יואל גבע
לא מתפשרים.

# 4 יח"ל

# רגרסיה

# פתרונות לבחינות בגרות

התוכנית החדשה במתמטיקה

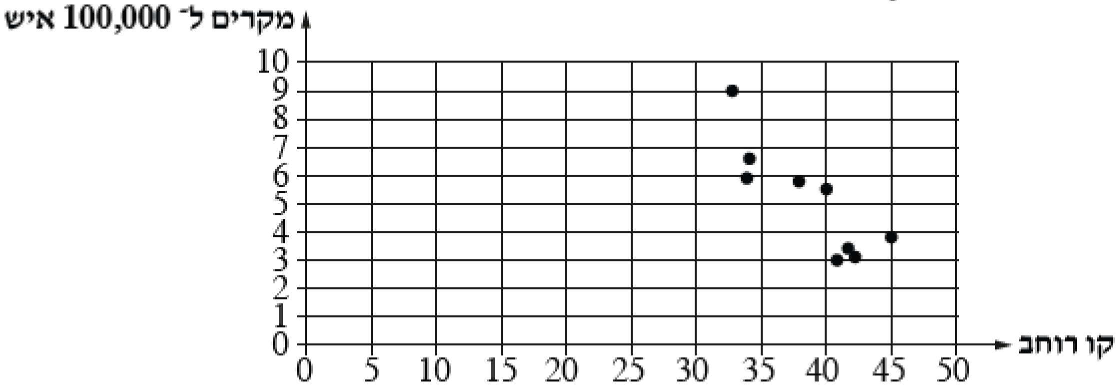
נפתר ע"י עפר ילין

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 ?

**א. נתוני המחקר שבדק את הקשר בין קו הרוחב שבו נמצא אזור מסוים ( $x$ ),**

**ובין שיעור מקרי סרטן העור (מלנומה) באוכלוסייה באותו אזור ( $y$ ) מוצגים בטבלה שלפנינו.**

**האוכלוסייה שנבדקה היא אזורים שונים בארצות הברית, במשך שלוש שנים.**

**נוסיף לטבלה עמודות שתעזורנה בחישוב סטיית התקן של שיעור מקרי המלנומה.**

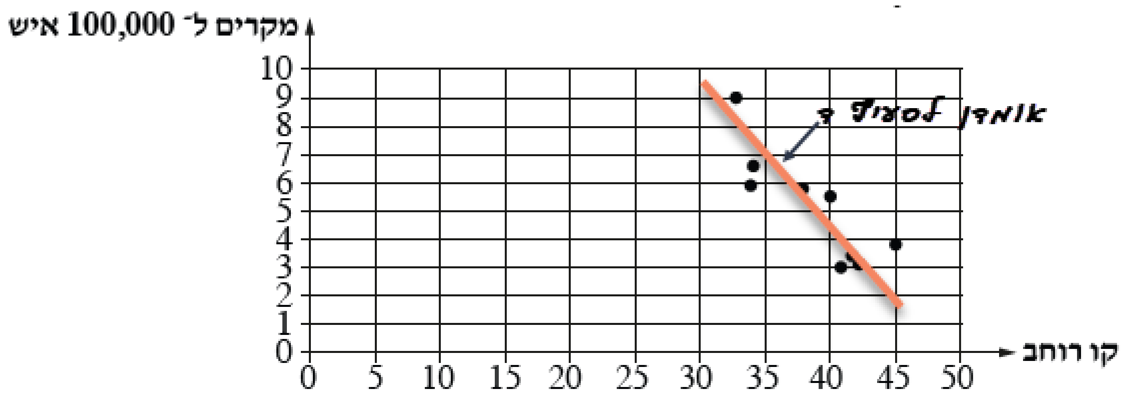
| $x$ | $y$ | $y-\bar{y}$ | $(y-\bar{y})^2$ |
|---|---|---|---|
| 32.8 | 9 | 3.88 | 15.0544 |
| 33.9 | 5.9 | 0.78 | 0.6084 |
| 34.1 | 6.6 | 1.48 | 2.1904 |
| 37.9 | 5.8 | 0.68 | 0.4624 |
| 40.0 | 5.5 | 0.38 | 0.1444 |
| 40.8 | 3.0 | −2.12 | 4.4944 |
| 41.7 | 3.4 | −1.72 | 2.9584 |
| 42.2 | 3.1 | −2.02 | 4.0804 |
| 45.0 | 3.8 | −1.32 | 1.7424 |
| $\bar{x}=38.71$ | $\bar{y}=5.12$ | | 31.7356 |
| $s_x=4.04$ | $s_y$ | | |

$$S_y=\sqrt{\frac{31.7356}{9}}=\sqrt{3.5262}$$

$$\boxed{S_y=1.8778\approx1.88}$$

**תשובה: הראינו כי סטיית התקן של שיעור מקרי המלנומה היא** $S_y=1.88$ **.**

ב. דיאגרמת הפיזור הבאה, מראה את $y$ כתלות של $x$.




ניתן לראות, די בבירור, שקיים קשר ליניארי די חזק ויורד בין קו הרוחב, לבין שיעור מקרי הסרטן.
הקו האדום שהוסף, מעין אומדן לקו הרגרסיה, יורד משמאל לימין.

בהתאם רק מקדם מתאם, $r$, שלילי – אפשרי, ונראה שיהיה סביב $-0.9 < r < -0.8$ (הערכה בלבד !).

כיוון שהנקודות בדיאגרמת הפיזור, בבירור, אינן ממוקמות על קו ישר,

אזי לא אפשרי $r = -1$, שהוא מקדם מתאר דטרמיניסטי.

עדיין, ניתן לראות שקיים קשר ליניארי די חזק, ולכן מקדם מתאם של $r = -0.857$ בהחלט מתאים.

תשובה: מקדם המתאם $r = -0.857$ מייצג את הקשר שבין הנתונים.

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

נמצא תחילה את שיפוע קו הרגרסיה: $m = r \cdot \frac{s_y}{s_x} = -0.857 \cdot \frac{1.88}{4.04} = -0.3988 \approx -0.399$.

נמצא את משוואת קו הרגרסיה, שעובר בנקודת הממוצעים:

$$y - \bar{y} = m(x - \bar{x})$$
$$y - 5.12 = -0.399(x - 38.71)$$
$$y - 5.12 = -0.399x + 15.45$$
$$\boxed{y = -0.399x + 20.57}$$

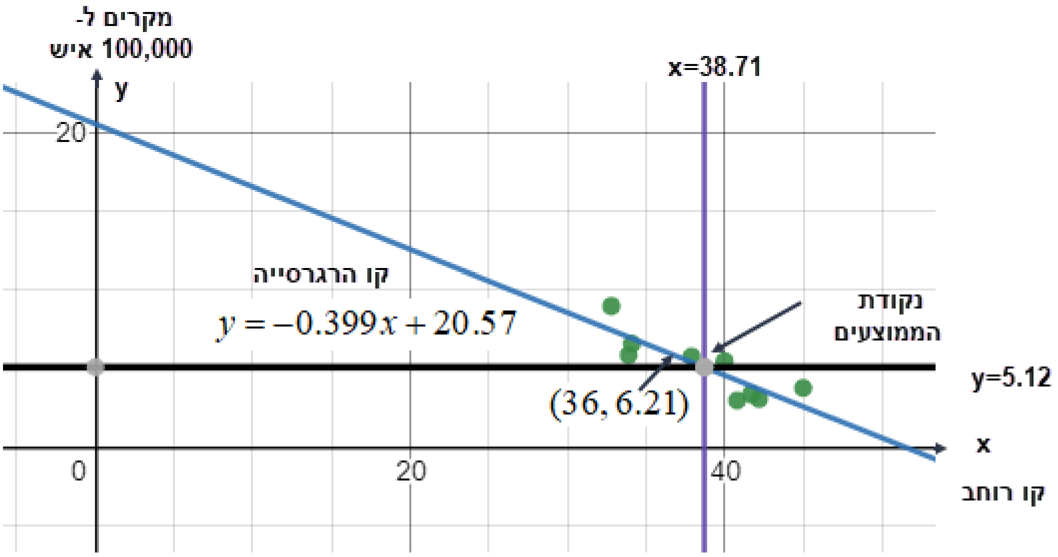
תשובה: משוואת קו הרגרסיה, לניבוי מקרי המלנומה $y$ על פי קו הרוחב $x$, היא $y = -0.399x + 20.57$.

ד. נמצא מהו הניבוי לשיעור מקרי המלנומה $y$ בקו רוחב $36$. (בשרטוט מעל, בחץ, אומדן ראשוני)

נציב $x = 36$, במשוואת קו הרגרסיה: $y = -0.399 \cdot 36 + 20.57 = 6.21$.

<u>תשובה</u>: הניבוי לשיעור מקרי המלנומה, בקו רוחב $36$, הוא בערך $6.21$ מקרים ל- $100,000$ תושבים.

# העשרה



מקרים ל-
100,000 איש
y
x=38.71
20
קו הרגרסייה
$y = -0.399x + 20.57$
נקודת
הממוצעים
y=5.12
(36, 6.21)
x
0
20
40
קו רוחב

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

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 $33°$ ?

**א. המחקר בדק את הקשר בין טמפרטורת המים ( $x$ ), ובין כמות המדוזות ( $y$ ) .**

**האוכלוסייה שנבדקה היא 30 זוגות נתונים.**

**התוצאות שהתקבלו הן: ממוצע טמפרטורת המים $\bar{x}=26.5°$ , עם סטיית תקן של $S_x=3.5°$,**

**וממוצע כמות המדוזות $\bar{y}=14$ , עם סטיית תקן של $S_y=2.5$ מדוזות.**

**הניבוי הטוב ביותר לכמות המדוזות, כאשר טמפרטורת המים היא הטמפרטורה הממוצעת ( $\bar{x}=26.5°$ ),**

**הוא הממוצע של כמות המדוזות ( $\bar{y}=14$ ).**

**נזכור, כמובן, שגם קו הרגרסיה עובר בנקודת הממוצעים.**

**תשובה: הניבוי לכמות המדוזות, כאשר טמפרטורת המים היא הטמפרטורה הממוצעת $26.5°$ ,**

**יהיה 14 מדוזות.**

**ב. נתון כי מקדם המתאם הוא $r=0.8$ .**

**זהו מקדם הקרוב למקדם הדטרמיניסטי $(r=1)$ , ולכן הניבוי יהיה ניבוי טוב.**

**כיוון שמקדם המתאר הוא חיובי, אז ככל שהמשתנה המנבא X יגדל, גם המשתנה Y יגדל.**

**ומכאן, שככל שהטמפרטורה תעלה, גם צפוי שכמות המדוזות תגדל.**

**תשובה: היגד (3) נכון – ככל שהטמפרטורה עולה, ננבא כמות גדולה יותר של מדוזות.**

**ג. נמצא את משוואת קו הרגרסיה לניבוי כמות המדוזות $y$ לפי הטמפרטורה $x$ .**

**נמצא תחילה את שיפוע קו הרגרסיה:** $m=r\cdot\frac{s_y}{s_x}=0.8\cdot\frac{2.5}{3.5}=\frac{4}{7}$ .

**נמצא את משוואת קו הרגרסיה, שעובר בנקודת הממוצעים:**

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

$$y-14=\frac{4}{7}(x-26.5)$$

$$y-14=\frac{4}{7}x-15\frac{1}{7}$$

$$\boxed{y=\frac{4}{7}x-1\frac{1}{7}}$$

**תשובה: משוואת קו הרגרסיה, לניבוי כמות המדוזות $y$ לפי הטמפרטורה $x$ , היא** $y=\frac{4}{7}x-1\frac{1}{7}$ .

**ד. נמצא מהו הניבוי לשיעור לכמות המדוזות $y$ ביום שבו טמפרטורת המים היא $33°$ .**

**נציב $x=33$ , במשוואת קו הרגרסיה:** $y=\frac{4}{7}\cdot 33-1\frac{1}{7}=17\frac{5}{7}\approx 18$ .

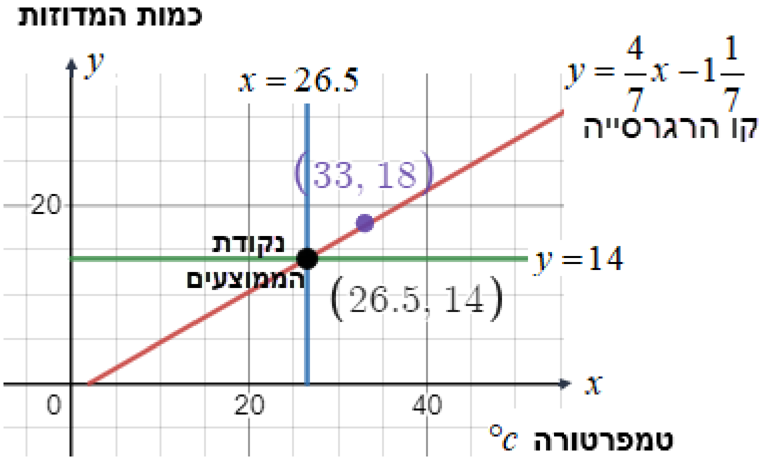
**<u>הערה</u> – בפתרון סעיף זה, חבויה הנחה שטמפרטורת מים של $33°$ הייתה בתחום,**

**שבו התקבלו נתונים על טמפרטורות המים.**

**לאור סטיית התקן של $3.5°$ , וממוצע של $26.5°$ , הרי שהנחה זו היא סבירה.**

**<u>תשובה</u>: הניבוי לכמות המדוזות, ביום שבו טמפרטורת המים היא $33°$, הוא בערך $18$ מדוזות.**

# העשרה



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:

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

**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) $\bar{X}$

(2) $s_x$

(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: $y = mx + 43.2$ .

