{X_1} valóban sűrűségfüggvény.
   (b) Határozzuk meg X_1 várható értékét és szórását.
   (c) Legyen $X = X_1 + \ldots + X_{100}$. Közelítsük alkalmas approximációval a $\mathbb{P}(X > 160)$ valószínűséget. (A valószínűség pontos értékét nem kell meghatározni.)
   Pótfeladat: Ha a (b)-t nem sikerült megoldani, tegyük fel, hogy $\mathbb{E}(X_1) = 1{,}6100$ *és* $\mathbb{D}(X_1) = 1/4$ *és oldjuk meg a (c)-t úgy.*

Megoldás.
(a) (5 pont)
(1 pont) Ahhoz, hogy f_{X_1} sűrűségfüggvény legyen, az kell, hogy nemnegatív legyen (ez itt elég világos, nem kell részletezni, csak megemlíteni)
(1 pont) és teljesüljön, hogy $1 = \int_{-\infty}^{\infty} f_{X_1}(x)\mathrm{d}x$
(1 pont) $= \int_1^2 \frac{3}{7}x^2\mathrm{d}x$ (az előzővel összevonható, ha ezt jól írja)
(1 pont) $= [\frac{1}{7}x^3]_1^2$
(1 pont) $= \frac{8}{7} - \frac{1}{7} = 1$, tehát tényleg teljesül. (Ha jól megadja a primitív függvényt, akkor itt már elég a végeredmény. Ha kihozza, hogy az eredmény 1, nem baj, ha nem kommentálja.)
(b) (8 pont)
(1 pont) $\mathbb{E}(X_1) = \int_{-\infty}^{\infty} x f_{X_1}(x)\mathrm{d}x = \int_1^2 \frac{3x^3}{7}\mathrm{d}x$ (itt csak 1 pont jár a két alakra együtt, ezért ha

a kettőből legalább az egyik megvan helyesen, az elég)
(1 pont) $= [\frac{3}{28}x^4]_1^2$
(1 pont) $= \frac{48}{28} - \frac{3}{28} = \frac{45}{28} (\approx 1{,}6071)$ (pontozás az előző integrálhoz hasonlóan, elég csak közönséges vagy csak tizedes tört alakban is).
(1 pont) $\mathbb{D}(X_1) = \sqrt{\mathbb{E}(X_1^2) - \mathbb{E}(X_1)^2}$ (itt arra, hogy a szórás a szórásnégyzet gyöke, nem jár külön pont, mert elég hosszú a feladat és sok mindenre kell pontot adni; erre és a $\mathbb{D}^2(X_1)$ képletére csak együtt jár pont).
(1 pont) $\mathbb{E}(X_1^2) = \int_{-\infty}^{\infty} x^2 f_{X_1}(x)\mathrm{d}x = \int_1^2 \frac{3x^4}{7}\mathrm{d}x$ (pontozás $\mathbb{E}(X_1)$-éhez hasonlóan)
(1 pont) $= [\frac{3}{35}x^5]_1^2$
(1 pont) $= \frac{96}{35} - \frac{3}{35} = \frac{93}{35} (\approx 2{,}6571)$ (pontozás az $\mathbb{E}(X_1)$ kiszámításának utolsó lépéséhez hasonlóan).
(1 pont) Tehát $\mathbb{D}(X_1) = \sqrt{\frac{93}{35} - (\frac{45}{28})^2} (\approx \sqrt{0{,}0742}) \approx 0{,}2725$.
(Ha már korábban megadta $\mathbb{E}(X_1)$ és $\mathbb{E}(X_1^2)$ értékét és $\mathbb{D}(X)$ általános képletét, akkor itt elég a végeredmény, de kell a numerikus értéke is. Az is OK, ha az általános képletet csak $\mathbb{E}(X_1)^2$ és $\mathbb{E}(X_1^2)$ konkrét értékeivel használja, feltéve, hogy azonosítható, hogy ezt a képletet alkalmazza.)

