# ביוסטטיסטיקה

## פרק 15 - מקדם המתאם ( מדד קשר ) הלינארי ומובהקותו

### תוכן העניינים


1. מקדם המתאם הלנארי ( פירסון) ..... 1
2. חישוב מקדם המתאם הלנארי (פירסון) ..... 12
3. בדיקת השערות על מקדם המתאם הלנארי ..... 17
4. ניתוח פלטים על מקדם המתאם הלנארי ..... 21

# מקדם המתאם (מדד קשר) הלינארי ומובהקותו

## מדד הקשר הלינארי (פירסון) – מבוא

מעוניינים לבדוק עד כמה קיים קשר מסוג קשר לינארי (קו ישר) בין שני משתנים.
שני המשתנים שאנו בודקים לגביהם קשר צריכים להיות משתנים כמותיים.
מבחינת סולמות מדידה כל משתנה נחקר צריך להיות מסולם רווחים או מנה.
בדרך כלל המשתנה המוצג כ- $Y$ הוא המשתנה התלוי והמשתנה המוצג ב- $X$ הוא המשתנה הבלתי תלוי.
תיאור גרפי לנתונים נעשה על ידי דיאגרמת פיזור. בדיאגרמת פיזור אנחנו מסמנים כל תצפית בנקודה לפי שיעור ה- $X$ ושיעור ה- $Y$ שלה.
דיאגרמת הפיזור נותנת אינדיקציה גרפית על הקשר בין שני המשתנים.

**דוגמה (פתרון בהקלטה) :**

בבניין 8 דירות בדקו לכל דירה את מספר החדרים שלה וכמו כן את מספר הנפשות הגרות בדירה. להלן התוצאות שהתקבלו :

| מספר חדרים בדירה | 2 | 2 | 3 | 2 | 3 | 3 | 4 | 4 |
|---|---|---|---|---|---|---|---|---|
| מספר הנפשות בדירה | 0 | 1 | 2 | 2 | 2 | 3 | 4 | 5 |

**1)** כמה תצפיות ישנן בדוגמה?

**2)** כמה משתנים ישנם בדוגמה, מי הם?

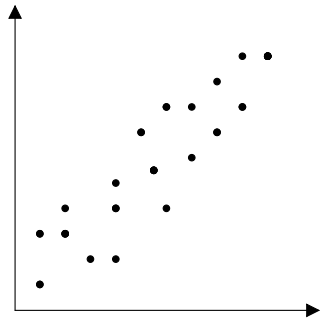
**3)** שרטטו לנתונים דיאגרמת פיזור.

**4)** מי המשתנה התלוי ומיהו המשתנה הבלתי תלוי?

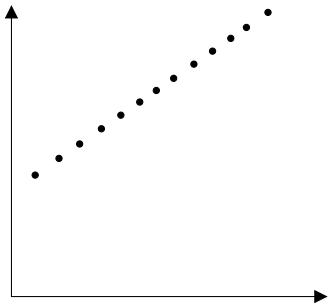

© כל הזכויות שמורות לברק קנדל

## דיאגרמות פיזור לקשר בין משתנים וניתוחם

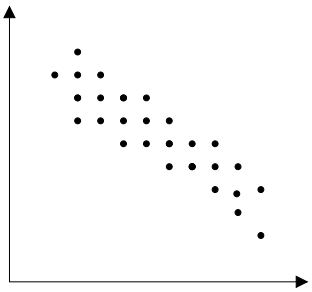
**קשר לנארי חיובי חלקי**



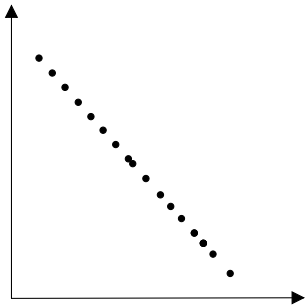
**קשר לנארי חיובי מלא**



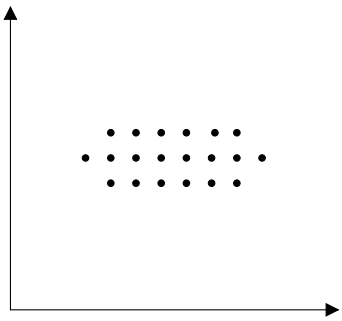
**קשר לינארי שלילי חלקי**



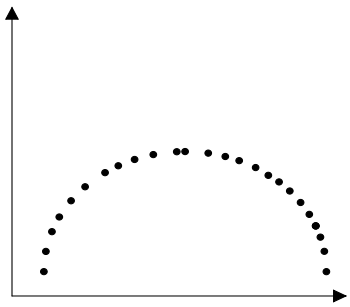
**קשר לינארי שלילי מלא**



**אין קשר**



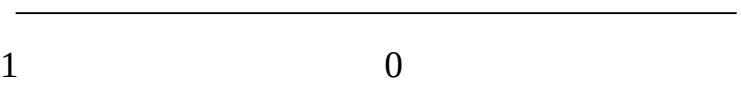
**אין קשר לינארי**



**משמעות מקדם המתאם:**

כדי לבדוק עד כמה קיים קשר לינארי בין שני המשתנים ישנו מדד קשר שנקרא גם מקדם המתאם הלינארי הידוע גם בשם מקדם המתאם של פירסון.

מקדם מתאם זה מקבל ערכים בין 1 ל-1.





© כל הזכויות שמורות לברק קנדל

מקדם מתאם 1-או 1 אומר שקיים קשר לינארי מלא בין המשתנים שניתן לבטאו על ידי נוסחה של קו ישר: $y = ax + b$.

**מתאם חיובי מלא (מקדם מתאם 1):**

קיים קשר לנארי מלא בו השיפוע $a$ יהיה חיובי ואילו מתאם שלילי (מקדם מתאם-1) מלא אומר שקיים קשר לנארי מלא בו השיפוע $a$ שלילי.

**מתאם חיובי חלקי:**

ככל שמשתנה אחד עולה לשני יש נטייה לעלות בערכו אבל לא קיימת נוסחה לינארית שמקשרת את $X$ ל- $Y$ באופן מוחלט ואילו מתאם שלילי חלקי אומר שככל שמשתנה אחד עולה לשני יש נטייה לרדת אבל לא קיימת נוסחה לינארית שמקשרת את $X$ ל- $Y$ באופן מוחלט. ככל שמקדם המתאם קרוב לאפס עוצמת הקשר יותר חלשה וככל שהמדד רחוק יותר מהאפס העוצמה יותר חזקה. לסיכום, מקדם המתאם בודק את עוצמת הקשר הלנארי, ואת כיוון הקשר.

מקדם המתאם הלינארי אינו מושפע מיחידות המדידה. כל שינוי ביחידות המדידה של המשתנים, לא ישנה את מקדם המתאם.

מדד הקשר הלינארי באוכלוסייה, שנקרא גם מקדם המתאם של פירסון או מדד הקשר של פירסון באוכלוסייה מסומן ב: $\rho$ - פרמטר המאפיין את עוצמת הקשר הלינארי באוכלוסייה וכיוונו בין שני המשתנים הנחקרים. כאשר:

$r$ - מדד הקשר הלינארי במדגם שמהווה אומד לפרמטר $\rho$.

קיומו של מתאם בין שני משתנים אינו מצביע על סיבתיות בהכרח. למשל, אם נמצא מתאם חיובי בין כמות הסוכרזית שאדם אוכל לבין במשקל שלו אין זה אומר שהסיבה להשמנה היא הסוכרזית. מדד הקשר של פירסון הוא מדד קשר סימטרי, כלומר אם נחליף את $X$ ב- $Y$ התוצאה תהיה זהה.