**ד.** **(1)** מצאו את הערך של $m$ .

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

**א. המחקר בדק את הקשר בין ציוני בחינת הבגרות במתמטיקה ( $x$ ),**

**ובין ציוני בחינת הבגרות במדעי המחשב ( $y$ ) .**

**האוכלוסייה שנבדקה היא ציוני תלמידים בשנה מסוימת.**

**התוצאות שהתקבלו הן: ממוצע הציונים במתמטיקה** $\bar{x}=64$ **, עם סטיית תקן של** $S_x=14$ **,**

**ממוצע הציונים במדעי המחשב** $\bar{y}=72$ **, עם סטיית תקן של** $S_y=9$ **,**

**ומקדם המתאם הוא** $r=0.77$ **(חיובי, ודי גבוה).**

**נמצא את משוואת קו הרגרסיה לניבוי הציון במדעי המחשב** $y$ **, על פי הציון במתמטיקה** $x$ **.**

**נמצא תחילה את שיפוע קו הרגרסיה:** $m=r\cdot\frac{s_y}{s_x}=0.77\cdot\frac{9}{14}=0.495$ .

**נמצא את משוואת קו הרגרסיה, שעובר בנקודת הממוצעים** $(64, 72)$ **:**

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

**תשובה: משוואת קו הרגרסיה, לניבוי הציון במדעי המחשב** $y$ **, על פי הציון במתמטיקה** $x$ **,**

**היא** $y=0.495x+40.32$ .

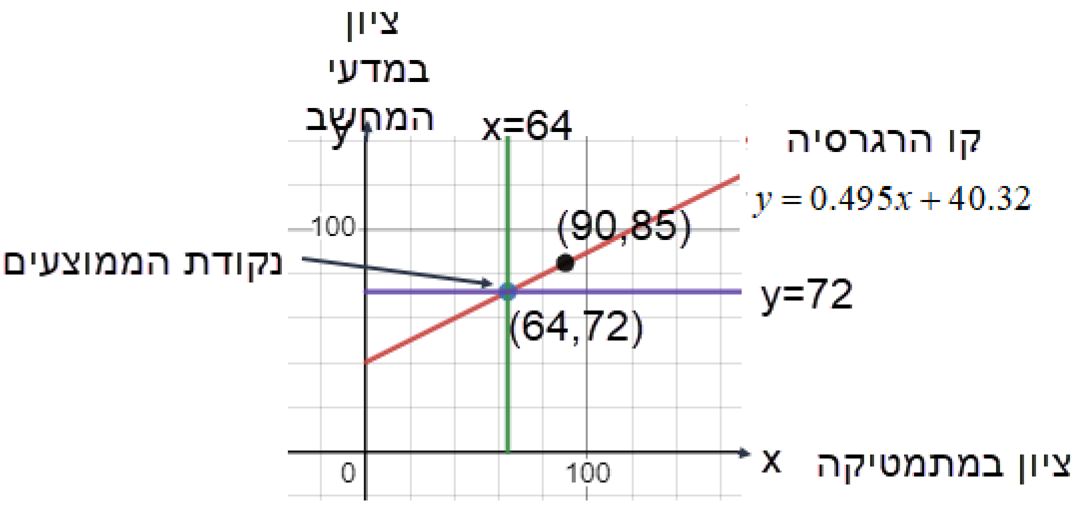
**ב. נמצא מהו הציון** $(y)$ **המשוער של דני בבחינת הבגרות במדעי המחשב,**

**כאשר נתון שציונו, באותה שנה בבחינת הבגרות במתמטיקה, היה** $90$ **.**

**נציב** $x=90$ **, במשוואת קו הרגרסיה:** $y=0.495\cdot 90+40.32=84.87\approx 85$ .

**תשובה: הציון המשוער של דני, בבחינת הבגרות במדעי המחשב, הוא בערך** $85$.

## **הערה**






יואל גבע

**ג. לאחר שהתברר שהבחינה במתמטיקה, באותה השנה הייתה קשה,**

**הוסיפו לכל אחד מן הציונים במתמטיקה $4$ נקודות.**

**(1) תוספת של קבוע, לכל אחד מן הנתונים, מעלה את הממוצע בדיוק באותו קבוע.**

**תשובה: ממוצע הציונים במתמטיקה עלה ב- $4$ נקודות (והגיע ל- $\overline{x}=68$).**

**(2) התוספת הקבועה, לכל אחד מן הנתונים, מזיזה את עקומת ההתפלגות ימינה,**

**אולם אינה משנה את הפיזור ובהתאם סטיית התקן אינה משתנה.**

**תשובה: $S_x$, סטיית התקן של ציוני הבחינה במתמטיקה, לא השתנתה, ונותרה $14$ ($S_x=14$).**

**(3) מקדם המתאם שווה לממוצע מכפלות ציוני התקן.**

$$r=\frac{1}{N\cdot S_x\cdot S_y}\left((x_1-\overline{x})(y_1-\overline{y})+\ldots+(x_N-\overline{x})(y_N-\overline{y})\right)$$

**עם השינוי בציוני הבחינה במתמטיקה, עלה גם הממוצע באותו שינוי של $4$ נקודות,**

**ולכן הסטייה $(x_N-\overline{x})$ נותרה ללא שינוי.**

**כיוון שבנוסף לא היה שינוי בציוני הבחינה במדעי המחשב, או במספר הנתונים,**

**הרי שמקדם המתאם $(r)$ נותר ללא שינוי.**

**תשובה: $r$, מקדם המתאם, לא השתנה, ונותר $0.77$ ($r=0.77$).**

**ד. בשנה שלאחר מכן התקבלו בשתי הבחינות אותם הממוצעים וסטיות התקן, כמפורט בתחילת השאלה:**

**$\overline{x}=64$, עם סטיית תקן של $S_x=14$, $\overline{y}=72$, עם סטיית תקן של $S_y=9$.**

**ידוע שמשוואת ישר הרגרסיה של הציונים בשנה זו הייתה $y=mx+43.2$.**

**(1) כיוון שקו הרגרסיה עובר בנקודת הממוצעים $(64,72)$, שלא השתנתה, נציב את שיעוריה במשוואה.**

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

$$28.8=64m \quad /:64$$

$$\boxed{m=0.45}$$

**תשובה: הערך של $m$ הוא $0.45$.**

**(2) נמצא את מקדם המתאם $(r)$, של הציונים בשנה זו.**

$$m=r\cdot\frac{s_y}{s_x}$$

$$0.45=r\cdot\frac{9}{14} \quad /:(\frac{9}{14})$$

$$\boxed{r=0.7}$$

**תשובה: מקדם המתאם של הציונים בשנה זו הוא $0.7$ ($r=0.7$).**

# 4

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:

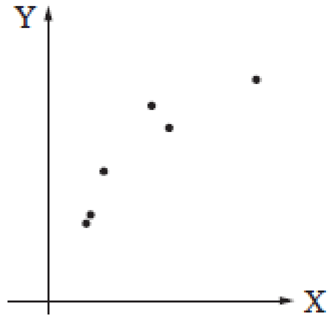
| | | | | | | |
|---|---|---|---|---|---|---|
| הוצאה על פרסום (X) | 9 | 10 | 13 | 24 | 28 | 48 |
| הכנסות ממכירות בחודש שלאחר מכן (Y) | 180 | 200 | 300 | 450 | 400 | 510 |

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

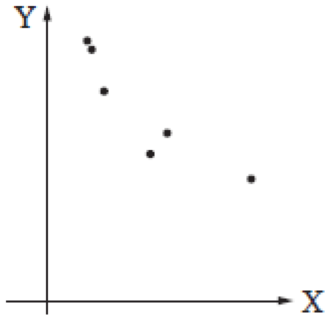
ממוצע ההכנסות ממכירות לחודש הוא $\overline{Y}=340$ וסטיית התקן היא $Sy=123.4$ .

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) $\overline{X}$

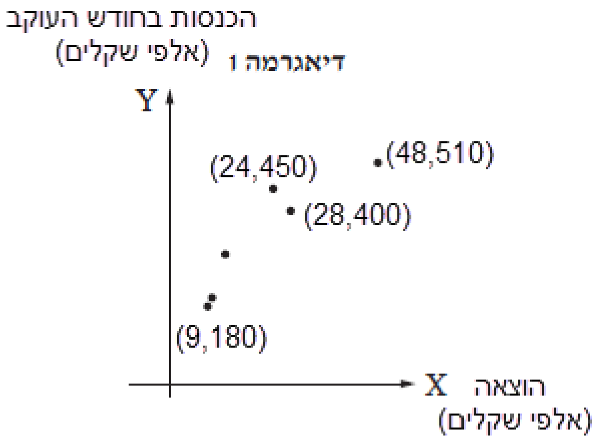
(2) $Sx$

(3) $r$

א. מנהל חברה בדק את הקשר בין ההוצאה החודשית של החברה על פרסום מוצריה ($x$),
ובין ההכנסות מן המכירות שלה בחודש שלאחר מכן ($y$). הנתונים הם באלפי שקלים.
האוכלוסייה שנבדקה היא נתונים מכמה חודשים רצופים.

| | | | | | | |
|---|---|---|---|---|---|---|
| ההוצאה על פרסום ($x$) | 9 | 10 | 13 | 24 | 28 | 48 |
| הכנסות ממכירות בחודש שלאחר מכן ($y$) | 180 | 200 | 300 | 450 | 400 | 510 |

התוצאות שהתקבלו הן: ממוצע ההוצאות לחודש על פרסום $\bar{x}=22$ , עם סטיית תקן של $S_x=13.6$,
ממוצע ההכנסות ממכירות לחודש הוא $\bar{y}=340$ , עם סטיית תקן של $S_y=123.4$,
דיאגרמה (1) מתאימה לנתונים שבטבלה (שיעורי נקודות הוספו לציור להמחשה),
שמראים כי כאשר ההוצאה גדלה, גדלות גם ההכנסות בחודש שלאחר מכן (למעט חריגה אחת).




תשובה: דיאגרמה (1) מתארת את הקשר בין שני המשתנים.

ב. כפי שאמרנו, ניתן לראות שקיימת מגמה של עלייה, ולכן נצפה לראות מקדם מתאם חיובי.
ניתן גם לראות שאין קו ישר, שעליו תהיינה כל הנקודות (יש גם חודש שבו הייתה ירידה),
ולכן מקדם המתאם אינו דטרמיניסטי $(r=1)$ , ומקדם מתאם $r=0.9$ נראה מתאים ביותר.
תשובה: $r=0.9$ הוא מקדם מתאם, שנראה כי מתאים לנתונים.

**ג. נמצא את משוואת קו הרגרסיה לניבוי ההכנסה בחודש העוקב $y$ , על פי ההוצאה על פרסום $x$.**

**נמצא תחילה את שיפוע קו הרגרסיה:** $m = r \cdot \frac{s_y}{s_x} = 0.9 \cdot \frac{123.4}{13.6} \approx 8.166$ .

**נמצא את משוואת קו הרגרסיה, שעובר בנקודת הממוצעים** $(22, 340)$ :

$$y - \bar{y} = m(x - \bar{x})$$
$$y - 340 = 8.166(x - 22)$$
$$y - 340 = 8.166x - 179.652$$
$$\boxed{y = 8.166x + 160.348}$$

**תשובה: משוואת קו הרגרסיה, לניבוי ההכנסות ממכירות , כתלות בהוצאה על פרסום,**
**היא** $y = 8.166x + 160.348$ .

**ד. נמצא מהי ההכנסה** $(y)$ **המשוערת (באלפי שקלים), בהתאם להוצאה של** $19{,}000$ **שקלים.**

**נשים לב שנתון של** $19$ **הוא בתחום של נתוני ההוצאות שנמדדו, ולכן ניתן להעריך את ההכנסה.**

**נציב** $x = 19$ **, במשוואת קו הרגרסיה:** $y = 8.166 \cdot 19 + 160.348 = 315.502 \approx 315.5$ .

**תשובה: ההערכה להכנסות ממכירות, בעבור הוצאה של** $19{,}000$ **שקלים לחודש על פרסום,**
**היא בערך** $315.5$ **אלפי שקלים.**

**ה. החברה המירה את ההוצאות וההכנסות בשקלים להוצאות והכנסות בדולרים,**
**כאשר המשמעות היא <u>שכל הנתונים שבטבלה</u> קטנו בערך פי** $3$ .

**(1) חילוק פי מספר קבוע, לכל אחד מן הנתונים, מקטינה את הממוצע בדיוק פי אותו קבוע.**

**תשובה: ממוצע ההוצאות לפרסום ירד פי** $3$ **(והגיע ל-** $\bar{x} = 7\frac{1}{3}$**).**

**(2) החילוק בקבוע, של כל אחד מן הנתונים, מקטינה את הפיזור של האוכלוסייה פי אותו קבוע.**

**תשובה:** $S_x$ **, סטיית התקן של ההוצאה לפרסום קטנה פי** $3$ **(**$S_x = 4\frac{8}{15}$**).**

**(3) מקדם המתאם שווה לממוצע מכפלות ציוני התקן.**

$$r = \frac{1}{N \cdot S_x \cdot S_y}\left((x_1 - \bar{x})(y_1 - \bar{y}) + \ldots + (x_N - \bar{x})(y_N - \bar{y})\right)$$

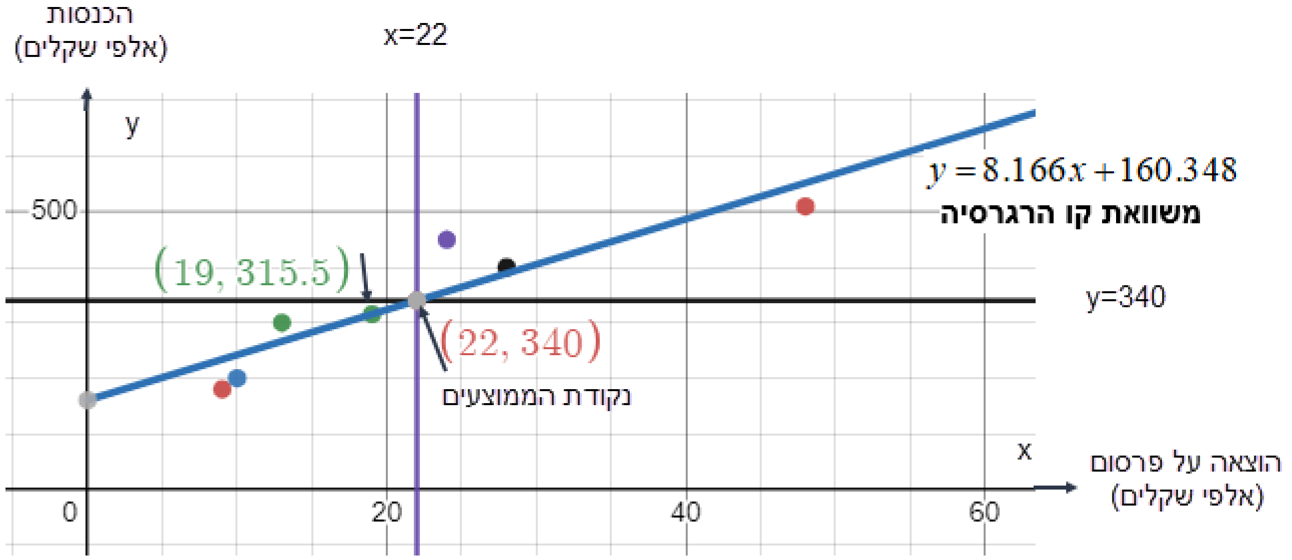
**עם השינוי בשיטת הרישום, קטנו גם הסטייה מהממוצע, של** $(x_N - \bar{x})$ **ו-** $(y_N - \bar{y})$ **פי** $3$,

**ובמקביל קטנו גם שתי סטיות התקן פי** $3$ **, ולכן הן המונה והן המכנה קטנו פי** $9$ **,**

**ומקדם המתאם** $(r)$ **נותר ללא שינוי.**

**תשובה:** $r$ **, מקדם המתאם, לא השתנה** **(**$r = 0.9$**).**

# הערשה


הכנסות
(אלפי שקלים)
x=22
y
500
$y = 8.166x + 160.348$
משוואת קו הרגרסיה
(19, 315.5)
y=340
(22, 340)
נקודת הממוצעים
x
הוצאה על פרסום
(אלפי שקלים)
0
20
40
60

# 5

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?

א. המחקר נועד לבדוק את הקשר בין צריכת הסיגריות היומית לפני קורס גמילה לעישון $(x)$,
ובין צריכת הסיגריות היומית לאחר שבוע מתחילת הקורס $(y)$.

האוכלוסייה שנבדקה היא 9 מעשנים שנכחו בקורס.

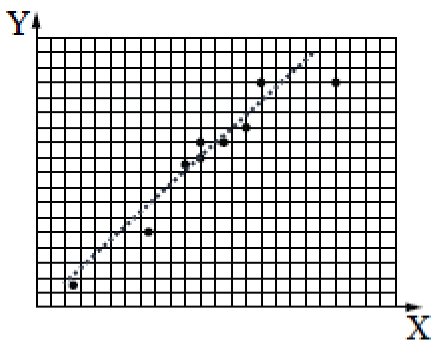
| צריכת הסיגריות היומיות לפני הקורס $(x)$ | 5 | 15 | 20 | 22 | 22 | 25 | 28 | 30 | 40 |
|---|---|---|---|---|---|---|---|---|---|
| צריכת הסיגריות היומיות לאחר שבוע מתחילת הקורס $(y)$ | 3 | 10 | 19 | 20 | 22 | 22 | 24 | 30 | 30 |

ממוצע הצריכה לפני הקורס הוא: $\bar{x}=\frac{40+30+28+25+22+22+20+15+5}{9}=\frac{207}{9}=23$

הממוצע לאחר שבוע מתחילת הקורס הוא: $\bar{y}=\frac{30+30+24+22+22+20+19+10+3}{9}=\frac{180}{9}=20$

תשובה: ממוצע צריכת הסיגריות למשתתף במחקר לפני הקורס הוא 23 סיגריות, $\bar{x}=23$.
ממוצע צריכת הסיגריות למשתתף לאחר שבוע מתחילת הקורס הוא 20 סיגריות, $\bar{y}=20$.

