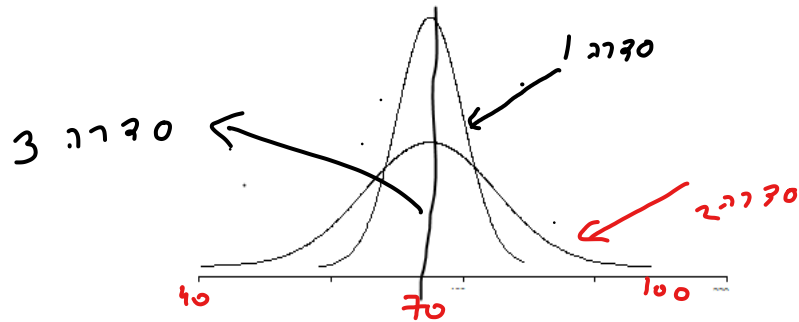


## מדדי פיזור - יחידה 3

דוגמה: מה ההבדל בין ההתפלגויות?



דוגמה: נתבונן על 3 סדרות של ציונים:

סדרה 1: 60, 60, 60, 70, 80, 80, 80	$\bar{X} = 70, Md = 70$
סדרה 2: 40, 50, 60, 70, 80, 90, 100	$\bar{X} = 70, Md = 70$
סדרה 3: 70, 70, 70, 70, 70, 70, 70	$\bar{X} = 70, Md = 70$

מרכז

א"ג

$$R_1 = 80 - 60 = 20$$

$$R_2 = 100 - 40 = 60$$

$$R_3 = 70 - 70 = 0$$

שלושת הסדרות אותו ממוצע אבל הם שונות אחת מהשנייה במידת השוני של הנתונים. לכן, כדי לסכם סידרה של נתונים לא מספיק מדד למיקום מרכזי, נצטרך להגדיר גם מדד לפיזור. המדדים לפיזור מתארים מידת השוני של הנתונים בסדרה, גודל ההבדלים בין הנתונים לבין עצמם או בין הנתונים לערך מרכזי מסוים. ברוב המקרים לוקחים את הפיזור סביב ממוצע.

א"ג

**הגדרה:** מדדים שמאפיינים את מידת ההטרוגניות של קבוצת נתונים (עד כמה הנתונים שונים/רחוקים מהמדד המרכזי בו בחרנו לתאר את הקבוצה). ככל שערך המדד גדול יותר הקבוצה הטרוגנית יותר (פיזור גדול) ככל שערך המדד נמוך יותר הקבוצה הומוגנית יותר (פיזור קטן).

**תכונות** המשותפות לכל מדדי הפיזור:

- מקבלים ערכים חיוביים (כולל 0) בלבד (בוחנים מרחק/סטייה/פיזור).
- כאשר כל התצפיות שוות, מדדי הפיזור מקבלים את הערך 0.
- בכל התפלגות נהוג להציג מדד מרכזי יחד עם מדד פיזור על מנת לקבל תמונה טובה יותר על צורת התפלגות הכוללת גם התייחסות אל הקצוות/פיזור של ההתפלגות.

**תחום (טווח):** ההפרש בין הערך המקסימאלי בנתונים לערך המינימאלי. (RANGE)  $R$

$$R = X_{\max} - X_{\min}$$

**תחום בין-רבעוני** (INTERQUARTILE RANGE - IQR): טווח הערכים של המשתנה בו

נמצאים 50% התצפיות שבמרכז ההתפלגות. מהווה מדד פיזור לחציון.

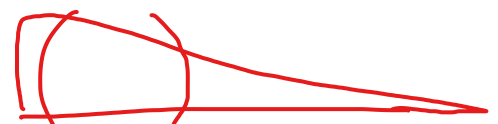
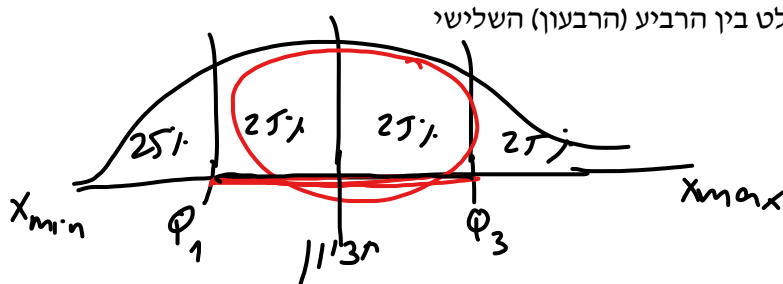
על מנת לחשב את התחום הבין רבעוני יש לחשב את הרבעונים של ההתפלגות:

Q1 - הרביע (הרבעון) הראשון: הערך שעד אליו 25% ומעליו 75% מהתצפיות בהתפלגות

Q3 - הרביע (הרבעון) השלישי: הערך שעד אליו 75% ומעליו 25% מהתצפיות בהתפלגות

התחום הבין רבעוני הוא ההפרש בערך מוחלט בין הרביע (הרבעון) השלישי

$$IQR = Q3 - Q1$$



$$p_3 - p_1 = 2 - 0 = 2$$

הטלות הבין רבזני זמ הקואטר מ<sub>2</sub> ל<sub>1</sub> קדק



שונות וסטיית תקן - מדד פ'י'ר ס'ק'ב הממוצע

הרעיון: נסתכל על הממוצע, ונראה עד כמה התצפיות שהתקבלו במדגם רחוקות ממנו. (מהממוצע) מדד למרחק חייב להיות חיובי. לכן נתבונן על נתוני סדרה 1. נחשב תחילה את הסטיות מהממוצע  $\bar{x} = 70$

80	80	80	70	60	60	60	$X_i$
10	10	10	0	-10	-10	-10	$X_i - \bar{X}$

סכום הסטיות מהממוצע הוא 0, ולכן גם הממוצע שלהם 0. תופעה זו נכונה באופן כללי, ולכן לא נוכל להשתמש בממוצע הסטיות כמדד לפיזור.

$$\sum (x_i - \bar{x}) = 0$$

כדי לפתור את בעיית של המרחקים השליליים, נסתכל על ממוצע ריבועי הסטיות מהממוצע.

### שונות (Variance) וסטיית תקן (Standard deviation)

דרך אחרת לפתור את בעיית של המרחקים השליליים, נסתכל על ממוצע ריבועי הסטיות מהממוצע. מגדירים:

$$s_x^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} = \frac{\sum_{i=1}^n X_i^2}{n} - \bar{X}^2$$

נוסחה לשונות בסדרת תצפיות בודדות:  $\sum X_i^2$  (נרשם על גביה)

לרוב קל יותר לחשב את השונות באמצעות הנוסחה האחרונה.

דוגמה: לפניך 5 ערכים: 2, 4, 6, 8, 10

$$\bar{x} = \frac{2+4+6+8+10}{5} = 6$$

חשבו שונות וסטיית תקן בעזרת שתי הנוסחאות.

$$s_x^2 = \frac{\sum (x_i - \bar{x})^2}{n}$$

$$s_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

## דוגמא נוספת

**בסידרה 1** נחשב את השונות לפי הנוסחה הראשונה

$$S^2 = \frac{(-10)^2 + (-10)^2 + (-10)^2 + 0 + 10^2 + 10^2 + 10^2}{7} = 85.71$$

היחידות של השונות הם ריבוע היחידות של המקור. כדי לחזור ליחידות המקוריות, אנו נוציא שורש (חיובי) של השונות, מדד זה מכונה **סטיית תקן**.

$$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}} \quad \text{סטיית התקן בסדרת תצפיות בודדות} :$$

$$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2 \cdot f(x_i)}{n}} = \sqrt{\frac{\sum_{i=1}^n X_i^2 \cdot f(x_i)}{n} - \bar{X}^2} \quad \text{בטבלת שכיחות} :$$

הערה:

- שונות וסטיית התקן מהווים מדד למידת ההטרוגניות/הומוגניות של הקבוצה
- כל האיברים בקבוצה נלקחים בחשבון, וככל שהתצפיות קרובות יותר לממוצע סטית תקן קטנה יותר.

דוגמה

לפניך טבלת שכיחות של מספר מכוניות שיש למשפחות באזור מסוים.

$X$ - מספר מכוניות	$f(x)$ - משפחות
0	6
1	4
2	2
3	2
4	1

חשבו שונות וסטיית תקן !