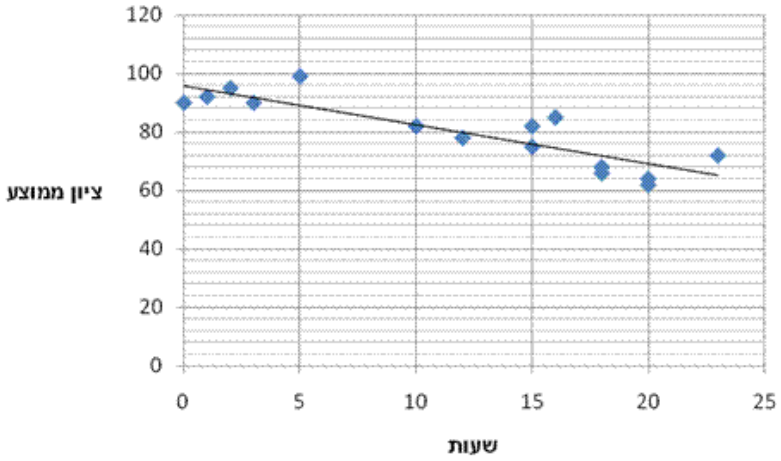
**דוגמה (פתרון בהקלטה):**

- מה ניתן להגיד על מקדם המתאם של שני המשתנים על סמך דיאגרמת הפיזור ששרטטנו?
- אם היינו משנים את השרטוט כך שבציר האנכי היה המשתנה "מספר החדרים" ובציר האופקי היה "מספר הנפשות", האם הדבר היה משפיע על מדד הקשר של פירסון?


© כל הזכויות שמורות לברק קנדל

## שאלות

**1)** חוקר רצה לאפיין את הקשר בין מספר השעות בשבוע שסטודנט מקדיש לבילויים לבין הציון הממוצע שלו בסוף הסמסטר. לשם כך הוא אסף נתונים של 15 סטודנטים ויצר דיאגרמת פיזור:




א. מיהו המשתנה הבלתי תלוי?

ב. מה ניתן לומר על כיוון הקשר בין מספר שעות הבילוי השבועיות לבין הציון הממוצע של הסמסטר? מה ניתן להגיד על עוצמת הקשר?

**2)** להלן טבלה המסכמת את מקדמי המתאם הלינארי בין ציוני מבחנים שונים שהתקבלו עבור תלמידים בכיתה מסוימת:

| | ספורט | לשון | מתמטיקה |
|---|---|---|---|
| ספורט | ? | -0.7 | ? |
| לשון | ? | ? | 0.6 |
| מתמטיקה | -0.1 | ? | ? |

א. השלימו את מקדמי המתאם שמסומנים בסימן שאלה בטבלה.

ב. בין אילו שני ציוני מקצועות שונים קיים מתאם בעל העוצמה החזקה ביותר?

**3)** במחקר נתבקשו לבדוק את הקשר בין מספר שעות התרגול של קורס לבין הציון הסופי שלו. להלן תוצאות מדגם שהתקבל:

| שעות תרגול | ציון סופי |
|---|---|
| 20 | 90 |
| 25 | 90 |
| 30 | 95 |
| 15 | 60 |
| 30 | 90 |
| 20 | 85 |
| 10 | 50 |

א. מיהו המשתנה התלוי ומיהו המשתנה הבלתי תלוי בדוגמה זו?

ב. שרטטו דיאגרמת פיזור לנתונים.

ג. מה ניתן לומר על הקשר בין המשתנים במדגם?

ד. מסתבר שבסופו של דבר נתנו פקטור של 5 נקודות לציון הסופי. כיצד הדבר היה משנה את מקדם המתאם של המדגם?

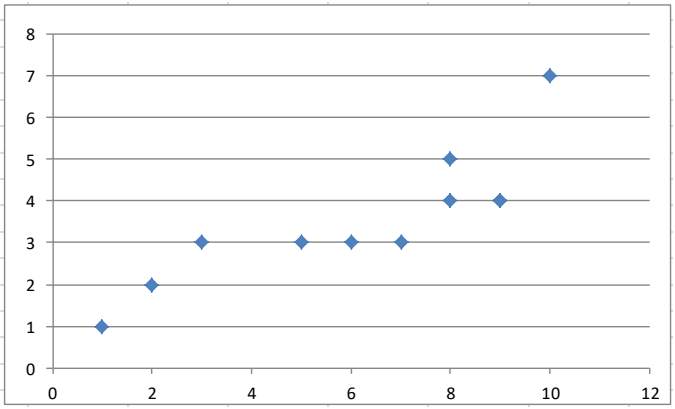

© כל הזכויות שמורות לברק קנדל

**4)** בתחנה המטאורולוגית רצו לבדוק את הקשר שבין הטמפרטורה במעלות צלזיוס לכמות המשקעים במ״מ. הם אספו נתונים על 10 ימים במהלך חודש ינואר. המתאם שהתקבל היה 0.8.

א. השלימו את המשפט:
בחודש ינואר ככל שהטמפרטורה היומית נוטה לרדת, כך כמות המשקעים נוטה _____.

ב. הוחלט להעביר את הטמפרטורה למעלות פרנהייט על מנת שיוכלו להשוות אותה לנתונים מארה״ב. נוסחת המעבר היא $F^0 = 32 + \frac{9}{5}C^0$.
כיצד הדבר ישפיע על מקדם המתאם בין הטמפרטורה במעלות פרנייט לכמות המשקעים במ״מ?

**5)** להלן דיאגרמת פיזור המראה קשר בין שני משנים:




א. השלימו: ניתן לראות שהקשר הוא לינארי _____ (מלא\ חלקי) כיוון הקשר הוא (חיובי \שלילי).

ב. השלימו: אם היינו מוסיפים תצפית שערך ה- $X$ שלה הוא 4 וערך ה- $Y$ שלה הוא 7, מקדם המתאם של פירסון היה _____ (גדל\ קטן\ לא משתנה).

**שאלות רב ברירה (יש לבחור את התשובה הנכונה):**

**6)** חוקר אקלים דגם כמה ימים בשנה ומדד את הטמפרטורה בטורונטו שבקנדה ואת הטמפרטורה בסידני שבאוסטרליה באותו היום. הוא חישב ומצא מקדם מתאם שלילי בין הטמפרטורה היומית בטורונטו לבין הטמפרטורה היומית בסידני. משמעות מקדם המתאם השלילי במדגם:

א. אין קשר בין הטמפרטורה בטורונטו לבין הטמפרטורה בסידני בימים שנדגמו.

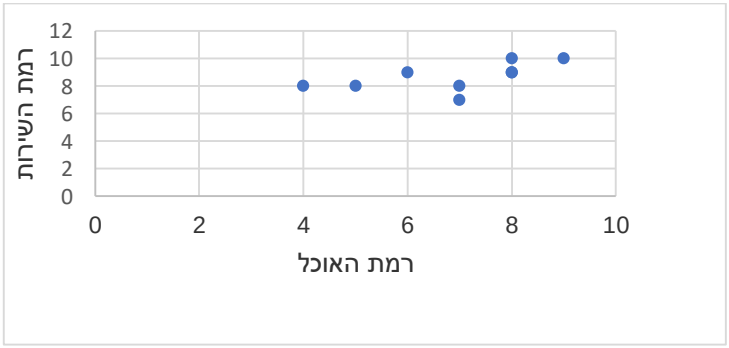
ב. במדגם, רוב הטמפרטורות בטורונטו היו שליליות.

ג. ההפרש בין הטמפרטורה בטורונטו לבין הטמפרטורה באוסטרליה, במדגם זה, הוא שלילי.

ד. במדגם יש נטייה שהטמפרטורה יורדת בטורונטו לטמפרטורה לעלות בסידני.