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




תשובה: הסרטוט מעל (הערה - כולל אפשרות לקו הניבוי הלא רגרסיבי, קו ציוני התקן).

ג. ניתן לראות שקיימת מגמה של עלייה, ולכן נצפה לראות מקדם מתאם חיובי.
הנקודות מפוזרות בצורה מהודקת סביב קשר ליניארי עולה,
ולכן מקדם המתאם צפוי להיות חיובי חזק (מעל חצי)
ניתן גם לראות שאין קו ישר, שעליו תהיינה כל הנקודות
(למשל, ניתן לראות שיש צריכה ראשונית 22 שעבורה שתי צריכות חדשות שונות,
ושתי צריכות חדשות 22, שעבורן שתי צריכות ראשוניות שונות),
ולכן מקדם המתאם אינו דטרמיניסטי $(r=1)$,

ומקדם מתאם $r=0.949$ נראה מתאים ביותר, כי הנקודות מסודרות מאוד יפה סביב קשר ליניארי עולה.

תשובה: $r=0.949$ הוא מקדם מתאם, שנראה כי מתאים לנתונים.

**ד. נמצא את משוואת קו הרגרסיה לניבוי צריכת הסיגריות היומית לאחר שבוע מתחילת הקורס $(y)$,**

**על פי צריכת הסיגריות היומית לפני קורס גמילה לעישון $(x)$,**

**התוצאות שהתקבלו הן: צריכת סיגריות יומית לפני הקורס: $\bar{x}=23$ , עם סטיית תקן של $S_x=9.226$,**

**צריכת סיגריות יומית לאחר שבוע מתחילת הקורס: $\bar{y}=20$ , עם סטיית תקן של $S_y=8.26$,**

**כאשר מקדם המתאם הוא $r=0.949$ .**

**נמצא תחילה את שיפוע קו הרגרסיה:** $m=r\cdot\frac{s_y}{s_x}=0.949\cdot\frac{8.26}{9.226}\approx0.85$ .

**נמצא את משוואת קו הרגרסיה, שעובר בנקודת הממוצעים** $(23,20)$ **:**

$$y-\bar{y}=m(x-\bar{x})$$
$$y-20=0.85(x-23)$$
$$y-20=0.85x-19.55$$
$$\boxed{y=0.85x+0.45}$$

**תשובה: משוואת קו הרגרסיה, לניבוי $y$ , על פי $x$, היא $y=0.85x+0.45$ .**

**ה. נמצא מהי צריכת הסיגריות המשוערת של אלעד לאחר שבוע מתחילת הקורס $(y)$**

**בהתאם לצריכת הסיגריות היומית שלו, $21$ סיגריות ביום.**

**נשים לב שנתון של $21$ סיגריות לפני הקורס הוא בתחום של הנתונים שנמדדו,**

**ולכן ניתן להעריך את צריכת הסיגריות לאחר שבוע מתחילת הקורס.**

**נציב $x=21$, במשוואת קו הרגרסיה:** $y=0.85\cdot21+0.45=18.3\approx18$ .

**<u>דרך פתרון נוספת</u> היא ללא חישוב משוואת קו הרגרסיה.**

**נחשב את ציון התקן של מס' הסיגריות של אלעד לפני הקורס:** $z=\frac{x-\bar{x}}{s_x}=\frac{21-23}{9.226}=-0.2168$ .

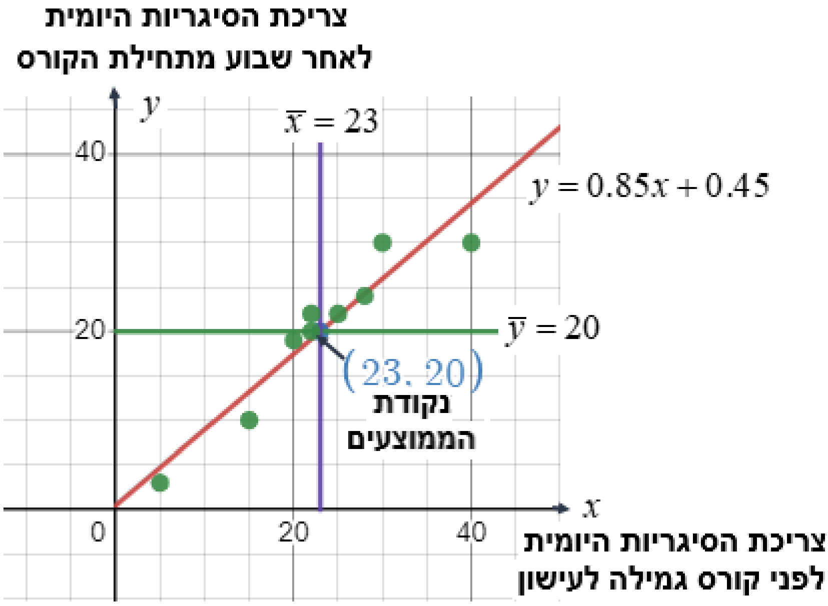
**נחשב את ציון התקן, לאחר הקורס,**

**על ידי הכפלת ציון התקן במקדם המתאם:** $z=-0.2168\cdot0.949=-0.2057$ .

**נציב מחדש במשוואה של ציון התקן:** $z=\frac{y-\bar{y}}{s_y}\ \to -0.2057=\frac{y-20}{8.26}\ \to y=18.3$ .

**תשובה: הניבוי לצריכת הסיגריות היומית של אלעד, לאחר שבוע מתחילת הקורס הוא $18.3\approx18$ סיגריות.**

# העשרה

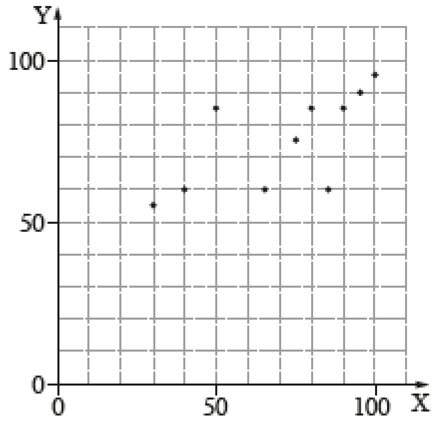


צריכת הסיגריות היומית
לאחר שבוע מתחילת הקורס
y
40
20
0
20
40
x
$\bar{x} = 23$
$y = 0.85x + 0.45$
$\bar{y} = 20$
(23, 20)
נקודת
הממוצעים
צריכת הסיגריות היומית
לפני קורס גמילה לעישון

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.

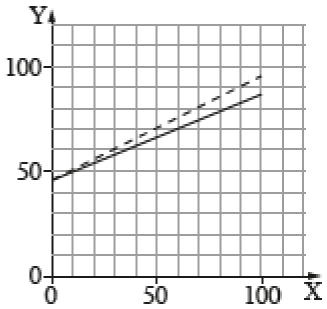
הוחלט להעלות את הציון הסופי של כל תלמיד ב־ 5 נקודות, ובעקבות העלאה זו התקבל ישר רגרסיה חדש.

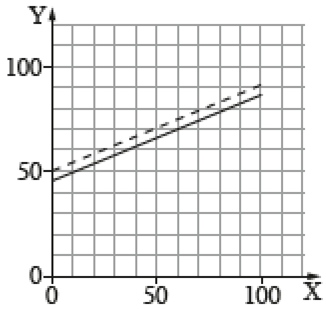
**ד.** **(1)** האם השתנתה סטיית התקן $S_Y$ לאחר העלאת הציונים?

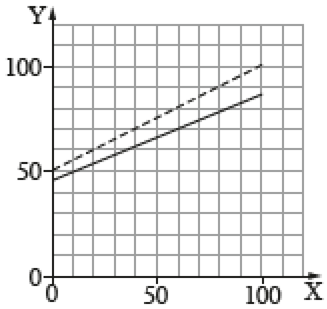
**(2)** אחד מן הגרפים I–III שבסוף השאלה מייצג את הישר הישן, שלפני העלאת הציון הסופי (מסורטט בקו מלא), ואת הישר החדש, שאחרי ההעלאה (מסורטט בקו מקווקו).

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?

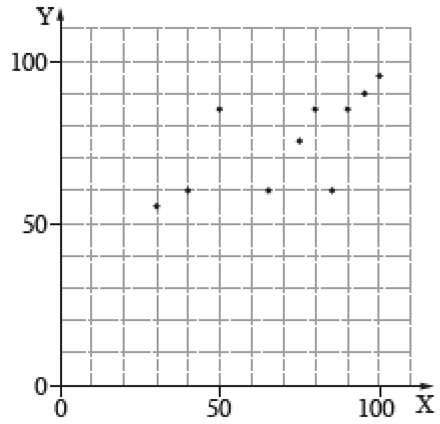

III


II


I

**א. המורה רצתה לבדוק את הקשר בין הציון על העבודה $(x)$, ובין הציון הסופי $(y)$ .**

**האוכלוסייה שנבדקה היא $10$ תלמידי כיתה י"ב.**




**ניתן לראות שישנם שלושה תלמידים שקיבלו ציון סופי של $60$ ,**

**למרות שקיבלו ציונים שונים על העבודה שהגישו.**

**תשובה: לא ניתן להסיק מן הדיאגרמה הנתונה שכל תלמיד, שקיבל על העבודה ציון גבוה יותר מתלמיד אחר,**

**קיבל בהכרח <u>ציון סופי</u> גבוה יותר מן התלמיד האחר.**

**ב. ניתן לראות שקיימת מגמה של עלייה, ולכן נצפה לראות מקדם מתאם חיובי.**

**הנקודות מפוזרות בצורה די מהודקת סביב קשר ליניארי עולה,**

**ולכן מקדם המתאם צפוי להיות חיובי חזק (מעל חצי).**

**ניתן גם לראות שאין קו ישר, שעליו תהיינה כל הנקודות, כפי שהסברנו בסעיף א,**

**ולכן מקדם המתאם אינו דטרמיניסטי $(r=1)$ או קרוב אליו.**

**לכן, מקדם מתאם $r=0.675$ נראה מתאים ביותר, כי הנקודות מסודרות די יפה סביב קשר ליניארי עולה.**

**תשובה: $r=0.675$ הוא מקדם מתאם, המתאים לקשר בין המשתנים.**

**ג. נמצא את משוואת קו הרגרסיה לניבוי הציון הסופי $(y)$, על פי הציון הסופי של העבודה $(x)$,**

**נתונים הממוצעים וסטיות התקן: $\bar{x}=71$ , $S_x=23$ , $\bar{y}=75$ , $S_y=14$ ,**

**כאשר מקדם המתאם הוא $r=0.675$ .**

**נמצא תחילה את שיפוע קו הרגרסיה: $m=r\cdot\frac{s_y}{s_x}=0.675\cdot\frac{14}{23}\approx 0.41$ .**

**נמצא את משוואת קו הרגרסיה, שעובר בנקודת הממוצעים $(71,75)$ :**

$$y-\bar{y}=m(x-\bar{x})$$
$$y-75=0.41(x-71)$$
$$y-75=0.41x-29.11$$
$$\boxed{y=0.41x+45.89}$$

**תשובה: משוואת קו הרגרסיה, לניבוי הציון הסופי על פי הציון הסופי של העבודה, היא $y=0.41x+45.89$ .**

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

**(1) בעקבות העלאה בקבוע של כל אחד מהציונים הסופיים $(y)$,**

**לא משתנה הפיזור של הציונים הסופיים, ולכן לא משתנה סטיית התקן $S_y$ של משתנה זה.**

**תשובה: סטיית התקן $S_y$ לא השתנתה, לאחר העלאת הציונים.**

**(2) מקדם המתאם הוא ממוצע של מכפלות ציוני התקן.**

**ההעלאה בקבוע של הציון הסופי, תעלה באותו קבוע גם את הממוצע של משתנה זה, ל- $\bar{y}=80$,**

**ולכן לא תשתנה הסטייה $y-\bar{y}$, וגם לא ישתנה ציון התקן, ומכאן שמקדם המתאם לא ישתנה.**

**(לסיכום, מקדם המתאם לא משתנה בהוספת קבוע, או כל שינוי ליניארי אחר.)**

**כתוצאה מכך, גם $m=r\cdot\frac{s_y}{s_x}$ לא ישתנה, ולמעשה נקבל תזוזה אנכית $5$ יחידות מעלה של קו הרגרסיה,**

**ונקבל קו רגרסיה המקביל לקו המקורי.**

**תשובה: גרף $\mathrm{II}$ מייצג את הישר הישן (מסורטט בקו מלא) ואת הישר החדש (מסורטט בקו מקווקו).**

**ה. משוואת קו הרגרסיה, לאחר העלאת הציון הסופי של כל העבודות, היא $y=0.41x+50.89$.**

**נבדוק מה צריך להיות הציון הסופי של תלמיד בכיתה זו, שהציון שלו על העבודה הוא $71$.**

**נשים לב שציון זה הוא בתחום של הנתונים שנמדדו, ולכן ניתן להעריך את הציון הסופי שלו.**

**כיוון שציון שלו על העבודה שווה לממוצע $\bar{x}=71$, אז הציון הסופי יהייה שווה ל- $\bar{y}=80$ החדש,**

**והנקודה המתאימה תהייה נקודת הממוצעים החדשה $(71,80)$.**

**<u>דרך פתרון נוספת</u>**

**נציב $x=71$, במשוואת קו הרגרסיה החדש: $y=0.41\cdot 71+50.89=80$,**

**או נציב $x=71$, במשוואת קו הרגרסיה הישן: $y=0.41\cdot 71+45.89=75$, ונוסיף $5$ נקודות עכשיו.**

**<u>דרך פתרון נוספת</u>, גם היא ללא חישוב משוואת קו הרגרסיה החדש.**

**ציון התקן של הציון שלו על העבודה הוא $0$, מתאים לנתון שבממוצע.**

**בהתאם גם ציון התקן של הציון הסופי הוא $0$.**

**נציב מחדש במשוואה של ציון התקן:** $z=\frac{y-\bar{y}}{s_y} \to 0=\frac{y-80}{14} \to y=80$.

**תשובה: הציון הסופי של התלמיד צריך להיות $80$,**

**כדי שהנקודה המייצגת את שני הציונים שלו תהיה על ישר הרגרסיה החדש.**

חוקרים בדקו את הקשר בין משקל של עכבר ($Y$ בגרמים) ובין משקל מנת המזון היומית שלו ($X$ בגרמים).

הם בדקו עשרה עכברים. משקלי העכברים ומשקל מנת המזון היומית של כל אחד מהם מוצגים בטבלה שלפניכם.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| משקל מנת המזון היומית ($X$ בגרמים) | 1 | 2 | 3 | 3 | 3 | 4 | 4 | 4 | 5 | 5 |
| משקל העכבר ($Y$ בגרמים) | 12 | 13 | 14 | 15 | 16 | 20 | 22 | 24 | 28 | 30 |

נתון כי המשקל הממוצע של מנת המזון היומית הוא 3.4 גרמים.

**א.** הראו כי סטיית התקן של משקל מנת המזון היומית היא 1.2 גרמים.

נתון כי המשקל הממוצע של עשרת העכברים הוא 19.4 גרמים, וסטיית התקן של משקלם היא 6.086 גרמים.

