סטטיסטיקה ורגרסיה – כל השאלות עם פתרונות מלאים

הכנה לבגרות במתמטיקה – 4 יחידות, כיתה י״ב (שאלון 472) • 8 שאלות

בגרות קיץ תשפ״ג (2023) – מועד ב׳ • שאלה 2

2. מרצה באוניברסיטה רצתה לבדוק אם היעדרות משיעורים בקורס שנתי (המשתנה x) קשורה לינארית לציון במבחן הסופי (המשתנה y).

נתון: ממוצע ההיעדרות מן השיעורים היה $\bar{x}=10$, הציון הממוצע היה $\bar{y}=70$, ומקדם המתאם היה שלילי ($r<0$).

המרצה מצאה את משוואת ישר הרגרסיה לניבוי y מ־x.

א. אחת מן המשוואות 1–4 שלפניכם היא המשוואה שמצאה המרצה. קבעו איזו מהן היא המשוואה, ונמקו את קביעתכם.

1. $y=10x-30$
2. $y=x+60$
3. $y=-2x+70$
4. $y=-2x+90$

המרצה חישבה את סטיות התקן בעבור ההיעדרות מן השיעורים ובעבור הציונים במבחן הסופי, וקיבלה: $Sx=4$, $Sy=10$.

ב. חשבו את מקדם המתאם r.

ג. מהו מספר ההיעדרויות שישר הרגרסיה מנבא בעבורו ציון 80?

דוד, מרצה אחר, רצה לערוך את אותה בדיקה בנוגע לתלמידיו. הוא מצא את משוואת ישר הרגרסיה לניבוי y מ־x, וגילה כי בעבור כל מספר של היעדרויות – הישר מנבא תמיד את הציון 65.

ד. מצאו את ערכו של כל אחד מן המדדים שבתת־סעיפים (1)–(2) בעבור התלמידים של דוד:

(1) שיפוע ישר הרגרסיה.

(2) הציון הממוצע ($\bar{y}$).

ה. על פי הנתונים שבשאלה, האם אפשר למצוא את הממוצע של היעדרויות התלמידים של דוד? נמקו את תשובתכם.

פתרון מלא

סעיף א

שתי תכונות של ישר הרגרסיה:

- הוא עובר תמיד בנקודת הממוצעים $(\bar{x},\bar{y})=(10,70)$.
- סימן השיפוע שלו הוא סימן מקדם המתאם. כאן $r<0$, לכן השיפוע שלילי.

משוואות 1 ו־2 – שיפוע חיובי, נפסלות. נבדוק את 3 ו־4 בנקודה $x=10$:

$$(3)\quad y=-2\cdot 10+70=50\neq 70$$

$$(4)\quad y=-2\cdot 10+90=70\ \checkmark$$

רק משוואה 4 עוברת בנקודת הממוצעים ושיפועה שלילי.

תשובה: משוואה 4: $y=-2x+90$

סעיף ב

שיפוע ישר הרגרסיה הוא $m=r\cdot\frac{S_y}{S_x}$. מסעיף א, $m=-2$:

$$-2 = r \cdot \frac{10}{4} \;\Rightarrow\; r = -2 \cdot \frac{4}{10} = -0.8$$

תשובה: $r = -0.8$

סעיף ג

נציב $y = 80$ במשוואת הרגרסיה:

$$80 = -2x + 90 \;\Rightarrow\; 2x = 10$$
$$\Rightarrow\; x = 5$$

תשובה: 5 היעדרויות

סעיף ד (1)

לכל מספר של היעדרויות הישר מנבא אותו ציון, 65. כלומר משוואת ישר הרגרסיה היא $y = 65$ – ישר אופקי, ושיפועו 0.

תשובה: $m = 0$

סעיף ד (2)

ישר הרגרסיה עובר תמיד בנקודת הממוצעים $(\bar{x}, \bar{y})$. הישר הוא $y = 65$, ולכן ערך ה־y של כל נקודה עליו – ובפרט של נקודת הממוצעים – הוא 65:

$$\bar{y} = 65$$

תשובה: $\bar{y} = 65$

סעיף ה

לא. הישר $y = 65$ עובר בנקודה $(\bar{x}, 65)$ בעבור כל ערך של $\bar{x}$, כך שהנקודה הזאת לא קובעת את $\bar{x}$.

גם משוואת השיפוע לא עוזרת: $0 = r \cdot \frac{S_y}{S_x}$ נותנת רק $r = 0$ (אין קשר לינארי בין המשתנים), ואין בה מידע על הממוצע של x.

תשובה: לא – הישר $y = 65$ אופקי ועובר ב־$(\bar{x}, 65)$ לכל $\bar{x}$.

בגרות קיץ תשפ"ג (2023) – מועד מיוחד • שאלה 2

2. חקלאית בעלת מטע עצי שזיף התעניינה בקשר הלינארי שבין קוטר השזיפים (המשתנה x) ובין משקלם (המשתנה y).
היא בחרה באקראי 4 שזיפים, מדדה את הקוטר (במילימטרים) של כל אחד מהם, ומדדה את המשקל (בגרמים) של כל אחד מהם.

נמצא כי $Sy < Sx$ (סטיית התקן של משתנה הקוטר גדולה מסטיית התקן של משתנה המשקל),
וכי מקדם המתאם הוא $r = 0.8$.

