

## סטודנטים יקרים

לפניכם ספר תרגילים בקורס מבוא לסטטיסטיקה והסתברות א'.  
הספר הוא חלק מקורס חדשני וראשון מסוגו בארץ בנושא זה,  
המועבר ברשת האינטרנט On-line.

הקורס באתר כולל פתרונות מלאים לספר התרגילים, וכן את  
התיאוריה הרלוונטית לכל נושא ונושא.

**הקורס כולו מוגש בסרטוני וידאו המלווים בהסבר קולי, כך שאתם  
רואים את התהליכים בצורה מובנית, שיטתית ופשוטה, ממש כפי  
שנעשה בשיעור פרטי, לדוגמה [לחצו כאן](#).**

את הקורס בנה מר ברק קנדל, מרצה מבוקש במוסדות אקדמיים  
שונים ובעל ניסיון עתיר בהוראת המקצוע.

אז אם אתם עסוקים מידי בעבודה, סובלים מלקויות למידה, רוצים  
להצטיין או פשוט אוהבים ללמוד בשקט בבית, אנחנו מזמינים אתכם  
לחויית לימודים יוצאת דופן וחדשה לחלוטין, היכנסו עכשיו לאתר  
[www.gool.co.il](http://www.gool.co.il).



אנו מאחלים לכם הצלחה מלאה בבחינות

צוות האתר GooL

**גול, בשביל התירגול...**

לפתרון מלא בסרטון וידאו היכנסו ל- [www.Gool.co.il](http://www.Gool.co.il)  
כתב ופתר - ברק קנדל ©

## תוכן

|                                                                      |    |
|----------------------------------------------------------------------|----|
| פרק 1 - סטטיסטיקה תיאורית - הקדמה                                    | 3  |
| פרק 2 - סטטיסטיקה תיאורית - הצגה של נתונים                           | 6  |
| פרק 3 - סטטיסטיקה תיאורית - סכימה                                    | 14 |
| פרק 4 - סטטיסטיקה תיאורית - מדדי מיקום מרכזי                         | 17 |
| פרק 5 - סטטיסטיקה תיאורית - מדדי פיזור: הטווח, השונות וסטיית התקן    | 27 |
| פרק 6 - סטטיסטיקה תיאורית - מדדי פיזור - טווח בין-רבעוני             | 32 |
| פרק 7 - סטטיסטיקה תיאורית - ממוצע משוקלל ושונות מצורפת               | 35 |
| פרק 8 - סטטיסטיקה תיאורית - מדדי מיקום יחסי - ציון תקן               | 38 |
| פרק 9 - סטטיסטיקה תיאורית - מדדי מיקום יחסי - אחוזונים במחלקות       | 41 |
| פרק 10 - סטטיסטיקה תיאורית - טרנספורמציה לינארית                     | 46 |
| פרק 11 - סטטיסטיקה תיאורית - שאלות מסכמות                            | 49 |
| פרק 12 - מדדי קשר - מדד הקשר הלינארי (פירסון)                        | 58 |
| פרק 13 - מדדי קשר - השפעת טרנספורמציה לינאריות על מדד הקשר של פירסון | 66 |
| פרק 14 - מדדי קשר - רגרסיה לינארית                                   | 69 |

## פרק 1 - סטטיסטיקה תיאורית - הקדמה

### רקע:

בסטטיסטיקה תיאורית אנו חוקרים קבוצה מסוימת.  
הקבוצה יכולה להיות קבוצת ילדים בגן, קבוצת מניות בתיק, כלל התושבים בעיר מסוימת וכולי.  
בין ישות לישות בקבוצה ישנם גורמים היכולים לקבל מספר ערכים. גורמים אלה נקראים משתנים.  
למשל, בין מניה למניה בתיק משתנה התשואה היומית של המניה, הוותק של המניה, תחום המניה וכדומה.

בסטטיסטיקה תיאורית אנחנו נתבונן בקבוצה מסוימת ובתוך הקבוצה הזו נאסוף נתונים לגבי משתנה מסוים ונלמד להציג את הנתונים ולנתח אותם מכל מיני אספקטים.

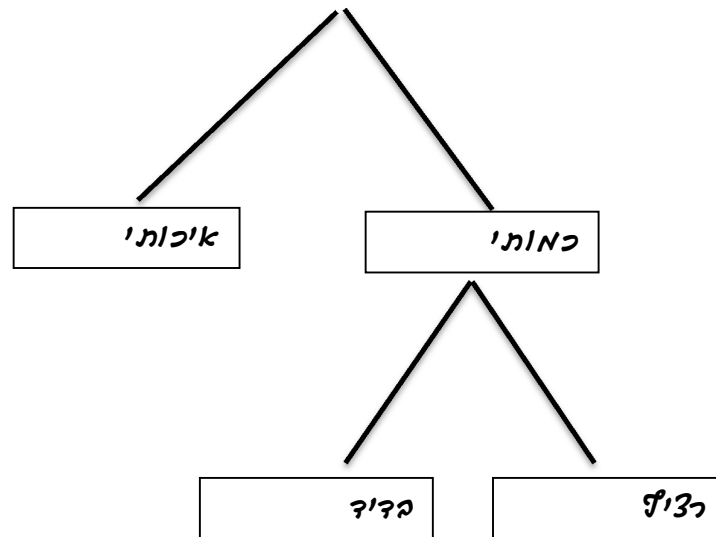
### דוגמה:

בתיק מניות 10 מניות. מנהל התיק פרסם את התשואה של כל מניה בשנת 2011.

מי הקבוצה הנחקרת?

מה גודל הקבוצה?

מה המשתנה הנחקר?

**סוגי משתנים:**

משתנה איכותי הוא משתנה שלערכיו אין משמעות של יותר או פחות, אין עניין כמותי לערכים המתקבלים.

כמו: מקום מגורים של אדם (רעננה, תל אביב, אשדוד..)

מין האדם (זכר, נקבה)

מצב משפחתי (רווק, נשוי, גרוש, אלמן)

משתנה כמותי הוא משתנה שערכיו הם מספרים להם יש משמעות כמותית כמו: גובה אדם בס"מ, ציון בבחינה וכדומה.

את המשתנה הכמותי נסווג לשני סוגים:

משתנה בדיד: משתנה שערכיו מתקבלים מתוך סידרה של ערכים אפשריים. כמו: מספר ילדים למשפחה (1,2,3..)

ציון בבחינה (מ 0 ועד 100 בקפיצות של 1)

הערה:

משתנה רציף: משתנה שערכיו מתקבלים מתוך אינסוף ערכים בתחום מסוים, הערכים מתקבלים ברצף וללא קפיצות של ערכים.

כמו: גובה בס"מ – אם למשל, הגובה הנמוך ביותר הוא 150 ועד 190 ס"מ בקבוצה הגבהים הם ברצף. גם בין 160 ל 161 ס"מ יש רצף אינסופי של ערכים אפשריים לגובה (160.33 ס"מ הוא גם גובה אפשרי)

משקל בק"ג, מהירות בקמ"ש וכולי.

### תרגילים:

1. סווג את המשתנים הבאים לפי: איכותי / כמותי בדיד / כמותי רציף:
  - א. מספר הדירות בבניין.
  - ב. גיל אדם בשנים.
  - ג. אחוז האבטלה בעיר.
  - ד. מקצוע לימוד מועדף.
2. להלן התפלגות מספר האיחורים לעבודה בחודש של העובדים בחברת "סטאר".  
בחברה 200 עובדים.

| מספר<br>העובדים | מספר האיחורים |
|-----------------|---------------|
| 17              | 0             |
| 23              | 1             |
| 85              | 2             |
| 50              | 3             |
| 25              | 4             |

- א. מהו המשתנה הנחקר כאן?
- ב. האם מדובר במשתנה איכותי או כמותי? אם הוא כמותי האם הוא בדיד או רציף?
3. לפניכם רשימה של משתנים כמותיים. ציין ליד כל משתנה אם הוא רציף או בדיד.
  - א. שכר עובד בש"ח.
  - ב. ציון בחינת בגרות.
  - ג. תוצאה בהטלת קובייה.
  - ד. מהירות ריצה בתחרות.
  - ה. שיעור התמיכה בממשלה.

## פרק 2 - סטטיסטיקה תיאורית - הצגה של נתונים

### רקע:

דרכים להצגת נתונים שנאספו :

#### א. רשימה של תצפיות:

התצפית היא הערך שנצפה עבור ישות מסוימת בקבוצה.  
רושמים את התצפיות שהתקבלו כרשומה, יעיל שיש מספר מועט של תצפיות.  
ההצגה הזו רלבנטית לכל סוגי המשתנים.  
למשל, להלן מספר החדרים בבניין בן 5 דירות :

3 4 3 5 4

#### ב. טבלת שכיחויות בדידה :

| שם המשתנה - X | שכיחות - $f(X)$        | שכיחות יחסית באחוזים       |
|---------------|------------------------|----------------------------|
| $X_1$         | $f_1$                  | $\frac{f_1}{N} \times 100$ |
| $X_2$         | $f_2$                  | $\frac{f_2}{N} \times 100$ |
| $X_3$         | $f_3$                  | $\frac{f_3}{N} \times 100$ |
| $\vdots$      | $\vdots$               | $\vdots$                   |
| $X_k$         | $f_k$                  | $\frac{f_k}{N} \times 100$ |
| סה"כ          | $N = \sum_{i=1}^k f_i$ | 100%                       |

רושמים את התצפיות בטבלה שבה עמודה אחת מבטאת את ערכי המשתנה והשנייה את השכיחות. יעיל עבור משתנה איכותי וכמותי בדיד וכשיש מספר רב של תצפיות.  
לא יעיל למשתנה כמותי רציף.

למשל, להלן התפלגות הציונים בכיתה מסוימת:

| הציון X- | מספר התלמידים – השכיחות f- | $F_i$ | $\frac{f_i}{n}$ |
|----------|----------------------------|-------|-----------------|
| 5        | 2                          | 2     | $0.08=2/25$     |
| 6        | 4                          | 6     | $0.16=4/25$     |
| 7        | 8                          | 14    | $0.32=8/25$     |
| 8        | 5                          | 19    | $0.2=5/25$      |
| 9        | 4                          | 23    | $0.16=4/25$     |
| 10       | 2                          | 25    | $0.08=2/25$     |

שכיחות מצטברת – צבירה של השכיחויות:  $F_i$  - השכיחות המצטברת נותנת כמה תצפיות קטנות או שוות לערך.

שכיחות יחסית (פרופורציה) – השכיחות מחולקת לכמות התצפיות הכללי:  $\frac{f_i}{n}$  - איזה חלק מהתצפיות בקבוצה שוות לערך.