**ב.** לפניכם 4 מספרים שונים: $0$ , $-0.123$ , $0.923$ , $1$ . אחד מן המספרים הוא מקדם המתאם $r$ בין משקל העכבר ובין משקל מנת המזון היומית שלו.

בחרו איזה מהם הוא מקדם המתאם, ונמקו את בחירתכם (אין צורך לחשב).

**ג.** מצאו את משוואת ישר הרגרסיה לניבוי משקל העכברים מתוך משקל מנת המזון היומית שלהם.

לאחר זמן מה התגלה כי המאזניים שבהם נשקלו העכברים לא היו מכוילים ויש להפחית 2 גרמים ממשקלו של כל עכבר (המאזניים שבהם נשקלה מנת המזון היומית היו מכוילים).

**ד.** מה תהיה משוואת ישר הרגרסיה החדש לאחר הכנסת התיקון במשקלי העכברים?

**ה.** על פי ישר הרגרסיה שמצאתם בסעיף ד, מהו הניבוי למשקל עכבר שמשקל מנת המזון היומית שלו הוא 3.5 גרם?

**א. חוקרים בדקו את הקשר בין משקל של עכבר ( $\mathrm{Y}$ בגרמים),**

**ובין משקל מנת המזון היומית שלו ( $\mathrm{X}$ בגרמים) .**

**האוכלוסייה שנבדקה, על ידי החוקרים, היא $10$ עכברים.**

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **משקל מנת המזון היומית ( $\mathrm{X}$ בגרמים)** | 1 | 2 | 3 | 3 | 3 | 4 | 4 | 4 | 5 | 5 |
| **משקל העכבר ( $\mathrm{Y}$ בגרמים)** | 12 | 13 | 14 | 15 | 16 | 20 | 22 | 24 | 28 | 30 |

**נתון כי המשקל הממוצע של מנת מזון היומית הוא $\bar{\mathrm{X}} = 3.4$ גרם.**

$$\mathrm{S} = \sqrt{\frac{(1-3.4)^2 \cdot 1 + (2-3.4)^2 \cdot 1 + (3-3.4)^2 \cdot 3 + (4-3.4)^2 \cdot 3 + (5-3.4)^2 \cdot 2}{10}}$$

$$\mathrm{S} == \sqrt{\frac{14.4}{10}} = \sqrt{1.44}$$

$$\boxed{\mathrm{S} = 1.2}$$

**תשובה: הראינו כי סטיית התקן, של משקל מנת המזון היומית, היא $1.2$ גרמים.**

**ב. נתון כי $19.4$ גרמים $\bar{\mathrm{Y}} =$ , משקל ממוצע של עכבר , עם סטיית תקן של $S_Y = 6.086$ גרמים,**

**כאשר המשקל הממוצע של מנת מזון היומית הוא $\bar{\mathrm{X}} = 3.4$ גרם, ומצאנו כי סטיית תקן היא $S_X = 1.2$ גרמים.**

**ניתן לראות שקיימת מגמה של עלייה במשקל העכבר, עם עליית משקל מנת המזון היומית,**

**ולכן נצפה לראות מקדם מתאם חיובי.**

**ניתן גם לראות שאין קו ישר, שעליו תהיינה כל הנקודות**

**(ישנן גם תצפיות המראות עלייה במשקל העכבר ללא שינוי במשקל מנת המזון היומית),**

**ולכן מקדם המתאם אינו דטרמיניסטי $(r = 1)$ , ומקדם מתאם $r = 0.923$ נראה מתאים ביותר.**

**תשובה: $r = 0.923$ הוא מקדם המתאם.**

**ג. נמצא את משוואת ישר הרגרסיה לניבוי משקל העכברים $\mathrm{Y}$ , מתוך משקל מנת המזון היומית שלהם $\mathrm{X}$.**

**נמצא תחילה את שיפוע קו הרגרסיה:** $m = r \cdot \frac{s_Y}{s_X} = 0.923 \cdot \frac{6.086}{1.2} \approx 4.681$ .

**נמצא את משוואת ישר הרגרסיה, שעובר בנקודת הממוצעים $(3.4, 19.4)$ :**

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

$$\mathrm{Y} - 19.4 = 4.681(\mathrm{X} - 3.4)$$

$$\mathrm{Y} - 19.4 = 4.681\mathrm{X} - 15.92$$

$$\boxed{\mathrm{Y} = 4.681\mathrm{X} + 3.484}$$

**תשובה: משוואת קו הרגרסיה, לניבוי משקל העכברים $\mathrm{Y}$ , מתוך משקל מנת המזון היומית שלהם $\mathrm{X}$,**

**היא $\mathrm{Y} = 4.681\mathrm{X} + 3.484$ .**

**ד. לאחר זמן מה התברר כי , עקב אי כיול המאזניים, יש להפחית $2$ גרמים ממשקלו של כל עכבר.**

**זאת, ללא שינוי בנתוני משקלי מנות המזון, שנשקלו במאזניים מכוילים היטב.**

**הפחתה בשיעור קבוע של הנתונים, מורידה בשיעור זה את הממוצע,**

**ללא שינוי בסטיית התקן כי הפיזור נשמר, ולכן $17.4$ גרמים $\bar{Y}=$ .**

**מכאן שיש תזוזה אנכית, $2$ יחידות מטה של ישר הרגרסיה.**

**לחילופין, נמצא את משוואת ישר הרגרסיה, שעובר בנקודת הממוצעים החדשה $(3.4, 17.4)$ :**

$$y-\bar{y}=m(x-\bar{x})$$
$$Y-17.4=4.681(X-3.4)$$
$$Y-17.4=4.681X-15.92$$
$$\boxed{Y=4.681X+1.484}$$

**תשובה: משוואת קו הרגרסיה החדש, לאחר הכנסת התיקון במשקלי העכברים,**

**היא $Y=4.681X+1.484$ .**

**ה. נמצא מהו הניבוי $(Y)$ למשקל עכבר שמשקל מנת המזון היומית שלו הוא $3.5$ גרם.**

**נשים לב שנתון של $3.5$ גרם הוא בתחום של נתוני משקלי מנות המזון שנצפו,**

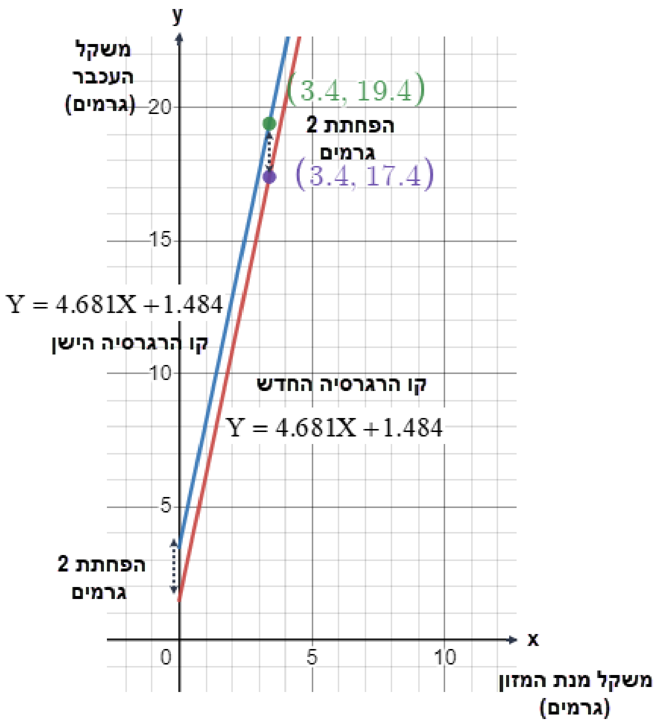
**ולכן ניתן להעריך את משקל העכבר.**

**נציב $X=3.5$ , במשוואת קו הרגרסיה החדש:** $Y=4.681\cdot3.5+1.484=17.8675\approx17.87$ .

**תשובה: הניבוי למשקל העבר, שמשקל מנת המזון היומית שלו הוא $3.5$ גרם ,**

**הוא בערך $17.8675\approx17.87$ גרם.**

# העשרה



חקלאית בעלת מטע עצי שזיף התעניינה בקשר הליניארי 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.

**א. חקלאית בעלת מטע עצי שזיף התעניינה בקשר הליניארי שבין קוטר השזיפים ( $x$ ), ובין משקלם ( $y$ ) .**

**האוכלוסייה שנבדקה היא $4$ שזיפים, כאשר הקוטר נתון במילימטרים, והמשקל בגרמים.**

**נמצא כי $s_y < s_x$ , וכי מקדם המתאם הוא $r = 0.8$ .**

**(1) נביע באמצעות $s_x$ ו- $s_y$ את שיפוע קו הרגרסיה.**

$$m = r \cdot \frac{s_y}{s_x} \quad \rightarrow m = 0.8 \cdot \frac{s_y}{s_x}$$

**תשובה: שיפוע קו הרגרסיה הוא** $0.8 \cdot \frac{s_y}{s_x}$ .

**(2) כיוון ש-** $s_y < s_x$ **, אז** $0 < \frac{s_y}{s_x} < 1$ **(סטיית תקן היא תמיד גודל חיובי).**

**מכאן ששיפוע קו הרגרסיה קטן מ-** $1$ **, ולמעשה** $0 < m < 0.8$ .

**תשובה: שיפוע קו הרגרסיה קטן מ-** $1$ .

**ב. נתון כי ישר הרגרסיה לניבוי משקל השזיפים לפי קוטרם הוא:** $y = \frac{3}{4}x + 15$ **, וכי** $\bar{x} = 80$ .

**לפנינו טבלה, ובה נתונים על המשקל של שלושה מתוך ארבעת השזיפים שנבחרו.**

| **שזיף** | **המשקל ( $y$ ) בגרמים** |
|---|---|
| **א** | 70 |
| **ב** | 70 |
| **ג** | 80 |
| **ד** | **?** |

**(1) ישר הרגרסיה עובר בנקודת הממוצעים** $(\bar{x}, \bar{y})$ **, כלומר בנקודה** $(80, \bar{y})$ .

**נציב** $\bar{x} = 80$ **במשוואת ישר הרגרסיה:** $y = \frac{3}{4} \cdot 80 + 15 = 75$ **, ומכאן ש-** $\bar{y} = 75$ **גרם.**

**סכום הנתונים שווה למכפלת הממצע במספר הנתונים, ולכן סך כל המשקלים הוא** $300$ **גרם** $= 75 \cdot 4$ .

**ומשקל שזיף ד הוא** $80$ **גרם** $= 300 - 70 - 70 - 80$ .

**אפשר כמובן גם:** $y = 80 \leftarrow 300 = 220 + y \leftarrow 75 = \frac{70 + 70 + 80 + y}{4}$ .

**תשובה: המשקל של שזיף ד הוא** $80$ **גרם.**

| שזיף | המשקל ( $y$ ) בגרמים |
|---|---|
| א | 70 |
| ב | 70 |
| ג | 80 |
| ד | 80 |

**(2) נמצא את סטיית התקן של משקל השזיפים.**

$$s_y=\sqrt{\frac{(y_1-\bar{y})^2\cdot f_1+(y_2-\bar{y})^2\cdot f_2+....+(y_n-\bar{y})^2\cdot f_n}{n}}$$

$$s_y=\sqrt{\frac{(70-75)^2\cdot 1+(70-75)^2\cdot 1+(80-75)^2\cdot 1+(80-75)^2\cdot 1}{4}}$$

$$s_y=\sqrt{\frac{25+25+25+25}{4}}=\sqrt{\frac{100}{4}}=\sqrt{25}$$

$$\boxed{s_y=5}$$

**הערה: נשים לב כי סטיית התקן קטנה או שווה למרחק הממוצע מנתוני הקצה (מעין בדיקה קטנה).**

**תשובה: סטיית התקן של משקל השזיפים היא $5$ גרם.**

**(3) נמצא את סטיית התקן של קוטר השזיפים.**

**בתת-סעיף א (1) מצאנו כי** $m=0.8\cdot\frac{s_y}{s_x}$ .

**על-פי ישר הרגרסיה** $y=\frac{3}{4}x+15$ **מתקיים** $m=\frac{3}{4}$ **, ובתת סעיף ב(2) מצאנו כי** $s_y=5$ .

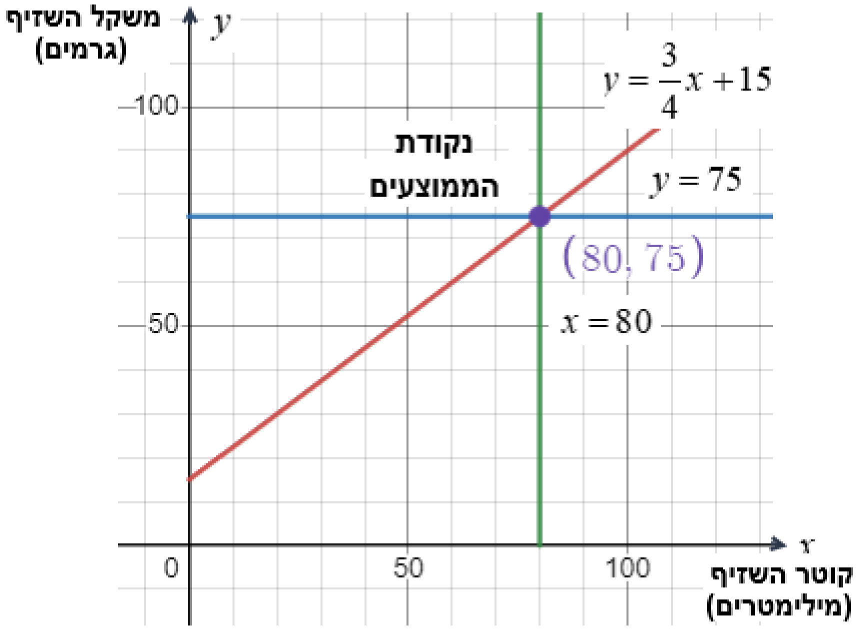
**מכאן ש-** $\frac{3}{4}=0.8\cdot\frac{5}{s_x}$.

$$\frac{3}{4}=\frac{4}{s_x}$$

$$\boxed{s_x=5\frac{1}{3}}$$

**תשובה: סטיית התקן של קוטר השזיפים היא $5\frac{1}{3}$ מילימטר.**

## העשרה


משקל השזיף
(גרמים)
y
100
50
0
50
100
x
קוטר השזיף
(מילימטרים)
$y=\frac{3}{4}x+15$
נקודת
הממוצעים
$y=75$
$(80, 75)$
$x=80$

מורה למתמטיקה לתלמידי כיתה י"א רצתה לבדוק את הקשר הלינארי בין ציוני תלמידיה בבחינת הבגרות במתמטיקה (המשתנה $x$) ובין ציוני ההגשה שלהם (המשתנה $y$).

בטבלה שלפניכם מוצגים הציונים של חמישה תלמידים שנבחנו בשנת 2022. ציון ההגשה של התלמיד החמישי אינו מוצג.

| ציון בחינת הבגרות ($x$) | ציון ההגשה ($y$) |
|---|---|
| 59 | 78 |
| 60 | 81 |
| 60 | 81 |
| 60 | 81 |
| 61 | ? |

**א.** מצאו את ממוצע ציוני בחינת הבגרות של תלמידים אלה.

המורה חישבה את ישר הרגרסיה לניבוי $y$ על פי $x$. נתון כי בעבור $x = 60$ מנובא הערך $y = 80$.
נתון גם כי שיפוע ישר הרגרסיה הוא $0.5$.

**ב.** (1) מצאו את משוואת ישר הרגרסיה.

(2) מצאו את ציון ההגשה של התלמיד החמישי. נמקו.