## ציון תקן

### דוגמא 1:

תלמיד נבחן בשני מקצועות. באנגלית קבל ציון 80, ובמתמטיקה קבל ציון 80.

	אנגלית	מתמטיקה
ציון	80	80

איפה הצליח יותר? נראה כי הצליח במידה שווה בשתי הבחינות, ואולם כאשר נוסיף יחידת מידע כמו הממוצע, מתקבלת תמונה שונה:

	אנגלית	מתמטיקה
ציון	80	80
ממוצע	70	90

### דוגמא 2:

תלמיד נבחן בשני מקצועות. באנגלית קבל ציון 80, ובמתמטיקה קבל ציון 80.

	אנגלית	מתמטיקה
ציון	80	80
ממוצע	70	70

איפה הצליח יותר? כאשר נוסיף יחידת מידע – סטית התקן, מתקבלת תמונה שונה:

	אנגלית	מתמטיקה
ציון	80	80
ממוצע	70	70
סטית תקן	5	10

ציון תקן זה **מדד למיקום יחסי** עבור משתנה כמותי, כשהציונים המתקבלים הם יחסית לממוצע  $\bar{X}$  ולסטית התקן של ההתפלגות  $S_x$  ובכך הוא "מנטרל" את יחידות המדידה ומתקבל "ציון טהור". ציון זה מבטא את המרחק של תצפית מסוימת מממוצע ההתפלגות ביחידות של סטית תקן. כלומר, מספר סטיות התקן שיש בין תצפית מסוימת (ציון הגלם) לבין הממוצע. ככל שהפזור בהתפלגות קטן יותר, משמעות הפער בניהם הולכת וגדלה, ונקבל ציון תקן גדול יותר (בערכים מוחלטים). ציון התקן של כל תצפית בכל משתנה יצביע רק על המיקום היחסי של התצפית ביחס להתפלגות אליה היא שייכת.

$$z_x = \frac{X - \bar{X}}{S_x}$$

### תרגיל:

- סטודנט נבחן בסטטיסטיקה וקיבל 85 וגם נבחן בשפת c וקיבל 80 .  
ידוע שהציון הממוצע בסטטיסטיקה הוא 82 וסטיית התקן היא 5 .  
וגם ידוע שהציון הממוצע בשפת c הוא 70 וסטיית התקן היא 3 .  
**באיזה משני המקצועות יצטיין יותר יחסית לבני כיתתו ?**

### תרגיל :

בטבלה הבאה נתונים ציוניו של יורם במבחני שליש ונתונים לגבי הממוצע הכיתתי וסטיית התקן של הציונים בכיתה :

	מתמטיקה	תנ"ך	ספרות	ביולוגיה
ציון	60	80	45	58
ממוצע כיתתי	55	85	45	56
סטיית תקן	15	10	10	8

- א. מהו המקצוע החזק של יורם ומהו המקצוע החלש שלו יחסית לבני כיתתו ? **(חזק במתמטיקה וחלש בתנ"ך)**
- ב. ציון התקן של יוסי במתמטיקה שווה לציון התקן של יורם בתנ"ך . מהו ציונו הגולמי של יוסי במתמטיקה ? **(47.5)**

## שאלות חזרה

### שאלה 1

להלן נתונים על מספר תוכניות החיסכון בהן חוסכים אנשים שנבחרו באקראי מתוך אוכלוסיה מסוימת שאת חסכוניותיהם מפקידים בבנק הפועלים.

מספר אנשים- $f(x)$	מספר תוכניות חיסכון- $X$
14	0
40	1
50	2
30	3
10	4
6	5

- א. תאר את הנתונים בעזרת דיאגרמת מקלות !
- ב. מצא את השכיח, החציון, הממוצע, אמצע טווח, שונות וסטיית תקן.
- ג. מהו אחוז האנשים שמספר תוכניות החיסכון שלהם הוא בין 1 ל-4 (כולל).
- ד. למדגם הנ"ל נוספו עוד 10 לקוחות מבנק אחר (סה"כ = 160) .  
חמישה אנשים מתוכם פתחו 5 תוכניות חסכון וחמשת האחרים לא היו מעוניינים באף תוכנית חסכון. כיצד זה ישפיע, אם בכלל, על הממוצע, החציון, השכיח וסטיית התקן ? (תן הסבר לגבי כיוון השינוי –אין צורך בחישוב נוסף).