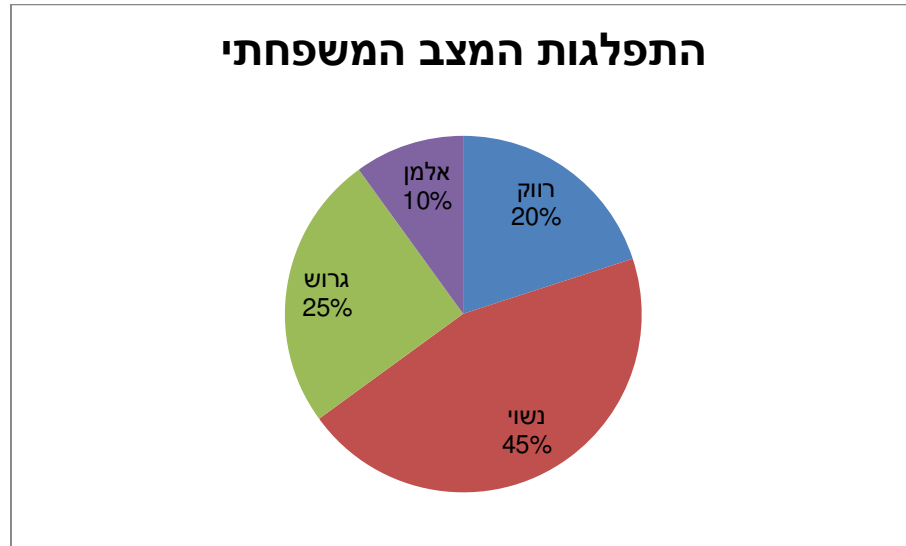
### ג. טבלת שכיחויות במחלקות:

משתמשים שהמשתנה כמותי רציף או כאשר יש מספר ערכים רב במשתנה הבדיד וטבלת שכיחויות תהיה ארוכה מידי.  
למשל, נתנו לקבוצת ילדים לבצע משימה מסוימת ובדקו את התפלגות זמן ביצוע המשימה בדקות.  
להלן ההתפלגות שהתקבלה:

| מספר הילדים | זמן בדקות |
|-------------|-----------|
| 20          | 0.5-3.5   |
| 18          | 3.5-9.5   |
| 14          | 9.5-19.5  |
| 8           | 19.5-29.5 |

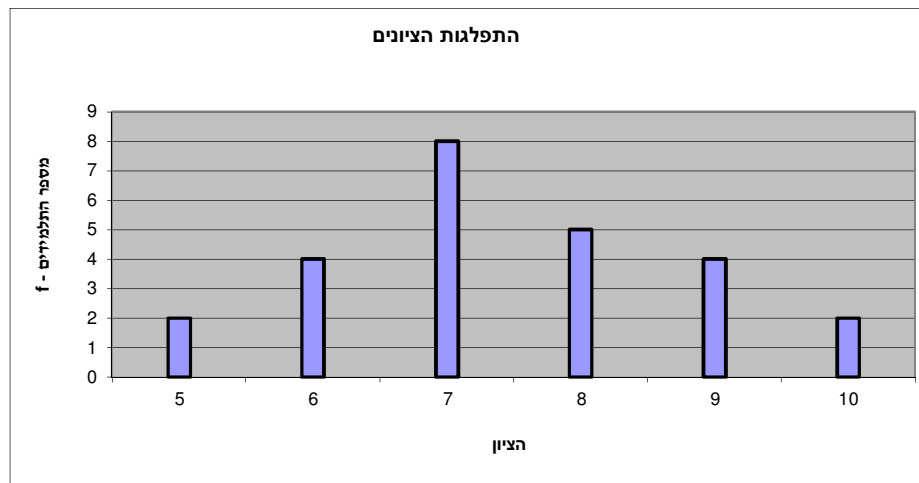
ד. דיאגרמת עוגה :

זהו התיאור הגרפי של משתנה איכותי. בדיאגרמת עוגה כל ערך במשתנה מקבל "נתח" יחסי מהעוגה. הנתח בעוגה פרופורציוני לשכיחות היחסית של ערך המשתנה בנתונים.



ה. דיאגרמת מקלות :

הציר האופקי הוא הציר של המשתנה הציר האנכי של השכיחות – הגובה של המקל מעיד על השכיחות .  
 רלבנטי למשתנה כמותי בדיד. לא נהוג להשתמש בתיאור למשתנה איכותי וכמו כן לא למשתנה כמותי רציף .  
 כמו כן בסולמות מדידה עבור משתנה מסולם סדר.



# 1. היסטוגרמה :

ההיסטוגרמה היא הדרך הגרפית כדי לתאר טבלת שכיחויות במחלקות.

רלבנטית למשתנה כמותי רציף.

בהיסטוגרמה ציר האופקי הוא הציר של המשתנה וציר האנכי הוא הציר של הצפיפות.

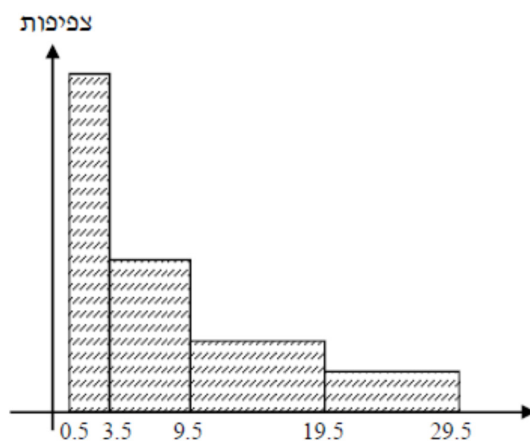
הצפיפות מחושבת בכל מחלקה על ידי חלוקת השכיחות ברוחב של כל המחלקה והיא נותנת

את מספר התצפיות הממוצע בכל מחלקה ליחידה.

אם המחלקות הן שוות ברוחב, ניתן לשרטט את ההיסטוגרמה לפי השכיחות ואין צורך

בצפיפות.

|        |        |        | X    |      |             |
|--------|--------|--------|------|------|-------------|
| צפיפות | מצטברת | שכיחות | אמצע | רוחב |             |
| 6.6667 | 20     | 20     | 2    | 3    | 0.5 - 3.5   |
| 3      | 38     | 18     | 6.5  | 6    | 3.5 - 9.5   |
| 1.4    | 52     | 14     | 14.5 | 10   | 9.5 - 19.5  |
| 0.8    | 60     | 8      | 24.5 | 10   | 19.5 - 29.5 |

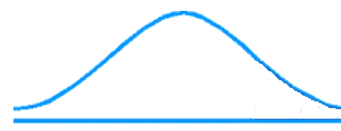


פוליגון-מצולעון : אם נחבר את אמצע קצה כל מלבן בקווים ישרים. נותן מראה חזותי לצורה של

התפלגות המשתנה.

### צורות התפלגות נפוצות

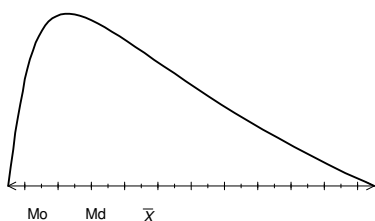
התפלגות סימטרית פעמונית- רוב התצפיות במרכז וככל שנתרחק מהמרכז יהיו פחות תצפיות באופן סימטרי. למשל, ציוני IQ.



ישנן התפלגויות סימטריות שאינן פעמוניות:

התפלגות אסימטרית ימנית ( חיובית ) –רוב התצפיות מקבלות ערכים נמוכים ויש מיעוט הולך וקטן של תצפיות שמקבלות ערכים גבוהים קיצוניים. למשל, שכר במשק.

#### התפלגות א-סימטרית ימנית או חיובית



התפלגות אסימטרית שמאלית ( שלילית ) רוב התצפיות מקבלות ערכים גבוהים ויש מיעוט הולך וקטן של תצפיות שמקבלות ערכים נמוכים קיצוניים. למשל, אורך חיים.

#### התפלגות א-סימטרית שמאלית או שלילית



### תרגילים:

1. בסקר צפייה בטלוויזיה התקבלו התוצאות הבאות: 25 צפו בערוץ הראשון, 25 צפו בערוץ 10, 75 צפו בערוץ השני, 50 צפו באחד מערוצי הכבלים ו - 25 לא צפו בטלוויזיה בזמן הסקר.  
א. רשמו את טבלת השכיחות ואת השכיחות היחסית.  
ב. תארו את הנתונים באופן גרפי.

2. להלן נתונים על התפלגות המקצוע המועדף של תלמידי שכבה ו' בבית הספר "מעוף":

| מספר התלמידים | המקצוע   |
|---------------|----------|
| 44            | מתמטיקה  |
| 20            | תנ"ך     |
| 12            | אנגלית   |
| 26            | היסטוריה |

- א. מהו המשתנה הנחקר?  
ב. מהי פרופורציית התלמידים שמעדיפים תנ"ך?

3. להלן התפלגות ההשכלה במקום עבודה מסוים:

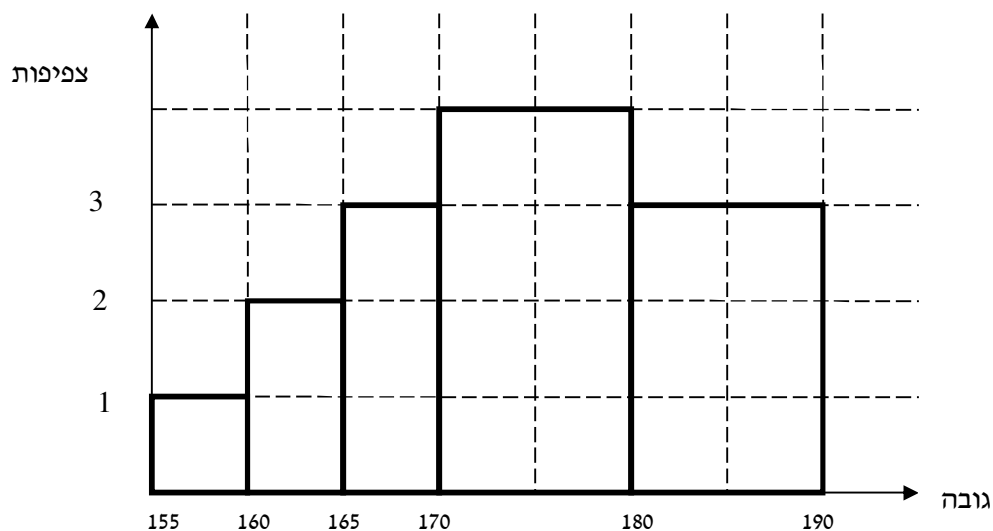
| מספר העובדים | השכלה   |
|--------------|---------|
| 60           | נמוכה   |
| 120          | תיכונית |
| 20           | אקדמאית |

- א. מהו המשתנה הנחקר? מאיזה סולם הוא?  
ב. תארו את הנתונים באופן גרפי.

4. להלן רשימת הציונים של 20 תלמידים שנבחנו במבחן הבנת הנקרא:  
7, 6, 8, 9, 10, 6, 4, 5, 8, 7, 6, 7, 6, 8, 9, 6, 7, 8, 5, 6

- א. מהו המשתנה? האם הוא בדיד או רציף?  
ב. תאר את הרשימה בטבלת שכיחויות.  
ג. הוסף שכיחויות יחסיות לטבלה.  
ד. תאר את הנתונים באופן גרפי.

5. להלן היסטוגרמה המתארת את התפלגות הגבהים בס"מ של קבוצה מסוימת :



- מהו המשתנה הנחקר? האם הוא בדיד או רציף?
- תאר את הנתונים בטבלת שכיחויות במחלקות.
- הוסף שכיחות יחסית לטבלה.
- הוסף את הצפיפות של כל מחלקה לטבלה.
- מהי צורת ההתפלגות של הגבהים?

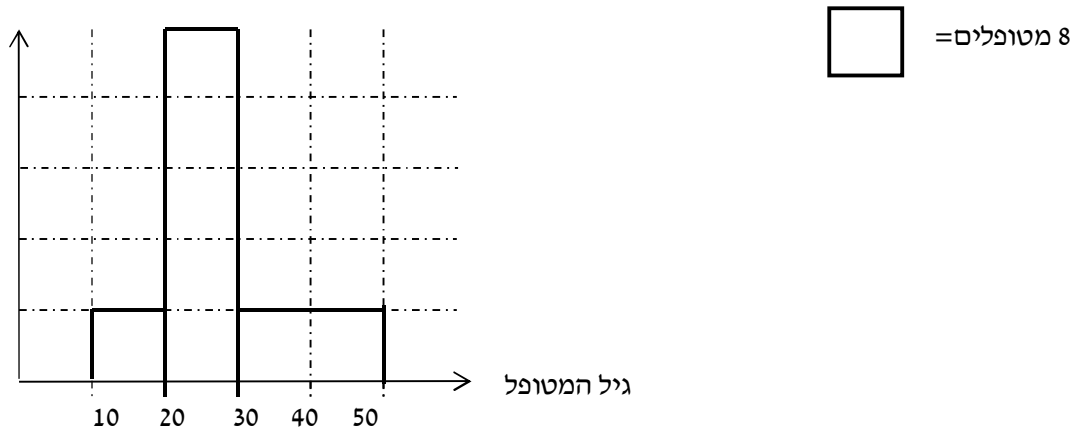
6. להלן התפלגות המשקל של קבוצה מסוימת בק"ג :

| מספר מקרים | משקל  |
|------------|-------|
| 10         | 40-45 |
| 20         | 45-50 |
| 30         | 50-60 |
| 20         | 60-65 |
| 10         | 65-70 |

- תאר את ההתפלגות באופן גרפי.
- מה ניתן להגיד על צורת ההתפלגות?

7. להלן גיל המטופלים של ד"ר שוורץ בשנים :

קנה מידה :



א. מה המשתנה הנחקר? האם הוא בדיד או רציף?

ב. מהי הקבוצה הנחקרת?

ג. תרגמו את ההסיטוגרמה לטבלת שכיחות.

ד. מהי הפרופורציה של המטופלים של ד"ר שוורץ בגילאים 20-30?