בעבור חמישה תלמידים שנבחנו בשנת 2021 חישבה המורה את מקדם המתאם בין שני המשתנים, ואז חישבה את ישר הרגרסיה לניבוי $y$ על פי $x$. היא גילתה שציון ההגשה של כל אחד מן התלמידים זהה **בדיוק** לציון ההגשה המנובא לו באמצעות ישר הרגרסיה.

**ג.** קבעו מה הן שתי הטענות האפשריות מבין הטענות (1)–(4) שלפניכם. נמקו.

(1) $0 < r < 1$

(2) $r = 1$

(3) $-1 < r < 0$

(4) $r = -1$

**א. מורה למתמטיקה לתלמידי כיתה י"א רצתה לבדוק את הקשר הליניארי**

**בין ציוני תלמידיה בבחינת הבגרות במתמטיקה $(x)$ ובין ציוני ההגשה שלהם $(y)$ .**

**האוכלוסייה שנבדקה, על ידי המורה, היא $5$ תלמידים, שנבחנו בשנת $2022$.**

| ציון בחינת הבגרות $(x)$ | ציון ההגשה $(y)$ |
|---|---|
| 59 | 78 |
| 60 | 81 |
| 60 | 81 |
| 60 | 81 |
| 61 | ? |

**נמצא את ממוצע ציוני הבגרות:** $\bar{x}=\frac{59+60+60+60+61}{5}=\frac{300}{5}=60$

**אפשר גם לראות שיש שלושה ציוני $60$, והשניים האחרים בקצוות באותו מרחק, ולכן זה הממוצע.**

**תשובה: ממוצע ציוני בחינת הבגרות של תלמידים אלו הוא $60$ .**

**ב. המורה חישבה את ישר הרגרסיה לניבוי $y$ על פי $x$ .**

**(1) בעבור $x=60$ מנובא הערך $y=80$ , כאשר בנוסף נתון כי שיפוע ישר הרגרסיה הוא $m=0.5$ .**

**נמצא את משוואת קו הרגרסיה.**

$$y-80=0.5(x-60)$$
$$y-80=0.5x-30$$
$$\boxed{y=0.5x+50}$$

**תשובה: משוואת ישר הרגרסיה היא $y=0.5x+50$.**

**(2) קו הרגרסיה עובר בנקודת הממוצעים, לכן:** $\bar{y}=0.5\cdot60+50=80$.

**אם ממוצע ציוני ההגשה הוא $80$ , אז סכום ציוני ההגשה הוא $80\cdot5=400$ .**

**מכאן שציון ההגשה של התלמיד החמישי הוא :** $400-(78+81+81+81)=79$ .

**תשובה: ציון ההגשה של התלמיד החמישי הוא $79$ .**

**ג. גם בשנת $2021$ בדקה המורה את הקשר הליניארי**

**בין ציוני תלמידיה בבחינת הבגרות במתמטיקה $(x)$ ובין ציוני ההגשה שלהם $(y)$ .**

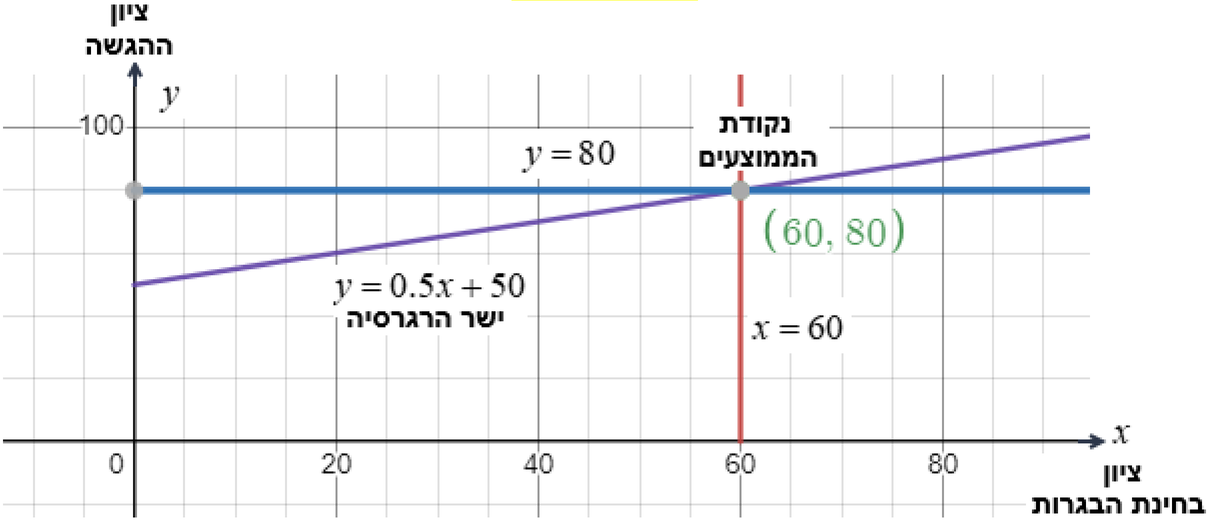
**היא גילתה שציון ההגשה של כל אחד מהתלמידים זהה <u>בדיוק</u> לציון ההגשה המנובא לו באמצעות ישר הרגרסיה.**

**מכאן שהקשר הוא קשר דטרמיניסטי, חיובי או שלילי,**

**ומקדם המתאם יכול להיות $r=1$ , או $r=-1$ .**

**תשובה: שתי הטענות האפשריות הן (2) $r=1$, או (4) $r=-1$ .**

# העשרה



ציון
ההגשה
y
100
y = 80
נקודת
הממוצעים
(60, 80)
y = 0.5x + 50
ישר הרגרסיה
x = 60
0
20
40
60
80
x
ציון
בחינת הבגרות

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.

א. מרצה באוניברסיטה רצתה לבדוק אם היעדרות משיעורים בקורס שנתי (המשתנה $x$),

קשורה ליניארית לציון במבחן הסופי (המשתנה $y$).

נתון: ממוצע ההיעדרויות מן השיעורים היה $\bar{x}=10$ , הציון הממוצע היה $\bar{y}=70$ .

בנוסף, מקדם המתאם היה שלילי $(r<0)$ .

המרצה מצאה את משוואת ישר הרגרסיה לניבוי $y$ מ- $x$ .

שיפוע ישר הרגרסיה נתון על פי הנוסחה $m=r\cdot\frac{s_y}{s_x}$ .

סטיות התקן לא יכולות להיות אפס, כי אז מקדם המתאם , שהוא ממוצע מכפלות ציוני התקן, לא מוגדר.

לכן סטיות התקן חיוביות, ומכאן שסימני שיפוע ישר הרגרסיה שווים לסימני מקדם המתאם, ולכן $m<0$ .

ישר הרגרסיה עובר בנקודת הממוצעים, שהיא $(10,70)$ .

לסיכום $m<0$ והישר עובר בנקודה $(10,70)$ , ולכן המשוואה המתאימה היא .4 $y=-2x+90$ .

תשובה: המשוואה .4 $y=-2x+90$ היא המשוואה שמצאה המרצה.

ב. המרצה חישבה את סטיות התקן, בעבור ההיעדרות מן השיעורים ובעבור הציונים במבחן הסופי.

סטיות התקן שהתקבלו הן: $s_x=4$ ו- $s_y=10$ .

נחשב את מקדם המתאם $(r)$ , כאשר על פי ישר הרגרסיה $y=-2x+90$ מתקבל ש- $m=-2$ .

$$m=r\cdot\frac{s_y}{s_x}$$

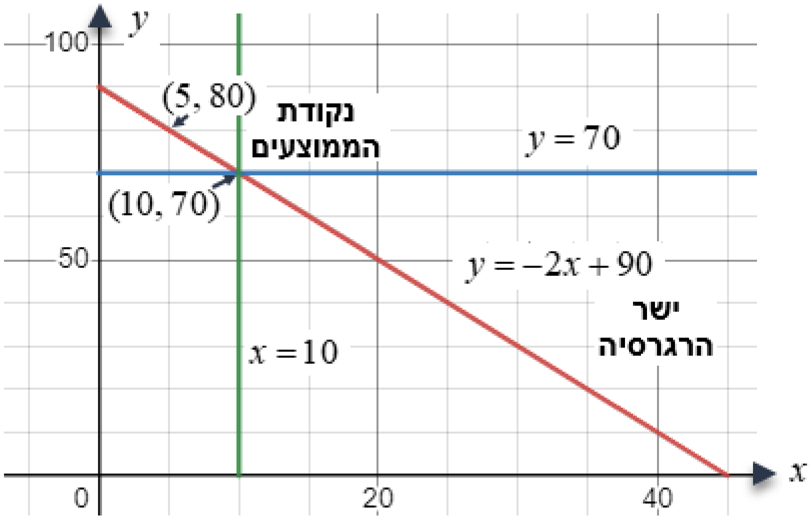
$$-2=r\cdot\frac{10}{4}$$

$$-2=2.5r \quad /:2.5$$

$$\boxed{r=-0.8}$$

תשובה: מקדם המתאם הוא $r=-0.8$.

**הערה**




ג. כדי למצוא את מספר ההיעדרויות ( $x$ ), שישר רגרסיה מנבא עבורו ציון $y=80$ ,

נציב $y=80$ במשוואת ישר הרגרסיה $y=-2x+90$ .

$$80=-2x+90 \quad /-90$$

$$-10=-2x \quad /:(-2)$$

$$\boxed{x=5}$$

תשובה: מספר ההיעדרויות, שישר הרגרסיה מנבא בעבורו ציון 80 , הוא 5.

ד. 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 $65$.

מסקנה: משוואת ישר הרגרסיה, שמצא דוד, היא $y=65$.

(1) כיוון שמשוואת ישר הרגרסיה, שמצא דוד, היא $y=65$, הרי ש- $m=0$.

תשובה: שיפוע ישר הרגרסיה הוא $0$.

(2) כיוון שהישר מנבא תמיד את הציון $65$, הרי שזה הציון הממוצע.

כל ישר רגרסיה חייב לעבור דרך נקודת הממוצעים ולכן אם הוא ישר אופקי,

הערך הקבוע של $y$ עליו הוא גם הממוצע של $y$.

תשובה: הציון הממוצע הוא $65$.

ה. אין כל דרך לחשב את ממוצע ההיעדרויות, כי לא נמצא כל קשר בין שני המשתנים.

למעשה, מקדם המתאם הוא $0$.

אין כל דרך להגיע לנתונים על המשתנה $x$, ולכן גם לא ניתן למצוא את הממוצע שלו.

תשובה: על פי הנתונים שבשאלה, לא ניתן למצוא את הממוצע של היעדרויות התלמידים של דוד.

## סעיף ד2: הרחבה והעשרה

שיפוע של קו רגרסיה יהיה $0$, רק כאשר:

- מקדם המתאם יהיה $0$ ובמצב כזה, אין למנבא $x$ שום משמעות בניבוי $y$,
  ולכן הערך היחיד שכדאי לנבא במצב כזה עבור כל ערך של $x$,
  הוא פשוט הממוצע של $y$ שהוא הניבוי הסטטיסטי הכי בטוח בהיעדר מידע נוסף.
- כל המדידות של $y$ במדגם זהות,
  ולכן סטיית התקן שלו היא $0$ ולכן ברור שזהו הערך היחיד שכדאי לנבא.
  מובן שבמקרה זה מקדם המתאם אינו מוגדר, כי סטיית התקן מופיעה במכנה

$$r=\frac{1}{N\cdot S_x\cdot S_y}\left((x_1-\bar{x})(y_1-\bar{y})+\ldots+(x_N-\bar{x})(y_N-\bar{y})\right).$$

ואומנם לא מתייחסים למקרה זה בניבוי על פי רגרסיה,

וממילא אין משמעות לניבוי $y$ על פי $x$.

בעל חנות המוכר טאבלטים בדק את הקשר הלינארי בין גודל המסך של טאבלט באינצ'ים (המשתנה $x$)

ובין מספר הדקות שנדרשו ללקוח להחליט לקנות את הטאבלט (המשתנה $y$).

ביום מסוים הוא מכר 8 דגמים שונים של טאבלטים.

לפניכם טבלה המתארת את הנתונים של שמונת הדגמים שהוא מכר באותו יום:

| גודל המסך באינצ'ים $(x)$ | מספר הדקות לקבלת ההחלטה לקנות את הטאבלט $(y)$ |
|---|---|
| 9 | 2 |
| 9 | 10 |
| 9 | 10 |
| 9 | 10 |
| 11 | 10 |
| 11 | 10 |
| 11 | 10 |
| 11 | 18 |

**א.** חשבו את הממוצעים ואת סטיות התקן של שני המשתנים, $x$ ו־ $y$.

**ב.** חשבו את מקדם המתאם $r$.

**ג.** מצאו את משוואת ישר הרגרסיה לניבוי מספר הדקות לקבלת החלטה כתלות בגודל המסך.

בעל החנות הזמין לחנותו דגם חדש של טאבלט, שגודל המסך שלו 10 אינצ'ים.

**ד.** על פי ישר הרגרסיה שמצאתם, מהו ניבוי מספר הדקות לקבלת ההחלטה בעבור דגם זה?

בעקבות העסקתו של מוכר חדש בחנות, התקצר ב־ $20\%$ זמן קבלת ההחלטה לקנות כל אחד מדגמי הטאבלטים.

**ה.** בעבור כל אחד מן המדדים שלפניכם קבעו אם ערכו יגדל, יקטן או לא ישתנה בעקבות השינוי הזה.

(1) מקדם המתאם $r$.

(2) סטיית התקן של המשתנה $y$.

(3) שיפוע ישר הרגרסיה לניבוי מספר הדקות לקבלת ההחלטה כתלות בגודל המסך.

