

תלמידים יקרים,

אני גאה להציג בפניכם חוברת זו, כחלק מהקורס במתמטיקה שאלון 471 (תכ"ל)

בנושא: **סטטיסטיקה - רגרסיה ליניארית, מתאם פירסון.**

חוברת זו מהווה חלק ממערך מקיף של חומרי לימוד, תרגול והכנה לקראת בחינות הבגרות במתמטיקה, הזמינים עבורכם באתר **OpenBook**.

באתר תוכלו ליהנות:

- **מקורסים אונליין** איכותיים וידידותיים, המאפשרים לכם ללמוד בקצב האישי שלכם, מכל מקום ובכל זמן.

הקורסים כוללים סרטוני הסברים עם פתרונות מלאים לבחינות הבגרות האחרונות, מותאמים ומעודכנים לתוכנית הלימודים החדשה.

אתם אף פעם לא לבד – אם לא הבנתם משהו בסרטון, תוכלו לשלוח אלינו הודעה ישירות בדף הסרטון באתר, ונשמח לעזור לכם!

- **ממרכז למידה** – בשיעורים מרחוק בזום, ובמפגשים פרונטליים.

החומרים שלנו מלווים בסרטוני הסברים מוקלטים, פתרונות מפורטים, ושלל אמצעי המחשה אינטראקטיביים, שנועדו להפוך את הלמידה שלכם לחוויה מהנה, יעילה ומעמיקה.

סימונים בחוברת:

✓ קיים פתרון מוקלט באתר – לחיצה על הסימן תוביל אתכם ישירות לדף הרלוונטי באתר.

מצאתם טעות או משהו לא ברור?

אנחנו כאן עבורכם במייל: info@OpenBook.co.il

בוואטסאפ: [050-5350554](https://wa.me/050-5350554)

מאחלים לכם לימוד מהנה ומעשיר,
ובהצלחה רבה!

רית הלפנבאום

קורסים אונליין OpenBook | מרכז למידה ריתמטיקה

תוכן עניינים

4	מדדי קשר
4	סוגי סולמות
5	מדדי קשר לסולמות רווח או מנה
7	דיאגרמת פיזור
7	מקדם המתאם של פירסון – r – מדד לקשר קווי
7	שונות משותפת
8	מקדם המתאם של פירסון
10	קו הרגרסיה הליניארית
11	רגרסיה פתרון תרגיל עם שימוש באקסל ✓
13	שרטוט דיאגרמה וקו מגמה ✓
13	חיזוי בעזרת ישר הרגרסיה ✓
14	תרגיל ✓
14	תרגיל ✓
15	שונות כללית, ניבויים וטעות
16	רגרסיה ליניארית ומקדם המיתאם במחשבוני Casio
17	שאלה 15
17	שאלה 16
18	שאלה 17
18	שאלה 18
19	שאלה 19
19	שאלה 22
20	שאלה 23
20	שאלה 24
21	שאלה 25
21	שאלה 26
22	שאלה 28
22	שאלה 29
23	שאלה 30
23	שאלה 31
24	שאלה 32
24	שאלה 33
25	שאלה 34
25	שאלה 35
26	שאלה 36

26	שאלה 37
27	שאלה 38
27	שאלה 46
27	שאלה 49
28	שאלה 51
28	שאלה 52
29	טענות
29	שאלה 20
29	שאלה 42
29	שאלה 43
29	שאלה 47
29	שאלה 63
30	שאלה 69
30	שאלה 75
30	שאלה 81
31	שאלה 83
32	שאלה 92

מדדי קשר

מהו קשר? אם ידיעת ערכו של משתנה אחד משפיעה על ידיעת ערכו של משתנה אחר, אזי ניתן לומר שקיים קשר סטטיסטי בין שני המשתנים. מבחינים בין 3 סוגי קשרים: קשר מלא, קשר חלקי וחוסר קשר מוחלט.

סוגי סולמות

1. סולם שמי – סדרת מספרים עם תכונה אחת: כל מספר מזהה מישור אחד מהאוקלסייה.

דוגמאות: מס' ת.ז, מס' אישי, קו אוטובוס, מקצוע, עיר מגורים, מין, מצב משפחה, מספר שחקן בקבוצת כדורגל, מספר רישוי של מכונית, בטופס מסוים, במקום נקבה רושמים 1 ובמקום זכר רושמים 0.

פעולות מותרות: השוואה. מבחינים: למדא וקרמר.

2. סולם סדר – הערכים המספריים מבטאים סדר טבעי של התכונה הנמדדת.

סדרת מספרים עם 2 תכונות: כל מספר מזהה מישור אחד מהאוקלסייה; לפי המספרים נראה מי יותר גבוה ומי יותר נמוך באוקלסייה.

לפי המספרים לא ניתן לדעת מהו המרחק בין הפריטים באוקלסייה. דוגמא: שביעות רצון, תאור מילולי של ציון, תאור מילולי גוף האדם, איכות. פעולות מותרות: בחינת סדר. מבחן ספירמן

3. סולם רווח – יש משמעות כמותית לערכים המספריים הנמדדים. יש משמעות להפרש אך לא ליחס שבין הערכים.

סדרת מספרים עם 3 תכונות: כל מספר מזהה מישור אחד מהאוקלסייה; לפי המספרים נראה מי יותר גבוה ומי יותר נמוך באוקלסייה; ניתן לראות מה המרחק (ההבדל) בין הפריטים באוקלסייה.

דוגמאות: מה הטמפרטורה כרגע בבית; איך היה השירות, בסולם של 1 עד 7; שנת לידה.

פעולות מותרות: חיבור וחיסור. (אין משמעות לאפס)

איך נזהה סולם רווחים: לפעולת חיבור או חיסור יש משמעות; יכולים להיות מספרים שליליים, ואפשר לקבוע נקודת אפס אחרת.

4. סולם מנה – יש משמעות להפרשים ולאפס המוחלט. יש משמעות כמותית לערכים המספריים הנמדדים. כולל ליחסים ביניהם.

סדרת מספרים עם 4 תכונות: כל מספר מזהה מישור אחד מהאוקלסייה; לפי המספרים נראה מי יותר גבוה ומי יותר נמוך באוקלסייה; לפי המספרים נראה מה המרחק (ההבדל) בין הפריטים באוקלסייה; בכל סולם שנמדד את הפריטים, למספר אפס תהיה אותה משמעות.

דוגמאות: מרחק, מהירות, מחירים, גובה, משקל, וותק בשנים. פעולות מותרות: כפל וחילוק.

איך נזהה סולם מנה: לפעולת חיבור או חיסור יש משמעות; לא יכולים להיות מספרים שליליים, ואי אפשר לקבוע נקודת אפס אחרת. רווח/מנה – פירסון, ניבוי/רגרסיה, אתא

מדדי קשר לסולמות רווח או מנה

התפלגות משותפת של שני משתנים X ו- Y ניתן לתאר בצורה גראפית בדיאגרמת פיזור. דיאגרמה זו מאפשר לנו לגלות אם קיים קשר בין 2 המשתנים ומהו כיוונו (מהו סוג הקשר) מתארת בצורה גרפית במערכת צירים את ההתפלגות המשותפת של שני משתנים X ו- Y . באמצעות הדיאגרמה ניתן לגלות אם קיים איזה קשר שהוא בין X לבין Y , ומהו אופיו?

סוג הקשר

1. קשר קווי (ליניארי, מתאמי): כשהנקודות נמצאות סביב קו ישר דמיוני, עולה או יורד.
2. קשר לא קווי: הנק' נמצאות סביב מבנה ברור מסוים, שאפשר לזהות אותו, אבל לא סביב קו ישר.
3. אין קשר: הנקודות מפוזרות בלי שום מבנה ברור.