### פרק 3 - סטטיסטיקה תיאורית - סכימה

#### רקע:

בסטטיסטיקה ישנה צורת רישום מקובלת כדי לרשום סכום של תצפיות:

$$\sum_{i=1}^n X_i$$

נסביר את צורת הרישום על ידי הדוגמה הבאה:

| i | $X_i$ |
|---|-------|
| 1 | 5     |
| 2 | 0     |
| 3 | 1     |
| 4 | 3     |
| 5 | 2     |

**תרגילים:**

1. בבניין 5 דירות, לכל דירה רשמו את מספר החדרים שיש בדירה (X) ומספר הנפשות החיות בדירה (Y).

| מספר דירה | X | Y |
|-----------|---|---|
| 1         | 2 | 1 |
| 2         | 3 | 1 |
| 3         | 2 | 2 |
| 4         | 4 | 3 |
| 5         | 3 | 2 |

חשבו :

$$\sum_{i=1}^3 X_i$$

$$\sum_{i=1}^5 Y_i$$

$$\sum_{i=1}^4 X_i$$

$$\left(\sum_{i=1}^4 X_i\right)^2$$

$$\sum X_i$$

$$\sum X_i Y_i$$

$$\sum (X_i) \sum (Y_i)$$

2. נתון לוח ערכי המשתנים  $x_i$  ו-  $y_i$  כאשר :  $i=1,2,\dots,6$

|       |   |   |   |    |    |   |
|-------|---|---|---|----|----|---|
| $i$   | 1 | 2 | 3 | 4  | 5  | 6 |
| $x_i$ | 3 | 2 | 4 | -2 | 1  | 4 |
| $y_i$ | 2 | 0 | 0 | 1  | -5 | 2 |

ונתונים הקבועים :  $a=2$   $b=5$  חשבו את הנוסחאות הבאות :

א.  $\sum_{i=1}^4 y_i$

ב.  $\sum_{i=1}^6 a$

ג.  $\sum_{i=1}^6 x_i y_i$

ד.  $\sum_{i=1}^6 (x_i + y_i)$

ה.  $\sum_{i=1}^6 x_i + a$

3. קבע לכל זהות אם היא נכונה :

א.  $\sum_{i=1}^n bX_i = b \cdot \sum_{i=1}^n X_i$

ב.  $\sum_{i=1}^n a = a \cdot n$

ג.  $(\sum_{i=1}^n X_i)^2 = \sum_{i=1}^n X_i^2$

## פרק 4 - סטטיסטיקה תיאורית - מדדי מיקום מרכזי

### רקע:

המטרה במדדי המיקום המרכזי למדוד את מרכז ההתפלגות של התצפיות.

### השכיח – MODE

השכיח הוא הערך הנפוץ ביותר בהתפלגות.

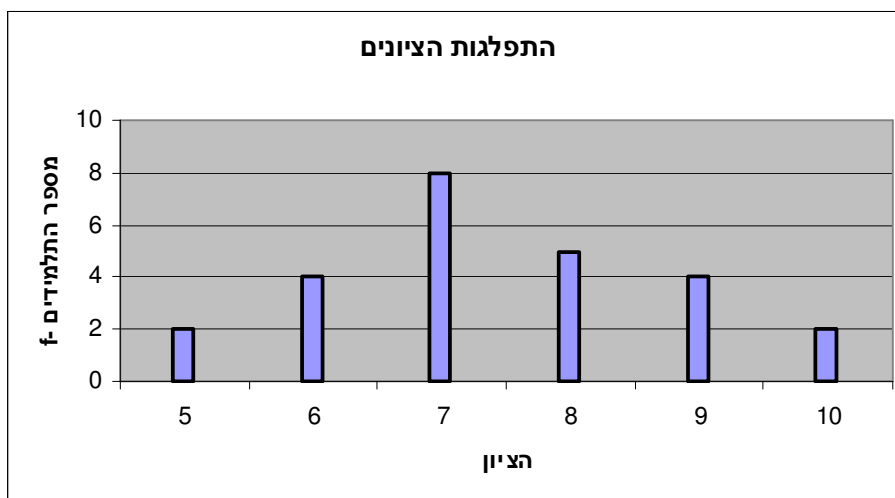
ברשימה : הערך החוזר על עצמו הכי הרבה פעמים .

7 9 4 8 4 10 6

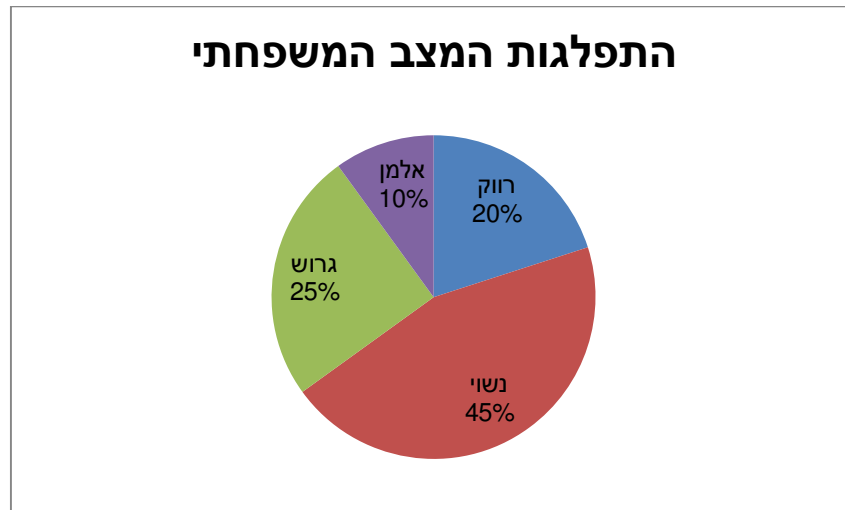
בטבלת שכיחויות בדידה : הערך שהשכיחות שלו היא הגבוהה ביותר.

| # תכניות החיסכון | $f(x)$ |
|------------------|--------|
| 0                | 100    |
| 1                | 75     |
| 2                | 25     |
| 3                | 25     |
| 4                | 25     |

בדיאגרמת מקלות : שיעור ה-  $X$  של המקל הגבוה ביותר.



בעוגה: הערך של הפלח הגדול ביותר.

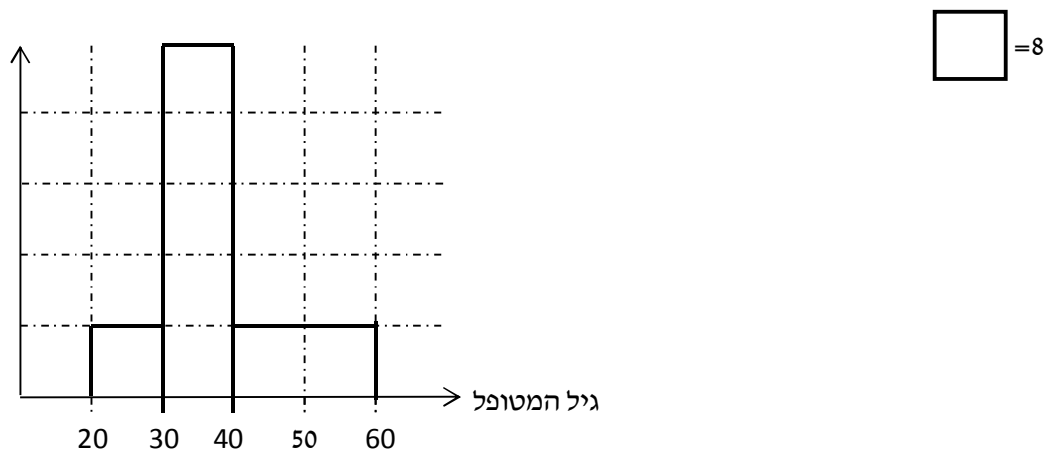


בטבלת שכיחויות במחלקות: אמצע המחלקה עם הצפיפות הגבוהה ביותר.  
התפלגות הציונים בכיתה.

| $f(x)$ | $X$    |
|--------|--------|
| 20     | 0-60   |
| 10     | 60-70  |
| 18     | 70-80  |
| 15     | 80-90  |
| 15     | 90-100 |

בהיסטוגרמה: שיעור ה-  $X$  של אמצע המחלקה הגבוהה ביותר.

להלן גיל המטופלים של די"ר שוורץ בשנים:



כללי: יתכן שלהתפלגות יותר משכיח אחד.  
השכיח הוא מדד הרלבנטי לכל סוגי המשתנים.

**MIDRANGE – (טווח)**

הממוצע בין התצפית הגבוהה ביותר לתצפית הנמוכה ביותר.

$$MR = \frac{X_{\min} + X_{\max}}{2}$$

**MEDIAN - החציון**

החציון הוא ערך שמחצית מהתצפיות קטנות או שוות לו ומחצית מהתצפיות גדולות או שוות לו .  
**ברשימה** : נסדר את התצפיות בסדר עולה.

אם יש מספר אי זוגי של איברים מקומו של החציון יהיה התצפית שמיקומה :  $\frac{n+1}{2}$

אם יש מספר זוגי של איברים החציון יהיה הממוצע של האיבר ה-  $\frac{n}{2}$  והאיבר ה-  $\frac{n}{2} + 1$

כלומר שיש מספר אי-זוגי של תצפיות החציון יהיה :  $md = X_{\frac{n+1}{2}}$

ושיש מספר זוגי של תצפיות החציון יהיה :  $md = \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2}$

**בטבלת שכיחויות בדידה** : נעשה תהליך דומה אך נעזר בשכיחות המצטברת.  
**דיאגרמת מקלות** : נמיר לטבלת שכיחויות בדידה במטרה למצוא את החציון.  
**בטבלת שכיחויות במחלקות** :

שלב א : נימצא את המחלקה החציונית שמיקומה יהיה  $\frac{n}{2}$ .

שלב ב : נציב בנוסחה הבאה :  $Md = L_0 + \frac{\frac{n}{2} - F(x_{m-1})}{f(x_m)} \cdot (L_1 - L_0)$

$F(x_{m-1})$  - שכיחות מצטברת של מחלקה אחת לפני המחלקה החציונית.

$f(x_m)$  - השכיחות של המחלקה החציונית.

$L_0$  - גבול התחתון של המחלקה.

$L_1$  - גבול העליון של המחלקה.

**היסטוגרמה** : החציון הוא הערך על ציר ה- $X$  שמחלק את ההיסטוגרמה לשני חלקים שווים בשטח.

כללי : החציון אינו רלבנטי למשתנה מסולם שמי ולא רלבנטי למשתנה איכותי.

**הממוצע**

הנו מרכז הכובד של ההתפלגות.

**ברשימה** :  $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$

**בטבלת שכיחויות** :  $\bar{x} = \frac{\sum x \cdot f}{n}$

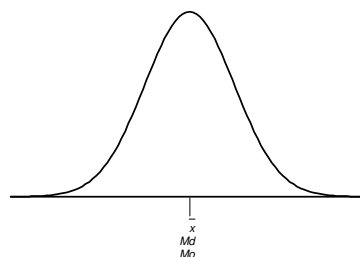
**במחלקות** : נשתמש באותה נוסחה רק נתייחס לאמצע המחלקה בתור ה- $X$ . הממוצע הזה יהיה ממוצע מקורב.

כללי : הממוצע רלבנטי רק למשתנה כמותי.

### מדדי המיקום המרכזי בהתפלגויות המיוחדות:

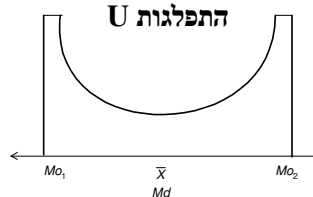
בהתפלגות סימטרית פעמונית כל מדדי המרכז שווים זה לזה :

#### התפלגות סימטרית



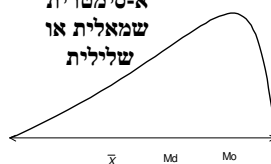
בהתפלגות סימטרית השכיח לא חייב להיות במרכז :

#### התפלגות U

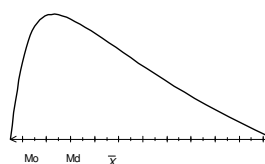


בהתפלגות אסימטרית

#### התפלגות א-סימטרית שמאלית או שלילית



#### התפלגות א-סימטרית ימנית או חיובית



### תרגילים:

1. להלן רשימת הציונים של 20 תלמידים שנבחנו במבחן הבנת הנקרא:  
 7, 6, 8, 9, 10, 6, 4, 5, 8, 7, 6, 7, 6, 8, 9, 6, 7, 8, 5, 6  
 חשב את החציון, השכיח, והממוצע של הציונים.