### תשובה לשאלה 1

- א. קל מאוד.
- ב. ממוצע=חציון=שכיח=2 אמצע טווח=2.5 שונות=1.46 סטיית תקן=1.21
- ג. 86%
- ד. הממוצע יגדל, חציון ושכיח לא ישתנו וסטיית התקן תגדל ל-1.33.

## שאלה 2

מתוך קהל הצופים במשחק הגמר במסגרת גביע אירופה לכדורסל, נבחרו **200 אנשים** ונשאלו בכמה מן המשחקים, מבין החמישה שכבר התקיימו, נכחו. להלן הנתונים שנאספו:

מספר המשחקים	מספר האנשים
0	13
1	43
2	45
3	50
4	45
5	4

- א. תאר את הנתונים באופן גרפי.
- ב. מצא את מספר המשחקים השכיח, החציוני והממוצע, ואת סטיית התקן של מספר המשחקים.

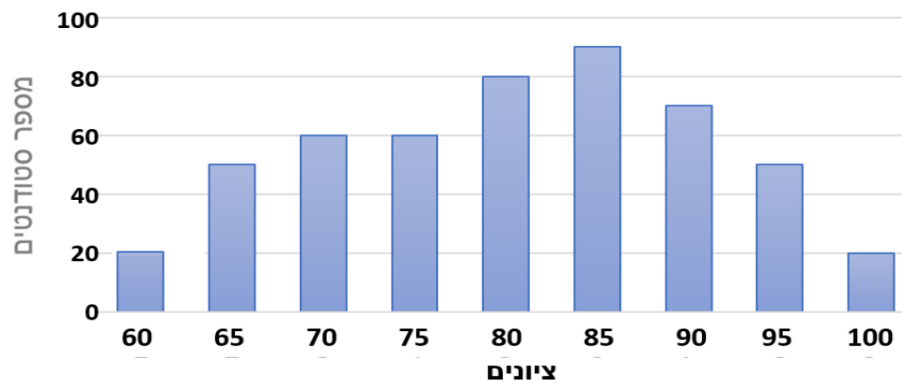
## תשובה לשאלה 2

א. דיאגרמת מקלות

ב. שכיח=3 ממוצע=2.415 סטיית תקן=1.2778 חציון=2

### שאלה 3

הגרף לפניכם מתאר את התפלגות הציונים שקיבלו סטודנטים בקורס סטטיסטיקה א' בשנת הלימודים הקודמת במכללת אשקלון .



א. חשבו את הציון השכיח, הממוצע והחציון

ב. חשבו את סטיית התקן של הציונים

### תשובות סופיות לשאלה 3

א. שכיח=85 , חציון=80 , ממוצע=80.5

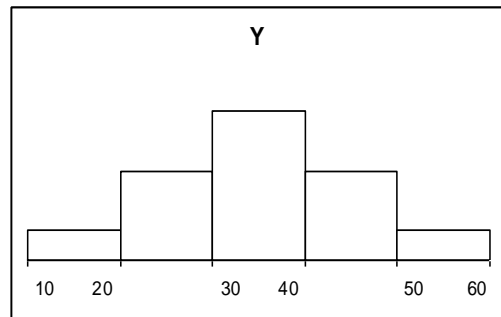
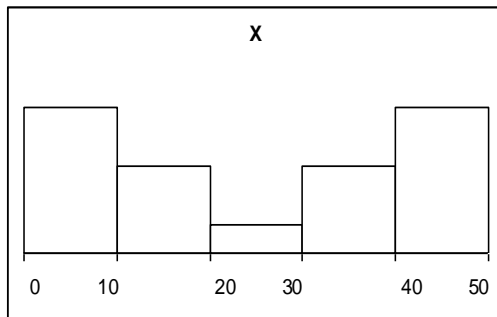
ב. סטיית תקן 10.5



## שאלות אמריקאיות

### שאלה 1

להלן 2 היסטוגרמות :



#### איזה מבין המשפטים הבאים נכון?

1. הממוצע וסטיית התקן של  $Y$  גדולים מהממוצע וסטיית התקן של  $X$ .
2. הממוצע וסטיית התקן של  $X$  גדולים מהממוצע וסטיית התקן של  $Y$ .
3. הממוצע של  $X$  גדול מהממוצע של  $Y$  וסטיית התקן של  $X$  קטנה מסטיית התקן של  $Y$ .
4. הממוצע של  $X$  קטן מהממוצע של  $Y$  וסטיית התקן של  $X$  גדולה מסטיית התקן של  $Y$ .