© כל הזכויות שמורות לברק קנדל

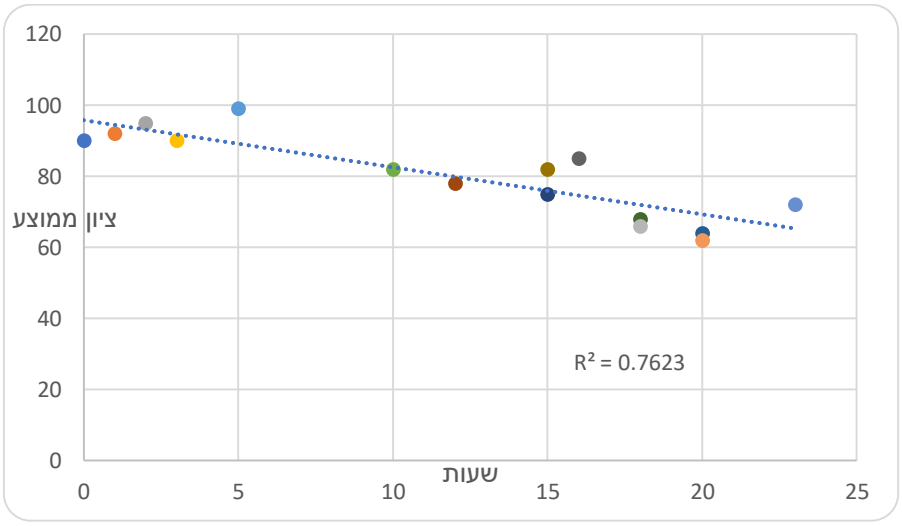
**7)** בסקר שביעות רצון שנערך בבית הקפה "פת לחם" התבקשו הלקוחות לדרג את מידת שביעות הרצון שלהם (בסולם 1-10) בשני נושאים: רמת האוכל ורמת השירות.




מה יהיה ערכו של מקדם המתאם ($r$)?

א. $r=-0.3$
ב. $r=0$
ג. $r=1.125$
ד. $r=0.593$

**8)** חוקר רצה לאפיין את הקשר בין מספר השעות בשבוע שסטודנט מקדיש לבילויים לבין הציון הממוצע שלו בסוף הסמסטר. לשם כך הוא אסף נתונים של 15 סטודנטים ויצר דיאגרמת פיזור.

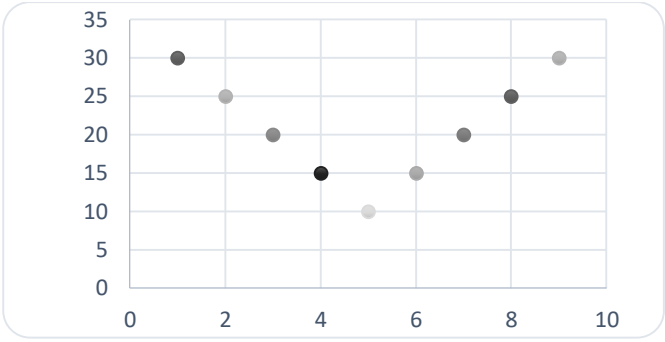



מה ניתן לומר על כיוון הקשר במדגם בין מספר שעות הבילוי השבועיות לבין הציון הממוצע של הסמסטר?

א. ככל שמבלים יותר הציון נוטה לרדת.
ב. אין קשר בין שעות הבילוי לציון.
ג. ככל שמבלים פחות הציון נוטה לרדת.
ד. ככל שהציון נוטה לרדת הסטודנט מבלה פחות.


© כל הזכויות שמורות לברק קנדל

**9)** התרשים הבא מתאר קשר בין שני משתנים, איזה מהמתאמים הבאים הוא המתאים ביותר לתיאור הקשר בין שני המשתנים?



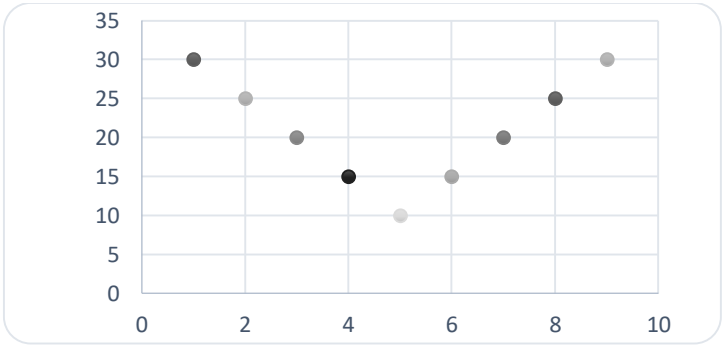

א. $r = 1$ היות ושני המשתנים יוצרים קוים ישרים.

ב. $r = 2$ היות ויש שני קוים בעלי קשר מושלם.

ג. $r = 0$ היות והקו יורד ואחר כך עולה באותו האופן.

ד. $r = \pm 1$ היות ויש קו עולה וגם קו יורד.

**10)** התרשים הבא מתאר דיאגרמת פיזור.




איזו טענה נכונה?

א. בתרשים מוצג הקשר בין שני משתנים.

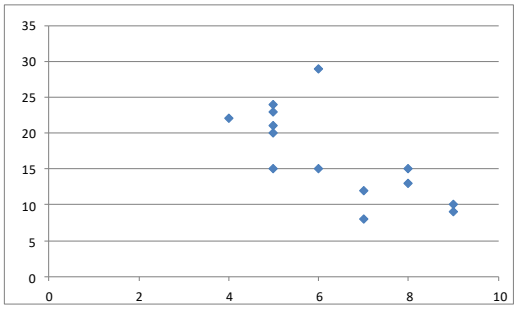
ב. בתרשים מוצג הקשר בין 9 משתנים.

ג. בתרשים מוצג הקשר בין 10 משתנים.

ד. אין לדעת כמה משתנים מוצגים בתרשים.


© כל הזכויות שמורות לברק קנדל

בגרף הבא מתוארת דיאגרמת פיזור של שני משתנים:




$X$ - (משתנה בלתי תלוי בציר האופקי)
ו- $Y$ (משתנה תלוי).
במדגם התקבל $r^2 = 0.52$.

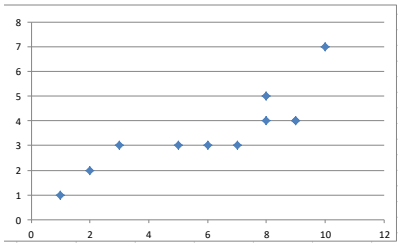
**11)** לאור הנתונים המופיעים בדיאגרמה, איזה מבין הערכים הבאים מתאים להיות התוצאה של $r$?

א. -0.52
ב. 0.72
ג. -0.72
ד. 0.52

**12)** אם מקדם המתאם בין שני משתנים הוא 1, אזי:

א. הערכים של המשתנים הם חיוביים.
ב. עבור כל תצפית ערך של משתנה אחד שווה לערך של המשתנה השני.
ג. הקשר הלינארי הוא בעוצמה חזקה.
ד. אף אחת מהתשובות לא בהכרח נכונה.

**13)** להלן דיאגרמת פיזור:




מה יהיה מקדם המתאם בין שני המשתנים?

א. 1
ב. 0.85
ג. 0.15
ד. 0

**14)** בבדיקת קשר בין שני משתנים התקבל: $r = -1$.

א. קיימת נוסחה לינארית הקושרת בין כל התצפיות.
ב. לא קיים קשר בין שני המשתנים.
ג. ככל שמשתנה אחד נוטה לרדת גם לשני יש נטייה לרדת.
ד. קיים קשר בין שני המשתנים, אך לא ניתן לדעת מאיזה סוג.