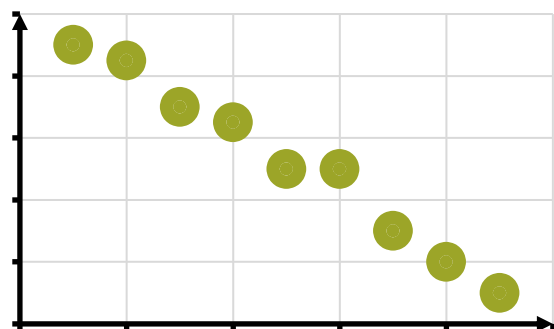
כיוון הקשר

1. קשר חיובי: כאשר X עולה, גם Y עולה (הקו עולה).
2. קשר שלילי: כאשר X עולה, Y יורד (הקו יורד).

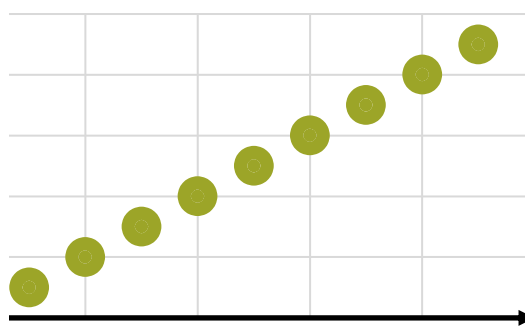
חוזק הקשר

1. קשר מושלם: כל הנקודות יושבות בדיוק על הקו הדימוני.
2. קשר חזק / טוב: הנקודות קרובות מאוד לקו הדימוני.
3. קשר בינוני: הנקודות נמצאות במרחק לא רב מהקו הדימוני.
4. קשר חלש: הנקודות רחוקות מהקו הדימוני.

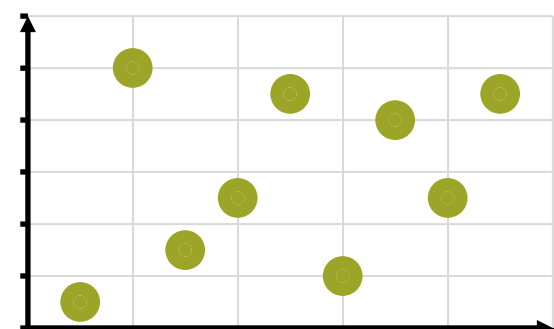
קשר קווי שלילי וחזק



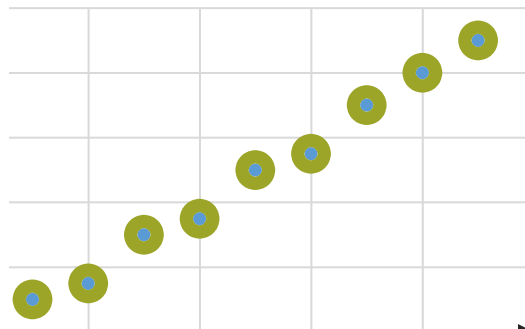
קשר קווי חיובי מושלם



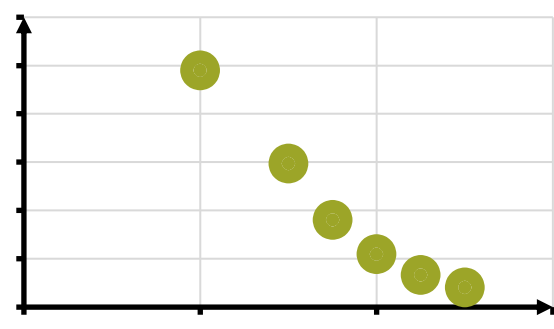
אין קשר



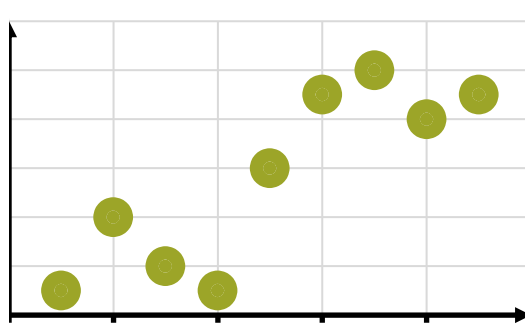
קשר קווי חיובי וחזק



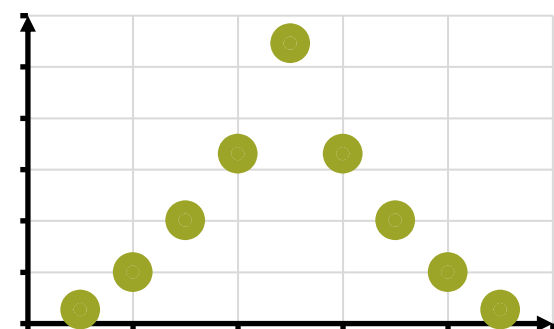
קשר לא קווי



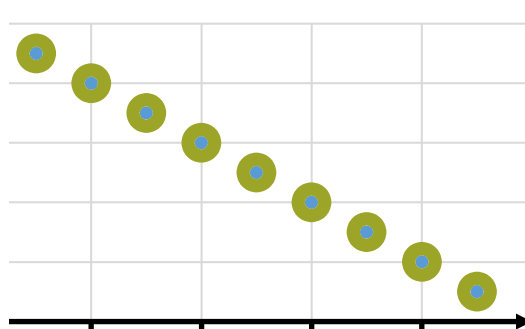
קשר קווי חיובי וחלש



קשר לא קווי



קשר קווי שלילי ומושלם





דיאגרמת פיזור

התפלגות משותפת של שני משתנים X ו- Y ניתן לתאר בצורה גרפית בדיאגרמת פיזור.

דיאגרמה זו מאפשר לנו לגלות אם קיים קשר בין 2 המשתנים ומהו כיוונו (מהו סוג הקשר)

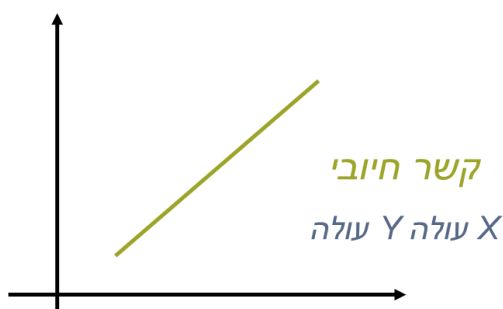


מקדם המתאם של פירסון – r – מדד לקשר קווי

מתאם פירסון נועד למדוד האם יש קשר ליניארי (קווי) בין שני משתנים שסולמם רווח או מנה.

בקשר ליניארי השתנות של משתנה אחד מופיעה עם השתנות כלשהי במשתנה השני ובקצב קבוע.

במתאם פירסון (r_p) יש ייצוג לעוצמת הקשר ולכיוון הקשר.



ניתן לראות בגרף פיזור ככל שהתצפיות בעלות ערכים מפוזרים יותר מהקו הליניארי, כך עוצמת הקשר פחותה יותר.

$$-1 \leq r_p \leq 1$$

$r = 1$ מציינ מתאם חיובי מלא בין המשתנים.

$r = -1$ מציינ מתאם שלילי מלא בין המשתנים.

$r = 0$ מציינ העדר מתאם ליניארי בין המשתנים, קרי אין קשר ליניארי בין המשתנים יתכן קשר מסוג אחר.



שונות משותפת

מתארת את כיוון הקשר של המשתנים (השינוי שחל בו זמנית בשני המשתנים), הפיזור של הנקודות בהתפלגות המשותפת של 2 משתנים.