2. בדקו את מספר החדרים לדירה בבניין בן 5 דירות והתקבל ממוצע 3.8  
 לגבי 4 דירות נמצא מספר חדרים: 4, 3, 4, 5.  
 א. כמה חדרים יש בדירה החמישית?  
 ב. מהו השכיח ומהו החציון?

3. להלן התפלגות מספר מקלטי הטלויזיה שנספרו עבור כל משפחה בשוב מסוים:

| מספר משפחות | מספר מקלטים |
|-------------|-------------|
| 22          | 0           |
| 28          | 1           |
| 18          | 2           |
| 22          | 3           |
| 10          | 4           |

ג. חשב את הממוצע, החציון והשכיח של ההתפלגות.  
 ד. הסבר ללא חישוב כיצד כל מדד שחישבת בסעיף א' היה משתנה אם חלק מהמשפחות (לא כולן) שלא היה להם עד היום טלויזיה היו רוכשים מקלט אחד.

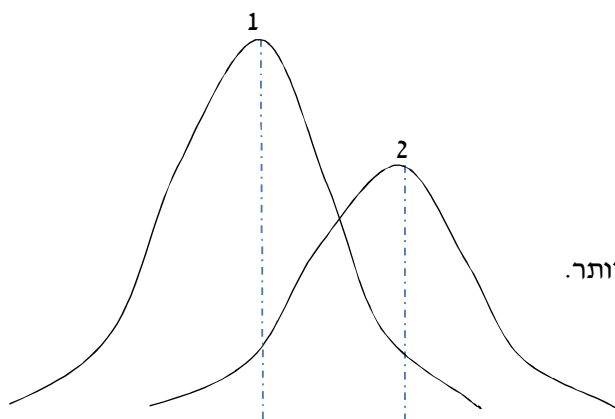
4. להלן התפלגות מספר המכוניות למשפחה בשוב "הגורן"

| מספר מכוניות למשפחה | 1  | 2   | 3   | 4   | 5  |
|---------------------|----|-----|-----|-----|----|
| שכיחות              | 65 | 150 | 220 | 140 | 55 |

א. כמה משפחות יש בשוב?  
 ב. מה אחוז המשפחות בשוב עם לכל היותר 2 מכוניות?  
 ג. חשבו את הממוצע, החציון והשכיח.  
 הקפידו להסביר לגבי כל סעיף מה משמעות התוצאה שקיבלתם!

5. מורה לימד 2 כיתות, הוא תיאר באותה מערכת צירים את התפלגות הציונים בכל כיתה.

בחר בתשובה הנכונה :



א. בכיתה 1 השכיח גבוה יותר מכיתה 2.

ב. בכיתה 2 השכיח גבוה יותר מכיתה 1.

ג. בשתי הכיתות אותו שכיח.

ד. לא ניתן לדעת באיזו כיתה השכיח גדול יותר.

6. ביישוב מסוים בדקו לכל משפחה את מספר הטלוויזיות שיש לה בבית. ביישוב גרות 200

משפחות. בממוצע יש למשפחה 1.5 טלוויזיות.

| מספר טלוויזיות | מספר משפחות |
|----------------|-------------|
| 0              | 28          |
| 1              | 62          |
| 2              |             |
| 3              |             |

א. השלימו את הטבלה.

ב. מהו השכיח, אמצע טווח והחציון.

ג. חלק מהמשפחות להן הייתה טלוויזיה אחת בדיוק הוציאו את הטלוויזיה מביתם, כיצד כל

מדד ישתנה (יגדל, יקטן או לא ישתנה) הסבירו ללא חישוב.

7. להלן התפלגות המשקל של קבוצה מסוימת בק"ג:

| משקל  | מספר מקרים |
|-------|------------|
| 40-45 | 10         |
| 45-50 | 20         |
| 50-60 | 30         |
| 60-65 | 20         |
| 65-70 | 10         |

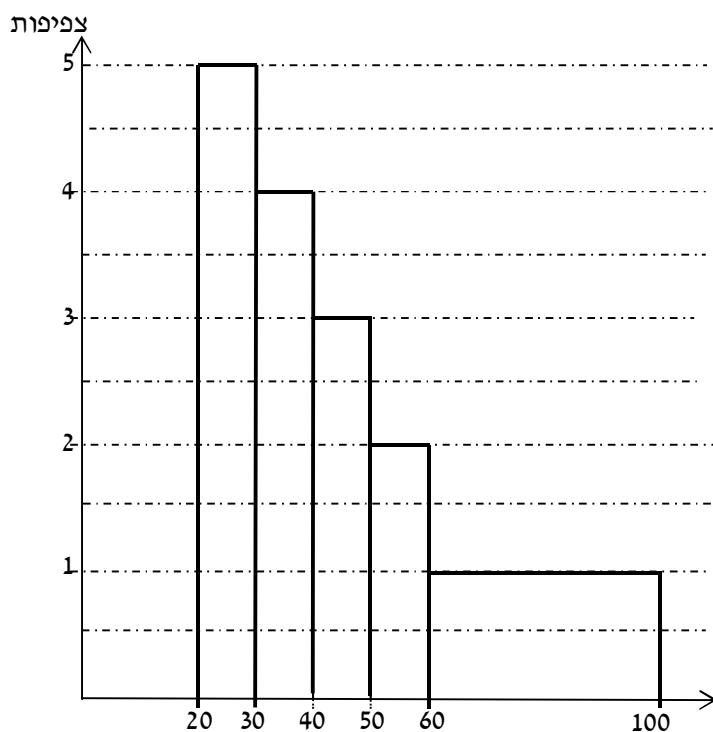
מה הממוצע והחציון של ההתפלגות?

8. להלן התפלגות הגבהים בס"מ בקבוצה מסוימת.

| שכיחות | גובה בס"מ |
|--------|-----------|
| 30     | 150-160   |
| 40     | 160-170   |
| 60     | 170-175   |
| 70     | 175-180   |
| 40     | 180-190   |

חשב את הממוצע, החציון והשכיח של הגבהים בקבוצה זו.

9. בפקולטה מסוימת בדקו לסטודנטים העובדים בה את השכר לשעת עבודה. להלן התוצאות:



א. מצא את השכיח בהתפלגות.

ב. מצא את החציון בהתפלגות.

ג. הסבירו ללא חישוב האם הממוצע גדול/קטן/שווה לחציון.

ד. הסתבר שיש להוציא מספר תלמידים במחלקה בין 20-30 שקלים כיצד הדבר ישפיע על הממוצע, החציון והשכיח? הסבירו ללא חישוב.

**פתרונות:****שאלה 1:**

החציון : 7

השכיח : 6

הממוצע : 6.9

**שאלה 2:**

א. 3

ב. שכיח : 3,4 חציון : 4

**שאלה 3:**

א. הממוצע : 1.7

החציון : 1.5

השכיח : 1

ב. הממוצע יגדל ויתר המדדים לא ישתנו.

**שאלה 4:**

א. 630

ב. 34.13%

ג. שכיח וחציון : 3

ממוצע : 2.952

**שאלה 5:**

תשובה : ב

**שאלה 6:**

ב חציון : 2 שכיח : 2 אמצע טווח : 1.5

**שאלה 7:**

חציון וממוצע : 55

## פרק 5 - סטטיסטיקה תיאורית - מדדי פיזור : הטווח, השונות וסטיית התקן

### רקע:

**המטרה :** למדוד את הפיזור של הנתונים כלומר כמה הם רחוקים זה מזה ושונים זה מזה.

### הטווח ותחום RANGE:

ההפרש בין התצפית הגבוהה ביותר לנמוכה ביותר :  $R = X_{\max} - X_{\min}$

### שונות וסטיית תקן:

השונות היא ממוצע ריבועי הסטיות מהממוצע וסטיית התקן היא שורש של השונות.

$$s_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2 \quad \text{עבור סדרת נתונים :}$$

דוגמה : נחשב את השונות של סדרת המספרים הבאה : 5,4,9

$$s_x^2 = \frac{\sum (x - \bar{x})^2 f}{n} = \frac{\sum x^2 \cdot f}{n} - \bar{x}^2 \quad \text{עבור טבלת שכיחויות :}$$

להלן התפלגות הציונים בכיתה מסוימת בה ממוצע הציונים הוא 7.44

| הציון X- | השכיחות f- | $x^2 \cdot f$ |
|----------|------------|---------------|
| 5        | 2          | 50            |
| 6        | 4          | 144           |
| 7        | 8          | 392           |
| 8        | 5          | 320           |
| 9        | 4          | 324           |
| 10       | 2          | 200           |
| סה"כ     |            | 1430          |

$$s_x^2 = \frac{\sum x^2 f(x)}{n} - \bar{x}^2 = \frac{1430}{25} - 7.44^2 = 1.8464$$

$$s = \sqrt{s_x^2} = \sqrt{1.8464} = 1.3588$$

כשיש מחלקות נעזר באמצע המחלקה כדי לחשב את השונות.

### תרגילים:

1. להלן רשימת הציונים של 20 תלמידים שנבחנו במבחן הבנת הנקרא:

7, 6, 8, 9, 10, 6, 4, 5, 8, 7, 6, 7, 6, 8, 9, 6, 7, 8, 5, 6

חשבו את השונות, סטיית התקן והטווח של הציונים.

2. להלן התפלגות מספר המכוניות למשפחה ב"הגורן"

| מספר מכוניות למשפחה | 1  | 2   | 3   | 4   | 5  |
|---------------------|----|-----|-----|-----|----|
| שכיחות              | 65 | 150 | 220 | 140 | 55 |

א. חשבו סטיית התקן.

ב. חשבו את הטווח של הנתונים.

הקפידו להסביר לגבי כל סעיף מה משמעות התוצאה שקיבלתם!

3. בחברה העוסקת בטלמרקטינג בדקו עבור כל עובד את מספר שנות הוותק שלו. התקבל שממוצע שנות הוותק הוא 4 שנים וסטיית התקן היא שנתיים.

א. האם הממוצע יגדל/יקטן/לא ישתנה וסטיית התקן תגדל/תקטן/לא תשנה כאשר יתווספו שני עובדים עם וותק של 4 שנים להתפלגות?

ב. האם הממוצע יגדל/יקטן/לא ישתנה וסטיית התקן תגדל/תקטן/לא תשנה כאשר יתווספו שני

עובדים אשר אחד עם וותק של 0 שנים והשני עם וותק של 8 שנים להתפלגות?

4. נתונה רשימה של 5 תצפיות, אך רק עבור 4 מהן נרשמו הסטיות שלהן מהממוצע:

2, 3, 2, 1. חשב את השונות של חמש התצפיות.

5. בשכונה בדקו בכל דירה את מספר החדרים לדירה. בשכונה 200 דירות.

| מספר חדרים | פרופורציה |
|------------|-----------|
| 1          | 0.1       |
| 2          | 0.2       |
| 3          | 0.4       |
| 4          | 0.15      |
| 5          |           |

א. מה הממוצע של מספר החדרים לשכונה בדירה?

ב. חשבו את סטיית התקן של מספר החדרים לדירה.

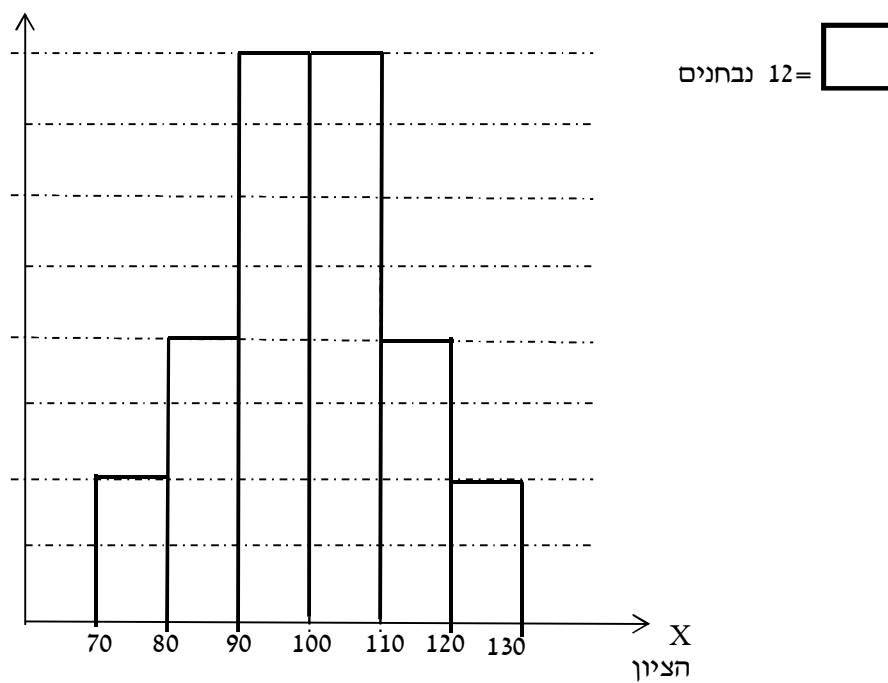
ג. חלק מבעלי הדירות בנות 2 החדרים הפכו את דירתם לדירת חדר. כיצד הדבר ישפיע (יקטין, יגדל, לא ישנה) כל מדד שחישבתם בסעיפים הקודמים.