© כל הזכויות שמורות לברק קנדל

**15)** לפי הפתגם "רחוק מהעין, רחוק מהלב", יש קשר ____ בין קרבה פיזית לקרבה נפשית.

א. חיובי

ב. שלילי

ג. אפסי

ד. לא ניתן לדעת.

**16)** מבחן אמי"ר הינו מבחן מיון באנגלית של המרכז הארצי לבחינות והערכה. הציון המינימלי בבחינה הינו 150 והמקסימלי הינו 250. בקורס הכנה למבחן השתתפו 19 תלמידים. להלן הציונים שלהם על פי פלט שהתקבל:

| | | |
|---|---|---|
| | | 159 |
| | | 170 |
| | | 180 |
| | | 185 |
| | | 204 |
| | | 224 |
| | | 236 |
| | | 212 |
| | | 168 |
| | | 189 |
| | | 195 |
| | | 163 |
| | | 187 |
| | | 206 |
| | | 201 |
| | | 223 |
| | | 242 |
| | | 203 |
| | | 205 |
| 197.47 | AVERAGE | |
| 536.25 | VARPA | |

יש להוסיף עמודה נוספת לצד עמודת הציונים שתראה לכל תלמיד כמה נקודות חסרות לו כדי להשלים לציון המקסימלי בבחינה.

מה יהיה מקדם המתאם בין שתי העמודות (כלומר, מקדם המתאם בין הציון לבין הנקודות החסרות)?

א. -1

ב. 1

ג. -0.5

ד. 0.5

**17)** מקדם המתאם בין שטחי דירה למחיר שלהם חושב ונמצא 1.2. מה נובע מכך?

א. ככל שהדירה גדולה יותר בשטחה כך היא יקרה יותר.

ב. ככל שהדירה קטנה יותר בשטחה כך היא זולה יותר.

ג. לא קיים קשר בין שטח הדירה למחיר הדירה.

ד. מצב כזה שמתואר הנתונים לא אפשרי.

**18)** אם ניקח 10 אנשים ונרשום לכל אדם את הגובה במטר וכמו כן את הגובה בס"מ. מה יהיה מקדם המתאם בין גובה האדם במטר לגובה האדם בס"מ?

א. 1

ב. 0

ג. -1

ד. לא ניתן לדעת.


© כל הזכויות שמורות לברק קנדל

**19)** נמצא מתאם חיובי בעוצמה גבוהה בין $X$ – ציון בבגרות בלשון ל $Y$ – ציון בבגרות במתמטיקה. אילו מהמשפטים הבאים נכון?

א. ניתן לומר שאחת מהסיבות להבדלים שיש לסטודנטים במתמטיקה נובעים מההבדלים שיש להם בלשון.

ב. קיימת נוסחה של קו ישר שקושרת בין ציון בבגרות במתמטיקה לציון בבגרות בלשון.

ג. ללא יוצא מן הכלל, ניתן להגיד שכל תלמיד שמצליח יותר מתלמיד אחר בלשון גם יצליח יותר מאותו תלמיד במתמטיקה.

ד. אף אחד מהטענות שהוצגו אינה בהכרח נכונה.

**20)** עבור סדרה של תצפיות מדדו את $X$ ואת $Y$. נמצא שעבור כל התצפיות שהערך של $Y$ ירד הערך של $X$ בהכרח ירד ללא יוצא מן הכלל. מקדם המתאם של פירסון יהיה בהכרח:

א. 1

ב. -1

ג. 0

ד. אף אחת מהתשובות.


© כל הזכויות שמורות לברק קנדל

## תשובות סופיות

**1)** א. שעות בילוי. ב. הקשר חלקי, כיוון הקשר שלילי .

**2)** א. להלן טבלה:

| | ספורט | לשון | מתמטיקה |
|---|---|---|---|
| ספורט | 1 | -0.7 | 0.1 |
| לשון | -0.7 | 1 | 0.6 |
| מתמטיקה | -0.1 | 0.6 | 1 |

ב. ספורט ולשון.

**3)** א. ב"ת- מס' שעות התרגול, תלוי- ציון. ב. ראה גרף בפתרון וידאו.
ג. קשר לינארי חיובי חלקי. ד. מקדם המתאם לא היה משתנה.

**4)** א. לעלות. ב. לא ישפיע על מקדם המתאם.

**5)** א. חלקי, חיובי. ב. קטן.

**6)** ד'. **7)** ד'. **8)** א'. **9)** ג'. **10)** א'.

**11)** ג'. **12)** ד'. **13)** ב'. **14)** א'. **15)** א'.

**16)** א'. **17)** ד'. **18)** א'. **19)** ד'. **20)** ד'.


© כל הזכויות שמורות לברק קנדל

## מדדי קשר – מדד הקשר הלינארי (פירסון) – רקע

המטרה היא לבדוק האם קיים קשר (קורלציה, מתאם) של קו ישר בין שני משתנים כמותיים. מבחינת סולמות המדידה קשר בין סולמות רווחים ומנה.
בדרך כלל, $X$ הוא המשתנה המסביר (הבלתי תלוי) ו- $Y$ הוא המשתנה המוסבר (התלוי).

**דוגמה:**

נרצה להסביר כיצד השכלה של אדם הנמדדת בשנות לימוד – $X$ מסבירה את ההכנסה שלו $Y$. במקרה זה שנות ההשכלה זהו המשתנה המסביר (או הבלתי תלוי) ואנחנו מעוניינים לבדוק כיצד שינויים בשנות ההשכלה של אדם יכולים להסביר את השינויים שלו בהכנסה, ולכן רמת ההכנסה זהו המשתנה המוסבר התלוי במשתנה המסביר אותו.

**שלב ראשון:** נהוג לשרטט דיאגרמת פיזור. זו דיאגרמה שנותנת אינדיקציה ויזואלית על טיב הקשר בין שני המשתנים.

**דוגמה:**

בבניין של 5 דירות בדקו את הנתונים הבאים:
$X$ -מס' חדרים בדירה. $Y$-מס' נפשות הגרות בדירה.
להלן התוצאות שהתקבלו:

| מס' דירה | $X$ | $Y$ |
|---|---|---|
| 1 | 3 | 2 |
| 2 | 2 | 2 |
| 3 | 4 | 3 |
| 4 | 3 | 3 |
| 5 | 5 | 4 |

נשרטט מנתונים אלה דיאגרמת פיזור (הדיאגרמה המלאה בסרטון). נתבונן בכמה מקרים של דיאגרמות פיזור וננתח אותן (הדיאגרמות המלאות בסרטון).

**שלב שני:** מחשבים את מקדם המתאם (מדד הקשר) שבודק עד כמה קיים קשר לינארי בין שני המשתנים. המדד (ניקרא גם מדד הקשר של פירסון) מכמת את מה שניראה בשלב הראשון רק בעין.
המדד בודק את כיוון הקשר (חיובי או שלילי) ואת עוצמת הקשר (חלש עד חזק).
מקדם מתאם זה מקבל ערכים בין 1- ל-1.
מקדם מתאם 1- או 1 אומר שקיים קשר לינארי מוחלט ומלא בין המשתנים שניתן לבטאו על ידי הנוסחה: $y = bx + a$.

**מתאם חיובי מלא (מקדם מתאם 1):**

קיים קשר לנארי מלא בו השיפוע $b$ יהיה חיובי ואילו מתאם שלילי מלא אומר שקיים קשר לנארי מלא בו השיפוע $b$ שלילי (מקדם מתאם 1-).