שונות חיובית (המשתנים מתפתחים באותו הכיוון- גם X וגם Y עולים/יורדים)



שונות שלילית (המשתנים מתפתחים בכיוונים מנוגדים- X עולה Y יורד או להיפך).



$$Cov(x, y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n}$$

$$Cov(x, y) = \frac{\sum x_i \cdot y_i}{n} - \bar{x} \cdot \bar{y}$$

$Cov(x, y) > 0$ יש קשר חיובי בין X ל- Y ✓

$Cov(x, y) < 0$ יש קשר שלילי בין X ל- Y ✓

$Cov(x, y) = 0$ אין קשר בין X ל- Y . שני המשתנים בלתי מתואמים ✓



מקדם המתאם של פירסון

נוסחאות לחישוב

המתאם שווה לממוצע מכפלות ציוני התקן של שני המשתנים.

$$r = \frac{1}{n} \sum_{i=1}^n z_{x_i} \cdot z_{y_i}$$

מנוסחת ציון תקן נראה שניתן לכתוב זאת גם כך:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n s_x s_y}$$

מנוסחת השונות משותפת של X ו- Y - Covariance(x, y) ניתן לכתוב כך:

$$r = \frac{cov(x, y)}{s_x s_y}$$

$$r = \frac{n \cdot \sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{[n \cdot \sum x_i^2 - (\sum x_i)^2][n \cdot \sum y_i^2 - (\sum y_i)^2]}}$$

$$r = \frac{\sum x_i y_i - n \cdot \bar{x} \cdot \bar{y}}{\sqrt{[\sum x_i^2 - n \cdot \bar{x}^2][\sum y_i^2 - n \cdot \bar{y}^2]}}$$

שלבי חישוב r

1. הכנה לחישוב:

א. $\bar{X}$ הממוצע של X ואת הממוצע של Y $\bar{Y}$

ב. חשב את $\sum x_i^2$ ואת $\sum y_i^2$

ג. $\sum x_i y_i$

$$2. r = \frac{\sum x_i y_i - n \cdot \bar{x} \cdot \bar{y}}{\sqrt{[\sum x_i^2 - n \cdot \bar{x}^2][\sum y_i^2 - n \cdot \bar{y}^2]}}$$

ניתן גם לחשב סטיות תקן COV ואז לחשב את r.

$$\text{cov}(x, y) = \frac{\sum x_i \cdot y_i}{n} - \bar{x} \cdot \bar{y} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n}$$

$$r = \frac{\text{cov}(x, y)}{s_x s_y} = \frac{\sum x_i y_i - n \cdot \bar{x} \cdot \bar{y}}{\sqrt{[\sum x_i^2 - n \cdot \bar{x}^2][\sum y_i^2 - n \cdot \bar{y}^2]}}$$

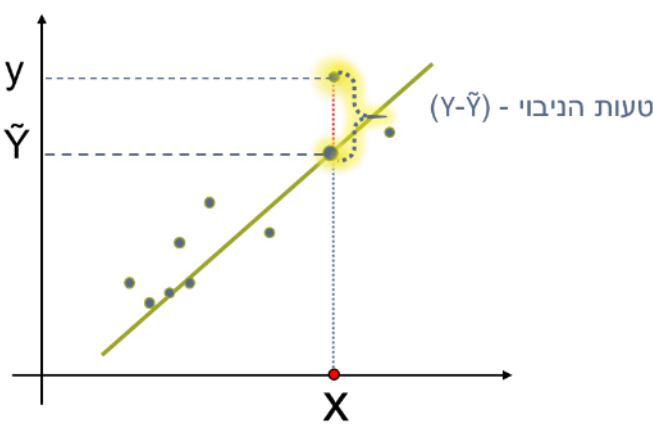


קו הרגרסיה הליניארית

אך מכון שלא כל הנקודות נופלות על קו הרגרסיה (שכן הקשר בין X ל- y אינו תמיד מלא) – יש טעויות ניבוי. הטעות, הסטייה היא טעות הניבוי - $(Y - \tilde{Y})$

עקרון הריבועים הפחותים

הקו המתאר בצורה הטובה ביותר את הנתונים הוא זה שממוצע ריבועי טעויות הניבוי שהוא יוצר יהיה מינימלי – קו הרגרסיה



קו התחזית ל- Y כתלות ב- X (נתון X , מבקשים את Y)

משוואת הקו: $\tilde{y} = b \cdot x + a$

ערך אמיתי - y_i , הניבוי $\tilde{y}_i$

$b = r \frac{S_y}{S_x}$ b – שיפוע הקו

$a = \bar{y} - b \cdot \bar{x}$ a – נק' החיתוך עם ציר ה-y:

קו התחזית ל- X כתלות ב- Y (נתון Y , מבקשים את X)

$b' = \frac{S_x}{S_y} \cdot r$ b' – שיפוע הקו:

$a' = \bar{x} - b' \cdot \bar{y}$ a' – נק' החיתוך עם ציר ה-X:

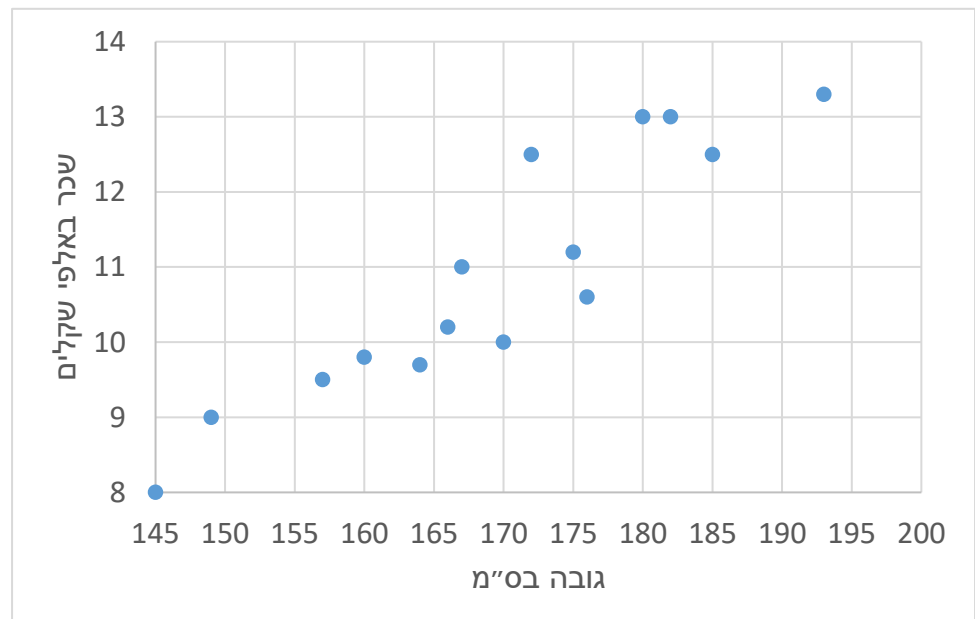
$\tilde{x} = b' \cdot y + a'$ משוואת הקו:

גרסיה פתרון תרגיל עם שימוש באקסל

האם יש קשר בין גובהו של אדם לשכר שהוא מקבל?

נתונים:

y	x
שכר באלפי שקלים	גובה בס"מ
13.3	193
12.5	185
13	182
13	180
11.2	175
10.6	176
12.5	172
10	170
11	167
10.2	166
9.7	164
9.8	160
9.5	157
9	149



משוואת קו הרגרסיה:

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

קו הרגרסיה עובר בנקודת הממוצעים.

שיפוע הישר הוא: $m = r \cdot \frac{s_y}{s_x}$, כאשר r זה מקדם המתאם ו- s_y ו- s_x הן הסטיות תקן של המשתנים y ו- x בהתאמה.

מכאן משוואת קו הרגרסיה היא:

$$y - \bar{y} = r \cdot \frac{s_y}{s_x} (x - \bar{x})$$

בתרגיל שלנו עלינו למצוא את הממוצעים, סטיות התקן ומקדם המתאם כדי למצוא את קו הרגרסיה (במידה והנתונים לא ידועים לנו)

בתרגיל שלנו נמצא את משוואת קו הרגרסיה:

נמצא את הממוצעים, סטיית התקן ומקדם המתאם.