לאחר מכן חישבה החקלאית את ישר הרגרסיה לניבוי משקלם של השזיפים לפי קוטרם.

א. **(1)** הביעו באמצעות Sx ו־ Sy את שיפוע ישר הרגרסיה.

(2) האם השיפוע של ישר הרגרסיה גדול מ־ 1 או קטן מ־ 1 ? נמקו את תשובתכם.

נתון כי ישר הרגרסיה לניבוי משקל השזיפים לפי קוטרם הוא $y = \frac{3}{4}x + 15$, וכי $\bar{x} = 80$.
לפניכם טבלה, ובה נתונים על המשקל של שלושה מתוך ארבעת השזיפים שנבחרו.

שזיף	המשקל (y)
א	70
ב	70
ג	80
ד	?

ב. **(1)** מצאו את המשקל של שזיף ד.

(2) מצאו את סטיית התקן של משקל השזיפים.

(3) מצאו את סטיית התקן של קוטר השזיפים.

פתרון מלא

סעיף א (1)

שיפוע ישר הרגרסיה של y על x (לפי הנוסחאון):

$$m = r \cdot \frac{S_y}{S_x}$$

נתון $r = 0.8$, לכן:

$$m = 0.8 \cdot \frac{S_y}{S_x}$$

תשובה: $m = 0.8 \cdot \frac{S_y}{S_x}$

סעיף א (2)

נתון $S_y < S_x$, ושתי סטיות התקן חיוביות, לכן:

$$\frac{S_y}{S_x} < 1$$

נכפול ב־0.8:

$$m = 0.8 \cdot \frac{S_y}{S_x} < 0.8 \cdot 1 = 0.8$$

כלומר $m < 0.8$, ובפרט $m < 1$.

תשובה: השיפוע קטן מ־1 (כי $m < 0.8$).

סעיף ב (1)

ישר הרגרסיה עובר תמיד בנקודת הממוצעים $(\bar{x}, \bar{y})$. נציב $\bar{x} = 80$:

$$\bar{y} = \frac{3}{4} \cdot 80 + 15 = 60 + 15 = 75$$

ממוצע המשקל של 4 השזיפים הוא 75 גרם, לכן סכום המשקלים:

$$4 \cdot 75 = 300$$

נסמן ב־y_4 את משקל שזיף ד:

$$70 + 70 + 80 + y_4 = 300$$
$$y_4 = 80$$

תשובה: המשקל של שזיף ד: 80 גרם

סעיף ב (2)

המשקלים: 70, 70, 80, 80, והממוצע $\bar{y} = 75$.

$$S_y = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n}}$$
$$S_y = \sqrt{\frac{(-5)^2 + (-5)^2 + 5^2 + 5^2}{4}}$$
$$S_y = \sqrt{\frac{100}{4}} = \sqrt{25} = 5$$

תשובה: $S_y = 5$ גרם

סעיף ב (3)

מישר הרגרסיה $y = \frac{3}{4}x + 15$ השיפוע הוא $m = \frac{3}{4}$. לפי סעיף א (1), ועם $S_y = 5$ מסעיף ב (2):

$$\frac{3}{4} = 0.8 \cdot \frac{5}{S_x}$$
$$\frac{3}{4} = \frac{4}{S_x}$$
$$S_x = \frac{16}{3} = 5\tfrac{1}{3} \approx 5.33$$

(בדיקה: $S_y = 5 < S_x = 5\frac{1}{3}$, כנתון.)

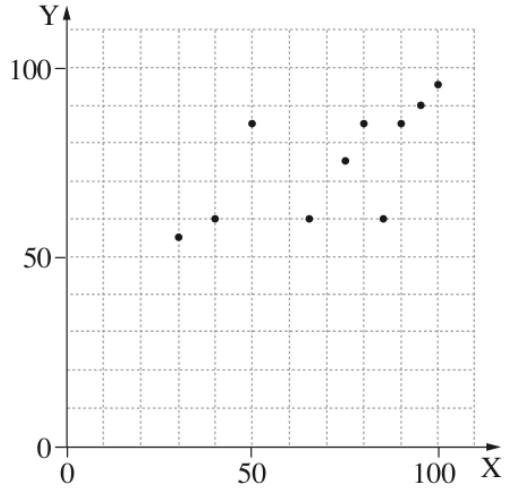
תשובה: $S_x = 5\frac{1}{3} \approx 5.33$ מילימטר

בגרות קיץ תשפ"ג (2023) – מועד א' • שאלה 2

2. תלמידי כיתה י"ב התבקשו לכתוב עבודה ולהגישה. הציון שקיבלו על העבודה שוקלל בציון הסופי של כל תלמיד.

המורה רצתה לבדוק את הקשר בין הציון על העבודה ובין הציון הסופי, ולשם כך סרטטה את דיאגרמת הפיזור של שני הציונים:

X – הציון על העבודה, Y – הציון הסופי. הדיאגרמה שהתקבלה מתוארת בתרשים שלפניכם.



א. האם אפשר להסיק מן הדיאגרמה הנתונה שכל תלמיד שקיבל על העבודה ציון גבוה יותר מתלמיד אחר קיבל **בהכרח** ציון סופי גבוה יותר מן התלמיד האחר? נמקו.

