2. vizsga
Pontozási útmutató

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. a) Legyen X egy valószínűségi változó. Mit nevezünk az X sztenderdizáltjának?

   b) Mit jelent az, hogy a $T(X_1,\dots,X_n)$ statisztika *torzítatlan becslés* a háttéreloszlás egy θ paraméterére? Mit jelent az, hogy ugyanezen statisztika *aszimptotikusan torzítatlan becslés* ugyanezen paraméterre? A tanult alapstatisztikák közül melyikről bizonyítottuk az előadáson, hogy aszimpotikusan torzítatlan, de nem torzítatlan becslés valamilyen paraméterre?

   Megoldás:

   a)

   (1 pont) Ha az X várható értéke és szórása véges, továbbá $\mathbb{D}(X) > 0$ is teljesül,

   (4 pont) akkor az $\dfrac{X-\mathbb{E}(X)}{\mathbb{D}(X)}$ valószínűségi változót az X sztenderdizáltjának nevezzük.

   b)

   (2 pont) A $T(X_1,\dots,X_n)$ statisztika *torzítatlan becslés* a θ paraméterre, ha

   $$\mathbb{E}(T(X_1,\dots,X_n)) = \theta.$$

   (2 pont) A $T(X_1,\dots,X_n)$ statisztika *aszimptotikusan torzítatlan* becslés θ-ra, ha

   $$\lim_{n\to\infty} \mathbb{E}(T(X_1,\dots,X_n)) = \theta.$$

   (1 pont) A tanult alapstatisztikák közül a tapasztalati szórásnégyzet aszimptotikusan torzítatlan, de nem torzítatlan becslés a szórásnégyzetre.

2. Választunk két számot egymástól függetlenül a $(0;1)$ intervallumon. Mi a valószínűsége, hogy a második választott szám legalább akkora, mint az első szám fele, de legfeljebb akkora, mint az első szám kétszerese?

Megoldás:

(1 pont) Jelölje az első választott számot x, a másodikat pedig y. Ekkor a két szám választása tekinthető egy (x, y) pont választásának az $N = (0; 1) \times (0; 1)$ egységnégyzeten, ez tehát az eseményterünk.

(2 pont) Az

$$A = \left\{ \frac{1}{2}x \leq y \leq 2x \right\}$$

esemény valószínűségét keressük.

(1 pont) Ez a valószínűség $\dfrac{T(A)}{T(N)}$, ahol $T(\cdot)$ jelöli a területet.

(1 pont) $T(N) = 1$,

(1 pont) továbbá az A esemény által leírt terület az $y = \frac{1}{2}x$ és $y = 2x$ egyenesek közötti hegyes szögű szögtartomány N-be eső része. (Ha ez rajzon szerepel, akkor is jár a pont.)

(2 pont) Ehelyett egyszerűbb a komplementer esemény, azaz az N és az A különbségének a területét kiszámolni, ugyanis ez két egybevágó derékszögű háromszög uniója, amelynek befogói 1 ill. $\frac{1}{2}$.

(1 pont) Tehát $T(N \setminus A) = 2 \cdot \dfrac{1 \cdot \frac{1}{2}}{2} = \dfrac{1}{2}$,

(1 pont) azaz $\mathbb{P}(A) = 1 - \mathbb{P}(N \setminus A) = 1 - \dfrac{1}{2} = \dfrac{1}{2}$.

(Az utolsó 4 pont megszerezhető úgy is, ha valaki direkt módon kiszámolja az A területét, pl. két háromszögre szétbontva azt.)

3. Béla minden héten lottózik, továbbá a kedvenc száma a 42. Elhatározza, hogy a következő 52 hét mindegyikén feljegyzi magának a húzott számokat, hogy különböző statisztikákat készíthessen. Többek között azt is meg akarja figyelni, hogy hányszor húzzák ki a kedvenc számát. Mi a valószínűsége annak, hogy a következő 52 hét alatt legalább 7-szer húznak 42-est?

Megoldás:

(0 pont) Jelölje A_i azt az eseményt, hogy az i-edik héten kihúzzák a 42-t.