(c) (7 pont)
Megoldás a (b)-ben kapott, négy értékes jegyre kerekített numerikus értékekkel dolgozva:
(1+1 pont) Sztenderdizálunk (ezt nem kell szövegesen odaírni):

$$\mathbb{P}(X > 160) = \mathbb{P}\Big(\frac{X - 100\,\mathbb{E}(X_1)}{\sqrt{100}\,\mathbb{D}(X_1)} > \frac{160 - 100\,\mathbb{E}(X_1)}{\sqrt{100}\,\mathbb{D}(X_1)} \approx \mathbb{P}\Big(\frac{X - 160{,}71}{\sqrt{100}\cdot 0{,}2725} > \underbrace{\frac{160 - 160{,}71}{\sqrt{100}\cdot 0{,}2725}}_{\approx -0{,}2606}\Big)$$

(ha a második képlet helyes és világos módon az elsőből származik, akkor a két lépés összevonható pontlevonás nélkül),
(1 pont) ami a centrális határeloszlás-tétel (elég: CHT) miatt (De Moivre–Laplace itt nem jó!)
(1 pont) közelítőleg $\mathbb{P}(Y > -0{,}2606)$, ahol $Y \sim \mathrm{N}(0;1)$.
(1 pont) $\mathbb{P}(Y > -0{,}2606) = \mathbb{P}(Y < 0{,}2606)$ (ugyanez kifejezhető Φ-értékekkel is, de nem kötelező)
(1 pont) $\approx \mathbb{P}(Y < 0{,}26) = \Phi(0{,}26)$ (itt sem kötelező a Φ jelölés)
(1 pont) $\approx 0{,}6026$.
A (b)-ben kiszámolt momentumok pontos értékeivel számolva kapott megoldásról:
$\frac{160 - \frac{4500}{28}}{\sqrt{100}\sqrt{\frac{93}{35} - (\frac{45}{28})^2}} \approx -0{,}2621$, tehát két tizedesjegyre való kerekítés után ugyanazt az eredményt kell kiolvasni a normális eloszlás táblázatból, mint a kerekített értékekkel számolva.
A pótmegoldásról:
Ebben az esetben a standardizálás után a jobb oldalon $\frac{160-161}{10\cdot 0{,}25} = -0{,}4$ áll, így a végeredmény $\mathbb{P}(Y < 0{,}4) \approx 0{,}6554$. Pontozás ezeket az adatokat figyelembe véve ugyanúgy, mint az első megoldás esetén.
Ha (c)-ben a pótmegoldással számol, akkor a (b)-re és (c)-re együtt max. 13 pont, akkor is, ha a (b)-ben esetleg jól kiszámolta a várható értéket és szórást, csak nem mert azokkal számolni. De ha a rendes megoldást és a pótmegoldást is jól megadja, akkor nem kell pontot levonni.

3. Legyen X és Y két független és a $(0;1)$ intervallumon egyenletes eloszlású valószínűségi változó, valamint $U = X + Y$ és $V = XY$.

   (a) Határozzuk meg az U és V valószínűségi változók kovarianciáját és a várható értékeiket, valamint U szórásnégyzetét.
   Segítség: A feladat integrálás nélkül is megoldható (a tanult azonosságok használatával).

   (b) Mi V regressziós egyenese U-ra?