**א. בעל חנות המוכר טאבלטים בדק את הקשר הליניארי**
**בין גודל המסך של טאבלט באינצ'ים (המשתנה $x$)**
**ובין מספר הדקות שנדרשו ללקוח להחליט לקנות את הטאבלט (המשתנה $y$).**

**האוכלוסייה שנבדקה, על ידי בעל החנות, היא מכירה של 8 טאבלטים.**

| גודל המסך באינצ'ים $(x)$ | מספר הדקות לקבלת החלטה לקנות את הטאבלט $(y)$ |
|---|---|
| 9 | 2 |
| 9 | 10 |
| 9 | 10 |
| 9 | 10 |
| 11 | 10 |
| 11 | 10 |
| 11 | 10 |
| 11 | 18 |

**נמצא את ממוצע גדלי המסך. קל לראות שהממוצע הוא 10 אינץ',**
**כי יש בדיוק ארבעה נתונים הקטנים ב- 1 ס"מ, ומספר זהה של נתונים (ארבעה) הגדולים ב- 1 ס"מ.**

**או:** $\bar{x} = \frac{9\cdot 4 + 11\cdot 4}{8} = \frac{80}{8} =$ **10 אינץ'**

**סטיית התקן היא:** $S_x = \sqrt{\frac{(9-10)^2\cdot 4 + (11-10)^2\cdot 4}{8}} = \sqrt{\frac{8}{8}} = 1$

**נמצא את ממוצע מספר הדקות לקבלת החלטה. קל לראות שהממוצע הוא 10 אינץ',**
**כי יש בדיוק נתון אחד הקטן ב- 8 דקות, ומספר זהה של נתונים (אחד) הגדול ב- 8 דקות.**

**או:** $\bar{y} = \frac{1\cdot 2 + 6\cdot 10 + 1\cdot 18}{8} = \frac{80}{8} =$ **10 דקות**

**סטיית התקן היא:** $S_y = \sqrt{\frac{(2-10)^2\cdot 1 + (10-10)^2\cdot 6 + (18-10)^2\cdot 1}{8}} = \sqrt{\frac{128}{8}} = \sqrt{16} =$ **4 דקות**

**תשובה:** $\bar{x} =$ **10 אינץ'**, $S_x =$ **1 אינץ'**, $\bar{y} =$ **10 דקות**, $S_y =$ **4 דקות**.

**ב. מצאנו כי:** $10$ **אינץ'** $\bar{x}=$ , $1$ **אינץ'** $S_x=$ , $10$ **דקות** $\bar{y}=$ , $4$ **דקות** $S_y=$ .

**נחשב את מקדם המתאם** $r$ .

**מקדם המתאם שווה לממוצע מכפלות ציוני התקן.**

$$r=\frac{1}{N\cdot S_x\cdot S_y}\left((x_1-\bar{x})(y_1-\bar{y})+\ldots+(x_N-\bar{x})(y_N-\bar{y})\right)$$

**ששה מציוני התקן של המשתנה** $y$ **הם** $0$ **, כי הנתון שווה לממוצע, ומכאן שניתן להתעלם מהם בחישוב.**

$$r=\frac{1}{8\cdot 1\cdot 4}\cdot\left((9-10)(2-10)+(11-10)(18-10)\right)$$

$$r=\frac{1}{32}\cdot\left((-1)\cdot(-8)+(1\cdot 8)\right)$$

$$r=\frac{1}{32}\cdot 16$$

$$\boxed{r=0.5}$$

**תשובה: מקדם המתאם הוא** $r=0.5$ .

**ג. נמצא את משוואת ישר הרגרסיה לניבוי** $y$ **על פי** $x$ .

$$m=r\cdot\frac{s_y}{s_x}\quad\rightarrow m=0.5\cdot\frac{4}{1}=2$$

**נמצא את משוואת קו הרגרסיה, שעובר בנקודת הממוצעים.**

$$y-10=2(x-10)$$

$$y-10=2x-20$$

$$\boxed{y=2x-10}$$

**תשובה: משוואת ישר הרגרסיה, לניבוי מספר הדקות להחלטה כתלות בגודל המסך, היא** $y=2x-10$ .

**ד. בעל החנות הזמין לחנותו דגם חדש של טאבלט, שגודל המסך שלו** $10$ **אינצ'ים, זהה לממוצע גודל המסך.**

**לכן, הניבוי למספר הדקות הוא בדיוק כממוצע של הדקות, כלומר** $10$ **דקות.**

**אפשר גם להציב במשוואת קו הרגרסיה:** $y=2\cdot 10-10=10$ .

**תשובה: ניבוי מספר הדקות לקבלת החלטה, כתלות בגודל המסך, היא** $10$ **דקות.**

**ה. בעקבות העסקתו של מוכר חדש בחנות,**

**התקצר ב-** $20\%$ **זמן קבלת ההחלטה לקנות כל אחד מדגמי הטאבלטים,**

**כלומר ירידה ל-** $0.8 = 80\%$ **מהזמן הקודם.**

**לכן, הממוצע החדש הוא** $8$ **דקות** $\bar{y} = 0.8 \cdot 10 =$ **, וסטית התקן היא** $3.2$ **דקות** $S_y = 0.8 \cdot 4 =$ **.**

**נבדוק את ההשפעה על כל אחד מהמדדים.**

**(1) מקדם המתאם** $r$ **.**

**מקדם המתאם נבנה בצורה כזו שהוא אינו רגיש לשינוי לינארי אם נעשה על אחד מ- או שני המשתנים.**

**(אלא אם אחד מהמשתנים מוכפל במקדם חיובי והשני בשלילי, ואז רק סימנו של מקדם המתאם מתהפך).**

**<u>הסבר נוסף</u>: מקדם המתאם שווה לממוצע מכפלות ציוני התקן.**

$$r = \frac{1}{N \cdot S_x \cdot S_y}\left((x_1 - \bar{x})(y_1 - \bar{y}) + \ldots + (x_N - \bar{x})(y_N - \bar{y})\right)$$

**גם הסטייה** $(y_1 - \bar{y})$ **קטנה ב-** $20\%$ **, וגם סטיית התקן** $S_y$ **קטנה ב-** $20\%$ **,**

**כך שמקדם המתאם לא משתנה.**

**תשובה: מקדם המתאם** $r$ **לא ישתנה.**

**(2) סטיית התקן של המשתנה** $y$ **.**

**החסרה של קבוע מכל דגימה משנה את הממוצע אך לא את סטיית התקן שכן הפיזור לא משתנה.**

**לעומת זאת, החסרה של אחוז קבוע מכל דגימה אינה בעצם החסרה של קבוע,**

**אלא הכפלה (כי אחוזים זהים משלמים שונים זה מזה).**

**לכן, במקרה של החסרה באחוזים, גם סטיית התקן מוכפלת באותו פקטור כי ההחסרה באחוזים מקטינה**

**יותר את הדגימות בעלות הערכים הגבוהים, מאשר את אלו בעלות הערכים הקטנים כך שהפיזור קטן.**

**(והפוך לגבי תוספת של אחוזים).**

**או לאחר חישוב, סטיית הקטן תקטן ב-** $20\%$ **, כי כל סטייה של נתון מהממוצע תקטן ב-** $20\%$ **.**

**כאשר מספר הנתונים לא השתנה.**

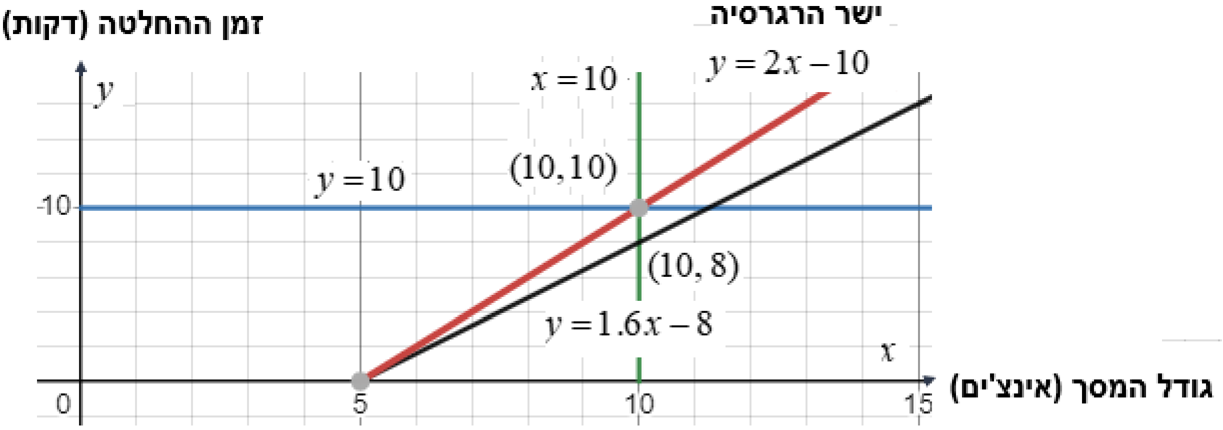
**תשובה: סטיית התקן של המשתנה** $y$ **תקטן.**

**(3) שיפוע ישר הרגרסיה .**

$m = r \cdot \frac{s_y}{s_x}$ **, וכאשר** $s_y$ **קטן ב-** $20\%$ **, בעוד ששאר המשתנים קבועים, הרי שהשיפוע קטן.**

**תשובה: שיפוע ישר הרגרסיה, לניבוי מספר הדקות לקבלת ההחלטה כתלות בגודל המסך, יקטן.**

# הערשה




לאחר הירידה ב- $20\%$ בזמן קבלת ההחלטה,

נקודת הממוצעים החדשה היא $(10,8)$ .

$m=0.5\cdot\frac{3.2}{1}=1.6$ (גם ירידה ב- $20\%$)

וקו הרגרסיה החדש הוא $y=1.6x-8$

דן ערך מחקר. הוא בדק את הקשר בין אחוז הגידול השנתי של האוכלוסייה ב־ 12 מדינות (משתנה $x$) ובין אחוז הילדים בני 0–14 באותן המדינות (משתנה $y$).

דן רצה למצוא ישר רגרסיה לניבוי $y$ לפי $x$. הוא חישב את הממוצע ואת סטיית התקן של אחוז הגידול השנתי של האוכלוסיות, וכן את מְקַדם המְתאם בין שני המשתנים, וקיבל את התוצאות האלה:

$\bar{x} = 0.465$ , $s_x = 0.683$ , $r = 0.871$ .

דן מצא כי משוואת ישר הרגרסיה היא: $y = 11.3x + 16.3$ .

**א.** (1) מצאו את הממוצע של אחוז הילדים באותן המדינות (הממוצע של המשתנה $y$).

(2) מצאו את סטיית התקן של אחוז הילדים באותן המדינות.

במדינה מסוימת נתון כי גודל האוכלוסייה נשאר קבוע (אין גידול שנתי באוכלוסייה שלה).

**ב.** על פי ישר הרגרסיה, מהו אחוז הילדים במדינה זו?

נתונה מדינה נוספת, שאחוז הגידול השנתי של האוכלוסייה שלה הוא 2 .

**ג.** האם אפשר להסיק כי אחוז הילדים במדינה זו הוא בדיוק 38.9 ? נמקו את תשובתכם.

**א. דן ערך מחקר, בו בדק את הקשר בין אחוז הגידול השנתי של האוכלוסייה (המשתנה $x$)**
**ובין אחוז הילדים בני $0-14$ באותן מדינות (המשתנה $y$).**
**האוכלוסייה שנבדקה היא אוכלוסייתן של $12$ מדינות.**

**דן קיבל את התוצאות הבאות: ממוצע אחוז הגידול הוא $\bar{x}=0.465$ , עם סטיית תקן $S_x=0.683$,**
**ומקדם מתאם $r=0.871$ (מקדם מתאם חיובי וחזק, $0.7<r<1$).**

**דן מצא כי משוואת ישר הרגרסיה היא: $y=11.3x+16.3$ .**

**(1) ישר הרגרסיה עובר בנקודת הממוצעים .**

$$\bar{y}=11.3\cdot 0.465+16.3=21.55$$

**תשובה: הממוצע של אחוז הילדים באותן מדינות הוא** $21.55$.

**(2)** $m=r\cdot\frac{s_y}{s_x}$**, כאשר על פי משוואת ישר הרגרסיה** $m=11.3$ .

$$11.3=0.871\cdot\frac{s_y}{0.683}$$
$$11.3=1.2753s_y \quad /:1.2753$$
$$\boxed{s_y=8.86}$$

**תשובה: סטיית התקן של אחוז הילדים באותן מדינות היא** $8.86$.

**ב. במדינה מסוימת נתון כי גודל האוכלוסייה נשאר קבוע (אין גידול שנתי באוכלוסייה <u>שלה</u>).**
**כלומר במדינה זו, $x=0$ , ועל פי ישר הרגרסיה $y=16.3$ .**

**הערה – נניח כי מדינה זו היא מתוך אוכלוסיית המדינות שהמדגם של $12$ המדינות שנבדקו נלקח מתוכה,**
**אם כי הדבר לא מצוין במפורש בשאלה.**

**תשובה: על פי ישר הרגרסיה, אחוז הילדים במדינה זו הוא** $16.3$ .

**ג. במדינה נוספת, אחוז הגידול השנתי של האוכלוסייה הוא $x=2$ .**

**הערה – נניח כי מדינה זו היא מתוך אוכלוסיית המדינות שהמדגם של $12$ המדינות שנבדקו נלקח מתוכה, אם כי הדבר לא מצוין במפורש בשאלה.**

**נציב $x=2$ במשוואת ישר הרגרסיה: $y=11.3\cdot 2+16.3=38.9$ .**

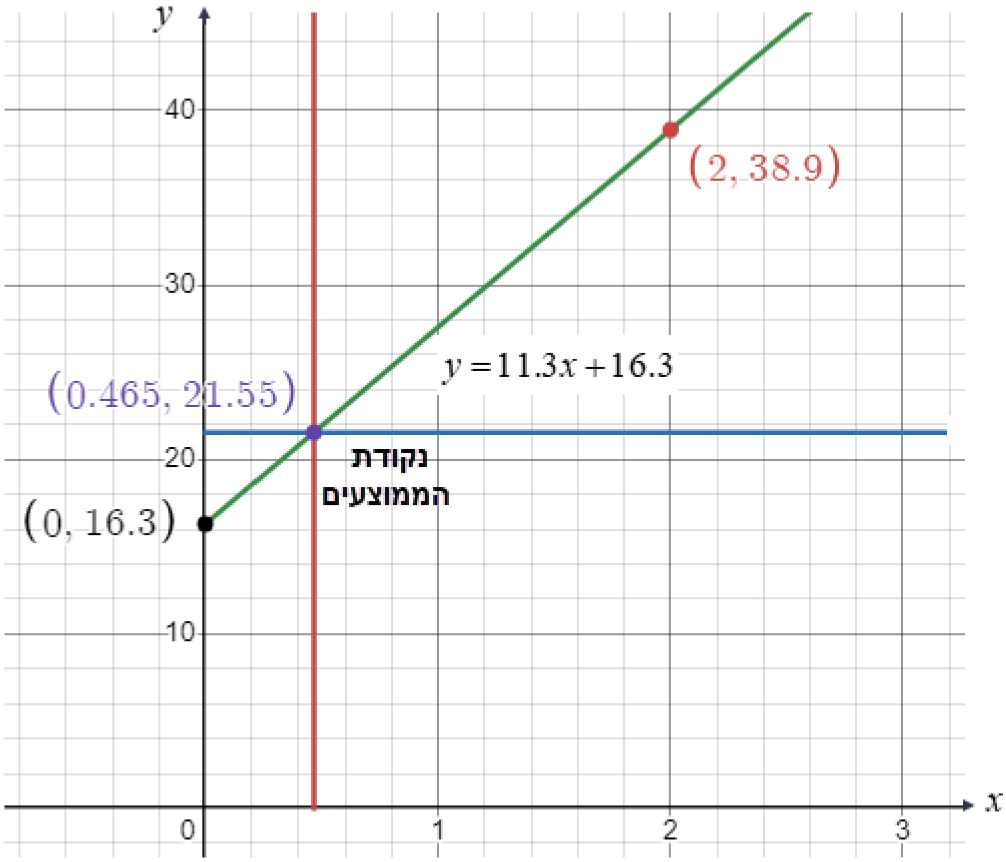
**אומנם מקדם המתאם הוא גבוה, אבל לא דטרמיניסטי ($r=0.871\neq 1$ ), ולכן זה רק ניבוי.**

**הסבר נוסף: הערך המנובא לפי משוואת הרגרסיה הוא ממוצע כל הערכים של המדינות, בהן שיעור ה- $x$ הוא זהה לזה שבדקנו.**

**כלומר, בכל המדינות באוכלוסיית המדינות הנתונה, בהן שיעור גידול האוכלוסייה הוא $2\%$ , הממוצע של אחוז הילדים הוא $38.9$ .**

**תשובה: לא ניתן להסיק כי אחוז הילדים במדינה זו הוא <u>בדיוק</u> $38.9$ .**

## העשרה



סטטיסטיקאית ערכה מחקר בקרב זוגות.

היא החליטה לבדוק את הקשר בין מספר הילדים שיש לזוג (המשתנה $x$) ובין ההוצאה החודשית על דלק של הזוג (המשתנה $y$).

לצורך כך היא דגמה 4 זוגות ממאגר הנתונים שלה:

– זוג ללא ילדים שההוצאה החודשית שלו על דלק היא 1,500 שקלים.

– זוג שיש לו ילד אחד, וההוצאה החודשית שלו על דלק היא 1,800 שקלים.

– זוג שיש לו שלושה ילדים, וההוצאה החודשית שלו על דלק היא 2,900 שקלים.