© כל הזכויות שמורות לברק קנדל

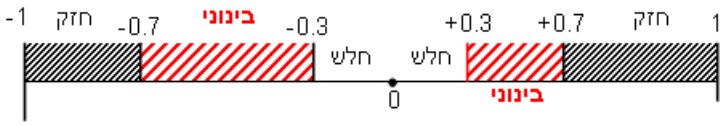
**מתאם חיובי חלקי:**

ככל שמשתנה אחד עולה לשני יש נטייה לעלות בערכו אבל לא קיימת נוסחה לינארית שמקשרת את $X$ ל- $Y$ באופן מוחלט.

**מתאם שלילי חלקי:**

ככל שמשתנה אחד עולה לשני יש נטייה לרדת אבל לא קיימת נוסחה לינארית שמקשרת את $X$ ל- $Y$ באופן מוחלט.

ככל שערך מקדם המתאם קרוב לאפס נאמר שעוצמת הקשר חלשה יותר וככל שמקדם המתאם רחוק מהאפס נאמר שעוצמת הקשר חזקה יותר:




מקדם המתאם יסומן באות $r$.

כדי לחשב את מקדם המתאם, יש לחשב את סטיות התקן של כל משתנה ואת השונות המשותפת.

שונות משותפת: $COV(x,y)=\frac{\sum(x-\bar{x})(y-\bar{y})}{n}=\frac{\sum xy}{n}-\bar{x}\cdot\bar{y}$

שונות של המשתנה $X$: $s_x^2=\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}=\frac{\sum_{i=1}^{n}x_i^2}{n}-\bar{x}^2$

שונות המשתנה $Y$: $S_Y^2=\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n}=\frac{\sum_{i=1}^{n}y_i^2}{n}-\bar{y}^2$

מקדם המתאם הלינארי: $r_{xy}=\frac{COV(x,y)}{S_x\cdot S_y}$


© כל הזכויות שמורות לברק קנדל

## שאלות

**1)** להלן נתונים לגבי שישה תלמידים שנגשו למבחן. בדקו לגבי כל תלמיד את הציון שלו בסוף הקורס וכמו כן את מספר החיסורים שלו מהקורס.

| מספר חיסורים | 2 | 1 | 0 | 2 | 3 | 4 |
|---|---|---|---|---|---|---|
| ציון | 80 | 90 | 90 | 70 | 70 | 50 |

א. שרטטו דיאגראמת פיזור לנתונים. מה ניתן להסיק מהדיאגרמה על טיב הקשר ביו מספר החיסורים של תלמיד לציונו? מיהו המשתנה הבלתי תלוי ומיהו המשתנה התלוי?

ב. חשבו את מדד הקשר של פירסון. האם התוצאה מתיישבת עם תשובתך לסעיף א'?

ג. הסבירו, ללא חישוב, כיצד מקדם המתאם היה משתנה אם היה מתווסף תלמיד שהחסיר 4 פעמים וקיבל ציון 80?

**2)** במחקר רפואי רצו לבדוק האם קיים קשר בין רמת ההורמון $X$ בדם החולה לרמת ההורמון $Y$ שלו. לצורך כך מדדו את רמת ההורמונים ההלו עבור חמישה חולים. להלן התוצאות שהתקבלו:

| $X$ | $Y$ |
|---|---|
| 10 | 12 |
| 14 | 15 |
| 15 | 15 |
| 18 | 17 |
| 20 | 21 |

א. מה הממוצע של כל רמת הורמון?

ב. מהו מקדם המתאם בין ההורמונים? ומה משמעות התוצאה?

**3)** נסמן ב- $X$ את ההכנסה של משפחה באלפי ₪. נסמן ב- $Y$ את ההוצאות של משפחה באלפי ₪. נלקחו 20 משפחות והתקבלו התוצאות הבאות:

$$\sum_{i=1}^{20} X_i = 240 \qquad \sum_{i=1}^{20} Y_i = 200$$

$$\sum_{i-1}^{20} (X_i - \bar{X})^2 = 76 \qquad \sum_{i-1}^{20} (Y_i - \bar{Y})^2 = 76$$

$$\sum_{i=1}^{20} (X_i - \bar{X})(Y - \bar{Y}) = 60.8$$

א. חשב את מדד הקשר הלינארי בין $X$ ל- $Y$. מיהו המשתנה התלוי?

ב. מה המשמעות של התוצאה שקיבלת בסעיף א?


© כל הזכויות שמורות לברק קנדל

**4)** נסמן ב- $X$ את ההכנסה של משפחה באלפי ₪. נסמן ב- $Y$ את ההוצאות של משפחה באלפי ₪. נלקחו 20 משפחות והתקבלו התוצאות הבאות:

$$\sum_{i=1}^{20} X_i = 240 \qquad \sum_{i=1}^{20} Y_i = 200$$

$$\sum_{i-1}^{20} X_i^2 = 2960 \qquad \sum_{i=1}^{20} Y_i^2 = 2080$$

$$\sum_{i=1}^{20} X_i Y_i = 2464$$

חשבו את מדד הקשר הלינארי בין $X$ ל- $Y$.

**5)** במוסד אקדמי ציון ההתאמה מחושב כך: מכפילים את הציון הממוצע בבגרות ב-3 ומפחיתים 2 נקודות. ידוע שעבור 40 מועמדים סטיית התקן של ממוצע הציון בבגרות הייתה 2.
מה מקדם המתאם בין ציון ההתאמה לציון הממוצע בבגרות שלהם?

**6)** להלן רשימת טענות, לגבי כל טענה קבעו נכון/לא נכון ונמקו.

א. מתווך דירות המיר מחירי דירות מדולר לשקל. נניח שדולר אחד הוא 3.5₪. אם מתווך הדירות יחשב את מדד הקשר של פירסון בין מחיר הדירה בשקלים למחיר הדירה בדולרים הוא יקבל 1.

ב. לסדרה של נתונים התקבל $\bar{X} = \bar{Y} = 6$ , $S_x = S_y = 1$.
לכן, מדד הקשר של פירסון יהיה 1.

ג. אם השונות המשותפת של $X$ ושל $Y$ הינה 0 אז בהכרח גם מקדם המתאם של פירסון יהיה 0.

**שאלות רב-ברירה:**

**7)** נמצא שקיים מקדם מתאם שלילי בין הציון בעברית לציון בחשבון בבחינה לכן:

א. הדבר מעיד שהציונים בכתה היו שליליים.

ב. ככל שהציון של תלמיד יורד בחשבון יש לו נטייה לרדת בעברית.

ג. ככל שהציון של תלמיד עולה בחשבון יש לו נטייה לרדת בעברית.

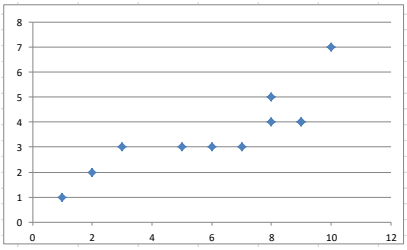
ד. אף אחת מהתשובות לא נכונה.


© כל הזכויות שמורות לברק קנדל

**8)** נלקחו 20 מוצרים ונבדק ביום מסוים המחיר שלהם בדולרים והמחיר שלהם בש"ח (באותו היום ערך הדולר היה-4.2₪).
מהו מקדם המתאם בין המחיר בדולר למחיר בש"ח?

א. 1
ב. 0
ג. 4.2
ד. לא ניתן לדעת.

**9)** להלן דיאגראמת פיזור:
מה יהיה מקדם המתאם בין שני המשתנים?




א. 1
ב. 0.85
ג. 0.15
ד. 0

## תשובות סופיות

**1)** א. משתנה תלוי: ציון, משתנה ב"ת: מס' חיסורים. ראה דיאגרמה בוידאו. ניתן להסיק שקיים קשר לינארי שלילי וחלקי בין מספר החיסורים לציון התלמיד.
ב. 0.9325-. ג. הקשר יישאר לינארי שלילי חלקי אך עוצמתו תחלש.