6. להלן התפלגות המשקל של קבוצה מסוימת בק"ג:

| משקל  | מספר מקרים |
|-------|------------|
| 40-45 | 10         |
| 45-50 | 20         |
| 50-60 | 30         |
| 60-65 | 20         |
| 65-70 | 10         |

מהי סטיית התקן של התפלגות המשקל?

7. להלן התפלגות הציונים במבחן אינטליגנציה:



- מה הממוצע ומה החציון של ההתפלגות?
- חשבו את סטיית התקן של הציונים.
- מסתבר שיש להוסיף 20 תצפיות לכל אחת משתי המחלקות 90-100 ו-100-110. כיצד הדבר ישתנה את כל אחד מהמדדים של הסעיפים הקודמים?

**פתרונות :****שאלה 1:**

השונות : 2.19

סטיית תקן : 1.48

טווח : 6

**שאלה 2:**

א. סטיית תקן : 1.106

ב. טווח 4

**שאלה 3:**

א. ממוצע לא ישתנה, סטיית התקן תקטן.

ב. ממוצע לא ישתנה, סטיית התקן תגדל.

**שאלה 4:**

10.8

**שאלה 5:**

א. 3.05

ב. 1.16

**שאלה 6:**

7.73

**שאלה 7 :**

א. 100

ב. 12.96

## פרק 6 - סטטיסטיקה תיאורית - מדדי פיזור - טווח בין-רבעוני

### רקע:

הטווח הבין-רבעוני נותן את הטווח בין הרבעונים בו נמצאים 50% מהתצפיות המרכזיות.

שלבים במציאת טווח בין-רבעוני במחלקות:

| F   | f<br>מספר<br>עובדים<br>(שכיחות) | $L_1 - L_0$<br>רוחב | מספר שנות ותק |
|-----|---------------------------------|---------------------|---------------|
| 56  | 56                              | 4                   | 0.5 – 4.5     |
| 106 | 50                              | 5                   | 4.5 – 9.5     |
| 154 | 48                              | 2                   | 9.5 – 11.5    |
| 190 | 36                              | 3                   | 11.5 – 14.5   |
| 200 | 10                              | 5                   | 14.5 – 19.5   |

שלב א: נימצא את הרבעון התחתון ( האחוזון ה-25 ) והרבעון העליון ( האחוזון ה-75 ).

מיקום הרבעון התחתון יהיה:  $\frac{n}{4}$

מיקום הרבעון העליון יהיה:  $\frac{3n}{4}$

נוסחאות הרבעונים יהיו:

$$Q_1 = L_0 + \frac{\frac{n}{4} - F(x_{m-1})}{f(x_m)} \cdot (L_1 - L_0) \quad ; \quad Q_3 = L_0 + \frac{\frac{3n}{4} - F(x_{m-1})}{f(x_m)} \cdot (L_1 - L_0)$$

שלב ב: נחסר את הרבעונים:

$$IQR = Q_3 - Q_1$$

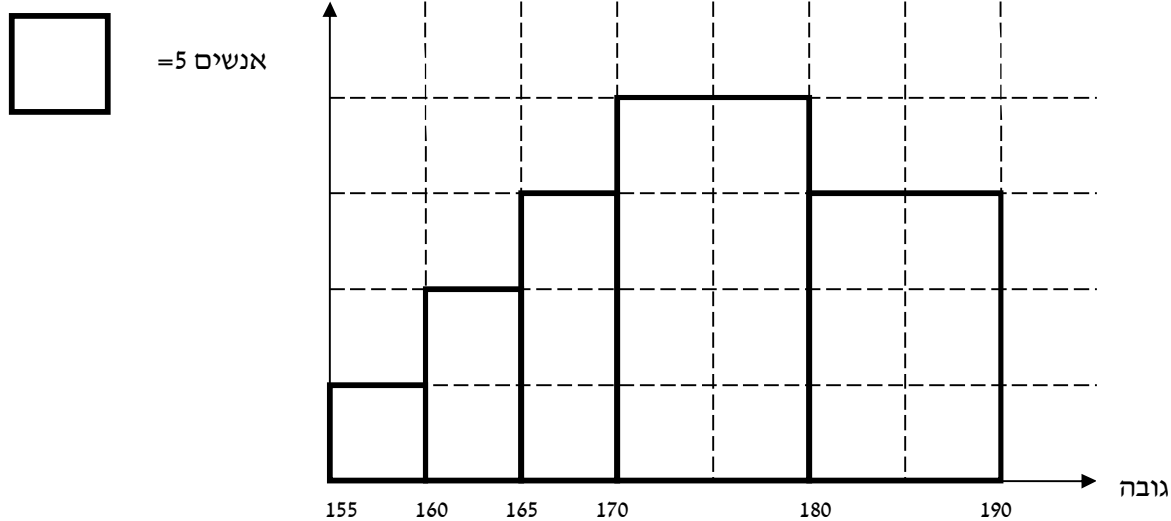
### תרגילים:

1. להלן התפלגות המשקל של קבוצה מסוימת בק"ג:

| משקל  | מספר מקרים |
|-------|------------|
| 40-45 | 10         |
| 45-50 | 20         |
| 50-60 | 30         |
| 60-65 | 20         |
| 65-70 | 10         |

מצא את הטווח הבין-רבעוני.

2. להלן היסטוגרמה המתארת את התפלגות הגבהים בס"מ של קבוצה מסוימת:



מצא את הטווח הבין-רבעוני.

**פתרונות:**

**שאלה 1:**

13.75

**שאלה 2:**

13.33

## פרק 7 - סטטיסטיקה תיאורית - ממוצע משוקלל ושונות מצורפת

### רקע:

מדובר על מצב שבו ישנן כמה קבוצות שרוצים לאחד אותן לקבוצה אחת גדולה.

מתעניינים בממוצע והשונות של הקבוצה הגדולה המתקבלת מאיחוד הקבוצות הקטנות.

$$n_j = \text{מס' התצפיות בקבוצת ה-} j$$

$j$  = אינדקס של הקבוצה.

$$N = \text{מס' התצפיות בכל הקבוצות יחד (סכום כל ה-} n_j)$$

$$\bar{x}_j = \text{הממוצע בקבוצה ה-} j$$

$$S_j^2 = \text{השונות בקבוצה ה-} j$$

הנוסחאות לממוצע משוקלל ושונות מצורפת:

$$\bar{\bar{x}} = \frac{\sum_{j=1}^k \bar{x}_j n_j}{N} ; \quad N = \sum_{j=1}^k n_j ; \quad s_c^2 = \frac{\sum_{j=1}^k n_j s_j^2}{N} + \frac{\sum_{j=1}^k n_j (\bar{x}_j - \bar{\bar{x}})^2}{N}$$

דוגמה: (פתרון בהקלטה)

בחברה שני אגפים, אגף א' מונה עשרים עובדים, השכר הממוצע שם הוא 6,000 ₪ וסטיית התקן היא 2,000 ₪. באגף ב' עשרה עובדים השכר הממוצע הוא 12,000 ₪ וסטיית התקן היא 3,000 ₪.

מהו השכר הממוצע ומהי סטיית התקן של שכר העובדים בחברה?

### תרגילים:

1. להלן נתונים לגבי ציונים במבחן באנגלית ב-3 כיתות מתוך שכבה י' בתיכון:

| כיתה | ממוצע | מס' תלמידים | סטיית תקן |
|------|-------|-------------|-----------|
| 1    | 76    | 40          | 12        |
| 2    | 68    | 20          | 15        |
| 3    | 82    | 30          | 10        |

א. חשב את הממוצע המשוקלל בשכבה.

ב. חשב את השונות המצורפת בשכבה.

2. נתונות שתי קבוצות:

בקבוצה I פי שתיים תצפיות מאשר בקבוצה II.

הממוצע בשתי הקבוצות הוא 70.

השונות בקבוצה I היא 100.

השונות בקבוצה II היא 400.

א. מצא את הממוצע של התצפיות לאחר שאוחדו שתי הקבוצות לקבוצה אחת.

ב. מצא את סטיית התקן של התצפיות לאחר שאוחדו שתי הקבוצות לקבוצה אחת.

**פתרונות:****שאלה 2**

א. 70

ב. 14.14

**שאלה 1**

א. 76.22

ב. 173.5

## פרק 8 - סטטיסטיקה תיאורית - מדדי מיקום יחסי - ציון תקן

### רקע:

המטרה למדוד איך תצפית ממוקמות יחסית לשאר התצפיות בהתפלגות.

### ציון תקן:

$$Z = \frac{X - \bar{X}}{S} : \text{הנוסחה לציון תקן של תצפית היא}$$

ציון התקן נותן כמה סטיות תקן סוטה התצפית מהממוצע.

כלומר, ציון התקן מעיד על כמה סטיות תקן התצפית מעל או מתחת לממוצע.

ציון תקן חיובי אומר שהתצפית מעל הממוצע.

ציון תקן שלילי אומר שהתצפית מתחת לממוצע.

ציון תקן אפס אומר שהתצפית בדיוק בממוצע.

### דוגמה: ( פתרון בהקלטה )

במקום עבודה מסוים ממוצע המשכורות 8 אלפי ₪ עם סטית תקן של 2 אלפי ₪ באותו מקום עבודה ההשכלה הממוצעת של העובדים הנה 14 שנים עם סטית תקן של 1.5 שנים. ערך מרויח במקום עבודה זה 11 אלף ₪ והשכלתו 16 שנים. מה ערך יותר באופן יחסי משכיל או משתכר ?

### תרגילים:

1. תלמידי כיתה ח' ניגשו למבחן בלשון ולמבחן במתמטיקה.  
להלן התוצאות שהתקבלו:

| המקצוע  | ממוצע | סטיית תקן |
|---------|-------|-----------|
| לשון    | 74    | 12        |
| מתמטיקה | 80    | 16        |

עודד קיבל: 68 בלשון ו70 במתמטיקה.

- א. באיזה מקצוע עודד טוב יותר באופן יחסי לשכבה שלו?  
ב. איזה ציון עודד צריך לקבל במתמטיקה כדי שיהיה שקול לציונו בלשון?

2. במפעל לייצור מצברים לרכב בדקו במשך 40 ימים את התפוקה היומית ( מספר מצברים במאות) ואת מספר הפועלים שעבדו באותו היום.  
להלן טבלה המסכמת את האינפורמציה שנאספה על שני המשתנים:

|           | תפוקה | מספר פועלים |
|-----------|-------|-------------|
| ממוצע     | 48    | 15          |
| סטיית תקן | 10    | 2           |

באחד הימים מתוך כלל הימים שנבדקו התפוקה הייתה 50 מאות מצברים ובאותו היום עבדו 13 פועלים.

מה יותר חריג באותו היום יחסית לשאר הימים שנבדקו נתוני התפוקה או כמות הפועלים?  
בחר בתשובה הנכונה.

א. התפוקה.

ב. כמות הפועלים.

ג. חריגים באותה מידה.

ד. חסרים נתונים כדי לדעת זאת.

3. הגובה הממוצע של המתגייסים לצבא הוא 175 סנטימטר עם סטיית תקן 10 סנטימטר. המשקל הממוצע 66 ק"ג עם סטיית תקן 8 ק"ג. ערן התגייס, גובהו 180 ס"מ ומשקלו 59 ק"ג.  
א. במה ערן חריג יותר ביחס לשאר המתגייסים- גובהו או משקלו?  
ב. כמה ערן אמור לשקול כדי שמשקלו יהיה שקול לגובהו?