(1 pont) Összesen $\binom{90}{5}$-féle húzás lehetséges,

(1 pont) ebből a jó esetek azok, amikben a 42 benne van, a többi 4 szám $\binom{89}{4}$-féleképp választható, tehát

$$\mathbb{P}(A_i) = \frac{\binom{89}{4}}{\binom{90}{5}} = \frac{1}{18}.$$

(1 pont) Legyen X azon hetek száma, amikor kihúzzák a 42-t, ekkor (mivel a húzások függetlenek, és 52-szer húznak) $X \sim B(52; \frac{1}{18})$ teljesül.

(1 pont) A kérdés tehát a $\mathbb{P}(X \geq 7)$ valószínűség. (Ez a pont csak akkor jár, ha ezt egy valószínűségi változó segítségével írja fel a megoldó.)

(2 pont) A fenti egyenlőtlenségben a várható értéket mindkét oldalból kivonva, majd a szórással mindkét oldalt leosztva a

$$\mathbb{P}\left(\frac{X - \mathbb{E}(X)}{\mathbb{D}(X)} \geq \frac{7 - \mathbb{E}(X)}{\mathbb{D}(X)} \right) = \mathbb{P}\left(\frac{X - 52 \cdot \frac{1}{18}}{\sqrt{52 \cdot \frac{1}{18} \cdot \frac{17}{18}}} \geq \frac{7 - 52 \cdot \frac{1}{18}}{\sqrt{52 \cdot \frac{1}{18} \cdot \frac{17}{18}}} \right) = \mathbb{P}\left(\frac{X - \frac{26}{9}}{\frac{1}{9}\sqrt{13 \cdot 17}} \geq \frac{37}{\sqrt{13 \cdot 17}} \right)$$

kifejezés adódik.

(3 pont) A de Moivre–Laplace-tétel szerint ez közelítőleg

$$1 - \Phi\left(\frac{37}{\sqrt{221}} \right) \approx 1 - \Phi(2{,}49)$$

(1 pont)

$$\approx 1 - 0{,}9936 = 0{,}0064.$$

Ha a megooldó nem hivatkozik a de Moivre–Laplace-tételre, csak felírja a formulát, akkor a megfelelő 3 pontból egy sem jár. Ha a de Moivre–Laplace-tétel helyett a centrális határeloszlás tételére hivatkozik, de nincs a tételnek megfelelő alakba átírva a formula, akkor szintén nem jár egy sem a 3 pontból.

Egy, a centrális határeloszlás tételét használó megoldás a következő:

Az első két pont ugyanúgy szerezhető meg.

(1 pont) Ha $\mathbb{1}_{A_i}$ az A_i eseményhez tartozó indikátor valószínűségi változó, akkor a hetek száma, amikor kihúzzák a 42-t: $\sum_{i=1}^{52} \mathbb{1}_{A_i}$.

(1 pont) A kérdés tehát a $\mathbb{P}(\sum_{i=1}^{52} \mathbb{1}_{A_i} \geq 7)$ valószínűség.

(1 pont) Mivel $\mathbb{E}(\mathbb{1}_{A_i}) = \frac{1}{18}$ és $\mathbb{D}(\mathbb{1}_{A_i}) = \sqrt{\frac{1}{18} \cdot \frac{17}{18}}$ teljesül minden i-re,

(1 pont) így a fenti összeget sztenderdizálva

$$\mathbb{P}\left(\frac{\sum_{i=1}^{52} \mathbb{1}_{A_i} - 52 \cdot \frac{1}{18}}{\sqrt{52 \cdot \frac{1}{18} \cdot \frac{17}{18}}} \geq \frac{7 - 52 \cdot \frac{1}{18}}{\sqrt{52 \cdot \frac{1}{18} \cdot \frac{17}{18}}}\right) = \mathbb{P}\left(\frac{\sum_{i=1}^{52} \mathbb{1}_{A_i} - \frac{26}{9}}{\frac{1}{9}\sqrt{221}} \geq \frac{37}{\sqrt{221}}\right)$$

adódik.