**2)** א. $\bar{y}=16$ , $\bar{x}=15.4$ ב. $r_{xy}=0.96$.

**3)** א. 0.8.

**4)** 0.8.

**5)** 1.

**6)** א. נכון. ב. לא נכון. ג. נכון.

**7)** ג'.

**8)** א'.

**9)** ב'.


© כל הזכויות שמורות לברק קנדל

## בדיקת השערות על מקדם המתאם הלינארי – רקע

מדד הקשר הלינארי באוכלוסייה, שנקרא גם מקדם המתאם של פירסון או מדד הקשר של פירסון באוכלוסייה מסומן ב: $\rho$ - פרמטר המאפיין את עוצמת הקשר הלינארי וכיוונו בין שני המשתנים הנחקרים באוכלוסייה. כאשר:
$r$ - מדד הקשר הלינארי במדגם שמהווה אומד לפרמטר $\rho$.

**השערת האפס:** תהיה שבאוכלוסייה לא קיים כלל קשר לינארי בין שני המשתנים $H_0:\rho=0$.

ההנחה שעליה אנו מתבססים בתהליך היא ששני המשתנים הנחקרים מתפלגים דו נורמלית.

סטטיסטי המבחן: $t=\dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}}\sim t(n-2)$

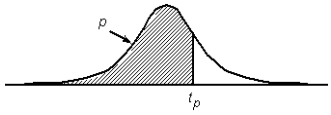
סטטיסטי זה מתפלג $t$ עם $n-2$ דרגות חופש.

| | | | |
|---|---|---|---|
| השערת האפס: | $H_0:\rho=0$ | $H_0:\rho=0$ | $H_0:\rho=0$ |
| השערת המחקר: | $H_1:\rho\neq 0$ | $H_1:\rho<0$ | $H_1:\rho>0$ |
| כלל ההכרעה:<br>אזור דחייה של השערת האפס | $t\geq t_{1-\alpha}$<br>או<br>$t\leq -t_{1-\alpha}$ | $t\leq -t_{1-\alpha}$ | $t\geq t_{1-\alpha}$ |


© כל הזכויות שמורות לברק קנדל

# טבלת ערכים קריטיים של t - נספח: טבלת התפלגות T

$P$




| דרגות חופש | 0.75 | 0.90 | 0.95 | 0.975 | 0.99 | 0.995 | 0.9995 |
|---|---|---|---|---|---|---|---|
| 1 | 1.000 | 3.078 | 6.314 | 12.709 | 31.821 | 63.657 | 636.619 |
| 2 | 0.816 | 1.886 | 2.920 | 4.303 | 6.965 | 9.925 | 31.598 |
| 3 | 0.765 | 1.638 | 2.353 | 3.182 | 4.541 | 5.841 | 12.941 |
| 4 | 0.741 | 1.533 | 2.132 | 2.776 | 3.747 | 4.604 | 8.610 |
| 5 | 0.727 | 1.476 | 2.015 | 2.571 | 3.365 | 4.032 | 6.859 |
| 6 | 0.718 | 1.440 | 1.943 | 2.447 | 3.143 | 3.707 | 5.959 |
| 7 | 0.711 | 1.415 | 1.895 | 2.365 | 2.998 | 3.499 | 5.405 |
| 8 | 0.706 | 1.397 | 1.860 | 2.306 | 2.896 | 3.355 | 5.041 |
| 9 | 0.703 | 1.383 | 1.833 | 2.262 | 2.821 | 3.250 | 4.781 |
| 10 | 0.700 | 1.372 | 1.812 | 2.228 | 2.764 | 3.169 | 4.587 |
| 11 | 0.697 | 1.363 | 1.796 | 2.201 | 2.718 | 3.106 | 4.437 |
| 12 | 0.695 | 1.356 | 1.782 | 2.179 | 2.681 | 3.055 | 4.318 |
| 13 | 0.694 | 1.350 | 1.771 | 2.160 | 2.650 | 3.012 | 4.221 |
| 14 | 0.692 | 1.345 | 1.761 | 2.145 | 2.624 | 2.977 | 4.140 |
| 15 | 0.691 | 1.341 | 1.753 | 2.131 | 2.602 | 2.947 | 4.073 |
| 16 | 0.690 | 1.337 | 1.746 | 2.120 | 2.583 | 2.921 | 4.015 |
| 17 | 0.689 | 1.333 | 1.740 | 2.110 | 2.567 | 2.898 | 3.965 |
| 18 | 0.688 | 1.330 | 1.734 | 2.101 | 2.552 | 2.878 | 3.922 |
| 19 | 0.688 | 1.328 | 1.729 | 2.093 | 2.539 | 2.861 | 3.883 |
| 20 | 0.687 | 1.325 | 1.725 | 2.086 | 2.528 | 2.845 | 3.850 |
| 21 | 0.686 | 1.323 | 1.721 | 2.080 | 2.518 | 2.831 | 3.819 |
| 22 | 0.686 | 1.321 | 1.717 | 2.074 | 2.508 | 2.819 | 3.792 |
| 23 | 0.685 | 1.319 | 1.714 | 2.069 | 2.500 | 2.807 | 3.767 |
| 24 | 0.685 | 1.318 | 1.711 | 2.064 | 2.492 | 2.797 | 3.745 |
| 25 | 0.684 | 1.316 | 1.708 | 2.060 | 2.485 | 2.787 | 3.725 |
| 26 | 0.684 | 1.315 | 1.706 | 2.056 | 2.479 | 2.779 | 3.707 |
| 27 | 0.684 | 1.314 | 1.703 | 2.052 | 2.473 | 2.771 | 3.690 |
| 28 | 0.683 | 1.313 | 1.701 | 2.048 | 2.467 | 2.763 | 3.674 |
| 29 | 0.683 | 1.311 | 1.699 | 2.045 | 2.462 | 2.756 | 3.659 |
| 30 | 0.683 | 1.310 | 1.697 | 2.042 | 2.457 | 2.750 | 3.646 |
| 40 | 0.681 | 1.303 | 1.684 | 2.021 | 2.423 | 2.704 | 3.551 |
| 60 | 0.679 | 1.296 | 1.671 | 2.000 | 2.390 | 2.660 | 3.460 |
| 120 | 0.677 | 1.289 | 1.658 | 1.980 | 2.358 | 2.617 | 3.373 |
| ∞ | 0.674 | 1.282 | 1.645 | 1.960 | 2.326 | 2.576 | 3.291 |


© כל הזכויות שמורות לברק קנדל

## שאלות

**1)** להלן נתונים על הוותק בעבודה (בשנים) ועל השכלה (בשנים) במדגם של 10 עובדים:

| נחקר | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $X$ -ווותק | 13 | 18 | 2 | 8 | 16 | 9 | 5 | 28 | 17 | 24 |
| $Y$ -השכלה | 12 | 14 | 17 | 8 | 11 | 12 | 13 | 8 | 12 | 15 |

מקדם המתאם חושב והתקבל: 0.31-.

א. האם קיים מתאם בין וותק העובד להשכלתו? בדקו ברמת מובהקות של 5%?

ב. אם הוותק של העובד היה נמדד בחודשים האם התשובה לסעיף א' הייתה משתנה?

**2)** מחקר התעניין לבדוק את הקשר בין גיל נשים בהריון לרמת ההמוגלובין שלהן בדם בזמן הריון. נדגמו 7 נשים והתקבלו התוצאות הבאות:

| נחקרת | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| המוגלובין | 14.7 | 13.5 | 9.7 | 12 | 10.8 | 13 | 10.3 |
| גיל | 39 | 34 | 30 | 29 | 28 | 26 | 23 |

במדגם חושב מדד הקשר של פירסון להיות 0.7.