**פתרונות:**

**שאלה 1:**

א. לשון

ב. 72

**שאלה 2:**

תשובה ב

**שאלה 3:**

א. משקל

ב. 70

## פרק 9 - סטטיסטיקה תיאורית - מדדי מיקום יחסי - אחוזונים במחלקות

### רקע:

האחוזון (המאון) ה- $p$  הוא הערך בנתונים המחלק את הנתונים בצורה כזאת שעד אליו יש  $p\%$  מהנתונים. מסמנים את האחוזון ה- $p$  ב- $X_p$ .

למשל, המאון ה-25 הוא האחוזון ה-25 או הרבעון התחתון : ערך ש-רבע מהתצפיות קטנות

ממנו והשאר גבוהות ממנו. מסומן :  $X_{0.25}$

מצאת מאון במחלקות :

שלב א : נימצא את המחלקה הרלבנטית שמיקומה יהיה  $\frac{np}{100}$ .

שלב ב : נציב בנוסחה הבאה :  $x_p = L_0 + \frac{\frac{n \cdot p}{100} - F(x_{m-1})}{f(x_m)} \cdot (L_1 - L_0)$

$F(x_{m-1})$  - שכיחות מצטברת של מחלקה אחת לפני המחלקה הרלבנטית.

$f(x_m)$  - השכיחות של המחלקה הרלבנטית.

$L_0$  - גבול התחתון של המחלקה.

$L_1$  - גבול העליון של המחלקה.

אם רוצים לחלץ את אחוז התצפיות שמתחת לערך מסוים נשתמש בנוסחה הבאה :

$$P_x = \left[ \frac{(x - L_0)}{(L_1 - L_0)} \cdot f(x_m) + F(x_{m-1}) \right] \cdot \frac{100}{n}$$

דוגמה : (פתרון בהקלטה)

להלן התפלגות השכר של עובדים בחברה מסוימת :

| שכר בש"ח    | $f(x)$ |
|-------------|--------|
| 4000-6000   | 140    |
| 6000-10000  | 128    |
| 10000-15000 | 60     |
| 15000-20000 | 54     |
| 20000-40000 | 18     |

א. מצאו את המאון ה-40.

ב. מהו אחוז העובדים שמשכרים מתחת ל-5,000 ₪?

### תרגילים:

1. להלן התפלגות השכר (באלפי שקלים) בחברה:

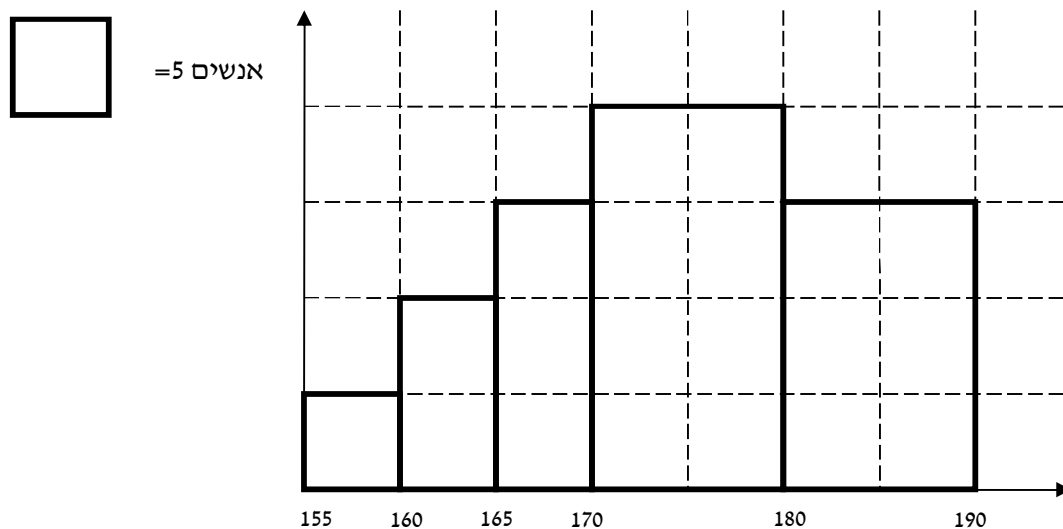
| שכר X | שכיחות מצטברת |
|-------|---------------|
| 6-10  | 48            |
| 10-15 | 100           |
| 15-20 | 120           |
| 20-30 | 132           |
| 30-60 | 136           |

- א. חשבו את המאון ה-60.
- ב. מהו העשירון העליון?
- ג. 20% מהמשכורות הגבוהות ביותר הן משכורות של הבכירים, מהי המשכורת המינימאלית לבכיר?
- ד. מה אחוז האנשים שמשתכרים מתחת ל-7000 ₪?
- ה. איזה אחוז מהעובדים משתכרים מעל ל-25,000 ₪?
- ו. איזה אחוז מהעובדים משתכרים בין 7000 ל-25,000 ₪?

2. למבחן ניגשו 400 נבחנים. נתון שהעשירון התחתון הוא הציון 60. הרבעון העליון הוא הציון 80. כמו כן ההתפלגות של הציונים היא סימטרית. מלאו את השכיחות החסרות.

| ציון - X | $f(X)$ |
|----------|--------|
| 50-60    |        |
| 60-70    |        |
| 70-80    |        |
| 80-90    |        |
| 90-100   |        |

3. להלן היסטוגרמה המתארת את התפלגות הגבהים בס"מ של קבוצה מסוימת:



א. העשירון התחתון.

ב. האחוזון ה-30.

ג. הגובה ש-20% מהתצפית גדולות ממנו.

ד. את אחוז התצפיות מתחת לגובה 158 ס"מ.

ה. את אחוז התצפיות מעל לגובה 185 ס"מ.

ו. את אחוז התצפיות בין גובה 170 ס"מ ל-185 ס"מ.

**פתרונות:****שאלה 1 :**

- א. 13.23
- ב. 22
- ג. 17.2
- ד. 8.82%
- ה. 7.36%
- ו. 83.82%

**שאלה 3:**

- א. 162.5
- ב. 170
- ג. 183.33
- ד. 3%
- ה. 15%
- ו. 55% ת

## פרק 10 - סטטיסטיקה תיאורית - טרנספורמציה לינארית

### רקע:

מצב שבו מבצעים שינוי מסוג הוספה של קבוע ( או החסרה ) והכפלה של קבוע ( או חילוק ) לכל התצפיות :  $y = a \cdot x + b$

וכך יושפעו המדדים השונים :

### מדדי המרכז :

$$MR_y = a \cdot MR_x + b$$

$$Mo_y = a \cdot Mo_x + b$$

$$\bar{y} = a \cdot \bar{x} + b$$

$$Md_y = a \cdot Md_x + b$$

### מדדי הפיזור :

$$R_y = |a| R_x$$

$$s_y = |a| s_x$$

$$s_y^2 = a^2 s_x^2$$

### מדדי המיקום היחסי :

$$Y_p = a \cdot X_p + b$$

$$Z_y = \frac{a}{|a|} Z_x$$

### שלבי העבודה :

1. נזהה שמדובר בטרנספורמציה לינארית ( שינוי קבוע לכל התצפיות ).
2. נרשום את כלל הטרנספורמציה לפי נתוני השאלה.
3. נפשט את הכלל ונזהה את ערכי  $a$  ו  $b$ .
4. נציב בנוסחאות שלעיל בהתאם למדדים שנשאלים.

### דוגמה: (פתרון בהקלטה)

השכר הממוצע של עובדים הנו 9000 ₪ וטווח 6000 ₪ חשבו את המדדים הללו לאחר שהעלו את כל המשכורות ב-10% ואחר כך קנסו אותם ב100 ₪.

**תרגילים:**

1. עבור סדרת נתונים התקבל:

$$\bar{X} = 80$$

$$S = 15$$

$$MO = 70$$

הוחלט להכפיל את כל התצפיות פי-4 ולהחסיר מהתוצאה 5. חשב את המדדים הללו לאחר השינוי.

2. בחברה מסוימת השכר הממוצע הוא 40 ₪ לשעה עם סטיית תקן של 5 ₪ לשעה. הוחלט להעלות את כל המשכורות ב-10%, אך זה לא סיפק את העובדים ולכן הם קיבלו לאחר מכן תוספת של 2 ₪ לשעה. מה הממוצע ומהי השונות של השכר לשעה לאחר כל השינויים.

3. במבחן הציון החציוני היה 73, טווח הציונים היה 40 נקודות. והעשירון העליון היה הציון 87. כיוון שהציונים בבחינה היו נמוכים, המורה החליט לתת פקטור של 4 נק' לכל התלמידים. חשבו את המדדים לאחר הפקטור.

4. דגמו מקו ייצור 50 קופסאות של גפרורים. בדקו בכל קופסא בה יש 40 גפרורים את כמות הגפרורים הפגומים. קבלו שבממוצע יש 3 גפרורים פגומים בקופסא. עם סטיית תקן של 1.5 גפרורים. מה יהיה הממוצע ומה תהיה סטיית התקן של מספר התקנים בקופסא?

5. חברת בזק הציעה את החבילה הבאה:  
שלושים שקלים דמי מנוי חודשיים קבועים. ובנוסף 10 אגורות לכל דקה של שיחה יוצאת, אדם בדק במשך שנה את דקות השיחות היוצאות שלו, וקיבל שבממוצע בחודש יש לו 600 דקות שיחות יוצאות. עם שונות 2500 דקות רבועות, כמו כן בחודש ינואר ציון התקן היה 2. חשבו את המדדים הללו עבור חשבון הטלפון החודשי של אותו אדם בשקלים אם היה משתמש בחבילה המוצעת לו על ידי בזק.

6. הוכח שאם כל התצפיות בהתפלגות עברו טרנספורמציה לינארית:

$$Y_i = a \cdot X_i + b$$

אזי הממוצע והשונות של כלל התצפיות לאחר הטרנספורמציה יהיו בהתאמה:

$$\bar{y} = a \cdot \bar{x} + b$$

$$s_y^2 = a^2 s_x^2$$

**פתרונות :****שאלה 1 :**

הממוצע : 315

סטיית התקן : 60

השכיח : 275

**שאלה 2 :**

הממוצע : 46

השונות : 30.25

**שאלה 3 :**

טווח : 40

חציון : 77

עשירון עליון : 91

**שאלה 4 :**

ממוצע : 37

סטיית תקן : 1.5

**שאלה 5 :**

ממוצע : 90

שונות : 25

ציון תקן : 2

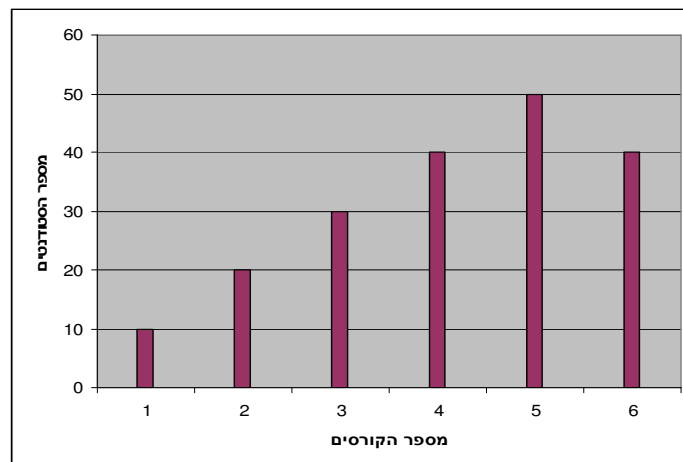
## פרק 11 - סטטיסטיקה תיאורית - שאלות מסכמות

1. בדקו עבור 5 תלמידים את המשקל שלהם :

| מספר תלמיד | משקל בק"ג |
|------------|-----------|
| 1          | 58        |
| 2          | 62        |
| 3          | 48        |
| 4          | 34        |
| 5          | 58        |

- א. מהו המשתנה הנחקר? האם הוא בדיד או רציף?
- ב. מהו המשקל החציוני, הממוצע והשכיח?
- ג. מה הטווח וסטיית התקן של המשקל?
- ד. לאותם תלמידים חישבו גם את הגובה בס"מ וקיבלו גובה ממוצע של 168 וסטיית תקן 6. במה תלמיד מספר 3 שגובהו 162 יותר חריג במשקל או בגובה?
- ה. הוסיפו עוד תלמיד השוקל 52 ק"ג בדיוק. הסבירו ללא חישוב כיצד הדבר ישפיע על הממוצע וסטיית התקן? ( יגדיל יקטין או לא ישנה)

2. בפקולטה להנדסה אספה מזכירות הסטודנטים נתונים לגבי מס' הקורסים שכל סטודנט סיים בשנה הראשונה ללימודיו בשנת 2008.  
להלן התוצאות שהתקבלו :



- א. מה המשתנה הנחקר? האם הוא בדיד או רציף?
- ב. מהי צורת ההתפלגות?
- ג. תאר את הנתונים בטבלת שכיחויות.
- ד. חשב את השכיח, החציון והטווח.