(3 pont) A de Moivre–Laplace-tétel szerint ez közelítőleg

$$1 - \Phi\left(\frac{37}{\sqrt{221}}\right) \approx 1 - \Phi(2{,}49)$$

(1 pont)

$$\approx 1 - 0{,}9936 = 0{,}0064.$$

4. Az X és Y valószínűségi változók együttes eloszlását tartalmazza az alábbi táblázat. Határozzuk meg az Y-nak az X-re vett lineáris regresszióját.

Y \ X	0	1
0	2/5	1/5
2	1/10	3/10

Megoldás:

(1 pont) Az X peremeloszlása:

$$\mathbb{P}(X = 0) = \frac{2}{5} + \frac{1}{10} = \frac{5}{10} = \frac{1}{2},$$

$$\mathbb{P}(X = 1) = \frac{1}{5} + \frac{3}{10} = \frac{5}{10} = \frac{1}{2}.$$

Az Y peremeloszlása:

$$\mathbb{P}(Y = 0) = \frac{2}{5} + \frac{1}{5} = \frac{3}{5},$$

$$\mathbb{P}(Y = 2) = \frac{1}{10} + \frac{3}{10} = \frac{4}{10} = \frac{2}{5}.$$

(1 pont) Így az X várható értéke:

$$\mathbb{E}(X) = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2},$$

illetve az Y várható értéke:

$$\mathbb{E}(Y) = 0 \cdot \frac{3}{5} + 2 \cdot \frac{2}{5} = \frac{4}{5}.$$

(1 pont) Az XY várható értékének kiszámolásánál csak egyetlen 0-tól különböző tag lesz:

$$\mathbb{E}(XY) = 1 \cdot 2 \cdot \frac{3}{10} = \frac{6}{10} = \frac{3}{5},$$

(1 pont) tehát az X és Y kovarianciája:

$$\mathrm{cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\,\mathbb{E}(Y) = \frac{3}{5} - \frac{1}{2} \cdot \frac{4}{5} = \frac{1}{5}.$$

(1 pont) Az Y-nak az X-re vett lineáris regressziója egy $\beta X + \alpha$ alakú változó,

(1 pont) ahol

$$\beta = \frac{\mathrm{cov}(X,Y)}{\mathbb{D}^2(X)}, \qquad \alpha = \mathbb{E}(Y) - \beta\,\mathbb{E}(X).$$

(Ha általános képlet nem szerepel, de hibátlan helyettesítés igen, akkor is jár a pont a képletekért.)

(1 pont) Mivel $X^2 = X$, így $\mathbb{E}(X^2) = \mathbb{E}(X) = \frac{1}{2}$,

(1 pont) azaz $\mathbb{D}^2(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$

(1 pont) $= \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$, (jó helyettesítésért képlet nélkül is jár a pont)

(1 pont) tehát

$$\beta = \frac{1/5}{1/4} = \frac{4}{5}, \qquad \alpha = \frac{4}{5} - \frac{4}{5} \cdot \frac{1}{2} = \frac{2}{5},$$

és a lineáris regresszió $\dfrac{4}{5}X + \dfrac{2}{5}$.

5. Legyenek $X \sim N(3;4)$ és $Y \sim N(4;5)$ azonos valószínűségi mezőn értelmezett, független valószínűségi változók. Határozzuk meg X és Y korrelációját, valamint $X+Y$ várható értékét és szórását. Mi az $X+Y$ eloszlása?

Megoldás:

(1 pont) Mivel független változók kovarianciája 0,

(1 pont) így $\mathrm{corr}(X,Y) = \dfrac{\mathrm{cov}(X,Y)}{\mathbb{D}(X)\,\mathbb{D}(Y)} = 0$.

(1 pont) A várható érték linearitása miatt $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$

(1 pont) $= 3 + 4 = 7$.

(2 pont) Mivel X és Y függetlenek, így $\mathbb{D}^2(X+Y) = \mathbb{D}^2(X) + \mathbb{D}^2(Y)$ teljesül. (Ha a megoldó nem hivatkozik ennek a tulajdonságnak a használatánál a függetlenségre, akkor ez a 2 pont nem jár.)

