

מבוא לסטטיסטיקה א



$$\{\sqrt{x}\}^2$$



תוכן העניינים

1. סטטיסטיקה תיאורית- סיווג משתנים וסולמות מדידה 1
2. סטטיסטיקה תיאורית- הצגה של נתונים 5
3. סטטיסטיקה תיאורית- סכימה 16
4. סטטיסטיקה תיאורית - מדדי מיקום מרכזי 20
5. סטטיסטיקה תיאורית - מדדי פיזור - הטווח, השונות וסטיית התקן 29
6. סטטיסטיקה תיאורית - מדדי פיזור - טווח בין רבעוני 32
7. סטטיסטיקה תיאורית-ממוצע משוקלל ושונות מצורפת 34
8. סטטיסטיקה תיאורית- מדדי מיקום יחסי-ציון תקן 36
9. סטטיסטיקה תיאורית-מדדי מיקום יחסי-אחוזונים במחלקות 38
10. סטטיסטיקה תיאורית - טרנספורמציה לינארית 41
11. סטטיסטיקה תיאורית- ניתוח פלטים 44
12. סטטיסטיקה תיאורית - שאלות מסכמות 46
13. סטטיסטיקה תיאורית שאלות אמריקאיות 53
14. מדדי קשר - מדד הקשר של קרמר 60
15. התפלגויות רציפות מיוחדות - התפלגות נורמלית (ללא ספר)

מבוא לסטטיסטיקה א

פרק 1 - סטטיסטיקה תיאורית- סיווג משתנים וסולמות מדידה

תוכן העניינים

1. כללי..... 1

סטטיסטיקה תיאורית – סיווג משתנים וסולמות מדידה:

רקע:

סטטיסטיקה תיאורית הוא ענף בו לומדים כיצד לאסוף נתונים, להציג אותם ולנתח אותם. בסטטיסטיקה תיאורית אנו פונים לקבוצה מסוימת, ובאותה קבוצה אנו אוספים נתונים על הישויות באותה קבוצה. משתנה – תכונה שיכולה לקבל מספר ערכים: דעה פוליטית, מקום מגורים, גובה של אדם וכדומה. חלוקה אחת של המשתנים הנמדדים היא לפי סולמות מדידה:

מיון משתנים לפי סולמות המדידה:

1. סולם שמי (נומינאלי) – משתנה שלערכיו יש משמעות רק מבחינת הזהות ואין עניין של יותר או פחות. לדוגמה: מצב משפחתי (רווק/נשוי/אלמן/גרוש), אזור מגורים. משתנה דיכוטומי (הינו מסולם שמי) אותם משתנים שיש להם רק שני ערכים אפשריות זכר/נקבה. מעשן/לא מעשן.
2. סולם סדר (אורדינאלי) – כאשר לערכים של המשתנה בנוסף לשם ישנה גם משמעות לסדר אבל אין משמעות לגודל ההפרש. למשל, דרגה בצבא.
3. סולם רווחים (אינטרוואלי) – משתנה שלערכים שלו בנוסף לשם ולסדר בניהם יש משמעות לרווחים בין הערכים אבל אין משמעות ליחס בין הערכים. למשל, קומה בבניין. סולם לא כל כך פופולרי.

סולם מנה/יחס:

משתנה שלערכיו בנוסף לשם, לסדר ולרווח יש משמעות גם ליחס בין הערכים. למשל, מספר מכוניות למשפחה, משקל אדם בק"ג. הדרך הקלה ביותר כדי לזהות עם הסולם הוא סולם מנה היא על ידי מבחן האפס. בסולם מנה האפס הוא מוחלט, אבסולוטי, ומייצג אין.

סוגי משתנים:

נבצע סיווג של המשתנים:

משתנה איכותי

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

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).

משתנה רציף: משתנה שערכיו מתקבלים מתוך אינסוף ערכים בתחום מסוים, הערכים מתקבלים ברצף – ללא קפיצות של ערכים. דוגמאות: גובה בס"מ – אם הגובה הנמוך ביותר הוא 150 ס"מ ועד 190 ס"מ – הגבהים בקבוצה הם ברצף. גם בין 160 ל-161 ס"מ יש רצף אינסופי של ערכים אפשריים (כמו 160.233 ס"מ, למשל).



שאלות:

- (1) באיזה סולם מדידה המשתנים הבאים נחקרים (שמי/סדר/רווחים/מנה):
- גובה (בס"מ).
 - מספר ילדים למשפחה.
 - מידת החרדה לפני מבחן.
 - שביעות רצון משירות לקוחות בסקלה מ-1 עד 7 (1 - כלל לא מרוצה עד 7 - מרוצה מאד)
 - השכלה.
 - מספר אוטובוס.
 - מקום מגורים.
 - מין (1=גבר ; 2=אישה).
 - מידת נעליים.