ב. אחד מן המספרים שלפניכם הוא מקדם המתאם המתאים לקשר בין שני המשתנים.

קבעו מיהו מבין המספרים האלה: 0.675 , 0 , 0.999 , -0.8 , 1.6 .

נתונים הממוצעים וסטיות התקן של שני המשתנים: $\overline{X}=71$, $S_X=23$, $\overline{Y}=75$, $S_Y=14$.

ג. מצאו את משוואת ישר הרגרסיה לניבוי הציון הסופי על פי הציון על העבודה.

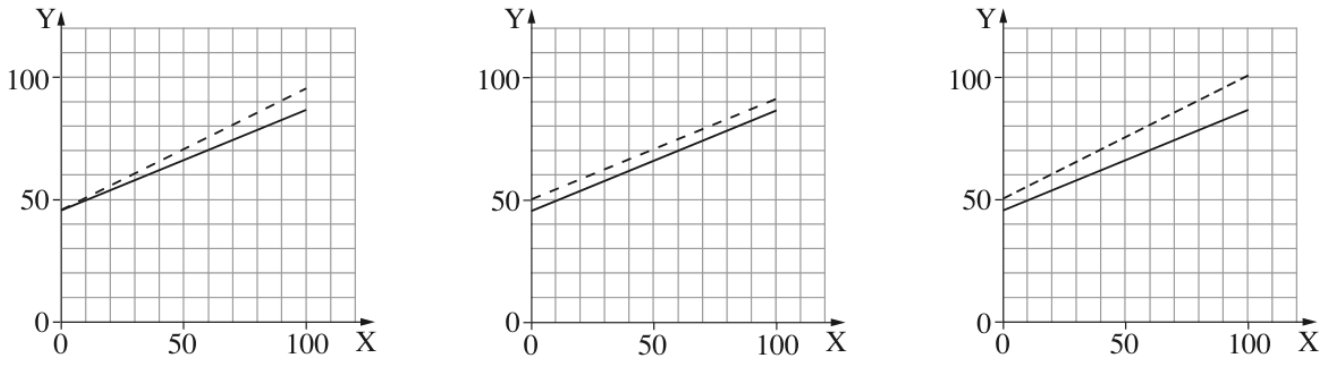
הוחלט להעלות את הציון הסופי של כל תלמיד ב־ 5 נקודות, ובעקבות העלאה זו התקבל ישר רגרסיה חדש.

ד. **(1)** האם השתנתה סטיית התקן S_Y לאחר העלאת הציונים?

(2) אחד מן הגרפים I–III שבסוף השאלה מייצג את הישר הישן, שלפני העלאת הציון הסופי (מסורטט בקו מלא), ואת הישר החדש, שאחרי ההעלאה (מסורטט בקו מקווקו).

קבעו מיהו הגרף, ונמקו.

ה. אם קיים בכיתה תלמיד שהציון שלו על העבודה הוא 71 , מה צריך להיות הציון הסופי שלו (לאחר ההעלאה בציון) כדי שהנקודה המייצגת את שני הציונים שלו תהיה על ישר הרגרסיה החדש?



III **II** **I**

פתרון מלא

סעיף א

לא. בדיאגרמה יש תלמיד שקיבל 50 על העבודה ו־85 בציון הסופי, ותלמיד אחר שקיבל 65 על העבודה (יותר) אבל רק 60 בציון הסופי (פחות).

הדיאגרמה מראה מגמה כללית של קשר חיובי, אבל לא כלל שמתקיים לכל זוג תלמידים.

תשובה: לא

למשל: תלמיד עם 50 על העבודה קיבל 85 בציון הסופי, ותלמיד עם 65 קיבל רק 60.

סעיף ב

נפסול את מה שלא מתאים:

- 1.6 – מקדם המתאם תמיד בין -1 ל־1, לכן לא ייתכן.
- -0.8 – הקשר בדיאגרמה חיובי (ככל ש־X גדל, Y נוטה לגדול), לכן המקדם חיובי.
- 0 – יש מגמה ברורה, לכן הקשר אינו אפסי.
- 0.999 – קשר כמעט מושלם, כלומר כל הנקודות כמעט על ישר אחד – אבל הנקודות פזורות למדי.

לכן מקדם המתאם הוא 0.675 – קשר חיובי בינוני.

תשובה: $r = 0.675$

סעיף ג

שיפוע ישר הרגרסיה של Y על פי X:

$$m = r \cdot \frac{S_y}{S_x} = 0.675 \cdot \frac{14}{23}$$

$$\approx 0.411$$

הישר עובר דרך נקודת הממוצעים $(\bar{x}, \bar{y}) = (71, 75)$:

$$y - \bar{y} = m(x - \bar{x})$$

$$y - 75 = 0.411(x - 71)$$

$$y = 0.411x + 45.83$$

שימו לב: בפתרון של גבע מעגלים קודם את השיפוע ל־0.41 ומקבלים $y = 0.41x + 45.89$. ההבדל נובע רק מהעיגול.

תשובה: $y = 0.411x + 45.83$ **(בעיגול השיפוע ל־0.41:** $y = 0.41x + 45.89$**)**

סעיף ד (1)

לא. כל הציונים הסופיים עלו באותו מספר נקודות (5), ולכן כל הנקודות זזות יחד, הממוצע עולה ב־5 והמרחק של כל ציון מהממוצע לא משתנה.