מקדם המתאם r הוא: $r = \frac{1}{n \cdot s_x \cdot s_y} ((x_1 - \bar{x})(y_1 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y}))$

$$r = 1.139$$

לכן השיפוע

$$m = r \cdot \frac{s_y}{s_x} = 1.139 \cdot \frac{1.436}{11.388} = 0.144$$

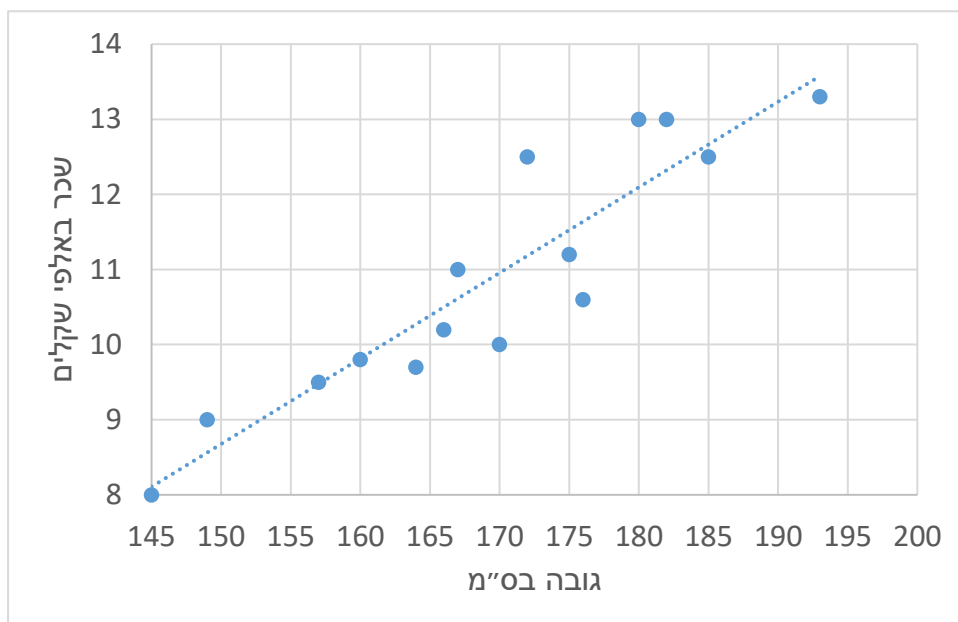
משוואת קו הרגרסיה:

$$y - 10.89 = 0.144(x - 169.4)$$

$$y - 10.89 = 0.144x - 24.3936$$

$$y = 0.144x - 13.5036$$

✓ שרטוט דיאגרמה וקו מגמה



מה צפוי להיות שכרו של אדם (באש"ח) שגובהו 180 ס"מ?

✓ חיזוי בעזרת ישר הרגרסיה

תלמידים נבחנו בשני מבחנים במתמטיקה. להלן התוצאות:

במבחן הראשון: $\bar{x} = 70, s_x = 5$

במבחן השני: $\bar{y} = 75, s_y = 8$

המתאם בין שני המבחנים הוא: $r=0.7$

א. מהי משוואת קו הרגרסיה לניבוי/חיזוי של הציון במבחן השני על פי הציון במבחן הראשון.

ב. תלמיד קיבל במבחן הראשון ציון של 80. העריכו מה יהיה הציון של אותו תלמיד במבחן השני.

ג. מה החיזוי של הציון במבחן השני עבור תלמיד שקיבל 70 במבחן הראשון.

תרגיל

נפלו לידיו תוצאות חלקיות ממחקר על קשר ליניארי בין שני משתנים:

$$\bar{x} = 72, s_x = 5$$

$$\bar{y} = 79, s_y = 8$$

$$r = -0.8$$

חשבו את קו הרגרסיה לניבוי y

האם הניבוי לפי קו הרגרסיה שחישבתם הוא ניבוי "טוב"?

תרגיל

בקרב 300 בני נוער, נבדק הקשר בין גיל לבין מספר הדקות המוקדשות לגלישה ברשתות חברתיות.

הגיל הממוצע של בני הנוער שהשתתפו בסקר: 15.4 שנים, סטיית התקן: 1.2

המספר הממוצע של דקות גלישה ברשתות חברתיות: 75, סטיית התקן: 24

$$r = 0.72$$

נמצא:

א. מצאו את קו הרגרסיה לניבוי מספר דקות הגלישה ברשתות החברתיות על פי גיל הנער או הנערה.

ב. על-פי קו הרגרסיה שמצאתם, מהו משך הזמן שמקדישה נערה בת 16 ו-3 חודשים לגלישה ברשתות החברתיות?

מוסיפים לסקר נתונים שמתייחסים לקבוצת גיל נוספת: 300 חיילים.

הגיל הממוצע של החיילים: 19.5.

המספר הממוצע של גלישה ברשתות החברתיות: 30.

ג. מהם הערכים של ממוצע הגיל וממוצע דקות הגלישה בקבוצה הכוללת 600 הנבדקים?

סטיות התקן בקבוצה הכוללת: סטיית התקן של הגיל: 1, סטיית התקן של דקות הגלישה: 18.

$$\hat{y}_x = \frac{7}{180}x + b$$

ידוע שקו הרגרסיה החדש הוא:

ד. (1) מצאו את הערך של b .

(2) מצאו את הערך של מקדם המתאם החדש.

(3) הסבירו את השינוי שחל במקדם המתאם.



שונות כללית, ניבויים וטעות

שונות הטעויות + שונות הניבויים = שונות כללית (של המשתנה המנובא)

$$S_Y^2 = S_{\hat{Y}}^2 + S_e^2$$

אחוז שונות הניבויים מתוך השונות הכללית (האחוז מהשונות של y שנובע מהשונות ב- X): $r^2 = \frac{\text{שונות הניבויים}}{\text{שונות כללית}}$

אחוז שונות הטעויות מתוך השונות הכללית (האחוז מהשונות של y שלא נובע מהשונות ב- X): $1 - r^2$

שונות הניבויים בניבוי Y מתוך X : $s_{\hat{y}}^2 = r_{y/x}^2 \cdot s_y^2$

שונות הניבויים בניבוי X מתוך Y : $s_{\hat{x}}^2 = r_{x/y}^2 \cdot s_x^2$

שונות הטעויות של Y : $s_{y-\hat{y}}^2 = (1 - r_{y/x}^2) \cdot s_y^2$

שונות הטעויות של X : $s_{x-\hat{x}}^2 = (1 - r_{x/y}^2) \cdot s_x^2$

השפעת טרנספורמציה על מקדם המתאם r

הערך של r לא משתנה אך הסימן (חיובי/שלילי) תלוי בשיפוע של הטרנספורמציה:

אם השיפוע של הטרנספורמציה של X ו- Y באותו סימן (שניהם חיוביים או שניהם שליליים) אז הסימן 

של r לא משתנה.

אם השיפוע של הטרנספורמציה של X ו- Y בסימנים מנוגדים (האחד שלילי והאחר חיובי ולהפך) אז 

הסימן של r יתהפך (מחיובי לשלילי ולהפך).

גרסיה ליניארית ומקדם המיתאם במחשבוני Casio

FC100v FC200v	FX82ES	FX-82MS	FX-82TL	
Stat ↓ exe	Mode 2 2	Mode 3 1		כניסה למצב רגרסיה
Shift stat 3 2	Shift 3 2 =	Shift Mode 1 =	Shift AC =	מחיקת כל הנתונים
		SHIFT 1 3 =	RCL HYP	בדיקת מספר הנתונים שבזיכרון
הקלדת כל ערכי X (אחד מתחת לשני), חזרה למעלה, ימינה והקלדת כל ערכי Y (אחד מתחת לשני). אחרי הקלדת כל המספרים לחץ על AC.		המספר הראשון (X), אחריו פסיק, אחריו המספר השני (Y) ובסוף M+		הקלדת זוג נתונים
Shift stat 7 2	Shift 1 7 2 =	Shift 2 →→2 =	Shift 8 =	b
Shift stat 7 1	Shift 1 7 1 =	Shift 2 →→1 =	Shift 7 =	a
Shift stat 7 3	Shift 1 7 3 =	Shift 2 →→3 =	Shift (=	r
		Mode 1		חזרה למצב רגיל
Shift stat 4 2 exe	Shift 1 4 2 =	Shift 1 2 =		$\sum X$
Shift stat 4 1 exe	Shift 1 4 1 =	Shift 1 1 =		$\sum X^2$
Shift stat 4 4 exe	Shift 1 4 4 =	Shift 1 → 2 =		$\sum Y$
Shift stat 4 3 exe	Shift 1 4 3 =	Shift 1 → 1 =		$\sum Y^2$
Shift stat 4 5 exe	Shift 1 4 5 =	Shift 1 → 3 =		$\sum X*Y$
Shift stat 5 2 exe	Shift 1 5 2 =	Shift 2 1 =	Shift 1 =	$\bar{X}$
Shift stat 5 5 exe	Shift 1 5 5 =	Shift 2 → 1 =	Shift 4 =	$\bar{Y}$
Shift stat 5 3 exe	Shift 1 5 3 =	Shift 2 2 =	Shift 2 =	S_x
Shift stat 5 6 exe	Shift 1 5 6 =	Shift 2 → 2 =	Shift 5 =	S_y