Megoldás.
(a) (15 pont)
Az azonosságok megnevezésére itt nem jár külön pont. Elég, ha részletesen megmutatja, hogy hogyan használja őket.
(0 pont) $\operatorname{cov}(U,V) = \operatorname{cov}(X+Y, XY)$
(1 pont) $\operatorname{cov}(X,XY) + \operatorname{cov}(Y,XY)$.
(1 pont) $\operatorname{cov}(X,XY) = \mathbb{E}(X^2Y) - \mathbb{E}(X)\,\mathbb{E}(XY)$. (Ez a pont csak akkor jár, ha a kovariancia általános képletét egy konkrét releváns helyzetben alkalmazza.)
(1 pont) $\mathbb{E}(X^2Y) = \mathbb{E}(X^2)\,\mathbb{E}(Y)$,
(1 pont) mert függetlenek.
(Nem kell részletezni, hogy ha X és Y függetlenek, akkor X^2 és Y is azok, és elég összesen egyszer hivatkozni arra, hogy függetlenség esetén a szorzat várható értéke a várható értékek szorzata).
(1 pont) $\mathbb{E}(Y) = \frac{1+0}{2} = \frac{1}{2}$ (elég a végeredmény)
(1 pont) és $\mathbb{E}(X^2) = \mathbb{D}^2(X) + \mathbb{E}(X)^2$
(1+1 pont) $= \frac{1}{12} + \frac{1}{4} = \frac{1}{3}$ (1 pont arra, ha valahol – a megoldás során bárhol – megjelenik, hogy $\mathbb{D}^2(X) = \mathbb{D}^2(Y) = \frac{1}{12}(= \frac{(1-0)^2}{12})$, indoklás nélkül is; sőt elég, ha az egyik jelenik meg és a másikat csak magyarázat nélkül írja oda; és 1 pont a végeredményre, a számolást nem kell nagyon részletezni).
(1 pont) $\mathbb{E}(XY) = \mathbb{E}(X)\,\mathbb{E}(Y) = \frac{1}{4}$ (ha megvan a szorzatra bontás, akkor már jár az 1 pont, és ha jól csinálja, akkor nem kell indokolni, hogy $\mathbb{E}(Y) = \mathbb{E}(X)$ [mert X és Y azonos eloszlásúak]).
(1 pont) Tehát $\operatorname{cov}(X,XY) = \frac{1}{3}\cdot\frac{1}{2} - \frac{1}{2}\cdot\frac{1}{4} = \frac{1}{6} - \frac{1}{8} = \frac{4-3}{24} = \frac{1}{24}$. (Ha minden korábbi részeredmény jó és a $\operatorname{cov}(X,XY) = \mathbb{E}(X^2Y) - \mathbb{E}(X)\,\mathbb{E}(XY)$ képlet is megvan, akkor elég a végeredmény is.)
(1 pont) A szimmeria miatt $\operatorname{cov}(Y,XY) = \operatorname{cov}(X,XY) = \frac{1}{24}$. (Indoklás nélkül is OK.)
(0 pont) Tehát $\operatorname{cov}(U,V) = \frac{1}{12}$.
(0 pont) $\mathbb{D}^2(U) = \mathbb{D}^2(X+Y)$
(1 pont) $= \mathbb{D}^2(X) + \mathbb{D}^2(Y)$
(0 pont) $= \frac{1}{12} + \frac{1}{12} = \frac{1}{6}$ (itt már nem jár pont a szórásnégyzetek értékére, mert korábban már adtunk rá),
(1 pont) mert függetlenek.
(1 pont) $\mathbb{E}(U) = \mathbb{E}(X) + \mathbb{E}(Y)$
(1 pont) $= \frac{1}{2} + \frac{1}{2} = 1$ (itt is elég a végeredmény).
(0 pont) $\mathbb{E}(V) = \mathbb{E}(X)\,\mathbb{E}(Y) = \frac{1}{4}$, ahogy már kiszámoltuk. (Erre a feladat során egyszer jár 1 pont és a várható érték szorzatra bomlására is összesen egyszer.)

Összefoglalva a végeredmények: $\operatorname{cov}(U,V) = \frac{1}{12}$, $\mathbb{D}^2(U) = \frac{1}{6}$, $\mathbb{E}(U) = 1$, $\mathbb{E}(V) = \frac{1}{4}$.

Jó úgy is, ha azt írja, hogy $\mathbb{D}^2(X+Y) = \mathbb{D}^2(X) + \mathbb{D}^2(Y) + 2\operatorname{cov}(X,Y)$, és utána veszi észre, hogy a függetlenség miatt a kovariancia 0. Ekkor a függetlenségre járó pont az utóbbi észrevételre jár.
Természetesen nem tilos a feladatot integrálással megoldani, ekkor a kovariancia kiszámításához kell X és Y együttes sűrűségfüggvénye, és ennek szorzat alakját is indokolni kell. Lehet egyes részfeladatokban is integrálni, ha nem megy másként, például ha $\mathbb{E}(X^2)$-et nem tudja a szórásnégyzet segítségével meghatározni, de integrálással igen.

(b) (5 pont)
(1 pont) A lineáris regresszió képletében szereplő mennyiségek: $\beta = \frac{\operatorname{cov}(U,V)}{\mathbb{D}^2(U)}$
(1 pont) $= \frac{\frac{1}{12}}{\frac{1}{6}} = \frac{1}{2}$ (itt már elég a végeredmény, ha korábban megvolt minden szükséges részletszámítás)
(1 pont) és $\alpha = \mathbb{E}(V) - \beta\,\mathbb{E}(U)$
(1 pont) $= \frac{1}{4} - \frac{1}{2}\cdot 1 = -\frac{1}{4}$ (az előzőhöz hasonlóan).
(1 pont) Tehát a regressziós egyenes $\{(x,y) \in \mathbb{R}^2 \colon y = \frac{1}{2}x - \frac{1}{4}\}$.