סטיית התקן מודדת את הפיזור סביב הממוצע, ולכן היא לא משתנה: $S_y = 14$.

תשובה: לא, S_y נשארת 14.

סעיף ד (2)

אחרי ההעלאה: $\bar{y} = 80$, ו־S_y, S_x, $\bar{x}$ לא משתנים. גם מקדם המתאם r לא משתנה (הוספת קבוע לכל הציונים לא משנה את צורת הפיזור).

לכן השיפוע $m = r \cdot \frac{S_y}{S_x}$ לא משתנה – הישר החדש מקביל לישר הישן, ומוזז 5 יחידות כלפי מעלה:

$$y = 0.411x + 50.83$$

בגרף II הקו המקווקו מקביל לקו המלא ונמצא מעליו (בהפרש קבוע של 5). בגרפים I ו־III הקווים אינם מקבילים.

תשובה: גרף II – השיפוע לא משתנה, הישר מוזז 5 יחידות למעלה.

סעיף ה

ישר רגרסיה תמיד עובר דרך נקודת הממוצעים. אחרי ההעלאה נקודת הממוצעים היא $(71, 80)$.

לכן לתלמיד שקיבל 71 על העבודה (בדיוק הממוצע), הנקודה תהיה על הישר אם הציון הסופי שלו הוא 80.
(בדיקה: $0.411 \cdot 71 + 50.83 \approx 80$.)

תשובה: 80

בגרות חורף תשפ"ג (2023) • שאלה 2

2. במחקר שנועד לבדוק את ההשפעה של קורס לגמילה מעישון, השתתפו 9 מעשנים שנכחו בקורס.

בעבור כל אחד מהם נמדדה צריכת הסיגריות היומית לפני הקורס (X), וצריכת הסיגריות היומית לאחר שבוע מתחילת הקורס (Y).

הנתונים שהתקבלו מפורטים בטבלה שלפניכם:

צריכת הסיגריות היומית לפני הקורס – X	5	15	20	22	22	25	28	30	40
צריכת הסיגריות היומית לאחר שבוע מתחילת הקורס – Y	3	10	19	20	22	22	24	30	30

א. חשבו את ממוצע צריכת הסיגריות היומית למשתתף במחקר לפני הקורס ואת ממוצע צריכת הסיגריות היומית למשתתף במחקר לאחר שבוע מתחילת הקורס.

ב. סרטטו דיאגרמת פיזור של Y כתלות ב־X (כל משבצת במחברת מייצגת 2 סיגריות ליום).

ג. על פי הדיאגרמה שסרטטתם, בחרו במקדם המתאם המתאים לנתונים מבין המספרים שלפניכם, ורשמו אותו:

$r = 1$, $r = 0.949$, $r = 0.212$, $r = -0.633$.

נתונות סטיות התקן $S_X = 9.226$, $S_Y = 8.26$.

ד. מצאו את משוואת ישר הרגרסיה לניבוי Y על פי X .

ה. אלעד מעשן 21 סיגריות ביום. הוא רוצה להשתתף בקורס לגמילה מעישון. על פי ישר הרגרסיה שמצאתם, מהו הניבוי לצריכת הסיגריות היומית שלו לאחר שבוע מתחילת הקורס?

פתרון מלא

סעיף א

ממוצע הצריכה לפני הקורס:

$$\Sigma x = 5 + 15 + 20 + 22 + 22$$
$$+25 + 28 + 30 + 40 = 207$$
$$\bar{x} = \frac{207}{9} = 23$$

ממוצע הצריכה אחרי שבוע:

$$\Sigma y = 3 + 10 + 19 + 20 + 22$$
$$+22 + 24 + 30 + 30 = 180$$
$$\bar{y} = \frac{180}{9} = 20$$

תשובה: לפני הקורס: 23 סיגריות ביום
אחרי שבוע: 20 סיגריות ביום.

סעיף ב

כל משתתף הוא נקודה (x, y): $(5, 3)$, $(15, 10)$, $(20, 19)$, $(22, 20)$, $(22, 22)$, $(25, 22)$, $(28, 24)$, $(30, 30)$, $(40, 30)$.

דיאגרמת פיזור של Y כתלות ב־X עם 9 נקודות במגמה עולה

תשובה: ראו דיאגרמה.

סעיף ג

הנקודות בדיאגרמה מסודרות במגמה עולה ברורה וקרובות לישר, אבל לא כולן על ישר אחד בדיוק.

- $r = 1$ – כל הנקודות על ישר אחד, וזה לא המצב.
- $r = -0.633$ – קשר שלילי, אבל המגמה עולה.
- $r = 0.212$ – קשר חיובי חלש מאוד, אבל הנקודות קרובות מאוד לישר.

לכן $r = 0.949$ – קשר חיובי חזק.

תשובה: $r = 0.949$

סעיף ד

$$m = r \cdot \frac{S_y}{S_x} = 0.949 \cdot \frac{8.26}{9.226}$$
$$\approx 0.85$$

הישר עובר דרך נקודת הממוצעים $(\bar{x}, \bar{y}) = (23, 20)$:

$$y - 20 = 0.85(x - 23)$$
$$y = 0.85x + 0.45$$

תשובה: $y = 0.85x + 0.45$

סעיף ה

נציב $x = 21$ במשוואת הרגרסיה:

$$y = 0.85 \cdot 21 + 0.45 = 18.3$$

תשובה: כ־18.3 סיגריות ביום (כ־18).