- (2) להלן התפלגות מספר האיחורים לעבודה בחודש של העובדים בחברת "סטאר":

מספר האיחורים	מספר העובדים
0	17
1	23
2	85
3	50
4	25

בחברה 200 עובדים.

- מהו המשתנה הנחקר כאן?
 - האם מדובר במשתנה איכותי או כמותי?
אם הוא כמותי האם הוא בדיד או רציף?
באיזה סולם מדידה המשתנה?
- (3) להלן רשימה של משתנים כמותיים. ציינו האם הוא משתנה רציף/בדיד:
- שכר ב-ש.
 - ציון בחינת בגרות.
 - תוצאה של הטלת קובייה.
 - מהירות ריצה בתחרות.
 - שיעור התמיכה בממשלה.

תשובות סופיות:

- (1) א. מנה. ב. מנה. ג. סדר.
ד. סדר. ה. מנה/ סדר. ו. שמי.
ז. שמי. ח. שמי. ט. סדר.
- (2) א. מספר האיחורים. ב. כמותי בדיד בסולם מנה.
- (3) א. רציף. ב. בדיד. ג. בדיד.
ד. רציף. ה. רציף.

מבוא לסטטיסטיקה א

פרק 2 - סטטיסטיקה תיאורית- הצגה של נתונים

תוכן העניינים

1. כללי 5

סטטיסטיקה תיאורית – הצגה של נתונים:

רקע:

דרכים להצגת נתונים שנאספו:

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: 4, 3, 5, 4, 3.

טבלת שכיחויות בדידה:

שם המשתנה- X	שכיחות – $f(x)$	שכיחות יחסית באחוזים
X_1	f_1	$\frac{f_1}{N} \cdot 100$
X_2	f_2	$\frac{f_2}{N} \cdot 100$
X_3	f_3	$\frac{f_3}{N} \cdot 100$
\vdots	\vdots	\vdots
X_k	f_k	$\frac{f_x}{N} \cdot 100$
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.

דוגמה:

להלן התפלגות הציונים בכיתה מסוימת:

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

שכיחות מצטברת – צבירה של השכיחויות.

השכיחויות F_i – השכיחות המצטברת נותנת כמה תצפיות קטנות או שוות לערך.

שכיחות יחסית (פרופורציה) – השכיחות מחולקת לכמות התצפיות הכללי:

$$\frac{f_i}{n} - \text{איזה חלק מהתצפיות בקבוצה שוות לערך.}$$

טבלת שכיחויות במחלקות:

משתמשים שהמשתנה כמותי רציף או כאשר יש מספר ערכים רב במשתנה הבדיד וטבלת שכיחויות תהיה ארוכה מידי.

דוגמה:

נתנו לקבוצת ילדים לבצע משימה, בדקו את התפלגות זמן הביצוע, 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

דיאגרמת עוגה:

זהו התיאור הגרפי של משתנה איכותי. בדיאגרמת עוגה כל ערך במשתנה מקבל "נתח", שהוא פרופורציונלי לשכיחות היחסית של ערך המשתנה בנתונים.

התפלגות המצב המשפחתי



דיאגרמת מקלות:

הציר האופקי הוא הציר של המשתנה והציר האנכי של השכיחות, כך שהגובה של המקל מעיד על השכיחות. רלבנטי למשתנה כמותי בדיד. לא נהוג להשתמש בתיאור למשתנה איכותי וכמו כן לא למשתנה כמותי רציף, וכן בסולמות מדידה עבור משתנה מסולם סדר.



היסטוגרמה:

היסטוגרמה היא הדרך הגרפית כדי לתאר טבלת שכיחויות במחלקות, והיא רלוונטית למשתנה כמותי רציף. בהיסטוגרמה הציר האופקי הוא הציר של המשתנה והציר 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.

צפיפות	מצטברת	שכיחות	אמצע	רוחב	X
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. לדוגמה, ציוני IQ.

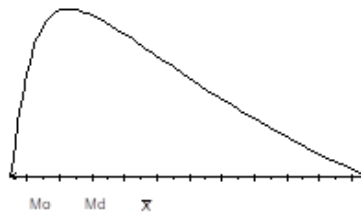


ישנן התפלגויות סימטריות שאינן פעמוניות, כגון:

התפלגות אסימטרית ימנית (חיובית)

רוב התצפיות מקבלות ערכים נמוכים ויש מיעוט הולך וקטן של תצפיות שמקבלות ערכים גבוהים קיצוניים. לדוגמה, שכר במשק.

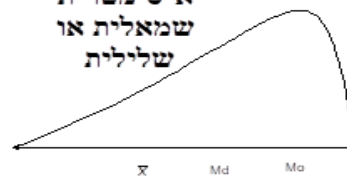
התפלגות א-סימטרית ימנית או חיובית



התפלגות אסימטרית שמאלית (שלילית)

רוב התצפיות מקבלות ערכים גבוהים ויש מיעוט הולך וקטן של תצפיות שמקבלות ערכים נמוכים קיצוניים. לדוגמה, אורך חיים.