Ha nem az egyenest adja meg, hanem csak a $\beta X + \alpha$ lineáris regressziót, akkor az utolsó pont nem jár.
Ha az egyenes helyes általános képletét ($\beta x + \alpha$-val) adja meg, akkor jár, még ha a számszerű értékeket nem is sikerült meghatározni.
A β és α jelölések nem kötelezőek, lehet helyettük más betűket is használni, és semmilyen betűt nem kell használni, ha anélkül is azonosítható, hogy melyik mennyiség honnan származik.

4. Feldobunk egy szabályos kockát, majd egy szabályos érmét annyiszor, amennyit a kocka mutat. Jelölje Y a dobott fejek számát és X a kockadobás értékét.

   (a) Várhatóan hány fejet dobunk, feltéve, hogy a kocka 3-at mutat (és miért)?

   (b) Mennyi $\mathbb{E}(Y|X)$?

   (c) Határozzuk meg Y várható értékét.

   (d) Lássuk be, hogy X és Y nem függetlenek.

A részfeladatokat tetszőleges sorrendben megoldhatjuk.
Megoldás.
(a) (4 pont)
(1 pont) Várhatóan $3/2 = 1{,}5$ fejet dobunk, ha a kocka 3-at mutat (ez az 1 pont indoklás nélkül is jár).
(3 pont) Megindokolja rendesen, hogy miért igaz, hogy $\mathbb{E}(Y|X = 3) = \frac{3}{2}$ (ehhez nem kötelező a feltételes várható értékes jelölést használni), ilyen esetben jár az előző 1 pont is.
Például: ha tudjuk, hogy a kocka 3-at mutat, akkor a fejek száma binomiális eloszlású $n = 3$ és $p = 1/2$ paraméterrel, és nekünk éppen ennek a várható értéke kell, ami $\frac{3}{2}$.
Ugyanez indikátorok összegével kifejezve: ha tudjuk, hogy a kocka 3-at mutat, minden dobás $1/2$ valószínűséggel lesz fej (tehát egy dobásból várható értékben $1/2$ fej lesz), így a várható érték linearitása miatt a fejek várható száma $3/2$. Ha itt a linearitásra nem hivatkozik és nem ír le olyan képletet sem, ami erre utalna, akkor a 3 pontból max. 2 pont.
Ha csak annyi indoklást ír, hogy átlagosan/várható értékben a dobások fele lesz fej, akkor tekintsük úgy, hogy az utóbbi megoldást próbálja leírni és adjunk rá 2 pontot a 3-ból. Ha csak annyit ír, hogy minden dobás $1/2$ valószínűséggel fej és ebből következik az eredmény, akkor 1 pontot adjunk a 3-ból.
(b) (4 pont)
(1 pont) Ha a kocka k-t mutat (ahol $k \in \{1, \dots, 6\}$), akkor várhatóan $k/2$ fejet dobunk. (Nem kell indokolni, ugyanúgy következik, mint az (a)-ban.)
(1 pont) Ez azt jelenti, hogy $\mathbb{E}(Y|X = k) = \frac{k}{2}$ ilyen k esetén. (Ha ez megvan, akkor jár az előző 1 pont is. Ha megnevezi, hogy ez regressziós függvény, elég $g(k)$-val jelölni.)
(1 pont) $\mathbb{E}(Y|X)$-et (definíció szerint) úgy kapjuk, hogy a $k \mapsto g(k) = \mathbb{E}(Y|X = k)$ regressziós függvénybe X-et írunk.
(1 pont) Tehát $\mathbb{E}(Y|X) = \frac{X}{2}$. Ez az előző lépéssel összevonható abban az esetben, ha a (b) első 2 pontjából legalább egyet megszerzett, különben a 3. pont nem jár.
(c) (7 pont)
(2 pont) A teljes várható érték tétele szerint (szintén OK: TVÉT, toronyszabály. Ha TVT-t ír, akkor 0 pont, kivéve, ha utána mégis megjelenik a jó képlet, mert akkor 1 pont)
(2 pont) $\mathbb{E}(Y) = \mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}(g(X))$ (elég a három közül az egyik jelölés, és ha már korábban **expliciten** szerepelt, hogy $g(k) = k/2$, akkor $\mathbb{E}(X/2)$ vagy $\mathbb{E}(X)/2$ is elég; egyébként $\mathbb{E}(X/2)$ vagy $\mathbb{E}(X)/2$ esetén csak 1 pont jár)
1. megoldás:
(1 pont) $= \sum_{k=1}^{6} \frac{k}{2} \mathbb{P}(X = k)$ (ez a pont akkor is jár, ha $\frac{k}{2}$ helyett csak $g(k)$-val szerepel a képlet)
(1 pont) $= \frac{1}{6} \sum_{k=1}^{6} \frac{k}{2}$ (úgy is jó, hogy $\frac{1}{6} \cdot \frac{1+2+\dots+6}{2}$)