בגרות קיץ תשפ"ב (2022) – מועד ב' • שאלה 2

2. מנהל חברה בדק את הקשר בין ההוצאה החודשית של החברה על פרסום מוצריה, ובין ההכנסות מן המכירות שלה בחודש שלאחר מכן. הוא בדק נתונים מכמה חודשים רצופים.

בטבלה שלפניכם מוצגים הנתונים על ההוצאות וההכנסות באלפי שקלים:

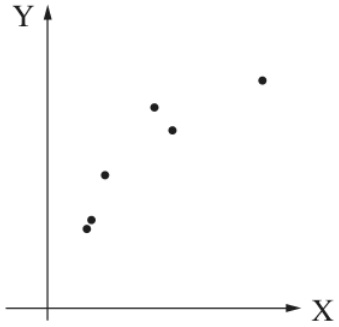
הוצאה על פרסום (X)	9	10	13	24	28	48
הכנסות ממכירות בחודש שלאחר מכן (Y)	180	200	300	450	400	510

המנהל חישב ומצא כי ממוצע ההוצאות לחודש על פרסום הוא $\overline{X} = 22$ וסטיית התקן היא $Sx = 13.6$.

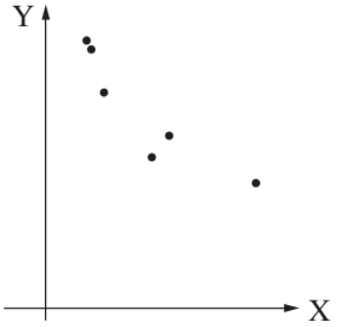
ממוצע ההכנסות ממכירות לחודש הוא $\overline{Y} = 340$ וסטיית התקן היא $Sy = 123.4$.

לפניכם שתי דיאגרמות פיזור (1)–(2). אחת מהן מתארת את הקשר בין שני המשתנים (X ו־ Y).

דיאגרמה 1



דיאגרמה 2



א. מבין שתי הדיאגרמות (1)–(2), מהי הדיאגרמה המתארת את הקשר בין שני המשתנים? נמקו את תשובתכם.

נתונים ארבעה מקדמי מתאם: $r = -0.7$, $r = -0.9$, $r = 1$, $r = 0.9$.

ב. אחד מארבעת מקדמי המתאם מתאים לנתונים. קבעו מיהו ונמקו את קביעתכם.

ג. מהי משוואת ישר הרגרסייה לניבוי ההכנסות ממכירות כתלות בהוצאה על פרסום?

ד. על פי ישר הרגרסייה שמצאתם, מהי ההערכה להכנסות ממכירות (באלפי שקלים) בעבור הוצאה של 19,000 שקלים לחודש על פרסום? נמקו את תשובתכם.

החברה המירה את ההוצאות וההכנסות בשקלים להוצאות והכנסות בדולרים (ולכן כל המספרים בטבלה קטנו בערך פי 3).

ה. מהי ההשפעה של המרת השקלים לדולרים על כל אחד מן הגדלים (1)–(3) שלפניכם (כלומר האם הוא גדל, קטן, או לא השתנה)?

נמקו את תשובותיכם.

(1) $\overline{X}$

(2) Sx

(3) r

פתרון מלא

סעיף א

בטבלה, כאשר ההוצאה על פרסום (X) גדלה, ההכנסות (Y) גדלות בדרך כלל: 180, 200, 300, 450, 400, 510. זה קשר חיובי – הנקודות עולות משמאל לימין.

בדיאגרמה 1 הנקודות עולות, ובדיאגרמה 2 הן יורדות. לכן דיאגרמה 1 מתארת את הקשר.

תשובה: דיאגרמה 1 – הקשר חיובי (כש־X גדל, Y גדל בדרך כלל).

סעיף ב

- הקשר חיובי (סעיף א), לכן $r > 0$: נפסלים $r = -0.7$ ו־$r = -0.9$.
- $r = 1$ פירושו שכל הנקודות נמצאות בדיוק על ישר אחד. זה לא המצב: למשל כאשר X עולה מ־24 ל־28, Y יורד מ־450 ל־400.

לכן $r = 0.9$ – קשר חיובי חזק אך לא מושלם.

(בדיקה בחישוב מהטבלה: $r \approx 0.90$.)

תשובה: $r = 0.9$

סעיף ג

שיפוע ישר הרגרסיה:

$$m = r \cdot \frac{S_y}{S_x} = 0.9 \cdot \frac{123.4}{13.6}$$

$$\approx 8.166$$

הישר עובר בנקודת הממוצעים $(22, 340)$:

$$y - 340 = 8.166\,(x - 22)$$

$$y = 8.166x + 160.348$$

תשובה: $y = 8.166x + 160.348$

סעיף ד

הנתונים בטבלה באלפי שקלים, לכן הוצאה של 19,000 שקלים היא $x = 19$:

$$y = 8.166 \cdot 19 + 160.348 \approx 315.5$$

ההערכה מתבססת על ישר הרגרסיה, כי מקדם המתאם $r = 0.9$ גבוה (קשר לינארי חזק), ו־19 נמצא בתחום הערכים של X בטבלה (בין 9 ל־48).