3. להלן התפלגות הציונים בבחינה בלשון שנעשתה עבור תלמידי כיתות ד'. השתתפו במחקר 150 תלמידים.

$$\text{ממוצע הציונים שהתקבל: } \bar{X} = 7 \frac{1}{15}$$

| ציון | מספר התלמידים |
|------|---------------|
| 4    | 12            |
| 5    | 16            |
| 6    |               |
| 7    | 38            |
| 8    |               |
| 9    | 14            |
| 10   | 10            |

- א. השלם את השכיחויות החסרות בטבלה.
- ב. חשב את הציון החציוני, השכיח.
- ג. חשב שונות וסטיית תקן להתפלגות הציונים.

4. חברה סלולארית דגמה 200 אנשים. עבור כל אדם נבדקה מידת שביעות הרצון של הלקוח מהחברה (1 – שביעות רצון נמוכה ועד 5 שביעות רצון גבוהה) להלן ההתפלגות שהתקבלה:

| שביעות רצון | מספר האנשים |
|-------------|-------------|
| 1           | 40          |
| 2           | 60          |
| 3           | 50          |
| 4           | 30          |
| 5           | 20          |

- א. מה אחוז האנשים עם רמת שביעות רצון נמוכה?
- ב. מה המשתנה הנחקר ומאיזה סוג הוא?
- ג. מהי הדרך הגרפית המתאימה ביותר לתיאור הנתונים?
- i. היסטוגרמה.
- ii. דיאגרמת מקלות.
- iii. דיאגרמת עוגה
- ד. חשבו את המדדים הבאים:
1. טווח
2. שכיח
3. חציון

5. להלן התפלגות מספר שעות העבודה לשבוע של העובדים בחברת "סטאר".  
בחברה 200 עובדים.

| שכיחות | שכיחות יחסית<br>(פרופורציה) | מספר שעות עבודה |
|--------|-----------------------------|-----------------|
|        | 15%                         | 10-20           |
|        | 20%                         | 20-30           |
|        | 30%                         | 30-40           |
|        | 20%                         | 40-50           |
|        |                             | 50-60           |

- א. השלם את הטבלה.
- ב. חשב את החציון, השכיח, והממוצע של התפלגות מס' שעות העבודה בחברה.
- ג. מהי סטיית התקן של מס' שעות העבודה?
- ד. מה העשירון העליון של ההתפלגות?
- ה. איזה אחוז מהעובדים עובדים מעל 45 שעות בשבוע?
- ו. מה ציון התקן של רינה שעובדת 30 שעות בשבוע?
- ז. כיצד ישתנה החציון, הממוצע וסטיית התקן אם מספר שעות העבודה המינימאלי אינו 10 אלא 15? הסבר.

6. חברה סלולארית דגמה 200 אנשים. עבור כל אדם נבדק מס' המסרונים ששלח במשך חודש. להלן ההתפלגות שהתקבלה:

| מספר המסרונים | מספר האנשים |
|---------------|-------------|
| 0-50          | 40          |
| 50-100        | 60          |
| 100-150       | 50          |
| 150-250       | 30          |
| 250-ומעלה     | 20          |

- א. מה אחוז האנשים ששלחו פחות מ-80 מסרונים בחודש?  
 ב. מה אחוז האנשים ששלחו בין 50 ל-120 מסרונים?  
 ג. הוחלט להעניק מתנה עבור  $\frac{1}{4}$  מהלקוחות שמשתמשים במספר הרב ביותר של מסרונים בחודש. החל מאיזה כמות של מסרונים תחולק המתנה?  
 ד. ציינו איזה מדד ניתן לחשב ואיזה לא ניתן. אם ניתן חשב:
1. ממוצע
  2. שכיח
  3. חציון
  4. שונות

7. נתנו לקבוצת ילדים לבצע משימה מסוימת ובדקו את התפלגות זמן ביצוע המשימה בדקות. להלן ההתפלגות שהתקבלה:

| זמן בדקות | מספר הילדים |
|-----------|-------------|
| 0.5-3.5   | 20          |
| 3.5-9.5   | 18          |
| 9.5-19.5  | 14          |
| 19.5-29.5 | 8           |

- א. שרטט היסטוגרמה לתיאור התפלגות זמן ביצוע המשימה.  
 ב. מתוך ההיסטוגרמה שבנית בסעיף א מהי צורת ההתפלגות?  
 ג. חשב את השכיח והחציון של ההתפלגות.  
 ד. הסבר, ללא חישובים, האם הזמן הממוצע לביצוע המשימה, קטן או גדול או שווה ביחס לשכיח ולחציון.

8. התפלגות ציוני מבחן אינטיליגנציה היא סימטרית .

| הציון   | מספר הנבחנים |
|---------|--------------|
| 50-70   |              |
| 70-90   |              |
| 90-100  |              |
| 100-110 |              |
| 110-130 |              |
| 130-150 |              |

נתון שהעשירון העליון הוא 130 והרבעון התחתון הוא 90.

נתון שלמבחן נגשו 500 מועמדים.

א. השלימו את הטבלה.

ב. מהו הממוצע והחציון של ההתפלגות?

ג. מהו הציון ש 40% מהתלמידים קיבלו מעליו? באיזה אחוזון מדובר?

ד. אם יוחלט להעלות את כל הציונים ב-10 נקודות . כיצד הדבר ישפיע על הממוצע וסטיית התקן של הציונים?

9. להלן מספר טענות , עבור כל טענה ציין אם היא נכונה או לא נכונה ונמקו .

א. בסדרה שבה כל התצפיות שוות זו לזו השונות הינה 0.

ב. ציון התקן של החציון תמיד יהיה 0.

ג. ציון התקן של האחוזון ה-70 בהתפלגות אסימטרית ימנית (חיובית) תמיד יהיה חיובי.

ד. אם נוסיף תצפיות לסדרה של תצפיות, הדבר בהכרח יגדיל את הממוצע של הסדרה.

ה. בסדרה החציון הינו 80. הוספו שתי תצפיות אחת 79 ואחת 100 לכן החציון יגדל.

ו. אם נוסיף את הערך 4 לכל התצפיות אז סטיית התקן לא תשתנה.

ז. אם נחלק את כל התצפיות בהתפלגות ב-2 אז השונות תקטן פי 2.

ח. אם נגדיל את ממוצע המשכורות של עובדים בחברה אז גם השונות תגדל.

**פתרונות:****שאלה 1:**

- א. המשתנה הנחקר כאן הוא משקל תלמיד בק"ג והוא משתנה כמותי רציף.  
 ב.

$$\bar{X} = 52$$

$$Md = X_{\frac{n+1}{2}} = X_3 = 58$$

השכיח הוא 58

$$R = 28 \quad \text{ג.}$$

$$s = 10.12$$

- ד. הוא חריג יותר בגובה כי שם ציון התקן בערך מוחלט יותר גבוה.  
 ה. הממוצע לא ישתנה אך סטיית התקן תקטן .

**שאלה 2:**

- א. מספר הקורסים. בדיד.  
 ב. התפלגות אסימטרית שמאלית  
 ד. השכיח : 5  
 הטווח : 5

**שאלה 3:**

- א. 20 תלמידים קיבלו ציון 6 ו-40  
 תלמידים קיבלו ציון 8.  
 החציון : 7  
 השכיח : 8  
 ג. השונות : 2.533  
 סטיית התקן : 1.592

**שאלה 4:**

- א. 20%  
 ב. שביעות רצון ( סדר )  
 ג. 2  
 ד. טווח : 4 שכיח : 2 חציון : 2.5  
 ה. חציון : 4

**שאלה 5:**

ב. החציון : 35

השכיח: 35

הממוצע: 35

ג. סטיית תקן : 12.65

ד. 53.333

ה. 25%

ו. -0.395

ז. חציון לא ישתנה, ממוצע יגדל סטיית התקן תקטן.

**שאלה 6:**

א. 38%

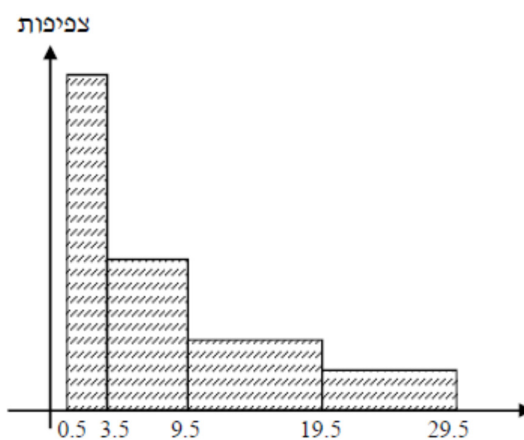
ב. 40%

ג. 150

ד. החציון : 100

**שאלה 7:**

א.



ב. ההתפלגות היא א-סימטרית ימנית .

ג. שכיח: 2 חציון: 6.83

ד. בהתפלגות א-סימטרית ימנית מתקיים  $Mo < Md < \bar{X} < MR$

**שאלה 8:**

א.

| ציון    | מספר הנבחנים |
|---------|--------------|
| 50-70   | 50           |
| 70-90   | 75           |
| 90-100  | 125          |
| 100-110 | 125          |
| 110-130 | 75           |
| 130-150 | 50           |

ב. 100

ג. 104

ד. הממוצע יעלה ב-10 נקודות אך סטיית התקן לא תשתנה .

**שאלה 9:**

א. נכון

ב. לא נכון

ג. לא נכון

ד. לא נכון

ה. לא נכון

ו. נכון

ז. לא נכון

ח. לא נכון

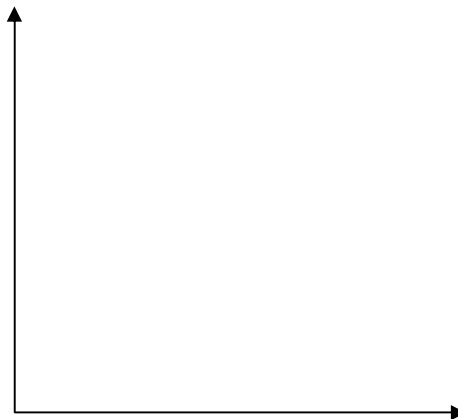
## פרק 12 - מדדי קשר - מדד הקשר הלינארי (פירסון)

### רקע:

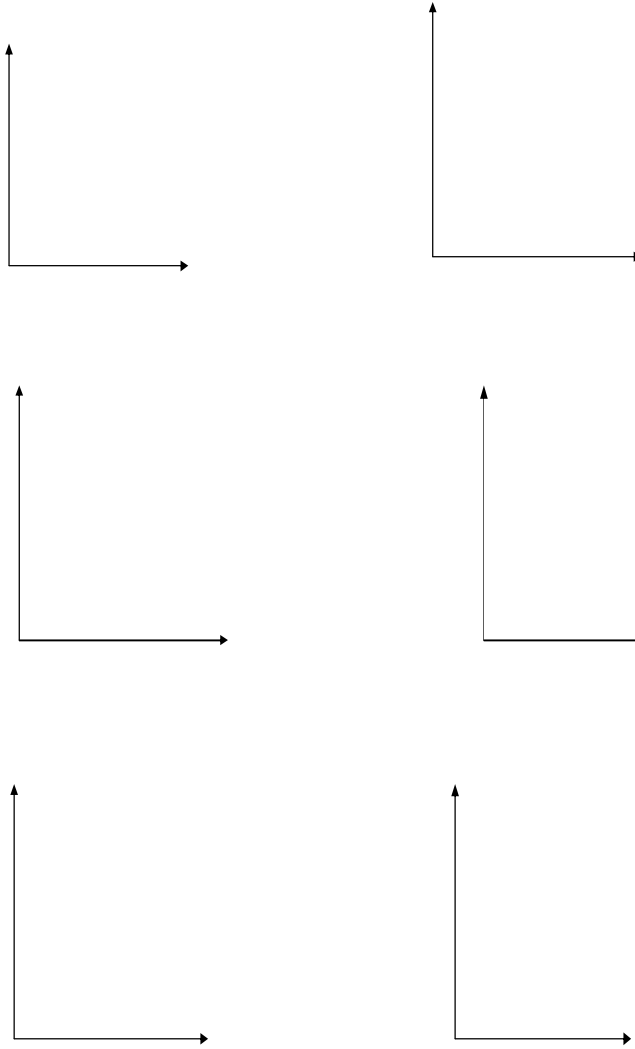
המטרה היא לבדוק האם קיים קשר (קורלציה, מתאם) של קו ישר בין שני משתנים כמותיים. מבחינת סולמות המדידה קשר בין סולמות רווחים ומנה. בדרך כלל,  $X$  הוא המשתנה המסביר (הבלתי תלוי) ו  $Y$  הוא המשתנה המוסבר (התלוי). למשל, נרצה להסביר כיצד השכלה של אדם הנמדדת בשנות לימוד  $X$  מסבירה את ההכנסה שלו  $Y$ . במקרה 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. בשלב הראשון, נהוג לשרטט דיאגרמת פיזור. זו דיאגרמה שנותנת אינדיקציה ויזואלית על טיב הקשר בין שני המשתנים. למשל, 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 הללו דיאגרמת פיזור:



נתבונן בכמה מקרים של דיאגרמות פיזור וננתח אותן :



בשלב השני, מחשבים את מקדם המתאם (מדד הקשר) שבדוק עד כמה קיים קשר לינארי בין שני המשתנים. המדד (ניקרא גם מדד הקשר של פירסון) מכמת את מה שנראה בשלב הראשון רק בעין.