– זוג שיש לו ארבעה ילדים, וההוצאה החודשית שלו על דלק היא 3,800 שקלים.

**א.** סרטטו את דיאגרמת הפיזור המתאימה לארבעת הזוגות שבמדגם. רשמו את ערכי הנקודות על הצירים.

הסטטיסטיקאית חישבה את סטיית התקן של המשתנה $y$ וקיבלה כי $S_y = \sqrt{835{,}000}$ .

**ב.** (1) מצאו את הממוצע של מספר הילדים שיש לזוג במדגם ואת ההוצאה החודשית הממוצעת על דלק של זוג במדגם.

(2) חשבו את מקדם המתאם בין שני המשתנים.

**ג.** מצאו את משוואת ישר הרגרסיה לניבוי ההוצאה החודשית על דלק לפי מספר הילדים.

בעקבות עליית מחירי הדלק, עלתה ב־ 6% ההוצאה החודשית של כל אחד מן הזוגות על דלק.

**ד.** בעבור כל אחד מן המדדים שלפניכם קבעו אם ערכו גדל, קטן או לא השתנה.

(1) סטיית התקן של המשתנה $y$ .

(2) מקדם המתאם $r$ .

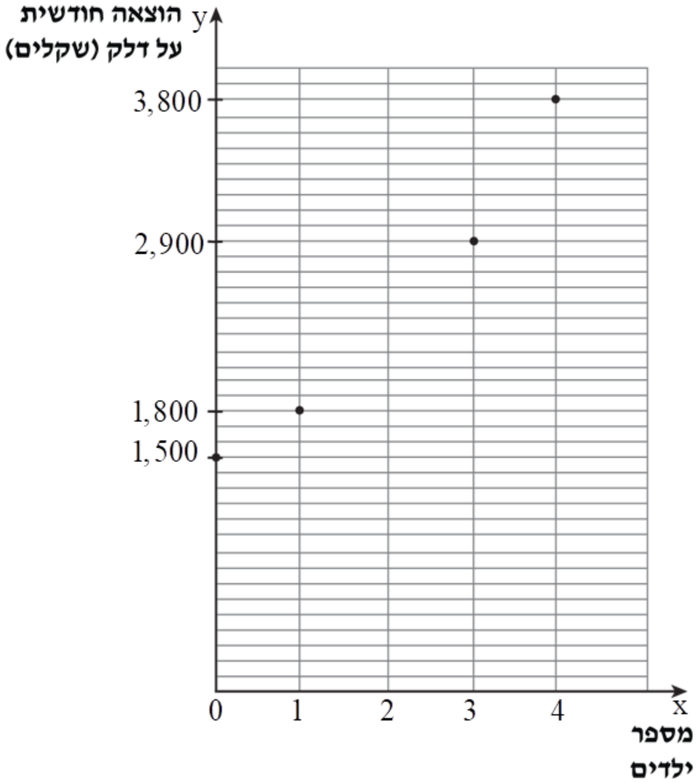
א. סטטיסטיקאית ערכה מחקר, בו בדקה את הקשר בין מספר הילדים שיש לזוג (המשתנה $x$)

האוכלוסייה שנבדקה $4$ זוגות ממאגר הנתונים שלה.

נרכז את הנתונים בטבלה מתאימה.

| מספר הילדים $x$ | 0 | 1 | 3 | 4 |
|---|---|---|---|---|
| הוצאה על דלק (שקלים) $y$ | 1,500 | 1,800 | 2,900 | 3,800 |

נסרטט דיאגרמת פיזור מתאימה, ונרשום את ערכי הנקודות על הצירים.




ניתן לראות שקיימת מגמה של עלייה, ולכן נצפה לראות מקדם מתאם חיובי.

הנקודות מפוזרות בצורה די מהודקת סביב קשר לינארי עולה,

ולכן מקדם המתאם צפוי להיות חיובי חזק (מעל $0.7$ ).

תשובה: הסרטוט מעל.

ב. הסטטיסטיקאית חישבה את סטיית התקן של המשתנה $y$ וקיבלה כי $S_y = \sqrt{835{,}000}$,

(1) נמצא את הממוצעים של שני המשתנים.

$$\bar{x} = \frac{0+1+3+4}{4} = \frac{8}{4} = 2$$

$$\bar{y} = \frac{1{,}500+1{,}800+2{,}900+3{,}800}{4} = \frac{10{,}000}{4} = 2{,}500$$

תשובה: הממוצע של מספר בילדים שיש לזוג במדגם הוא $2$ ילדים.

ההוצאה הממוצעות החודשית על דלק של זוג במדגם הוא $2{,}500$ שקלים.

(2) נחשב את מקדם המתאם $r$ בין שני המשתנים.

מקדם המתאם שווה לממוצע מכפלות ציוני התקן.

נחשב תחילה את סטיית התקן של המשתנה $x$ , את $S_x$.

סטיית התקן היא:

$$S_x=\sqrt{\frac{(0-2)^2\cdot 1+(1-2)^2\cdot 1+(3-2)^2\cdot 1+(4-2)^2\cdot 1}{4}}=\sqrt{\frac{4+1+1+4=10}{4}}=\sqrt{\frac{10}{4}}=\sqrt{2.5}$$

מצאנו כי: $\bar{x}=2$ ילדים, $S_x=1.58$ ילדים, $\bar{y}=2{,}500$ שקלים, $S_y=\sqrt{835{,}000}$ שקלים.

$$r=\frac{1}{N\cdot S_x\cdot S_y}\left((x_1-\bar{x})(y_1-\bar{y})+\ldots+(x_N-\bar{x})(y_N-\bar{y})\right)$$

$$r=\frac{(0-2)(1{,}500-2{,}500)+(1-2)(1{,}800-2{,}500)+(3-2)(2{,}900-2{,}500)+(4-2)(3{,}800-2{,}500)}{4\cdot\sqrt{2.5}\cdot\sqrt{835{,}000}}$$

$$r=\frac{5{,}700}{4\cdot\sqrt{2.5}\cdot\sqrt{835{,}000}}$$

$$\boxed{r=0.986}$$

זהו מקדם מתאם חיובי חזק מאוד ($0.7<r<1$ נחשב למקדם מתאם חזק), קרוב לדטרמיניסטי,

כפי שניתן היה לחוש את זה מדיאגרמת הפיזור,

כאשר נראה שהנקודות "ממש" מסודרות על קו ישר עולה.

תשובה: מקדם המתאם הוא $r=0.986$ .

ג. נמצא את משוואת ישר הרגרסיה לניבוי $y$ על פי $x$ .

$$m=r\cdot\frac{s_y}{s_x}\quad\rightarrow m=0.986\cdot\frac{\sqrt{835{,}000}}{\sqrt{2.5}}=570$$

נמצא את משוואת קו הרגרסיה, שעובר בנקודת הממוצעים.

$$y-2{,}500=570(x-2)$$

$$y-2{,}500=570x-1{,}140$$

$$\boxed{y=570x+1{,}360}$$

תשובה: משוואת ישר הרגרסיה, לניבוי ההוצאה החודשית על דלק לפי מספר הילדים,

היא $y=570x+1{,}360$ .

ד. בעקבות עליית מחירי הדלק, עלתה ב- $6\%$ ההוצאה החודשית של כל אחד מן הזוגות על דלק.

(1) העלאה ב- $6\%$ , כלומר העלאה פי $1.06$ של כל אחד מערכי המשתנה $y$ ,

מגדילה את פיזור הנתונים בדיוק פי $1.06$,

וסטיית התקן החדשה היא $968.6$ שקלים $=\sqrt{835,000}\cdot 1.06$.

הרחבה

- ניתן לחשב את סטיית התקן מחדש ולראות (לא רשום ללא חישוב).
- סטיית התקן (בניגוד לממוצע) אינה מושפעת מתוספת של קבוע זהה לכל המדידות,
  אך לעומת זאת אם כופלים את כל המדידות באותו קבוע,
  סטיית התקן מוכפלת גם היא באותו קבוע.

תשובה: ערכה של סטיית התקן של המשתנה $y$ גדלה .

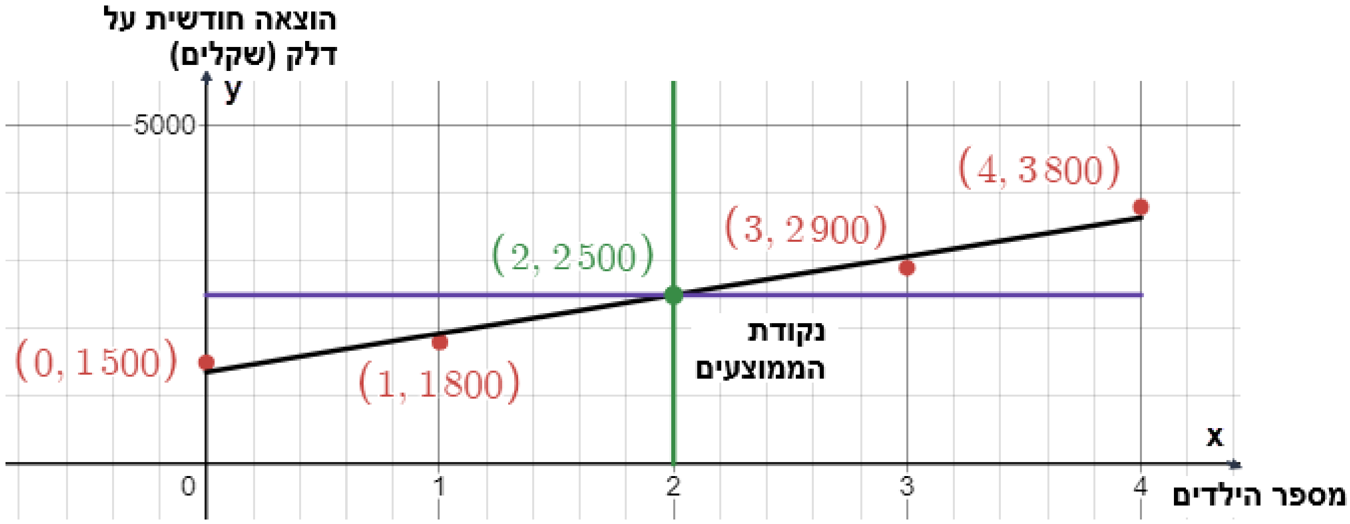
(2) שינוי ליניארי זהה של כל אחד מערכי אחד מהמשתנים, או שניהם, אינו משנה את מקדם המתאם.

הרחבה

- מקדם המתאם לא מושפע משינוי ליניארי על כל אחד מהמשתנים,
  גם אם השינוי הוא שינוי אחד למשתנה $x$ ושינוי אחר למשתנה $y$.
- אם בשינוי של המשתנה $x$,
  המקדם המכפיל שונה בסימנו מהמקדם המכפיל בשינוי של המשתנה $y$,
  אז מקדם המתאם מתחלף בסימנו (אך לא בערכו המוחלט).

תשובה: ערכו של מקדם המתאם $r$ לא השתנה.

## הערה



לפניכם שלוש טבלאות שבהן מוצגים ערכים של שני משתנים, שנמדדו בתצפיות שונות: משתנה $x$ ומשתנה $y$.

טבלה 1

| x | y |
|---|---|
| 1 | 2 |
| 2 | 4 |
| 3 | 6 |
| 4 | 8 |
| 5 | 10 |

טבלה 2

| x | y |
|---|---|
| 4 | 9 |
| 5 | 11 |
| 6 | 19 |
| 7 | 22 |
| 8 | 17 |

טבלה 3

| x | y |
|---|---|
| 16 | 35 |
| 17 | 34 |
| 18 | 33 |
| 19 | 32 |
| 20 | 31 |

נסמן את מקדם המתאם בין $x$ ל־ $y$ ב־ $r$.

**א.** התאימו כל אחד מן ההיגדים I–III שלפניכם לטבלאות 1–3:

I. $0 < r < 1$

II. $r = 1$

III. $r = -1$

הנתונים בטבלה 2 מתייחסים לקבוצה של 5 ספורטאים.

בטבלה מתואר הקשר בין מספר הפעמים בשבוע שכל אחד מן הספורטאים מתאמן (המשתנה $x$) ובין מספר השעות בשבוע שהוא מתאמן (המשתנה $y$).

נתון כי משוואת ישר הרגרסיה לניבוי $y$ על פי $x$ היא $y = 2.7x + b$, $b$ הוא פרמטר.

**ב.** מצאו כמה פעמים בשבוע בממוצע מתאמן ספורטאי בקבוצה זו, וכמה שעות בשבוע בממוצע הוא מתאמן.

**ג.** מצאו את ערך הפרמטר $b$.

נתון כי היחס בין סטיות התקן של הנתונים בטבלה 2 הוא $\frac{s_y}{s_x} = 3.45$.

**ד.** חשבו את הערך של $r$.

**א. לפנינו $3$ טבלאות שבהן מוצגים ערכים של שני משתנים, שנמדדו בתצפיות שונות: משתנה $x$ ומשתנה $y$ .**

**טבלה $1$**

| $x$ | $y$ |
|---|---|
| 1 | 2 |
| 2 | 4 |
| 3 | 6 |
| 4 | 8 |
| 5 | 10 |

**טבלה $2$**

| $x$ | $y$ |
|---|---|
| 4 | 9 |
| 5 | 11 |
| 6 | 19 |
| 7 | 22 |
| 8 | 17 |

**טבלה $3$**

| $x$ | $y$ |
|---|---|
| 16 | 35 |
| 17 | 34 |
| 18 | 33 |
| 19 | 32 |
| 20 | 31 |

**נסמן את מקדם המתאם בין $x$ ל- $y$ ב- $r$ .**

**נתאים כל אחד מן ההיגדים I - III שלפנינו לטבלאות $1-3$ .**

- **ניתן לראות שבטבלה $1$ כאשר $x$ גדל ב- $1$ אז $y$ גדל ב- $2$ ,**

  **ולפנינו קשר ליניארי דטרמיניסטי (מושלם) חיובי (ישר עולה),**

  **שאפשר לראות שמיוצג על ידי הישר $y=2x$ .**

  **מכאן שהביטוי II. $r=1$ מתאים לטבלה $1$.**

- **ניתן לראות שבטבלה $2$ כאשר $x$ גדל ב- $1$ אז $y$ גדל בהפרשים משתנים, ופעם אחת יורד.**