(1 pont) $= \frac{7}{4} = 1{,}75$ (elég az egyik alak és részletszámítások sem kellenek).
2. megoldás:
(1 pont) $= \mathbb{E}(X)/2$ (ha ez az alak megvan, akkor jár az előző 2 pont is).
(2 pont) Kockadobás várható értéke $\mathbb{E}(X) = 3{,}5 = \frac{7}{2}$ (ezt elég közölni, nem kell újra kiszámolni), tehát $\mathbb{E}(Y) = \frac{7}{4} = 1{,}75$ (megint elég az egyik alak).
(d) (5 pont)
1. megoldás: a függetlenség definíciója alapján.
Megad egy olyan (k, ℓ) párt, amelyre $\mathbb{P}(X = k, Y = \ell) \neq \mathbb{P}(X = k)\,\mathbb{P}(Y = \ell)$.
Itt 1 pont jár arra, ha egyértelműen ezt akarja csinálni, 1 pont arra, ha jó párt (k, ℓ) választ, és 1–1 pont a három valószínűség értékére, ezekhez nem kell indoklás. A legegyszerűbb olyan példát választani, ahol a jobb oldalon az egyik tényező 0, ekkor a másik tényező értékét nem is kell meghatározni és arra is jár az 1 pont. Például ha $\ell > k$ (több fejet dobunk, mint ahányat a kockával dobtunk, azaz ahány érmével dobunk), akkor mindig ez a helyzet.
Ugyanez a fajta megoldás feltételes valószínűségekkel és együttes/marginális eloszlásfüggvényekkel kifejezve is OK.
2. megoldás: a regresszió tulajdonságai alapján.
(2 pont) Ha X és Y függetlenek lennének, akkor $\mathbb{E}(Y|X) = \mathbb{E}(Y)$ teljesülne.
(3 pont) Ez viszont itt nem teljesül, mert $\mathbb{E}(Y)$ egy konstans, míg $\mathbb{E}(Y|X) = X/2$ [elég az egyik alak] nem (1 valószínűséggel) konstans. Vagy: mert $\mathbb{E}(Y) = 1{,}75$ a (c) alapján, és $\mathbb{E}(Y|X) = X/2$ nem (1 valószínűséggel) ezzel az értékkel egyenlő.

5. 7 darab, a BME campusán élő idősebb nyárfa magasságát megmértük és (méterben mérve) az alábbi eredményeket kaptuk:

$$x_1 = 23{,}1 \qquad x_2 = 21{,}5 \qquad x_3 = 22{,}0 \qquad x_4 = 22{,}8 \qquad x_5 = 21{,}2 \qquad x_6 = 20{,}9 \qquad x_7 = 22{,}5.$$

Egy kertészeti szakkönyv szerint az idősebb nyárfák magassága jó közelítéssel normális eloszlású, 1 méter szórással. Ezt az állítást elfogadva

(a) adjunk 0,98 szintű konfidenciaintervallumot a minta alapján a mintaelemek várható értékére,

(b) döntsük el $\varepsilon = 0{,}04$ terjedelmű próbával, hogy eléri-e a mintaelemek várható értéke a 22,5 métert. Adjuk meg H_0-t, H_1-et, a próbastatisztika értékét, azt, hogy mi alapján döntünk, és magát a döntést is.

(c) Határozzuk meg a minta empirikus mediánját.

Megoldás. (a) (8 pont)
(1 pont) arra, hogy az ismert szórás esetére vonatkozó (avagy u-kvantilises, u-próbánál szereplő stb.) konfidenciaintervallum képlete kell. Ha jó képletet használ, akkor nem baj, ha ezt nem írja oda.
(1 pont) $n = 7$ (ez a mintaelemszám),
(1 pont) $\overline{x_7} = \frac{1}{7}\sum_{i=1}^{7} x_i = 22$ (csak 1 pont jár rá, ezért OK indoklás nélkül is, $\overline{x_n}$ jelöléssel is jó),
(1 pont) $\sigma = 1$ (ez a mintaelemek szórásának felel meg, de ezt nem kell odaírni. Ha viszont azt állítja, hogy σ a szórásnégyzet, akkor nem jár a pont)
(1 pont) a szükséges u-kvantilis pedig $u_{\varepsilon/2} = u_{0{,}01} = \Phi^{-1}(0{,}99)$ (az ε jelölés nem kell, és elég ha az utóbbi két alakból az egyik szerepel, de derüljön ki, hogy ε-nak itt 0,02 felel meg)
(1 pont) $= 2{,}33$.
(1+1 pont) A képletbe behelyettesítve a konfidenciaintervallum:

$$\left[\overline{x_n} - \frac{\sigma u_{\varepsilon/2}}{\sqrt{n}}, \overline{x_n} + \frac{\sigma u_{\varepsilon/2}}{\sqrt{n}}\right] = \left[22 - \frac{1 \cdot 2{,}33}{\sqrt{7}}, 22 + \frac{1 \cdot 2{,}33}{\sqrt{7}}\right] = [21{,}1193, 22{,}8807].$$

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.

Javítóknak: Ha a t-próbás konfidenciaintervallum képletét akarja használni, akkor max. 7 pont. Ebből a 7 pontból 3 jár a korrigált empirikus szórásnégyzet kiszámolására (az előző vizsga 5-ös feladatához hasonlóan), és a kvantilisre járó pont csak akkor jár, ha a szabadsági fok ($n-1=6$) is helyes. Így a mintamegoldásban a σ értékére járó ponton kívül megszerezhető 7 pontból a korrigált empirikus szórás nélkül maximum 4 pont jár, további hibák esetén ebből vonunk le, és ha 0 pontra csökken az összpontszám, akkor az (a)-ra 0 pont és nem kell az (a) javítását folytatni.
Ha a jó képletet akarja használni, de azt hiszi, hogy σ a korrigált empirikus szórást jelöli, akkor s_7^* kiszámolására nem jár pont, a többi 7 pont megszerezhető. Ha az ismeretlen szórásos képletet akarja használni és s_n^* helyére beírja σ értékét, akkor a fentiek szerint max. 4 pont.

(b) (9 pont)
(1 pont) A feladat szövege alapján (nem kell indokolni) az alternatíva $H_0\colon \mu \geq 22{,}5$ vs. $H_1\colon \mu < 22{,}5$ (ahol μ jelöli a mintaelemek várható értékét – ezt sem kell odaírni).
(1 pont) Tehát ez egy egyoldali, egymintás próba (ha ezt nem írja oda, de ennek megfelelő képlettel számol, akkor ezt az 1 pontot meg lehet adni),
(1 pont) és mivel σ ismert, ezért u-próba kell. (Ha ezt nem írja oda, a pontot abban az esetben lehet megadni, ha a későbbiekben egyértelmű, hogy u-próbával számol és világos, hogy tudja, hogy mi σ.)
(1+1 pont) A próbastatisztika értéke $u(=u(x_1,\ldots,x_7)) = \frac{\overline{x_n}-\mu_0}{\sigma}\sqrt{n} = \frac{22-22{,}5}{1}\sqrt{7} = \frac{-\sqrt{7}}{2} \approx -1{,}3229$.
Itt 1 pont jár arra, hogy a helyes általános képletet értelmesen használja a már kiszámolt mennyiségekkel, ideértve, hogy tudja, hogy μ_0 a H_0 szerinti várható értéket jelenti. Elég, ha konkrét számokkal szerepel a képlet, ha azonosítható, hogy jó képletet használ. 1 pont pedig a helyes numerikus végeredményre jár, részletszámítások nem kellenek.
(1 pont) H_0-t akkor fogadjuk (nem utasítjuk) el, ha $u \geq -u_\varepsilon = -u_{0{,}04}$
(1 pont) $= -\Phi^{-1}(1-0{,}04) = -\Phi^{-1}(0{,}96)$ (ez az előzővel összevonható, ha helyesen szerepel)
(1 pont) $\approx -1{,}75$. (Ha jó eredményt olvas ki a táblázatból, akkor elég az $u_{\varepsilon/2}$ vagy $u_{0{,}02}$ jelölés, nem kell Φ^{-1}-zel kifejezni.)
(1 pont) Tehát H_0-t elfogadjuk/nem utasítjuk el (vagyis azt mondjuk, hogy a nyárfák magasságának várható értéke eléri a 22,5 métert – de itt nem kell szöveges magyarázat).