שאלה 15

לבדיקת הקשר בין השכר לבין מספר ימי ההיעדרות מהעבודה בחודש. התקבלו התוצאות הבאות:

$$\sum_{i=1}^8 x_i = 28; \sum_{i=1}^8 y_i = 53,500; \sum_{i=1}^8 x_i \cdot y_i = 171,250; \sum_{i=1}^8 x_i^2 = 130; \sum_{i=1}^8 y_i^2 = 368,990,000$$

כאשר Y – משכורת העובד, X – מספר ימי ההיעדרות.

- א. חשב את עצמת הקשר הליניארי בין השכר לבין מספר ימי ההיעדרות.
- ב. מהו השכר הצפוי לעובד שנעדר בחודש 3 ימים?
- ג. מצא את קו הניבוי לניבוי מספר ימי ההיעדרות לפי משכורת העובד
- ד. מהי שונות הטעויות בניבוי משכורת העובד לפי מספר ימי ההיעדרות בחודש?

שאלה 16

א. כדי לבדוק את הקשר בין מספר החוגים אליהם הולך ילד (X) ומספר השעות בו הוא צופה בטלוויזיה במשך שבוע (Y), נבדקו 8 ילדים. מספר החוגים אליהם הולך כל אחד מהילדים ומספר שעות הצפייה נתון בטבלה הבאה:

4	3	3	2	2	1	1	0	X
8	10	12	16	16	17	18	20	Y

האם קיים קשר לינארי בין מספר החוגים אליהם הולך ילד (X) ומספר השעות בו הוא צופה בטלוויזיה במשך שבוע (Y) ומהי עצמתו?

- ב. מהי שונות הטעויות בניבוי מספר שעות הצפייה בטלוויזיה לפי מספר החוגים?
- ג. מהי שונות הניבויים בניבוי מספר החוגים לפי מספר שעות הצפייה בטלוויזיה?

שאלה 17

על מנת לחקור את הקשר בין ציוני המבחן הפסיכומטרי (Y) לציון הממוצע בבחינות הבגרות (X) נלקח מדגם של 30 תלמידים והתקבלו התוצאות הבאות:

$$\sum_{i=1}^{30} x_i = 240, \sum_{i=1}^{30} x_i^2 = 1,950$$

$$\sum_{i=1}^{30} y_i = 16,800, \sum_{i=1}^{30} y_i^2 = 9,516,000, \sum_{i=1}^{30} x_i y_i = 135,840$$

- (א) חשב את השונות המשותפת בין X ל-Y.
- (ב) חשב את מקדם המתאם הקווי (פירסון) בין X ל-Y.
- (ג) מהי משוואת קו הניבויים לניבוי הציון הפסיכומטרי מממוצע ציוני הבגרות?
- (ד) מהי תחזית הציון במבחן הפסיכומטרי של תלמיד שממוצע ציוני הבגרות שלו הוא 9.28?
- (ה) מהי תחזית ממוצע ציוני הבגרות של תלמיד שציונו במבחן הפסיכומטרי הוא 690?

שאלה 18

על סמך מדגם של 25 עובדים נמצא אומדן לשכר יום עבודה (Y) כפונקציה של מספר שנות הלימוד (X) והוא:

$$\tilde{Y} = 20 + 10X$$

כמו כן ידוע כי שונות השכר ליום עבודה היא 2,500 וממוצע וסטיית התקן של שנות הלימוד הם 8 ו-4 בהתאמה. על-סמך נתונים אלה חשב:

- (א) ממוצע השכר היומי.
- (ב) השונות המשותפת בין X ל-Y.
- (ג) מקדם המתאם של פירסון.
- (ד) משוואת קו הניבויים שבו שכר יום העבודה מנבא את מספר שנות הלימוד.
- (ה) שכרו של מועמד לעבודה שהוא בוגר תיכון (12 שנות לימוד).

שאלה 19

במחקר לבדיקת הקשר בין X – הציון הממוצע בבגרות, לבין Y – הציון במבחן הפסיכומטרי התקבלו התוצאות הבאות במדגם של 50 תלמידים:

$$\sum_{i=1}^{50} x_i^2 = 3250 \sum_{i=1}^{50} x_i = 400$$

$$\sum_{i=1}^{50} y_i^2 = 15,860,000 \sum_{i=1}^{50} y_i = 28,000$$

$$\sum_{i=1}^{50} x_i y_i = 226,550$$

א. האם קיים קשר בין X לבין Y ? נמק.

ב. אם הציון הממוצע בבגרות של שמעון היה 9.5, מה יהיה הניבוי לציוני במבחן הפסיכומטרי?

ג. מהי שונות הטעויות בניבוי הציון במבחן הפסיכומטרי לפי הציון הממוצע בבגרות?

ד. מהי שונות הניבויים בינוי הציון הממוצע בבגרות לפי הציון במבחן הפסיכומטרי?

שאלה 22

העבודות הסמינריונות של סטודנטים נבדקות ע"י שני בודקים.

לבדיקת הקשר בין ציוני שני הבודקים נרשמו ציוני 6 עבודות כפי שניתנו ע"י שני הבודקים:

ציוני בודק א'	ציוני בודק ב'
90	87
99	80
88	82
93	85
86	81
84	95

א. חשב את עצמת הקשר הלינארי בין הציונים.

ב. מהו הניבוי לציון שייתן בודק ב' לעבודה שקיבלה ציון 91 אצל בודק א'?

ג. מהי שונות הניבויים בניבוי ציון שייתן בודק ב' לפי הציון שנתן בודק א'?

שאלה 23

המרכז לבקרת מחלות מעוניין לבדוק את הקשר בין עישון ובין שתיית קפה כחלק מהגורמים המגדילים את הסיכוי למחלות לב.

לשם כך אספו נתונים מ-25 חולי לב על כמות הסיגריות היומית שהם מעשנים (X) ועל מספר כוסות הקפה שהם שותים ביום (Y).

התוצאות שהתקבלו הן:

$$\sum x_i = 450 ; \sum y_i = 88 ; s_x^2 = 175 ; s_y^2 = 2.5 ; \sum x_i \cdot y_i = 2001$$

א. מהי עוצמת הקשר בין X לבין Y ?

ב. מהו הניבוי לכמות הצפויה של כוסות קפה, ששותה אדם חולה לב, אם הוא מעשן ביום 15 סיגריות?

ג. מהי שונות הניבויים בניבוי כמות הקפה היומית לפי מספר הסיגריות?

שאלה 24

במפעל בו מרכיבים מוצרים החליטו לבדוק את הקשר בין הזמן בדקות שלוקח להרכיב מוצר (y) לבין מספר הברגים שעל העובד להבריג בעת הרכבת המוצר (x).

במדגם של 30 מוצרים התקבלו הממצאים הבאים:

$$s_x = 4 ; s_y = 2.5 ; \bar{y} = 9.5 ; \tilde{y} = 0.6x + 3.5$$

א. מה ממוצע מספר הברגים הנדרש להרכבת מוצר ($\bar{x}$)?

ב. האם יש קשר בין מספר הברגים לבין זמן ההרכבה? נמק בעזרת חישוב מקדם המתאם הלינארי.

ג. בהנחה שהזמן בדקות שלוקח להרכיב מוצר מתפלג בקירוב נורמלית עם ממוצע וסטיית תקן כפי שהתקבלו במדגם, מצא:

1. מהו אחוז המוצרים שמשך זמן ההרכבה שלהם עולה על 12 דקות?