### שאלה 2

בבחינה במועד ב' השתתפו 10 תלמידים . ממוצע ציוני הבחינה שלהם היה 72. בטעות אבדה מחברת הבחינה של אחד התלמידים . אם הציונים הידועים הם: 97, 48, 71, 79, 95, 45, 57, 75, 83: מה ציונו של התלמיד שמחברת הבחינה שלו אבדה?

- א. 72.2
- ב. 70.0
- ג. 69.0
- ד. 72.0

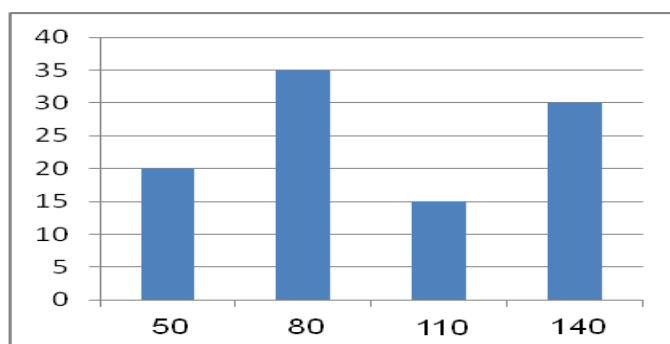
### שאלה 3

בסדרה המונה 6 תצפיות, נרשמו לגבי 5 מהן הסטיות מממוצע הסדרה :  
-22, -16, -8, 14, 10, שונות הסדרה היא :

- א. 240
- ב. 183.33
- ג. 200
- ד. 220
- ה. לא ניתן לחשב את שונות הסדרה ללא ידיעת התצפיות.

#### שאלה 4

להלן דיאגרמה המתארת את התפלגות מספר שעות חופשה (ציר X) שנוצלו ע"י עובדים במהלך החודשים אוגוסט וספטמבר בחברה מסוימת .  
(ציר Y הוא ציר השכיחות)



**מהי סטיית התקן בקירוב ומהו החציון ?**

א. סטיית תקן שווה ל-33.5 והחציון שווה ל-80

ב. סטיית תקן שווה ל-1122.75 והחציון שווה ל-80

ג. סטיית תקן שווה ל-33.5 והחציון שווה ל-95

ד. סטיית תקן שווה ל-96.5 והחציון שווה ל-50

#### שאלה 5

- ממוצע משכורות של עובדים בארגון מסוים הוא 6500 ₪ עם סטית תקן 800 ₪ .  
התקבלו לעבודה 3 עובדים חדשים עם משכורות 6500 ₪ , 4500 ₪ ו- 7000 ₪ .  
כיצד ישפיע השינוי על הממוצע :
- א. ממוצע המשכורות ישאר ללא שינוי.
  - ב. ממוצע המשכורות יגדל.
  - ג. ממוצע המשכורות יקטן.
  - ד. לא ניתן לדעת אם יהיה שינוי בממוצע.

## שאלה 6

מועמדים למשרה נבחנו בבחינה של יכולת זיכרון. להלן מספר התשובות הנכונות של כל נבחן :

4, 7, 7, 9, 10, 11, 13, 15, 15, 15, 17, 17, 19, 20, 21. על פי הערכים חושבו מדדי המרכז ומדדי פיזור.

לאחר החישוב התגלה כי נעשתה טעות ונבחן אחד שנרשמו לו 21 תשובות נכונות ענה נכון על 17 שאלות.

לאחר תיקון הטעות המדדים שלא ישתנו הם :

- א. החציון
- ב. הממוצע והשונות
- ג. הממוצע והשכיח
- ד. השכיח והשונות