Pontozás hibás alternatívák/próbatípusok esetén: Ha kétoldali u-próbát hajt végre $H_0\colon \mu = 22{,}5$ vs.$H_1\colon \mu \neq 22{,}5$ alternatívával, az nem felel meg a feladatnak, hibátlan végrehajtás esetén max. 7 pont. Ha egyoldali u-próbát hajt végre, de $H_0\colon \mu \leq 22{,}5$ vs. $H_1\colon \mu > 22{,}5$ alternatívával, az még kevésbé logikus, max. 8 pont.
Ha az alternatíva jó, de erre (egyoldali, egymintás) t-próbát alkalmaz, akkor max. 7 pont. Az ismert vs. ismeretlen szórás miatt ugyan csak egy pontot kellene levonni, de $t_{f,0{,}04}$ nincs a táblázatban, és ennek fel kellene tűnnie. Hasonlóan, ha kétoldali t-próbát hajt végre, akkor nemcsak a szórással és az alternatívával van probléma, hanem azzal is, hogy $t_{f,0{,}02}$ nincs a táblázatban, ekkor max. 5 pont. Végül ha egyoldali t-próbát hajt végre a $H_0\colon \mu \leq 22{,}5$ vs. $H_1\colon \mu > 22{,}5$ alternatívára, akkor is fennállnak ugyanezek a problémák és az alternatíva még kevésbé logikus, max. 4 pont.
Általában induljunk ki a hallgató által megadott alternatívából és a megoldás további részeinek helyességét az alapján ítéljük meg, hogy az alternatívának megfelel-e. Ha nem egyértelmű az sem, hogy mi az alternatíva, akkor legfeljebb arra járnak részletpontok, hogy a táblázat-

ban/képletekben szereplő egyes kifejezéseket az adott szituációban értelmezi és behelyettesíti/kiolvassa.

(c) (3 pont)
(1 pont) Páratlan mintaelemszám esetén
(1 pont) az empirikus medián a mintának a nagyság szerinti (növekvő/csökkenő) sorrendben a középső eleme,
(1 pont) ami itt 22,0. (Ha odaírja, hogy ez x_3, az nem baj, mert nincsenek egyenlő mintaelemek, de kell a számszerű érték is).
Ha csak a helyes eredményt közli, 1 pont. Természetesen jó rendezett mintás jelöléssel is, de akkor is hivatkozni kell arra, hogy a mintaelemszám páratlan. Ha a második pont nem jár, akkor az első sem.

6. * $N \geq 2$ ember felírja a nevét egy-egy cetlire és ezeket egy kalapba rakják, majd mindenki véletlenszerűen húz egy cetlit a kalapból (visszatevés nélkül).

   (a) Az N ember között szerepel Anna és Béla. Mi a valószínűsége, hogy ők ketten egymás nevét húzzák?

   (b) Mi a várható értéke a *transzpozíciók* számának, azaz azon párok számának, akik egymás nevét húzzák?

Megoldás.
(a) (6 pont)
(1 pont) Az eseménytér modellezhető az N név/elem (vagy az $\{1, \ldots, N\}$ halmaz) permutációinak/sorbarendezéseinek halmazával,
(1 pont) a valószínűségi mező pedig klasszikus (vagy: kedvező esetek száma/összes eset száma képlettel lehet számolni).
Az első pontért nem muszáj a permutációkra hivatkozni, ha később kiderül a megoldásból, hogy az összes eset száma $N!$. A második pontért viszont kell, hogy legyen legalább annyi a megoldásban, hogy „kedvező/összes” vagy „klasszikus valószínűség(i mező)”, még akkor is, ha láthatóan így számol.
(1 pont) A kedvező esetek azok, ahol Anna és Béla egymás nevét húzzák, ekkor a maradék $N-2$ ember közül mindenki e közül az $N-2$ ember közül húzza ki valakinek a nevét.
(1 pont) Így tehát $(N-2)!$ a kedvező esetek száma (mivel ha tudjuk, hogy Anna és Béla egymás nevét húzták, akkor a többi ember húzásai egyértelműen meghatározzák a kimenetelt – ezt nem kell leírni).
Ha ez szerepel helyesen és legalább valami minimális indoklást próbál rá adni, – ha nem is olyan részletesen, mint ami itt a mintamegoldásban az előző pontnál szerepelt – akkor jár az előző pont is.
(1 pont) Az összes lehetséges kimenetel száma $N!$,
(1 pont) így a keresett valószínűség $\frac{(N-2)!}{N!} = \frac{(N-2)!}{(N-2)!(N-1)N} = \frac{1}{N(N-1)}$.
(Ha egyértelmű, hogy kedvező/összes képlettel számol, akkor elég a végeredmény. A végeredményt előbb-utóbb egyszerűsíteni kell, de elég a (b)-ben megtenni ezt.)