2. מהו משך הזמן שרק עבור 10% מהמוצרים לוקח יותר זמן להרכיב אותם?

שאלה 25

חוקר החליט לבדוק את הקשר בין ציון הבחינה בסטטיסטיקה Y – לבין מספר השעות שהקדיש התלמיד ללימוד לבחינה X . – במדגם של 50 סטודנטים התקבלו התוצאות הבאות:

$$\sum_{i=1}^{50} y_i = 4,000; \sum_{i=1}^{50} x_i^2 = 175,000; \sum_{i=1}^{50} y_i^2 = 420,000$$

כמו כן חושב קו הניבוי לניבוי Y לפי X : $\tilde{y} = 1.4x + 10$

א. מה מספר השעות הממוצע שהקדישו התלמידים שנדגמו ללימוד לבחינה?

ב. חשב את עצמת הקשר הלינארי בין X לבין Y .

ג. מהו קו הניבוי לניבוי X לפי Y ?

ד. מהי שונות הטעויות בניבוי הציון לפי מספר השעות שהקדיש התלמיד ללימוד לבחינה?

שאלה 26

חוקר שווקים בדק את הקשר שבין כמות המכירות בחנות Y (בעשרות אלפי ₪)

לבין רמת ההוצאות לקידום מכירות X (באלפי ₪).

לצורך זאת הוא בדק 16 חנויות ב-16 ערים שונות, וקיבל את קו הניבוי הבא: $\tilde{y}_i = 12.7 + 0.7x_i$

כמו כן התקבלו התוצאות הבאות:

$$\bar{x} = 11, \bar{y} = 20.4$$

$$\sum_{i=1}^{16} x_i^2 = 2198.44, \sum_{i=1}^{16} y_i^2 = 6854.56$$

א. על סמך התוצאות מצא את ערכו של מקדם המתאם הלינארי בין X לבין Y .

ב. חשב את שונות הניבויים בניבוי מספר ימי כמות המכירות על סמך ההוצאות. נמק.

ג. חשב את שונות הטעויות בניבוי ההוצאות על סמך כמות המכירות. נמק.

שאלה 28

העבודות הסמינריות של סטודנטים נבדקות ע"י שני בודקים. לבדיקת הקשר בין ציוני שני הבודקים נרשמו ציוני 6 עבודות כפי שניתנו ע"י שני הבודקים:

ציוני בודק א'	ציוני בודק ב'
90	87
99	80
88	82
93	85
86	81
84	95

חשב את עצמת הקשר הלינארי בין הציונים.

שאלה 29

סטודנט לסטטיסטיקה החליט לבדוק אם קיים קשר בין הוותק (בשנים) של מורה פרטי - X ובין השכר שלו לשעת עבודה - Y לצורך כך הוא דגם וראיין 25 מורים פרטיים המופיעים במאגר של אגודת הסטודנטים. להלן התוצאות שקיבל:

$$\bar{x} = 5.2 ; \bar{y} = 62.5 ; \sum_{i=1}^{25} x_i^2 = 1,150 ; \sum_{i=1}^{25} y_i^2 = 102,500 ; \sum_{i=1}^{25} x_i \cdot y_i = 9,350$$

- א. מהו הניבוי לשכר לשעה של מורים פרטיים בעלי ותק של 8 שנים? נמק.
- ב. מהו גודלה של שונות הניבויים בניבוי Y לפי X ?
- ג. מהו קו הניבוי לניבוי X לפי Y ?

שאלה 30

כדי לבדוק אם קיים קשר בין התפוקה היומית (Y) למספר העובדים (X) נלקח מדגם של 60 תצפיות והתקבלו התוצאות הבאות:

$$\sum_{i=1}^{60} x_i = 300 ; \sum_{i=1}^{60} y_i = 720 ; \sum_{i=1}^{60} x_i \cdot y_i = 4032$$

$$\sum_{i=1}^{60} x_i^2 = 1740 ; \sum_{i=1}^{60} y_i^2 = 9600$$

- חשב את עצמת הקשר הלינארי בין התפוקה היומית לבין מספר העובדים.
- מהי התחזית לתפוקה היומית במפעל שבו מועסקים 8 עובדים?
- בהנחה שהתפוקה היומית מתפלגת בקירוב נורמלית עם ממוצע וסטיית תקן כפי שהתקבלו במדגם, מצא:
 - מהו אחוז המקרים שבהם התפוקה היומית היא בין 10 לבין 17 יחידות?
 - מהו הרבעון העליון של התפוקה היומית?

שאלה 31

במחקר לבדיקת הקשר בין X – גיל נישואי האישה, לבין Y – מספר ילדיה, התקבלו התוצאות הבאות לגבי 10 נשים נשואות:

$$\bar{x} = 24 ; \bar{y} = 3 ; \sum_{i=1}^{10} (x_i - \bar{x})^2 = 36 ; \sum_{i=1}^{10} (y_i - \bar{y})^2 = 25$$

- קו הריבועים הפחותים לניבוי מספר הילדים לפי גיל נישואי האם, חושב על סמך התוצאות: $\hat{y} = 21 - 0.75x$
- האם קיים קשר ליניארי בין X ל- Y ? נמק.
 - מה יהיה הניבוי למספר הילדים שיהיו לאישה שנישאה בגיל 25?

שאלה 32

במחקר לבדיקת הקשר בין Y – ההוצאה החודשית על מסעדות ובילויים במשפחה, לבין X – מס הילדים במשפחה, התקבלו התוצאות הבאות במדגם שכלל 10 משפחות:

$$\sum_{i=1}^{10} x_i = 18 ; \sum_{i=1}^{10} y_i = 2220 ; \sum_{i=1}^{10} x_i \cdot y_i = 4720$$

$$\sum_{i=1}^{10} x_i^2 = 48 ; \sum_{i=1}^{10} y_i^2 = 534,200$$

- האם קיים קשר ליניארי בין X ו- Y ? נמק.
- אם ההוצאה החודשית על מסעדות ובילויים במשפחת ישראלי היתה 250 ש"ח, מה יהיה הניבוי למספר הילדים?
- במשפחת כרמל יש 4 ילדים, מהו הניבוי להוצאה החודשית על מסעדות ובילויים משפחתיים של משפחת כרמל?

שאלה 33

בבדיקת הקשר הליניארי בין X ל- Y נתקבלו הנתונים הבאים:

$$\bar{x} = 10, \bar{y} = 50, S_x = 1, S_y = 2, r = 0.8$$

- מצא את קו הניבוי לניבוי Y מתוך X .
- מצא את קו הניבוי לניבוי X מתוך Y .
- חשב את גודל שונות הניבויים של Y .
- חשב את גודל שונות הטעויות בניבוי X .

שאלה 34

בקורס מסוים נבדק מספר המטלות שהגישו הסטודנטים והתקבלה התפלגות השכיחויות הבאה:

מספר מטלות	מספר הסטודנטים
0	22
1	36
2	56
3	44
4	30
5	12

א. מהם השכיח, החציון והממוצע של התפלגות מספר המטלות?

ב. מהו ממוצע הסטיות המוחלטות מהחציון בהתפלגות הנתונה, ומהי סטיית התקן?

ג. לבדיקת הקשר בין מספר המטלות והציון הסופי בקורס נבדק בנוסף למספר המטלות שהגישו גם הציון הסופי של הסטודנטים בקורס.

מתוך הנתונים חושב קו הניבוי לניבוי הציון הסופי בקורס לפי מספר המטלות שהוגשו והתקבל:

$$\hat{y}_i = 13x_i + 44.1 \quad \text{כאשר } y - \text{הציון הסופי בקורס ו- } x \text{ מספר המטלות. כמו כן נמצא כי } s_y = 22.$$

מהי עוצמת הקשר הליניארי בין x לבין y ?

