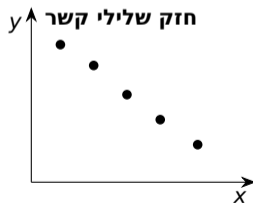
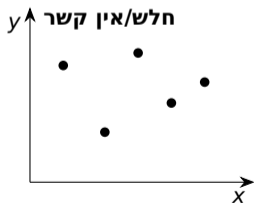


קו רגרסיה ומקדם מתאם r – שאלון 471 (4 יח"ל)

- נלמד לתאר **קשר ליניארי** בין שני משתנים כמותיים: x (מסביר) ו- y (מוסבר).
 - נכיר את **מקדם המתאם r** ואת **קו הרגרסיה** לניבוי y מתוך x .
- רעיון מרכזי:** עובדים עם **ענן נקודות** + מדדים מספריים: r , ומשוואת קו הרגרסיה.
- דגש:** בכל שאלה: קודם מבינים את הכיוון (חיובי/שלילי), ואז עוברים לחישובים.

ענן נקודות: מה מחפשים?

- מציירים נקודות (x_i, y_i) במערכת צירים.
- בודקים אם יש **מגמה**:
 - **מגמה עולה** \Rightarrow קשר חיובי.
 - **מגמה יורדת** \Rightarrow קשר שלילי.
 - **בלי מגמה** \Rightarrow אין קשר ליניארי ברור.
- ככל שהנקודות **קרובות יותר לקו ישר**, הקשר הליניארי חזק יותר.



מקדם המתאם r : תחום ערכים ומשמעות

- r מודד את **עוצמת** הקשר הליניארי ואת **כיוונו**.

- תחום הערכים:

$$-1 \leq r \leq 1$$

- פירוש:

- $r \approx 1$: קשר ליניארי **חיובי חזק**.

- $r \approx -1$: קשר ליניארי **שלילי חזק**.

- $r \approx 0$: **אין קשר ליניארי** (ייתכן קשר לא ליניארי).

רעיון מרכזי: הסימן של r נותן כיוון, והערך המוחלט $|r|$ נותן חוזק.

שגיאה נפוצה: גם אם r קטן, זה לא אומר "אין קשר בכלל" — אולי הקשר אינו ליניארי.

נוסחאות (כמו בדף הנוסחאות לבגרות 4 יח"ל)

ממוצע עם שכיחויות: אם הערכים x_1, \dots, x_n מופיעים עם שכיחויות f_1, \dots, f_n , אז:

$$N = f_1 + f_2 + \dots + f_n, \quad \bar{x} = \frac{f_1 x_1 + f_2 x_2 + \dots + f_n x_n}{N}$$

סטיית תקן עם שכיחויות:

$$S_x = \sqrt{\frac{f_1(x_1 - \bar{x})^2 + f_2(x_2 - \bar{x})^2 + \dots + f_n(x_n - \bar{x})^2}{N}}$$

מתאם: עבור N זוגות (x_i, y_i) :

$$S_{xy} = \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + \dots + (x_N - \bar{x})(y_N - \bar{y})}{N}, \quad r = \frac{S_{xy}}{S_x \cdot S_y}$$

דגש: ב- S_{xy} עובדים על כל הזוגות (x_i, y_i) לפי הסדר הנתון.

מה משפיע על r , ומה לא משפיע על r ?

● משפיע על r :

- מידת ה"התיישרות" של הנקודות סביב קו (ככל שיותר קרוב לקו — $|r|$ גדול).
- כיוון המגמה (עולה/יורד) — קובע את הסימן של r .
- נקודות חריגות עלולות לשנות מאוד את r .

● לא משפיע על r :

- הוספת קבוע לכל ערכי x או לכל ערכי y (הזזה).
- כפל במספר **חיובי** קבוע את כל ערכי x או את כל ערכי y (שינוי יחידות).