(b) (14 pont)
(0 pont) A transzpozíciók számát jelölje X, ekkor $\mathbb{E}(X)$-et kell meghatározni.
(3 pont) Egy tetszőleges $\{i, j\}$ párra, ahol $1 \leq i \leq N$ és $1 \leq j \leq N$ és $i \neq j$ (elég, ha az szerepel, hogy $i \neq j$), vezessük be az $X_{i,j} = \mathbb{1}_{\{i \text{ és } j \text{ transzpozíciót alkot}\}}$ valószínűségi változót/indikátorváltozót.
Más szavakkal megfogalmazva (elég az egyik megfogalmazás): legyen $X_{i,j} = 1$, ha i és j transzpozíciót alkot és $X_{i,j} = 0$ egyébként.
VAGY: az $X_{i,j}$-ket tekintsük csak $1 \leq i < j \leq N$ esetén.

VAGY: $X_{i,j}$-k helyett $X_{\{i,j\}}$-ket vezessünk be, ahol $1 \leq i \leq N$ és $1 \leq j \leq N$ és $i \neq j$ (és nyilván $\{i,j\} = \{j,i\}$).
(2 pont) Ekkor $X = \sum_{\{i,j\}:\, 1 \leq i \leq N, 1 \leq j \leq N} X_{i,j}$, VAGY: $X = \sum_{1 \leq i < j \leq N} X_{i,j}$.
(VAGY: $X = \sum_{\{i,j\}:\, 1 \leq i \leq N, 1 \leq j \leq N} X_{\{i,j\}}$; ezt a fajta megoldást a továbbiakban nem részletezem, mert majdnem ugyanaz, mint az első.)
(1 pont) A várható érték linearitása miatt (képlettel is kifejezhető, elég az egyszerű val. változókra vonatkozó verziójára hivatkozni, de nem elég a konkrét esetre felírni)
(2 pont) $\mathbb{E}(X)$ az összes $\mathbb{E}(X_{i,j})$ összegével egyezik meg (képlettel is jó persze).
(1 pont) $\binom{N}{2} = \frac{N(N-1)}{2}$, a feltételeknek megfelelő $\{i,j\}$ pár van.
(Ezt itt nem kell jobban indokolni. Ha az $1 \leq i < j \leq N$ alakú (i,j) rendezett párokat nézzük, akkor ezeknek a száma pontosan az N elem ismétlés nélküli kombinációinak száma, ami $\binom{N}{2}$. Ha pedig az $\{i,j\}$ alakú párokat számláljuk le, ahol i és j is 1 és N között lehet és különbözőek, akkor az N elem másodosztályú, ismétlés nélküli variációinak számát kell 2-vel osztani, mert $\{i,j\}$ és $\{j,i\}$ mindig megegyezik egymással, ez tehát $\frac{N(N-1)}{2}$ pár.)
(2 pont) Az (a) feladat alapján bármely két elem $\frac{1}{N(N-1)}$ valószínűséggel alkot transzpozíciót. (Ha csak annyit ír, hogy ez az Annából és Bélából álló párra igaz, arra 1 pont jár a 2-ből. Nem kell hivatkozni a szimmetriára/azonos eloszlásra.)
(2 pont) Ez alapján $\mathbb{E}(X_{i,j}) = \frac{1}{N(N-1)}$ minden vizsgált i,j esetén, mert indikátorváltozó várható értéke az indikátornak megfelelő esemény valószínűsége. (1 pont az indoklásra, 1 pont az eredményre. Természetesen jó úgy is, hogy a várható érték definíciója szerint kiszámolja $X_{i,j}$ várható értékét, ezt nem kell nagyon részletezni, de 1 pontért legyen ott az általános képlet vagy annak a konkrét esetre vonatkozó alakja, a 0-t el lehet hagyni, a másik pont az eredményre jár.)
(1 pont) Tehát $\mathbb{E}(X) = \frac{N(N-1)}{2} \frac{1}{N(N-1)} = \frac{1}{2}$ a keresett várható érték. (Ha minden részeredmény megvolt már, akkor elég a végeredmény, de az legyen kiegyszerűsítve: 1/2 vagy 0,5.)

Ha következetesen rendezett párokkal számol, akkor a (b)-re max. 12 pont. (A rendezett párok száma éppen az ismétlés nélküli variációk száma, azaz $N(N-1)$.)