  **מכאן שניתן להבין שקיים קשר ליניארי חיובי (ישר עולה), אבל לא מושלם.**

  **מכאן שהביטוי I. $0<r<1$ מתאים לטבלה $2$ .**

- **ניתן לראות שבטבלה $3$ כאשר $x$ גדל ב- $1$ אז $y$ יורד ב- $1$ ,**

  **ולפנינו קשר ליניארי דטרמיניסטי (מושלם) שלילי (ישר יורד),**

  **שאפשר לראות שמיוצג על ידי הישר $y=-x+51$ .**

  **מכאן שהביטוי III. $r=-1$ מתאים לטבלה $3$.**

**תשובה: הביטוי I. מתאים לטבלה $2$ . הביטוי II. מתאים לטבלה $1$. הביטוי III. מתאים לטבלה $3$.**

**טבלה** $2$

| $x$ | $y$ |
|---|---|
| 4 | 9 |
| 5 | 11 |
| 6 | 19 |
| 7 | 22 |
| 8 | 17 |

**ב. הנתונים שבטבלה** $2$ **מתייחסים לקבוצה של** $5$ **ספורטאים.**

**בטבלה מתואר הקשר בין מספר הפעמים בשבוע שכל אחד מן הספורטאים מתאמן (המשתנה** $x$ **)**

**ובין מספר השעות בשבוע שהוא מתאמן (המשתנה** $y$**).**

**נתון כי משוואת ישר הרגרסיה לניבוי** $y$ **על פי** $x$ **היא** $y=2.7x+b$ **(** $b$ **הוא פרמטר).**

**ישר הרגרסיה עובר בנקודת הממוצעים, לכן אם נמצא את ממוצעי המשתנים, נוכל למצוא את ערכו של** $b$.

**נמצא את הממוצעים של שני המשתנים.**

$$\bar{x}=\frac{4+5+6+7+8}{5}=\frac{30}{5}=6$$

$$\bar{y}=\frac{9+11+19+22+17}{5}=\frac{78}{5}=15.6 .$$

**תשובה: ספורטאי בקבוצה זו מתאמן בממוצע** $6$ **פעמים בשבוע, ובממוצע מתאמן** $15.6$ **שעות בשבוע.**

**ג. כאמור ישר הרגרסיה עובר בנקודת הממוצעים,**

**ולכן נציב את שיעורי נקודת הממוצעים** $(6,15.6)$ **במשוואת ישר הרגרסיה.**

$$15.6=2.7\cdot 6+b$$
$$15.6=16.2+b .$$
$$\boxed{b=-0.6}$$

**ומשוואת קו הרגרסיה היא** $y=2.7x-0.6$ .

**תשובה:** $b=-0.6$ .

**ד. נתון כי היחס בין סטיית התקן של הנתונים בטבלה** $2$ **הוא** $\frac{S_y}{S_x}=3.45$,

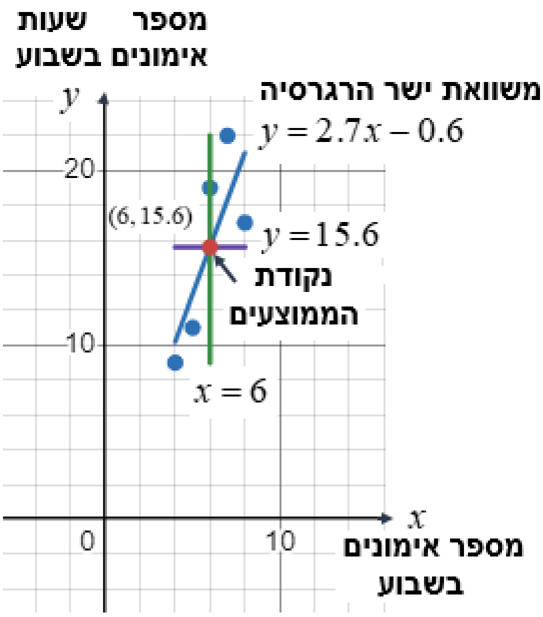
**שיפוע קו הרגרסיה** $(m=2.7)$ **נתון על ידי הנוסחה** $m=r\cdot\frac{s_y}{s_x}$ .

**נציב את הנתונים בנוסחה:** $\boxed{r=0.7826} \rightarrow \quad 2.7=r\cdot 3.45$ .

**וזה קשר ליניארי חזק,** $0.7<r<1$**, התואם את טבלה** $2$ **ואת התשובה שלנו לסעיף א.**

**תשובה:** $r=0.7826$ .

# הערשה



מספר שעות
אימונים בשבוע
משוואת ישר הרגרסיה
$y = 2.7x - 0.6$
(6, 15.6)
$y = 15.6$
נקודת
הממוצעים
$x = 6$
20
10
0
10
y
x
מספר אימונים
בשבוע

## 15

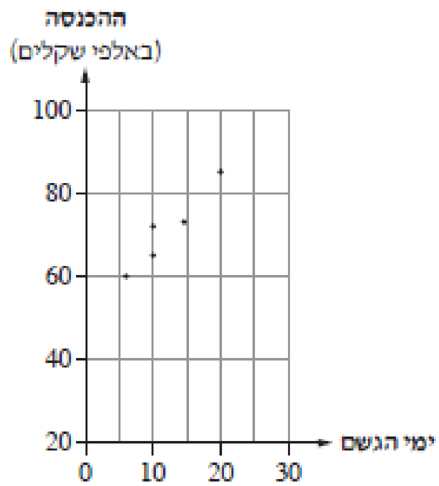
בבית מרקחת א' בעיר מסוימת בדקה המנהלת במשך 5 חודשים את הקשר בין מספר ימי הגשם בחודש (המשתנה $x$) ובין ההכנסה החודשית של בית המרקחת (המשתנה $y$) באותו החודש.

בטבלה שלפניכם מוצגים הנתונים שאספה המנהלת.

| החודש | ימי הגשם (המשתנה $x$) | ההכנסה (באלפי שקלים) (המשתנה $y$) |
|---|---|---|
| נובמבר | 6 | 60 |
| דצמבר | 10 | 72 |
| ינואר | 20 | 85 |
| פברואר | 14 | 73 |
| מרץ | 10 | 65 |

**א.** מצאו את ממוצע ימי הגשם בחודש ואת סטיית התקן.

לפניכם דיאגרמת פיזור המתארת את $y$ כתלות ב־$x$.




ממוצע ההכנסות החודשי (באלפי שקלים) של בית מרקחת א' הוא 71, וסטיית התקן היא 8.46.

אחד מן המספרים: 0.959, 1, $-0.959$ — הוא מקדם המתאם $r$.

**ב.** קבעו איזה מהם הוא מקדם המתאם $r$, ונמקו את קביעתכם.

**ג.** (1) מצאו את שיפוע ישר הרגרסיה לניבוי $y$ על פי $x$.

(2) מצאו את משוואת ישר הרגרסיה לניבוי $y$ על פי $x$.

**ד.** על פי ישר הרגרסיה, מהו ניבוי ההכנסה בחודש שבו מספר ימי הגשם יהיה 15?

גם בבית מרקחת ב' באותה העיר בדקה המנהלת במשך אותם 5 חודשים את הקשר בין מספר ימי הגשם בחודש ובין ההכנסה החודשית של בית המרקחת.

המנהלת מצאה שבכל חודש הייתה ההכנסה של בית מרקחת ב' קטנה ב־10 אלף שקלים מן ההכנסה של בית מרקחת א'.

**ה.** כתבו את משוואת ישר הרגרסיה לניבוי ההכנסה החודשית (באלפי שקלים) של בית מרקחת ב' על פי מספר ימי הגשם בחודש.

**א. בבית מרקחת א' בעיר מסוימת בדקה המנהלת במשך 5 חודשים את הקשר**
**בין מספר ימי הגשם בחודש (המשתנה $x$ ) ובין ההכנסה החודשית של בית המרקחת (המשתנה $y$ ).**

| החודש | ימי הגשם (המשתנה $x$ ) | ההכנסה (באלפי שקלים) (המשתנה $y$ ) |
|---|---|---|
| **נובמבר** | 6 | 60 |
| **דצמבר** | 10 | 72 |
| **ינואר** | 20 | 85 |
| **פברואר** | 14 | 73 |
| **מרץ** | 10 | 65 |

**נמצא את ממוצע ימי הגשם בחודש ואת סטיית התקן:**

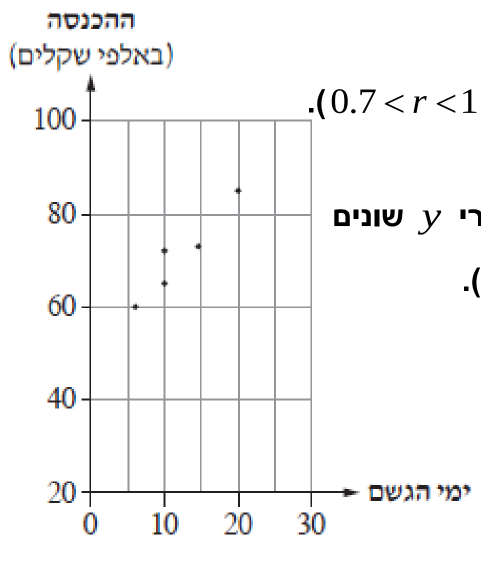
$$\bar{x}=\frac{6+10+20+14+10}{5}=\frac{60}{5}=12 \text{ ימי גשם}$$

**סטיית התקן היא:**

$$S_x=\sqrt{\frac{(6-12)^2\cdot 1+(10-12)^2\cdot 1+(20-12)^2\cdot 1+(14-12)^2\cdot 1+(10-12)^2\cdot 1}{5}}=\sqrt{\frac{112}{5}}=4.733$$

**תשובה: ממוצע ימי הגשם בחודש הוא** $\bar{x}=12$ **ימי גשם , וסטיית התקן היא** $S_x=4.733$ **ימי גשם,**

**ב. לפנינו דיאגרמת פיזור המתארת את $y$ כתלות ב- $x$ .**




**אחד מן המספרים** $0.959$ , $1$ , $-0.959$ **הוא מקדם המתאם** $r$ .

- **ניתן לראות בדיאגרמה שהקשר הוא קשר <u>חיובי</u>, די חזק (** $0.7<r<1$**).**
- **ניתן לראות שהקשר אינו מושלם, דטרמיניסטי,**

**לדוגמה יש שתי תצפיות עם אותם ערכי** $x=10$ **, אך שיעורי $y$ שונים**
**(הכנסות שונות, למרות שאותו מספר של ימי גשם בחודש).**

**תשובה: מקדם המתאם הוא** $r=0.959$ .

ג. ממוצע ההכנסות החודשי (באלפי שקלים) של בית המרקחת הוא $\bar{y}=71$ ,

וסטיית התקן היא $8.46$ ימי גשם $S_y=$ .

(1) נמצא את שיפוע ישר הרגרסיה לניבוי $y$ על פ $x$ .

$$m=r\cdot\frac{S_y}{S_x}=0.959\cdot\frac{8.46}{4.733}=1.714$$

תשובה: שיפוע ישר הרגרסיה לניבוי $y$ על פ $x$ הוא $1.714$

(2) נמצא את משוואת ישר הרגרסיה לניבוי $y$ על פ $x$ .

ישר הרגרסיה עובר בנקודת הממוצעים $(\bar{x},\bar{y})=(12,71)$.

$$y-71=1.714(x-12)$$
$$y-71=1.714x-20.568$$
$$\boxed{y=1.714x+50.432}$$

תשובה: משוואת ישר הרגרסיה לניבוי $y$ על פ $x$ היא $y=1.714x+50.432$ .

ד. כדי למצוא את ההכנסה המנובאת בחודש ( $y$ ), שבו מספר ימי הגשם יהיה $x=15$ ,

נציב $x=15$ (שנמצא בתוך טווח הנתונים) במשוואת ישר הרגרסיה $y=1.714x+50.432$.

$$y=1.714\cdot 15+50.432$$
$$\boxed{y=76.142}$$

תשובה: על פי ישר הרגרסיה, ניבוי ההכנסה בחודש שבו מספר ימי הגשם יהיה $15$,

הוא $76.142$ אלפי שקלים ( $76,142$ שקלים).

**ה. גם בבית מרקחת ב' <u>באותה העיר</u> בדקה המנהלת באותם 5 <u>חודשים</u> את הקשר בין מספר ימי הגשם בחודש (המשתנה $x$ ) ובין ההכנסה החודשית של בית המרקחת (המשתנה $y$ ).**

**המנהלת מצאה <u>שבכל חודש</u> הייתה ההכנסה של בית מרקחת ב' קטנה ב- $10,000$ שקלים (10 אלפי שקלים), כלומר מדובר בהחסרת קבוע (10) מכל המדידות של משתנה $y$ ולכן הממוצע שלו יקטן ב-10 כאשר סטיית התקן שלו לא משתנה.**

**כיוון שמדובר באותה עיר ובאותם חודשים, אז מספר ימי הגשם זהה,**

**ו- 4.733 ימי גשם $S_x =$ ללא שינוי.**

**מקדם המתאם אינו משתנה, וגם לא שיפוע קו הרגרסיה.**

**נקודת הממוצעים החדשה היא $(\bar{x},\bar{y})=(12,61)$ ,**

**ולמעשה יש הזזה אנכית של 10 יחידות כלפי מטה של ישר הרגרסיה,**

**ונקבל שהמשוואה חדשה היא $y=1.714x+40.432$ .**

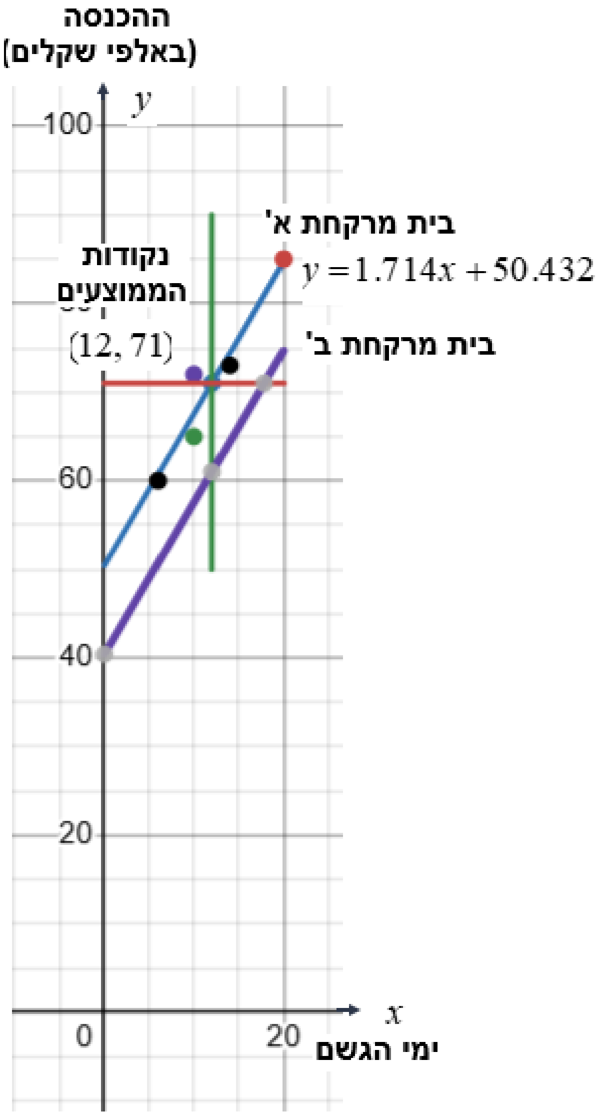
**תשובה: משוואת ישר הרגרסיה לניבוי ההכנסה החודשית (באלפי שקלים) של בית מרקחת ב'**
**על פי מספר ימי הגשם בחודש היא $y=1.714x+40.432$ .**

**הערה**

- **סטיית התקן של ההכנסה לא משתנה, כי זו טרנספורמציה לינארית, שלא משפיעה על הפיזור**
  **כיוון שמדובר באותה עיר ובאותם חודשים, אז מספר ימי הגשם זהה,**
  **ו- 4.733 ימי גשם $S_x =$ ללא שינוי.**
- **מקדם המתאם אינו משתנה, כי מקדם המתאם אינו משתנה בהפעלת טרנספורמציה לינארית,**
  **אא"כ מקדמי הכפל של המשתנים שונים בסימנם),**
- **וגם שיפוע קו הרגרסיה ( $m=r\cdot\frac{S_y}{S_x}$ ) אינו משתנה כי אין שינוי באף גורם.**
- **ניתן גם לחשב מחדש את משוואת קו הרגרסיה**
  **ישר הרגרסיה עובר בנקודת הממוצעים בחדשה $(\bar{x},\bar{y})=(12,61)$.**

$$y-61=1.714(x-12)$$
$$y-61=1.714x-20.568$$
$$\boxed{y=1.714x+40.432}$$

# העשרה




את הקשר החיובי בין המשתנים ניתן לראות במיקומן של הנקודות בדיאגרמת הפיזור,

כאשר אם נקודת הממוצעים היא ראשית הצירים, אז הנקודות ברובן המוחלט ברביע הראשון והשלישי.