תשובה: כ־315.5 אלפי שקלים (כ־315,500 ש"ח)

סעיף ה (1)

כל הערכים בטבלה קטנים בערך פי 3. אם כל הערכים של X מחולקים ב־3, גם הממוצע מתחלק ב־3:

$$\bar{X} \approx \frac{22}{3} \approx 7.33$$

תשובה: $\bar{X}$ קטן (בערך פי 3).

סעיף ה (2)

המרחקים של הערכים מהממוצע קטנים גם הם פי 3, ולכן גם סטיית התקן קטנה פי 3:

$$S_x \approx \frac{13.6}{3} \approx 4.53$$

תשובה: S_x קטנה (בערך פי 3).

סעיף ה (3)

חלוקת כל הערכים במספר חיובי משנה רק את קנה המידה של הצירים. צורת דיאגרמת הפיזור ועוצמת הקשר הלינארי לא משתנות, ולכן מקדם המתאם נשאר $r = 0.9$.

תשובה: r לא משתנה.

בגרות קיץ תשפ"ב (2022) – מועד א' • שאלה 2

2. כדי לבדוק את הקשר בין ציוני בחינת הבגרות במתמטיקה (X) ובין ציוני בחינת הבגרות במדעי המחשב (Y), חושבו בשנה מסוימת הממוצעים, סטיות התקן ומקדם המתאם של הציונים בבחינות האלה. תוצאות החישובים היו:

$\overline{\mathrm{X}} = 64$, $\overline{\mathrm{Y}} = 72$, $\mathrm{s_x} = 14$, $\mathrm{s_y} = 9$, $\mathrm{r} = 0.77$.

א. מצאו את משוואת ישר הרגרסיה לניבוי הציון במדעי המחשב על פי הציון במתמטיקה.

באותה השנה דני נבחן במתמטיקה ובמדעי המחשב, והציון שקיבל במתמטיקה היה 90 .

ב. על פי ישר הרגרסיה שמצאתם, מהו הציון המשוער של דני בבחינת הבגרות במדעי המחשב?

לאחר שהתברר שהבחינה במתמטיקה באותה השנה הייתה קשה, הוסיפו ל**כל אחד** מן הציונים במתמטיקה 4 נקודות.

ג. מהי ההשפעה של התוספת לציונים על כל אחד מן הגדלים (1)–(3) שלפניכם (כלומר, האם הוא גדל, קטן או לא השתנה)?

נמקו את תשובותיכם.

(1) $\overline{\mathrm{X}}$

(2) $\mathrm{s_x}$

(3) r

בשנה שאחרי כן התקבלו בבחינות במתמטיקה ובמדעי המחשב אותם הממוצעים ואותן סטיות התקן כמפורט בתחילת השאלה.

ידוע שמשוואת ישר הרגרסיה של הציונים בשנה זו הייתה: $\mathrm{y = mx + 43.2}$.

ד. **(1)** מצאו את הערך של m .

(2) מצאו את מקדם המתאם של הציונים בשנה זו.

פתרון מלא

סעיף א

הנתונים: $\bar{x} = 64,\ S_x = 14$ (מתמטיקה), $\bar{y} = 72,\ S_y = 9$ (מדעי המחשב), $r = 0.77$.

שיפוע קו הרגרסיה של y על פי x:

$$m = r \cdot \frac{S_y}{S_x} = 0.77 \cdot \frac{9}{14} = 0.495$$

קו הרגרסיה עובר בנקודת הממוצעים $(64,\ 72)$:

$$y - \bar{y} = m(x - \bar{x})$$
$$y - 72 = 0.495(x - 64)$$
$$y - 72 = 0.495x - 31.68$$
$$y = 0.495x + 40.32$$

תשובה: $y = 0.495x + 40.32$

סעיף ב

נציב את הציון של דני במתמטיקה, $x = 90$, במשוואת קו הרגרסיה:

$$y = 0.495 \cdot 90 + 40.32$$

$$y = 84.87 \approx 85$$

תשובה: הציון המשוער: $84.87 \approx 85$

סעיף ג (1)

כשמוסיפים לכל ציון את אותו מספר, גם הממוצע עולה באותו מספר:

$$\bar{x}_{new} = 64 + 4 = 68$$

הממוצע **גדל** ב־4 נקודות.

תשובה: גדל ב־4: $\bar{x} = 68$

סעיף ג (2)

תוספת קבועה מזיזה את כל הציונים ימינה באותו מרחק, אבל המרחקים של הציונים מהממוצע ($x_i - \bar{x}$) לא משתנים. לכן הפיזור לא השתנה.

סטיית התקן **לא השתנתה**: $S_x = 14$.

תשובה: לא השתנתה: $S_x = 14$

סעיף ג (3)

מקדם המתאם תלוי בסטיות $x_i - \bar{x}$, $y_i - \bar{y}$ ובסטיות התקן:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N \cdot S_x \cdot S_y}$$

כל ציון במתמטיקה וגם הממוצע עלו ב־4, ולכן הסטיות $x_i - \bar{x}$ לא השתנו. גם S_x וציוני מדעי המחשב לא השתנו.