א. האם ניתן לומר שבמדגם אם אישה היא יותר מבוגרת אזי בהכרח יש לה יותר המוגלובין בדם?

ב. האם ניתן לומר, ברמת מובהקות של 5%, שקיים מתאם בין גיל האישה שבהריון לבין רמת ההמוגלובין שלה בדם?

**3)** בתחנה המטאורולוגית רצו לבדוק את הקשר שבין הטמפרטורה במעלות צלזיוס לכמות המשקעים במ"מ. הם אספו נתונים על 10 ימים במהלך חודש ינואר. המתאם שהתקבל היה 0.8-.

א. בדקו ברמת מובהקות של 2.5% האם קיים קשר לינארי שלילי בחודש ינואר בין הטמפרטורה במעלות צלזיוס לבין המשקעים במעלות צלזיוס.

ב. כיצד הייתה משתנה התשובה לסעיף א אם הינו מוסיפים עוד תצפיות למדגם?

ג. על סמך טבלת T המצורפת עבור אילו רמות מובהקות ניתן להחליט שקיים קשר לינארי שלילי מובהק?

**4)** מתווך דירות חישב את מקדם המתאם בין שטח דירה במרכז תל אביב לבין המחיר של הדירה עבור 17 דירות. מקדם המתאם שקיבל היה 0.6.

א. בדוק ברמת מובהקות של 5% האם ניתן להגיד שקיים קשר ישר עולה בין שטח הדירה לבין מחיר הדירה במרכז תל אביב?

ב. מהי מובהקות התוצאה לבדיקת ההשערה שקיים קשר ישר עולה בין שטח הדירה לבין מחיר הדירה בתל אביב.


© כל הזכויות שמורות לברק קנדל

## תשובות סופיות

**1)** א. לא נדחה את $H_0$ .
ב. לא תשתנה.

**2)** א. לא
ב. לא נדחה את $H_0$ .

**3)** א. נדחה את $H_0$ .
ב. לא ניתן לדעת.
ג. לפחות 0.005.

**4)** א. נדחה את $H_0$ .
ב. $0.005 < P_V < 0.01$ .


© כל הזכויות שמורות לברק קנדל

## מדד הקשר הלינארי – ניתוח פלטים – רקע

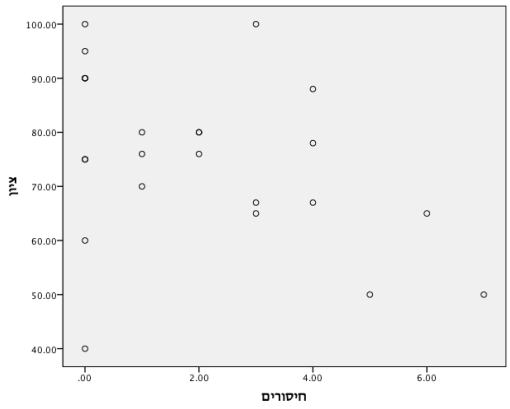
מדד הקשר הלינארי באוכלוסייה, שנקרא גם מקדם המתאם של פירסון או מדד הקשר של פירסון באוכלוסייה מסומן ב: $\rho$ - פרמטר המאפיין את עוצמת הקשר הלינארי באוכלוסייה וכיוונו בין שני המשתנים הנחקרים. כאשר:
$r$ - מדד הקשר הלינארי במדגם שמהווה אומד לפרמטר $\rho$.

השערת האפס: תהיה שבאוכלוסייה לא קיים כלל קשר לינארי בין שני המשתנים: $H_0: \rho = 0$.

ההנחה שעליה אנו מתבססים בתהליך היא ששני המשתנים הנחקרים מתפלגים דו נורמלית.

**דוגמה (פתרון בהקלטה):**

הדיקן ביקש לדגום סטודנטים כדי לבדוק את הקשר בין ציון הסטודנט בקורס למספר הפעמים שהוא החסיר שיעור בקורס. דיאגרמת הפיזור שהתקבלה במדגם שבוצע:




א. מיהו המשתנה התלוי ומיהו המשתנה הבלתי תלוי במחקר?
ב. מה ניתן לראות לגבי הקשר הלינארי בין המשתנים שהתקבל במדגם?

**Correlations**

| | | חיסורים | ציון |
|---|---|---|---|
| חיסורים | Pearson Correlation | 1 | -.389 |
| | Sig. (2-tailed) | | .060 |
| | N | 24 | 24 |
| ציון | Pearson Correlation | -.389 | 1 |
| | Sig. (2-tailed) | .060 | |
| | N | 24 | 24 |

ג. מהו מקדם המתאם שהתקבל במדגם? מה המשמעות שלו?
ד. האם ניתן להגיד ברמת מובהקות של 5% שקיים מתאם לינארי שלילי בין מספר החיסורים של הסטודנטים מהקורס לבין הציון של הסטודנטים בקורס?


© כל הזכויות שמורות לברק קנדל

## שאלות

**1)** מחקר רפואי התעניין לבדוק האם קיים קשר לינארי בין גיל האישה בהריון לרמת ההמוגלובין שלה. להלן תוצאות מדגם שהתקבלו, עבור נשים בהריון:

**Correlations**

| | | age | hemoglobin |
|---|---|---|---|
| age | Pearson Correlation | 1 | .565 |
| | Sig. (2-tailed) | | .005 |
| | N | 23 | 23 |
| hemoglobin | Pearson Correlation | .565 | 1 |
| | Sig. (2-tailed) | .005 | |
| | N | 23 | 23 |

א. מהי אוכלוסיית המחקר?

ב. מהן השערות המחקר?

ג. מהו המשתנה הבלתי תלוי ומהו המשתנה התלוי במחקר?

ד. מהי מסקנת המחקר ברמת מובהקות של 5%?