המדד בודק את כיוון הקשר (חיובי או שלילי).

ואת עוצמת הקשר (חלש עד חזק).

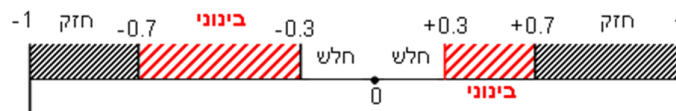
מקדם מתאם זה מקבל ערכים בין -1 ל 1.

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

מתאם חיובי חלקי אומר שככל שמשנתנה 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 שמשנתנה 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.

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

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

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

$$r_{xy} = \frac{COV(x, y)}{s_x \cdot s_y} : \text{מקדם המתאם הלינארי}$$

### תרגילים:

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

- א. שרטט דיאגרמת פיזור לנתונים. מה ניתן להסיק מהדיאגרמה על טיב הקשר בין מספר החיסורים של תלמיד לציונו? מיהו המשתנה הבלתי תלוי ומיהו המשתנה התלוי?
- ב. חשב את מדד הקשר של פירסון. האם התוצאה מתיישבת עם תשובתך לסעיף א'?
- ג. הסבר 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 :

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

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

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

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

ב. מה המשמעות של התוצאה שקיבלת בסעיף א?

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

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

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

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

חשב 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 קבע נכון/לא נכון ונמק!

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 אמריקאיות:**

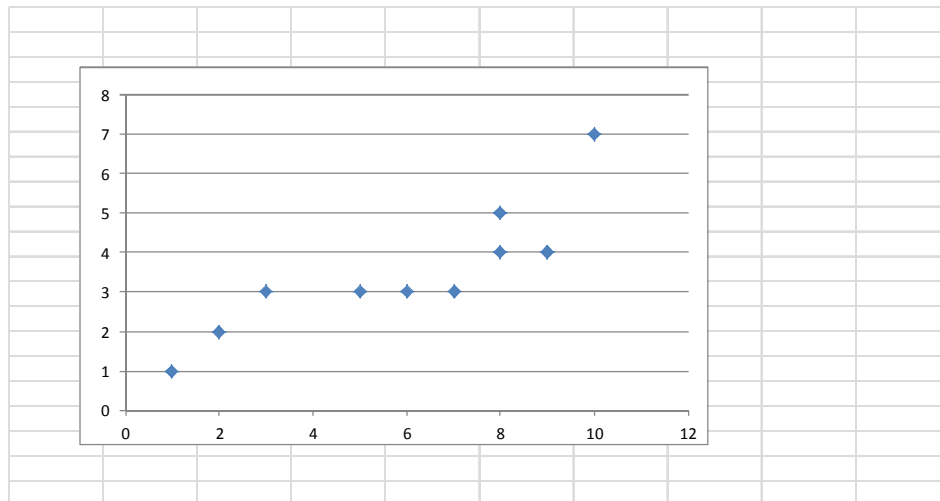
7. נמצא שקיים מקדם מתאם שלילי בין הציון בעברית לציון בחשבון בבחינה לכן :

- 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.

8. נלקחו 20 מוצרים וניבדק 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 דיאגרמת פיזור :



מה יהיה מקדם המתאם בין שני המשתנים?

- א. 1
- ב. 0.85
- ג. 0.15
- ד. 0

**פתרונות:****שאלה 1:**

א. בהקלטה

ב. -0.9325

**שאלה 2:**א.  $\bar{y} = 16$  $\bar{x} = 15.4$ ב.  $r_{xy} = 0.96$ **שאלה 3:**

א : 0.8

**שאלה 4:**

0.8

**שאלה 5:**

1

**שאלה 6:**

א. נכון

ב. לא נכון

ג. נכון

**שאלה 7:**

התשובה : ג

**שאלה 8:**

התשובה : א

**שאלה 9:**

התשובה : ב

## פרק 13 - מדדי קשר - השפעת טרנספורמציה לינאריות על מדד הקשר של פירסון

### רקע:

טרנספורמציה לינארית בין אם נעשית על  $X$  ובין אם נעשית על  $y$ , או בין אם נעשית על שניהם, אינה משנה את עוצמת הקשר. היא עלולה רק לשנות את כיוונו אם השיפועים של שתי הטרנספורמציות שוני סימן.

$$r_{[(aX+b)(cY+d)]} = \begin{cases} r_{x,y} & \text{if } a \cdot c > 0 \\ -r_{x,y} & \text{if } a \cdot c < 0 \end{cases}$$

### תרגילים:

1. מבחן בנוי מחלק כמותי ומילולי.
  - א. אם יעלו את כל הציונים בחלק המילולי ב-20%, מה יהיה מקדם המתאם בין הציון המילולי החדש לציון הכמותי ובין הציון המילולי הישן לציון המילולי החדש?
  - ב. נגדיר משתנה חדש  $W$  להיות המרחק של הציון בחשיבה מילולית מהציון המקסימאלי בבחינה-150. מצא את מקדם המתאם בין הציון המילולי ל- $W$  ובין  $W$  ל-ציון הכמותי.
2. מקדם המתאם בין ההכנסה לבין ההוצאה של 10 משפחות חושב והתקבל 0.7. אם חל גידול של 5% בהכנסת האוכלוסייה כולה וגידול של 7% בהוצאה שלה, אז מה יהיה מקדם המתאם בין ההכנסה החדשה להוצאה החדשה?
3. חברת "לק" המייצרת גלידה החליטה לערוך מחקר לבדיקת הקשר בין מספר חבילות הגלידה הנמכרות ביום לבין הטמפרטורה באותו יום. נבדקו 10 ימים והתקבל מתאם לינארי 0.85. חברת "לק" דואגת להתחיל כל יום עם מלאי של 150 חבילות גלידה. בנוסף, מעוניינים כי הטמפרטורה תבוטא במעלות פרנהייט במקום במעלות צלסיוס. מה ערכו של מקדם המתאם בין מספר חבילות הגלידה שנשארות בסוף היום לבין הטמפרטורה במעלות פרנהייט?
 

הקשר בין מעלות צלסיוס ( $C^\circ$ ) למעלות פרנהייט ( $F^\circ$ ) נתון ע"י  $F = \frac{9}{5}C + 32$ .

בחר בתשובה הנכונה:

  - א. 0.85
  - ב. -0.85
  - ג. 1
  - ד. לא ניתן לדעת.

4. מקדם המתאם בין  $X$  ל- $Y$  הנו 0.4 כל ערכי ה- $X$  הוכפלו ב-2 לכן מקדם המתאם החדש בין

שני המשתנים יהיה :

בחר בתשובה הנכונה:

- א. 0.8
- ב. 0.4
- ג. -0.4
- ד. לא ניתן לדעת.

**פתרונות :****שאלה 1:**

א. בין הציון המילולי הישן לחדש 1:

בין הציון המילולי החדש לכמותי: 0.9

ב. בין  $W$  ל ציון המילולי : -1

בין  $W$  לציון הכמותי: -0.9

**שאלה 2:**

0.7

**שאלה 3:**

התשובה: ב

**שאלה 4:**

התשובה: ב

## פרק 14 - מדדי קשר - רגרסיה ליניארית

### רקע:

במידה וקיים קשר חזק בין שני המשתנים הכמותיים נהוג לבצע ניבויי. לבנות קו ניבויים הנקרא גם קו רגרסיה המנבא משתנה אחד על סמך האחר.

מדובר בקו שמנבא את  $Y$  על סמך  $X$ . השיטה למציאת הקו הנ"ל נקראת שיטת הריבועים הפחותים והקו המתקבל נקרא קו הרגרסיה או קו הניבויים או קו הריבועים הפחותים.

a - בעצם נותן את ערך  $Y$  כאשר  $X$  הנו אפס על גבי קו הניבויים. הוא ניקרא החותך של הקו.  
 b - הוא שיפוע הקו נותן בכמה בעצם  $Y$  משתנה כאשר  $X$  גדל ביחידה אחת על גבי קו הניבויים.  
 להלן המשוואות למציאת הפרמטרים של קו הרגרסיה:

$$\tilde{Y} = bX + a$$

$$b = r \frac{S_y}{S_x}$$

$$a = \bar{Y} - b\bar{X}$$

אם נרצה לבנות קו ניבויים לניבוי  $X$  על סמך  $Y$  נצטרך לעדכן את הנוסחאות בהתאם.

**תרגילים:**

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

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

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

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

- א. חשב את מדד הקשר הלינארי בין  $X$  ל- $Y$ . מיהו המשתנה התלוי?  
 ב. מצא את קו הרגרסיה לניבוי ההוצאה של משפחה על סמך הכנסה שלה. הסבר את משמעות הפרמטרים של קו הרגרסיה.  
 ג. משפחת כהן הכניסה 15,000 ₪, מה ההוצאה הצפויה שלה?

2. נסמן ב- $X$  את ההשכלה של אדם בשנות למוד. נסמן ב- $Y$  את הכנסתו באלפי ₪. במחקר התקבלו התוצאות הבאות:

$$S_y = 5 \quad S_x = 2$$

$$\bar{Y} = 8 \quad \bar{X} = 14$$

$$COV(X, Y) = 7.5$$

- א. חשב את מדד הקשר של פירסון בין ההשכלה להכנסה.  
 ב. מה ההכנסה הצפויה לאדם שהשכלתו 12 שנים?  
 ג. מה ההשכלה הצפויה לאדם שהכנסתו 10,000 ₪?

3. חוקר רצה לחקור את הקשר הקווי שבין הציון המבחן בסטטיסטיקה לבין מספר שעות ההכנה של הסטודנטים למבחן. במדגם של 100 סטודנטים שנבחנו בקורס נרשמו התוצאות הבאות: הציון הממוצע של הסטודנטים היה 65 עם סטיית תקן של 27. מספר שעות ההכנה הממוצע היה 30 עם סטיית תקן של 18. מקדם המתאם בין הציון לשעות ההכנה היה 0.8.

- א. על פי משוואת הרגרסיה שעת הכנה נוספת משפרת את ציון המבחן?  
 ב. על פי משוואת הרגרסיה תלמיד שייגש למבחן ללא שעות הכנה כלל יקבל ציון?  
 ג. מהו קו הרגרסיה לניבוי הציון לפי שעות ההכנה?

4. נתונים 2 משתנים  $Y, X$ . כמו כן נתון:  $X$  ממוצע = 1.5, שונות  $X$  = שונות  $Y$  = 4, וכן שקו הרגרסיה של  $Y$  על בסיס  $X$  הינו  $Y = -0.2X + 0.5$ . חשב מהו מקדם המתאם בין  $X$  ל- $Y$ ?

**פתרונות:****שאלה 1:**

א. 0.8

ב.  $\tilde{Y} = 0.8X + 0.4$

ג. 12.4

**שאלה 2:**

א. 0.75

ב. 4.25 אלפי ש"ח

ג. 14.6 שנים

**שאלה 3:**

א. 1.2

ב. 29

ג.  $y = 1.2x + 29$

**שאלה 4:**

-0.2