לכן מקדם המתאם **לא השתנה**: $r = 0.77$.

תשובה: לא השתנה: $r = 0.77$

סעיף ד (1)

הממוצעים לא השתנו, ולכן קו הרגרסיה $y = mx + 43.2$ עובר בנקודת הממוצעים $(64,\ 72)$. נציב:

$$72 = m \cdot 64 + 43.2$$

$$28.8 = 64m$$

$$m = 0.45$$

תשובה: $m = 0.45$

סעיף ד (2)

סטיות התקן לא השתנו ($S_x = 14,\ S_y = 9$). מהקשר בין השיפוע למקדם המתאם:

$$m = r \cdot \frac{S_y}{S_x}$$

$$0.45 = r \cdot \frac{9}{14}$$

$$r = 0.45 \cdot \frac{14}{9} = 0.7$$

תשובה: $r = 0.7$

בגרות חורף נבצרים תשפ"ב (2022) • שאלה 2

2. במחקר על הקשר בין טמפרטורת המים X לבין כמות המדוזות Y , נאספו 30 זוגות נתונים והתקבלו התוצאות הבאות: $\bar{X} = 26.5^\circ$, $S_X = 3.5^\circ$, $\bar{Y} = 14$, $S_Y = 2.5$

נערך ניבוי בעזרת ישר רגרסיה לכמות המדוזות לפי טמפרטורת המים.

א. מה יהיה הניבוי לכמות המדוזות כאשר טמפרטורת המים היא הטמפרטורה הממוצעת 26.5° ?

נתון בנוסף כי מקדם המתאם הוא $r = 0.8$.

ב. לפניך שלושה היגדים (1)–(3). קבע אלו מההיגדים נכונים. נמק.

(1) מקדם המתאם לא מאפשר ניבוי טוב כי הוא קטן מ- 1 .

(2) ככל שהטמפרטורה יורדת, ננבא כמות גדולה יותר של מדוזות.

(3) ככל שהטמפרטורה עולה, ננבא כמות גדולה יותר של מדוזות.

ג. מצא את ישר הרגרסיה לניבוי כמות המדוזות לפי הטמפרטורה.

ד. על פי ישר הרגרסיה, מה יהיה הניבוי לכמות המדוזות ביום שבו טמפרטורת המים היא 33° ?

פתרון מלא

סעיף א

קו הרגרסיה תמיד עובר בנקודת הממוצעים $(\bar{x},\ \bar{y})$.

לכן כאשר הטמפרטורה היא הטמפרטורה הממוצעת $\bar{x} = 26.5^\circ$, הניבוי לכמות המדוזות הוא הממוצע שלהן: $\bar{y} = 14$.

תשובה: 14 מדוזות

סעיף ב

(1) מקדם המתאם תמיד בין -1 ל־1, ולכן ההיגד „קטן מ־1" לא אומר שהניבוי לא טוב. $r = 0.8$ קרוב ל־1, כלומר יש מתאם קווי חיובי חזק, והניבוי טוב. ההיגד **אינו נכון**.

(2) $r = 0.8 > 0$ – מתאם חיובי: ככל ש־X קטן, גם Y צפוי לקטון. כלומר כשהטמפרטורה יורדת נצפה כמות **קטנה** יותר של מדוזות. ההיגד **אינו נכון**.

(3) מתאם חיובי: ככל שהטמפרטורה עולה, ננבא כמות גדולה יותר של מדוזות. ההיגד **נכון**.

תשובה: רק היגד (3) נכון.

סעיף ג

שיפוע קו הרגרסיה:

$$m = r \cdot \frac{S_y}{S_x} = 0.8 \cdot \frac{2.5}{3.5} = \frac{4}{7}$$
$$\approx 0.571$$

הקו עובר בנקודת הממוצעים $(26.5,\ 14)$:

$$y - \bar{y} = m(x - \bar{x})$$
$$y - 14 = \frac{4}{7}(x - 26.5)$$
$$y - 14 = \frac{4}{7}x - 15\frac{1}{7}$$
$$y = \frac{4}{7}x - 1\frac{1}{7}$$

(בעיגול: $y \approx 0.571x - 1.143$.)

תשובה: $y = \frac{4}{7}x - 1\frac{1}{7} \approx 0.571x - 1.143$

סעיף ד

נציב $x = 33$ במשוואת קו הרגרסיה:

$$y = \frac{4}{7} \cdot 33 - 1\frac{1}{7}$$

$$y = 18\frac{6}{7} - 1\frac{1}{7} = 17\frac{5}{7} \approx 17.71$$

כלומר ננבא כ־18 מדוזות.