שאלה 35

במחקר שנעשה על הקשר בין מספר ימי היעדרות מהעבודה בשנה (y) לבין מספר שנות הוותק (x) התקבלו התוצאות הבאות:

$$\bar{x} = 7.6$$

$$\bar{y} = 12$$

$$S_x = 2.2$$

$$S_y = 5$$

$$\text{cov}(x, y) = -8.36$$

$$n = 25$$

1. חשב את שונות הניבויים בניבוי מספר ימי היעדרות מהעבודה על סמך מספר שנות הוותק. נמק.

2. חשב את שונות הטעויות בניבוי מספר שנות הוותק על סמך מספר ימי היעדרות מהעבודה.

שאלה 36

במחקר שנעשה על הקשר בין מספר ימי היעדרות מהעבודה בשנה (y) לבין מספר שנות הוותק (X) התקבלו התוצאות הבאות:

$cov(x, y) = -8.36$
 $n = 25$

$S_x = 2.2$
 $S_y = 5$

$\bar{x} = 7.6$
 $\bar{y} = 12$

- מהו הניבוי למספר שנות ותק של עובד שנעדר מעבודתו 9 ימים? נמק.
- חשב את שונות הניבויים בניבוי מספר ימי היעדרות מהעבודה על סמך מספר שנות הוותק. נמק.
- חשב את שונות הטעויות בניבוי מספר שנות הוותק על סמך מספר ימי היעדרות מהעבודה. נמק.

שאלה 37

להלן ציוני חמישה אנשים שנבחנו במבחן זריזות ידיים, X ציון ביד ימין, ו-Y ציון ביד שמאל (Y).

X	Y
2	1
5	8
4	2
7	4
6	5

ענו על הסעיפים הבאים:

- אנא חשבו מתאם פירסון בין הציונים ביד ימין לשמאל.
- ציונו של אדם ביד ימין הוא 6, מהו ציונו המנובא ביד שמאל?
- ציונו של אדם ביד שמאל הוא 3, מהו ציונו המנובא ביד ימין?
- מהו אחוז השונות המוסברת בניבוי Y ע"פ X, ולהיפך?
- מהי שונות הניבויים בניבוי Y ע"פ X?
- מהי שונות הטעויות בניבוי Y ע"פ X?
- מהי שונות הניבויים בניבוי X ע"פ Y?
- מהי שונות הטעויות בניבוי X ע"פ Y?

שאלה 38

כדי לבדוק את יעילות העבודה במפעל נלקחו 5 עובדים ונבדק לגביהם השכר והוותק:

15	15	20	25	30	שכר חודשי (אלפי שקלים)
3	4	2	5	10	מספר שנות ותק

- א. האם יש קשר לינארי בין הוותק לשכר? נמק בעזרת חישוב המדד המתאים
- ב. מהו השכר הצפוי לעובד בעל 8 שנות ותק?

שאלה 46

מכון למחקר רפואי בדק את הקשר בין עישון לשתיית קפה בקרב חולי לב.

במחקר נאספו נתונים מ-25 חולי לב על כמות הסיגריות היומית שהם מעשנים (X) ועל מספר כוסות הקפה שהם שותים ביום (Y).

להלן ריכוז התוצאות שהתקבלו:

$\sum_{i=1}^{25} x_i = 450$, $\sum_{i=1}^{25} y_i = 88$, $\sum_{i=1}^{25} x_i^2 = 12,476$, $\sum_{i=1}^{25} y_i^2 = 372$, $\sum_{i=1}^{25} x_i y_i = 2001$

א. מהי עוצמת הקשר בין X לבין Y ? נמק.

- ב. מהו הניבוי לכמות הצפויה של כוסות קפה, ששותה אדם חולה לב, אם הוא מעשן ביום 15 סיגריות?
- ג. חשב את שונות הניבויים בניבוי כמות הקפה היומית לפי מספר הסיגריות.
- ד. חשב את שונות הטעויות בניבוי כמות הסיגריות לפי מספר כוסות הקפה.

שאלה 49

כדי לבדוק את יעילות העבודה במפעל נלקחו 5 עובדים ונבדק לגביהם השכר והוותק:

15	15	20	25	30	שכר חודשי (אלפי שקלים)
3	4	2	5	10	מספר שנות ותק

האם יש קשר לינארי בין הוותק לשכר? נמק בעזרת חישוב המדד המתאים

שאלה 51

חוקר התעניין בקשר שבין מספר שעות העבודה השבועיות – X של סטודנט באוניברסיטה הפתוחה, לבין הישגיו בקורס סטטיסטיקה – Y .

הוא רשם עבור 10 סטודנטים שסיימו את לימודיהם, את מספר שעות העבודה שהם עובדים בשבוע, ואת ציונם בקורס. החוקר קיבל את הנתונים האלה:

$$\sum_{i=1}^{10} x_i = 320, \quad \sum_{i=1}^{10} y_i = 830, \quad \sum_{i=1}^{10} x_i^2 = 14,700, \quad \sum_{i=1}^{10} y_i^2 = 84,025, \quad \sum_{i=1}^{10} x_i y_i = 20,810$$

א. האם קיים קשר ליניארי בין X לבין Y ? נמק.

ב. מהו הניבוי לציון בסטטיסטיקה של סטודנט העובד בשבוע 30 שעות?

ג. סטודנט קיבל בבחינה בסטטיסטיקה 85. לאור התוצאות, מהו הניבוי למספר שעות העבודה שלו בשבוע?

שאלה 52

חוקר התעניין בקשר שבין מספר שעות העבודה השבועיות – X של סטודנט באוניברסיטה הפתוחה, לבין הישגיו בקורס סטטיסטיקה – Y .

הוא רשם עבור 10 סטודנטים שסיימו את לימודיהם, את מספר שעות העבודה שהם עובדים בשבוע, ואת ציונם בקורס. החוקר קיבל את הנתונים האלה:

$$\sum_{i=1}^{10} x_i = 320, \quad \sum_{i=1}^{10} y_i = 830, \quad \sum_{i=1}^{10} x_i^2 = 14,700, \quad \sum_{i=1}^{10} y_i^2 = 84,025, \quad \sum_{i=1}^{10} x_i y_i = 20,810$$

א. האם קיים קשר ליניארי בין X לבין Y ? נמק.

ב. מהו הניבוי לציון בסטטיסטיקה של סטודנט העובד בשבוע 30 שעות?

ג. סטודנט קיבל בבחינה בסטטיסטיקה 85. לאור התוצאות, מהו הניבוי למספר שעות העבודה שלו בשבוע?

ד. על סמך תוצאות המחקר, מהי שונות הניבויים בניבוי ההישגים בקורס סטטיסטיקה לפי מספר שעות העבודה?

טענות



שאלה 20

עבור 25 תלמידים נרשמו ציוני התנ"ך וציוני ההיסטוריה שלהם ונמצא כי ממוצע שני המבחנים זהה והוא 75 וכי סטיית התקן של שני המבחנים זהה והיא 12. מכאן ניתן לומר ש:

א. ערכו של המתאם הליניארי בין שני הציונים הוא 1.

ב. ערכו של המתאם הליניארי בין שני המבחנים קטן מ-1.

ג. ערכו של המתאם הליניארי בין שני המבחנים חיובי.

ד. ערכו של המתאם הליניארי בין שני המבחנים שלילי.

ה. לא ניתן לדעת מה ערכו של מקדם המתאם.



שאלה 42

יבול העגבניות (בטונות) במשק מסוים נדגם במשך 10 שנים. סטיית התקן של היבול הייתה 2 טון. ההכנסה (Y) מהעגבניות תלויה ביבול (X) לפי הקשר הבא: $Y = 3X - 2$. שונות ההכנסה מהעגבניות במדגם הנ"ל היא:



שאלה 43

יבול העגבניות (בטונות) במשק מסוים נדגם במשך 10 שנים. סטיית התקן של היבול הייתה 2 טון. ההכנסה (Y) מהעגבניות תלויה ביבול (X) לפי הקשר הבא: $Y = 3X - 2$. ערכו של מקדם המתאם הליניארי בין ההכנסה והיבול הוא:



שאלה 47

חוקר בדק קשר בין X – הציון בבחינה בסטטיסטיקה לבין Y – הציון בבחינה בכלכלה. קו הניבוי לניבוי Y לפי X שהתקבל הוא: $\hat{y} = 1.6x - 32$. כמו כן קו הניבוי לניבוי X לפי Y שהתקבל הוא: $\hat{x} = 0.4y + 38$. לאור הנתונים שהתקבלו, ערכו של מקדם המתאם הליניארי בין X לבין Y הוא:



שאלה 63

השכר הממוצע של עובדי חברה מסוימת הוא 7,500 ₪ וסטיית התקן היא 700 ₪. בחברה הוחלט להעלות את משכורות כל העובדים ב-10%.