© כל הזכויות שמורות לברק קנדל

**2)** במדגם שנעשה נבדקו מספר משתנים על התצפיות שנדגמו. להלן פלט שהופק על המדגם?

**Correlations**

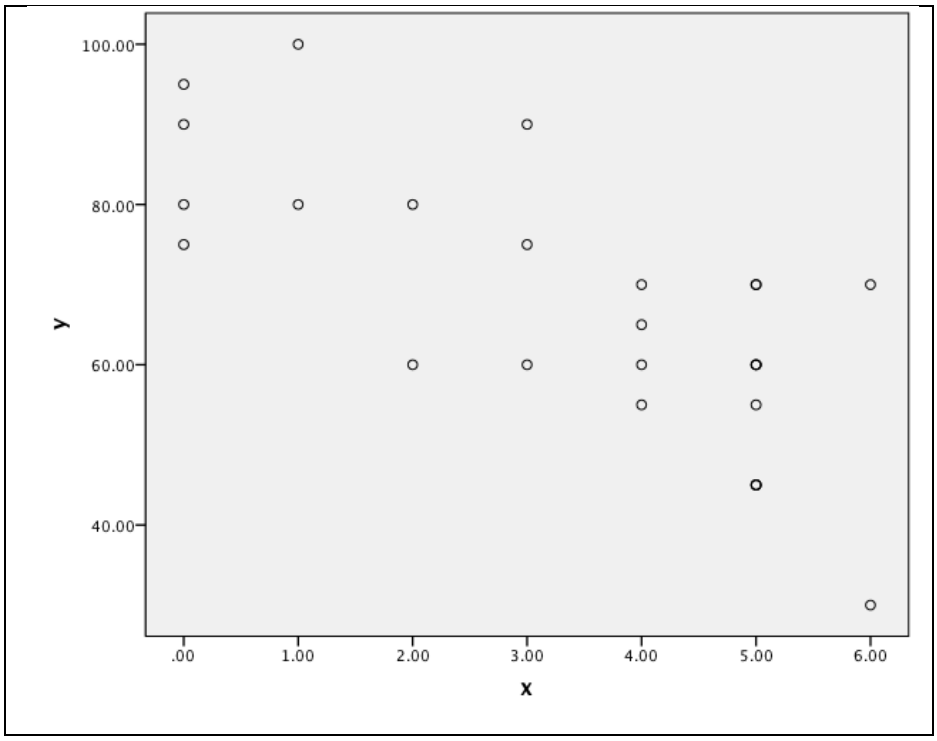
| | | x | y | z | w |
|---|---|---|---|---|---|
| x | Pearson Correlation | ??? | -.682 | .134 | .176 |
| | Sig. (2-tailed) | | .005 | .634 | .530 |
| | N | 15 | 15 | 15 | 15 |
| y | Pearson Correlation | -.682 | 1 | ??? | -.555 |
| | Sig. (2-tailed) | .005 | | .544 | .032 |
| | N | 15 | 15 | 15 | 15 |
| z | Pearson Correlation | .134 | .170 | 1 | -.247 |
| | Sig. (2-tailed) | ??? | .544 | | .374 |
| | N | 15 | 15 | 15 | 15 |
| w | Pearson Correlation | .176 | -.555 | -.247 | 1 |
| | Sig. (2-tailed) | .530 | .032 | .374 | |
| | N | 15 | 15 | 15 | 15 |

א. בין אילו שני משתנים שונים הקשר הלינארי במדגם נמצא עם העוצמה הכי חזקה?

ב. ברמת מובהקות של 5%, אילו שני משתנים בעל קשר לינארי מובהק?

ג. השלימו את המספרים המסומנים בפלט בסימני שאלה.

**3)** נדגמו מספר תלמידים בכיתה יב' ובדקו לכל תלמיד: $X$ - מספר שעות שבועיות שהתלמיד צופה בטלוויזיה ביום $Y$ - ציון הבגרות שלו במתמטיקה.
להלן התוצאות שהתקבלו במחקר:




**Correlations**

| | | x | y |
|---|---|---|---|
| x | Pearson Correlation | 1 | -.741** |
| | Sig. (2-tailed) | | .000 |
| | N | 26 | 26 |
| y | Pearson Correlation | -.741** | 1 |
| | Sig. (2-tailed) | .000 | |
| | N | 26 | 26 |

**. Correlation is significant at the 0.01 level (2-tailed).

א. מהו המשתנה התלוי ומהו המשתנה הבלתי תלוי?
ב. מהו כיוון הקשר שהתקבל במדגם ומהו עוצמתו?
ג. האם ניתן להגיד שבאופן מובהק ככל שתלמיד צופה יותר בטלוויזיה הוא מצליח פחות בבגרות במתמטיקה?
ד. בהמשך לסעיף הקודם, האם ניתן להגיד שהסיבה להצלחה או אי הצלחה בבגרות במתמטיקה היא זמן הצפייה בטלוויזיה?


© כל הזכויות שמורות לברק קנדל

**4)** נדגמו ילדים בגיל 8 ונבדק עבור כל ילד גובהו בס״מ ומשקלו בק״ג. להלן הפלט שהתקבל עבור תוצאות המדגם:

**Correlations**

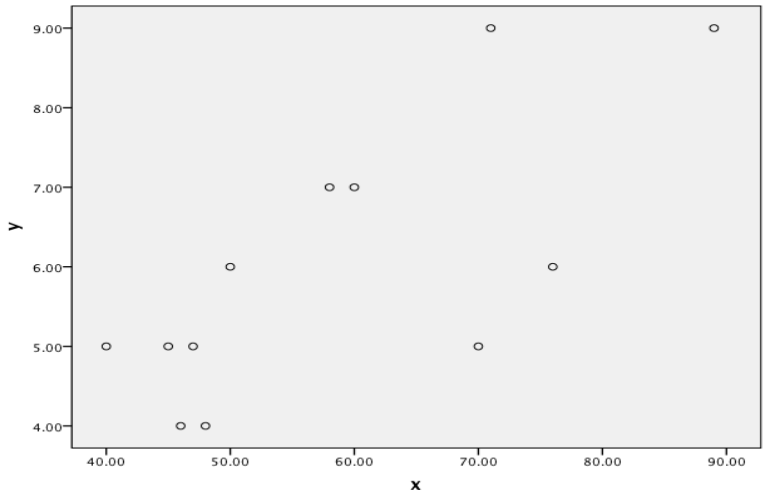
| | | גובה | משקל |
|---|---|---|---|
| גובה | Pearson Correlation | 1 | .552 |
| | Sig. (2-tailed) | | .062 |
| | N | 12 | 12 |
| משקל | Pearson Correlation | .552 | 1 |
| | Sig. (2-tailed) | .062 | |
| | N | 12 | 12 |

א. בדקו ברמת מובהקות של 5% האם קיים קשר לנארי חיובי בין המשקל והגובה.

ב. באילו רמות מובהקות ניתן לקבוע שקיים קשר לנארי חיובי בין במשקל והגובה?

ג. כיצד התשובה לסעיף הקודם הייתה משתנה אם היו מתווספות עוד 3 תצפיות למדגם?

**5)** בתהליך כימי מסוים חוקר בדק את הקשר בין הטמפרטורה בתהליך ( $X$ ) לבין אחוז החומר ( $Y$ ) בתהליך. דיאגרמת הפיזור שהתקבלה היא:




**Correlations**

| | | x | y |
|---|---|---|---|
| x | Pearson Correlation | 1 | .732** |
| | Sig. (2-tailed) | | .007 |
| | N | 12 | 12 |
| y | Pearson Correlation | .732** | 1 |
| | Sig. (2-tailed) | .007 | |
| | N | 12 | 12 |

**. Correlation is significant at the 0.01 level (2-tailed).

א. מה ניתן להגיד על סמך הפלט על הקשר שנימצא במדגם בין הטמפרטורה בתהליך לאחוז החומר?

ב. האם הקשר בין הטמפרטורה בתהליך לבין אחוז החומר הוא קווי חיובי מובהק? בדקו ברמת מובהקות של 5%.

ג. מה היה קורה למקדם המתאם במדגם ומובהקות התוצאה אם הייתה מתווספת תצפיות שבה הטמפרטורה היא 40 ואחוז החומר 9?


© כל הזכויות שמורות לברק קנדל

## תשובות סופיות

**1)** א. נשים בהריון. ב. $H_0 : p = 0$ $H_1 : p \neq 0$

ג. משתנה תלוי – רמת ההמוגלובין, משתנה בלתי תלוי- גיל.

ד. קיים קשר לינארי בין גיל האישה בהריון לרמת ההמוגלובין שלה בדם.

**2)** א. בין $X$ ל- $Y$. ב. $X$ ו- $Y$. כמו כן, $W$ ו- $Y$. ג. ראה וידאו.

**3)** א. משתנה תלוי – ציון בבגרות במתמטיקה, משתנה בלתי תלוי- שעות צפייה.

ב. כיוון שלילי ועוצמה של 0.741. ג. כן. ד. לא.

**4)** א. נדחה את $H_0$ . ב. לפחות 0.032. ג. לא ניתן לדעת.

**5)** א. קיים קשר לינארי חיובי וחלקי שעוצמתו : 0.732. ב. נדחה את $H_0$ .

ג. מקדם המתאם קטן ומובהקות התוצאה גדלה.


© כל הזכויות שמורות לברק קנדל