תשובה: $17\frac{5}{7} \approx 17.71$**, כלומר כ־18 מדוזות.**

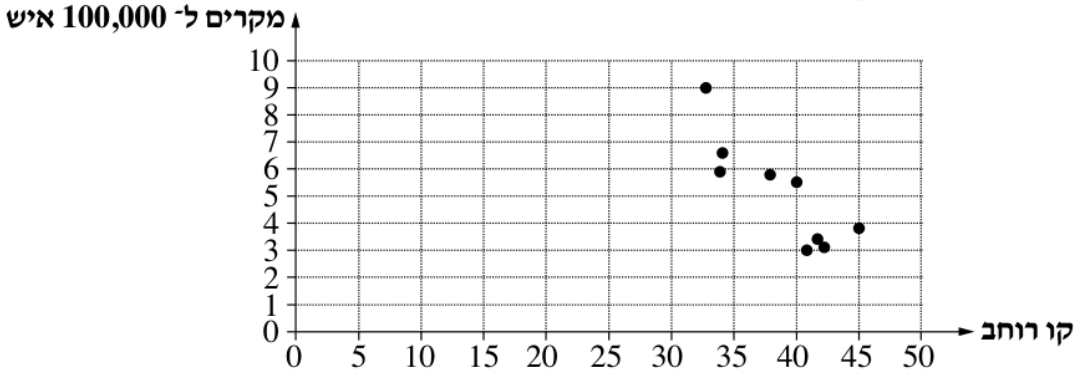
בגרות חורף תשפ"ב (2022) • שאלה 2

2. חוקרים בדקו אם יש קשר בין קו הרוחב שבו נמצא אזור מסוים ובין שיעור מקרי סרטן העור (מלנומה) באוכלוסייה באותו אזור. המחקר נערך בתשעה אזורים שונים בארצות הברית, במשך שלוש שנים. בכל אחד מן האזורים נבדקו 100,000 תושבים. נתוני המחקר מוצגים בטבלה שלפניך.

	קו רוחב – x	**שיעור מקרי המלנומה – y** (מקרים ל־ 100,000)
	32.8	9
	33.9	5.9
	34.1	6.6
	37.9	5.8
	40.0	5.5
	40.8	3.0
	41.7	3.4
	42.2	3.1
	45.0	3.8
ממוצע	$\bar{x} = 38.71$	$\bar{y} = 5.12$
סטיית תקן	$s_x = 4.04$	s_y

א. הראה כי סטיית התקן של שיעור מקרי המלנומה היא $s_y = 1.88$.

לפניך דיאגרמת הפיזור של y כתלות ב־ x .



ב. אחד מבין מקדמי המתאם (1)–(4) שלפניך מייצג את הקשר בין הנתונים. קבע איזה מהם מייצג את הקשר, ונמק את קביעתך.

(1) $r = 1$

(2) $r = -0.857$

(3) $r = 0.651$

(4) $r = -1$

ג. מצא את משוואת קו הרגרסיה לניבוי מקרי המלנומה y על פי קו הרוחב x .

ד. מהו הניבוי לשיעור מקרי המלנומה y בקו רוחב 36 ?

פתרון מלא

סעיף א

נשתמש בממוצע הנתון $\bar{y} = 5.12$ ונחשב את סכום ריבועי הסטיות:

y_i	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
9	3.88	15.0544
5.9	0.78	0.6084
6.6	1.48	2.1904
5.8	0.68	0.4624
5.5	0.38	0.1444
3	2.12-	4.4944
3.4	1.72-	2.9584
3.1	2.02-	4.0804
3.8	1.32-	1.7424

$$\sum (y_i - \bar{y})^2 = 31.7356$$

$$S_y = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n}} = \sqrt{\frac{31.7356}{9}}$$

$$S_y = \sqrt{3.526} \approx 1.88$$

תשובה: $S_y \approx 1.88$ **(מקרים ל־100,000 תושבים)**

סעיף ב

בדיאגרמת הפיזור הנקודות „יורדות" משמאל לימין: ככל שקו הרוחב גדל, שיעור מקרי המלנומה קטן. זה קשר ליניארי שלילי, לכן מקדם המתאם שלילי – נשארות האפשרויות (2) ו־(4).

$r = -1$ פירושו שכל הנקודות נמצאות בדיוק על קו ישר. בדיאגרמה הנקודות אינן על קו ישר אחד, ולכן $r \neq -1$.

הקשר בכל זאת חזק למדי (הנקודות קרובות לקו יורד), ולכן מתאים $r = -0.857$.

תשובה: (2) $r = -0.857$

סעיף ג

שיפוע קו הרגרסיה:

$$m = r \cdot \frac{S_y}{S_x} = -0.857 \cdot \frac{1.88}{4.04}$$

$$m \approx -0.399$$

הקו עובר בנקודת הממוצעים $(38.71,\ 5.12)$:

$$y - \bar{y} = m(x - \bar{x})$$

$$y - 5.12 = -0.399(x - 38.71)$$

$$y - 5.12 = -0.399x + 15.445$$

$$y = -0.399x + 20.57$$

תשובה: $y = -0.399x + 20.57$

סעיף ד

נציב $x = 36$ במשוואת קו הרגרסיה:

$$y = -0.399 \cdot 36 + 20.57$$

$$y = -14.364 + 20.57 \approx 6.21$$

שימו לב: התוצאה רגישה לעיגול. בלי לעגל את השיפוע ואת החותך מקבלים $y \approx 6.20$; בפתרון של גבע מופיע 6.21 ובתשובות של מלומד 6.19 (וחותך 20.56) – כולם „בערך 6.2".

תשובה: בערך 6.2 מקרים ל־100,000 תושבים ($y \approx 6.21$**)**

הפתרונות נכתבו בעזרת בינה מלאכותית ונבדקו, אך ייתכנו טעויות – ללא אחריות. מומלץ להשוות לפתרון הרשמי.