(1 pont) Azaz $\mathbb{D}^2(X+Y) = 4 + 5 = 9$,

(1 pont) tehát $\mathbb{D}(X+Y) = 3$.

(1 pont) Egy, az előadáson szerepelt tétel szerint független, normális eloszlású valószínűségi változók összegének eloszlása normális,

(1 pont) a fent meghatározott paramétereket felhasználva tehát $X + Y \sim N(7;9)$.

6. Tegyük fel, hogy egy nagyvárosban a hétköznapokon történt közlekedési balesetek napi száma normális eloszlást követ, melynek szórása ismeretlen. Egymás után 10 hétköznapon feljegyeztük a balesetek számát, melyek a következők: 32, 27, 37, 42, 28, 19, 30, 26, 35, 37. Elfogadható-e 95%-os megbízhatósági szinten az a hipotézis, hogy egy hétköznap átlagosan 30 baleset történik (azaz a balesetek számának várható értéke 30). Döntsünk ugyanerről a hipotézisről 98%-os megbízhatósági szint mellett is.

Megoldás:

(2 pont) A várható értékre vonatkozó nullhipotézis eldöntéséhez (mivel a háttéreloszlás normális ismeretlen szórással) a $t(x_1, \ldots, x_{10}) = \sqrt{n} \cdot \dfrac{\overline{x} - \mu_0}{s^*}$ statisztikát kell kiszámolni, ahol $n = 10$ a minta elemszáma, $\overline{x}$ a mintaátlag, s^* a korrigált tapasztalati szórás, illetve $\mu_0 = 30$.

(1 pont) A mintaátlag:

$$\overline{x} = \frac{32 + 27 + 37 + 42 + 28 + 19 + 30 + 26 + 35 + 37}{10} = 31{,}3.$$

(1 pont) A korrigált tapasztalati szórásnégyzet $s^{*2} = \frac{n}{n-1}s^2$, ahol $s^2 = \overline{x^2} - \overline{x}^2$ a tapasztalati szórásnégyzet.

(1 pont) Itt

$$\overline{x^2} = \frac{32^2 + 27^2 + 37^2 + 42^2 + 28^2 + 19^2 + 30^2 + 26^2 + 35^2 + 37^2}{10} = 1020{,}1,$$

(1 pont) tehát $s^{*2} = \frac{10}{9}(1020,1 - 979{,}69) = 44{,}9$, és így

$$s^* = \sqrt{s^{*2}} = \sqrt{44{,}9} \approx 6{,}7007.$$

(1 pont) Ezért tehát $t(x_1, \ldots, x_{10}) \approx \sqrt{10} \cdot \frac{31{,}3-30}{6{,}7007} \approx 0{,}6135$.

Ha az általános képletek nem szerepelnek, de a hibátlan helyettesítés igen, akkor a képletekért járó pont is megadandó. Ha a megoldó számológép segítségével számolja az átlagot és a korrigált tapasztalati szórást közvetlenül a mintából (tehát a fenti helyettesítéseket nem végzi el), a korrigált tapasztalati szórás és szórásnégyzet, valamint a tapasztalati szórásnégyzet képletének (vagy egy összevont képletnek), illetve az $\overline{x^2}$ értelmezésének szerepelnie kell. Ezek bármelyikének hiányáért egyenként 1 pont, de összesen maximum 3 pont levonás jár.

(1 pont) Az elfogadási tartomány $(-t_{\varepsilon/2}(n-1); t_{\varepsilon/2}(n-1))$, ahol $t_{\varepsilon/2}(n-1)$ az $n-1$ szabadságfokú Student-eloszlás $1-\varepsilon/2$ kvantilise. Mivel $\varepsilon = 0,05$ (és $n-1=9$), így ez

$$(-2{,}262; 2{,}262).$$

(1 pont) Ebbe a statisztika értéke beleesik, tehát a nullhipotézist elfogadjuk.

(1 pont) Ha a megbízhatósági szint nő, akkor az elfogadási tartomány is: $\varepsilon = 0,02$ mellett ez a

$$(-2{,}821; 2{,}821)$$

intervallum, tehát ebben az esetben is elfogadjuk a nullhipotézist.