רעיון מרכזי: r חסין לשינויי יחידות/הזזה (כל עוד הכפל חיובי). **שגיאה נפוצה:** כפל ב-1 — הופך את סימן r .

קו רגרסיה לניבוי y לפי x (כמו בדף הנוסחאות) שיפוע ישר הרגרסיה:

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

משוואת ישר הרגרסיה:

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

דגש: בדיקה מהירה: הקו עובר דרך (\bar{x}, \bar{y}) , וסימן m כמו סימן r .

כתיבה בצורה $\hat{y} = ax + b$
 מהמשוואה $\hat{y} - \bar{y} = m(x - \bar{x})$ נקבל:

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

לכן:

$$a = m, \quad b = \bar{y} - a\bar{x}$$

רעיון מרכזי: אחרי שמוצאים m , מוצאים b בעזרת (\bar{x}, \bar{y}) .

איך משתמשים בקו הרגרסיה לניבוי?

- מציבים ערך x במשוואת הקו ומקבלים \hat{y} (ניבוי).
- \hat{y} הוא ערך צפוי לפי המגמה הליניארית.
- ניבוי אמין יותר כאשר x בתוך תחום הנתונים (לא אקסטרפולציה רחוקה).

דגש: אם $|x|$ קטן, הניבוי פחות אמין.

השאלה במלואה (1/2)

בעל חנות המוכר טאבלטים בדק את הקשר הליניארי בין גודל המסך של טאבלט באינצ'ים (המשתנה x) ובין מספר הדקות שנדרשו ללקוח להחליט אם לקנות את הטאבלט או לא. (משתנה y)
 ביום מסויים הור מכר שמונה דגמים שונים של טאבלטים.
 לפניכם טבלה המתארת את הנתונים של שמונת הדגמים שהוא מכר באותו היום:

אערכי המשתנה : 9, 9, 9, 9, 11, 11, 11, 11

עערכי המשתנה : 2, 10, 10, 10, 10, 10, 10, 18

לכל ערך של x מתאים ערך של y באותו מקום סידורי שכתבתי.
 בבקשה תציג את הנתונים של x ו y בטבלה מסודרת.

(א) חשבו את הממוצעים ואת סטיות התקן של שני המשתנים x ו y
 (ב) חשבו את מקדם המתאם r

השאלה במלואה (2/2)

ג) מוצאו את משוואת קו הרגרסיה לניבוי מספר הדקות לקבלת החלטה כתלו בגודל המסך.

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

ד) על פי ישר הרגרסיה שמצאתם, מהו מספר הדקות לקבלת החלטה בעבור דגם זה?

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

ה) בעבור כל אחד מהמדדים שלפניכם קבעו אם ערכו יגדל, יקטן או לא ישתנה בעקבות שינוי זה.

1) מקדם המתאם r

2) סטיית התקן של y

3) שיפוע ישר הרגרסיה

הצגת הנתונים בטבלה מסודרת

דגם	1	2	3	4	5	6	7	8
x	9	9	9	9	11	11	11	11
y	2	10	10	10	10	10	10	18

דגש: מכאן נוח לראות: ל- x יש שני ערכים עם שכיחויות 4 ו-4; ל- y יש 2 ו-18 פעם אחת, ו-10 שש פעמים.

פתרון א: \bar{x} , S_x בעזרת שכיחויות f

ל- x : הערכים 9, 11 עם שכיחויות 4, 4. לכן $N = 8$.

$$\bar{x} = \frac{4 \cdot 9 + 4 \cdot 11}{8} = \frac{80}{8} = 10$$

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

רעיון מרכזי: $\bar{x} = 10$, $S_x = 1$.

פתרון א: \bar{y} , S_y בעזרת שכיחויות f
 ל- y : הערכים 2, 10, 18 עם שכיחויות 1, 6, 1. לכן $N = 8$.