ציון התקן של משכורתו של דני לפני ההעלאה היה 1.5 .

לאחר ההעלאה ציון התקן של המשכורת של דני לא ישתנה.

שאלה 69

במחקר לבדיקת הקשר בין X – גיל הנשואין של האישה, לבין Y – מספר ילדיה התקבלו התוצאות הבאות לגבי 10 נשים

$$\sum_{i=1}^{10} (x_i - \bar{x})^2 = 36, \quad \sum_{i=1}^{10} (y_i - \bar{y})^2 = 25$$

כמו כן חושב קו החיזוי לניבוי מספר הילדים לפי גיל נישואי האם והתקבל: $\tilde{y}_i = 21 - 0.75x_i$

לכן מקדם המתאם הליניארי בין X לבין Y ערכו -0.9 .

שאלה 75

חוקר שווקים בדק את הקשר שבין כמות המכירות בחנות Y (בעשרות אלפי ₪)

לבין רמת ההוצאות לקידום מכירות X (באלפי ₪).

לצורך זאת הוא בדק 16 חנויות ב- 16 ערים שונות, וקיבל את קו הניבוי הבא: $\tilde{y}_i = 12.7 + 0.7x_i$

כמו כן התקבלו התוצאות הבאות:

$$\bar{x} = 11, \quad \bar{y} = 20.4$$

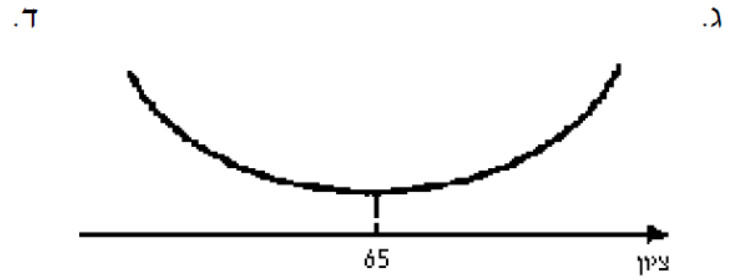
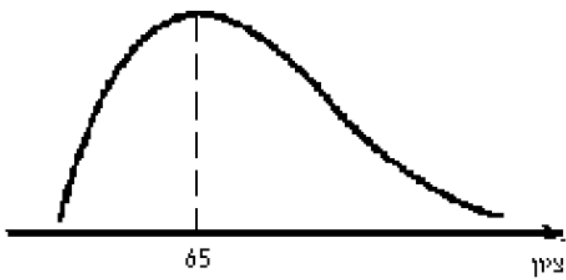
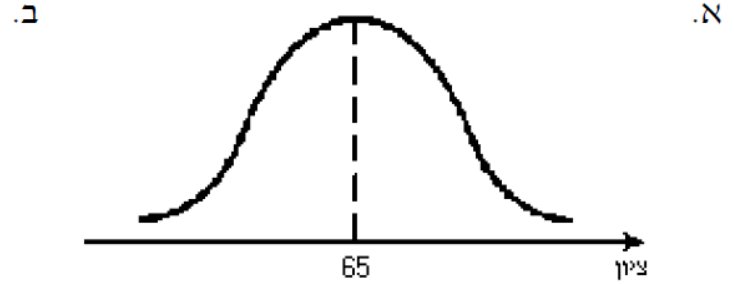
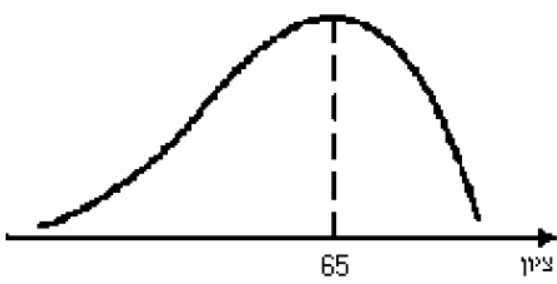
$$\sum_{i=1}^{16} x_i^2 = 2198.44, \quad \sum_{i=1}^{16} y_i^2 = 6854.56$$

לאור הנתונים שהתקבלו, ערכו של מקדם המתאם הליניארי בין X לבין Y הוא:

שאלה 81

עבור התפלגות הציונים בבחינת מיון באנגלית ידוע כי $Mo=65$ ו- $\bar{X} < Mo$.

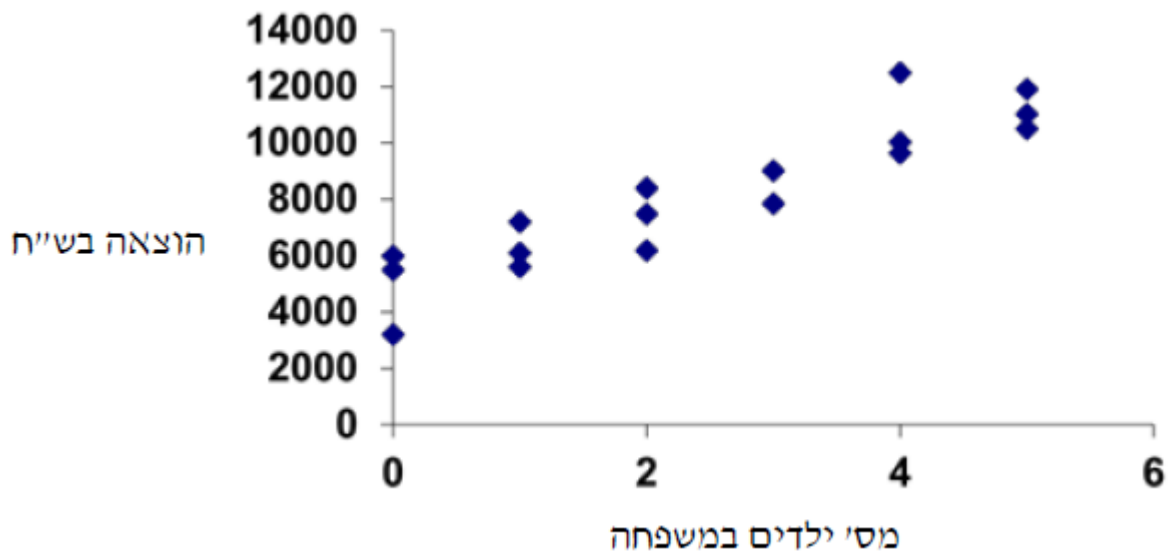
עקומה אפשרית של התפלגות הציונים היא:



שאלה 83

במסגרת מפקד האוכלוסין נאספו נתונים לגבי ההוצאה החודשית ומספר הילדים של משפחות.

בדיאגרמת הפזור הבאה מובאים נתונים המתייחסים ל- 17 משפחות.



להלן 3 טענות על מקדם המתאם של פירסון (בין מספר הילדים במשפחה ובין ההוצאה החודשית שלה.

1. $r=1$.

2. $r>0$.

3. אם נתוני ההוצאה החודשית היו נמדדים באלפי ש"ח ולא בש"ח מקדם המתאם לא היה משתנה.

הטענות הנכונות הן:



שאלה 92

לבדיקת הקשר בין הציון בכלכלה (X) לבין הציון בסטטיסטיקה (Y) נבדקו ציוני 40 סטודנטים, שנבחנו בשני המקצועות.

התבלו התוצאות האלה:

$$s_x = 8, s_y = 12, cov(x, y) = 69.12$$

לכן שונות הניבויים בניבוי הציון בסטטיסטיקה לפי הציון בכלכלה היא 46.08.