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

$$S_y = \sqrt{\frac{1(2 - 10)^2 + 6(10 - 10)^2 + 1(18 - 10)^2}{8}} = \sqrt{\frac{64 + 0 + 64}{8}} = \sqrt{16} = 4$$

רעיון מרכזי: $\bar{y} = 10$, $S_y = 4$

פתרון ב: חישוב r לפי דף הנוסחאות

נתון: $\bar{x} = 10$, $\bar{y} = 10$, $S_x = 1$, $S_y = 4$, $N = 8$

נחשב:

$$S_{xy} = \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + \dots + (x_N - \bar{x})(y_N - \bar{y})}{N}$$

נחשב את סכום מכפלות הסטיות $\sum (x_i - 10)(y_i - 10)$ בצורה מרוכזת:

● עבור $x = 9$ (4 פעמים): $(x - 10) = -1$, $(y - 10) = -8, 0, 0, 0 \Rightarrow y : 2, 10, 10, 10$

$$(-1)(-8) + (-1)(0) + (-1)(0) + (-1)(0) = 8$$

● עבור $x = 11$ (4 פעמים): $(x - 10) = 1$, $(y - 10) = 0, 0, 0, 8 \Rightarrow y : 10, 10, 10, 18$

$$(1)(0) + (1)(0) + (1)(0) + (1)(8) = 8$$

לכן:

$$\sum (x_i - 10)(y_i - 10) = 8 + 8 = 16 \Rightarrow S_{xy} = \frac{16}{8} = 2$$

ולבסוף:

$$r = \frac{S_{xy}}{S_x \cdot S_y} = \frac{2}{1 \cdot 4} = 0.5$$

רעיון מרכזי: $r = 0.5$ — קשר חיובי בינוני.

פתרון ג: משוואת קו הרגרסיה \hat{y} על x

$$m = r \cdot \frac{S_y}{S_x} = 0.5 \cdot \frac{4}{1} = 2$$

$$\hat{y} - \bar{y} = m(x - \bar{x}) \Rightarrow \hat{y} - 10 = 2(x - 10) \Rightarrow \hat{y} = 2x - 10$$

רעיון מרכזי: קו הרגרסיה: $\hat{y} = 2x - 10$.

פתרון ד: ניבוי עבור $x = 10$

$$\hat{y} = 2x - 10 \Rightarrow \hat{y} = 2 \cdot 10 - 10 = 10$$

רעיון מרכזי: מספר הדקות החזוי לדגם בגודל 10 אינך' הוא 10 דקות.

פתרון ה: קיצור ב-20% בזמני ההחלטה

לאחר השינוי: $y' = 0.8y$.

1 **מקדם המתאם r** : כפל כל ערכי y במספר חיובי קבוע $\Rightarrow r$ לא ישתנה.

2 **סטיית התקן של y** : סטיית תקן מוכפלת באותו קבוע:

$$S_{y'} = 0.8S_y$$

לכן S_y יקטן.

3 **שיפוע ישר הרגרסיה: $m = r \frac{S_y}{S_x}$** . אם r לא משתנה ו- S_y מוכפל ב-0.8, אז גם m יקטן (מוכפל ב-0.8).

רעיון מרכזי: תשובות: r — לא ישתנה; S_y — יקטן; שיפוע — יקטן.

לסיכום:

- תחום המתאם: $-1 \leq r \leq 1$. הסימן נותן כיוון, ו- $|r|$ נותן חוזק.
- סטיית תקן (עם שכיחויות):

$$S_x = \sqrt{\frac{\sum f(x - \bar{x})^2}{N}}$$

- מתאם:

$$S_{xy} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{N}, \quad r = \frac{S_{xy}}{S_x S_y}$$

- רגרסיה:

$$m = r \frac{S_y}{S_x}, \quad \hat{y} - \bar{y} = m(x - \bar{x})$$

מומלץ: לפני חישובים — לזהות כיוון קשר מהשרטוט/הנתונים.