התפלגות א-סימטרית שמאלית או שלילית



שאלות:

- 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":

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

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

- 3) להלן התפלגות ההשכלה במקום עבודה מסוים:

השכלה	מספר העובדים
נמוכה	60
תיכונית	120
אקדמאית	20

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

- 4) להלן רשימת הציונים של 20 תלמידים שנבחנו במבחן הבנת הנקרא:
- 6, 5, 8, 7, 6, 7, 6, 8, 7, 8, 5, 4, 6, 10, 9, 8, 6, 7.
- א. מהו המשתנה? האם הוא בדיד או רציף?
- ב. תארו את הרשימה בטבלת שכיחויות.
- ג. הוסיפו שכיחויות יחסיות לטבלה.
- ד. תארו את הנתונים באופן גרפי.

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



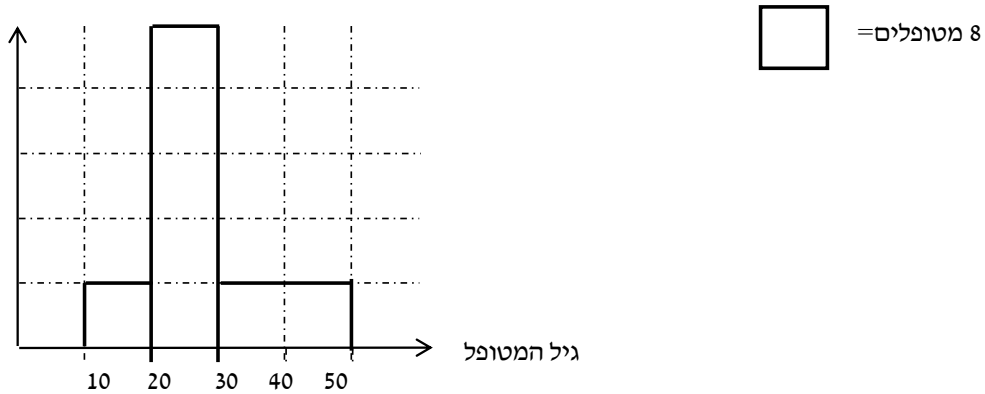
- מהו המשתנה הנחקר? האם הוא בדיד או רציף?
- תארו את הנתונים בטבלת שכיחויות במחלקות.
- הוסיפו שכיחות יחסית לטבלה.
- הוסיפו את הצפיפות של כל מחלקה לטבלה.
- מהי צורת ההתפלגות של הגבהים?

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

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

7) להלן גיל המטופלים של ד"ר שוורץ בשנים :
 קנה מידה :



- מה המשתנה הנחקר? האם הוא בדיד או רציף?
- מהי הקבוצה הנחקרת?
- תרגמו את ההסיטוגרמה לטבלת שכיחות.
- מהי הפרופורציה של המטופלים של ד"ר שוורץ בגילאים 20-30?

תשובות סופיות:

ב. עיין גרף מלא בסרטון הוידאו.

1) א. להלן טבלה:

%	$\frac{f(x)}{n}$	$f(x)$	x
12.5%	$\frac{25}{200}$	25	ערוץ 1
12.5%	$\frac{25}{200}$	25	ערוץ 10
37.5%	$\frac{75}{200}$	75	ערוץ 2
25%	$\frac{50}{200}$	50	כבלים
12.5%	$\frac{25}{200}$	25	לא צפו
100%	1	200	סה"כ

ב. 19.6%.

2) א. מקצוע מועדף.

ב. עיין גרף מלא בסרטון הוידאו.

3) א. משתנה נחקר: השכלה, סוג: סדר.

4) א. המשתנה : ציון, משתנה בדיד.
ד. עיין גרף מלא בסרטון הוידאו.

ב+ג. להלן טבלה :

%	$\frac{f(x)}{n}$	$f(x)$	x
5%	$\frac{1}{20}$	1	4
10%	$\frac{2}{20}$	2	5
30%	$\frac{6}{20}$	6	6
20%	$\frac{4}{20}$	4	7
20%	$\frac{4}{20}$	4	8
10%	$\frac{2}{20}$	2	9
5%	$\frac{1}{20}$	1	10
100%	20	20	סה"כ

5) א. גובה בס"מ, רציף.

ב+ג+ד. להלן טבלה : ה. אסימטרית.

d	%	$\frac{f(x)}{n}$	$f(x)$	x
1	5%	$\frac{5}{100}$	5	155-160
2	10%	$\frac{10}{100}$	10	160-165
3	15%	$\frac{15}{100}$	15	165-170
4	40%	$\frac{40}{100}$	40	170-180
3	30%	$\frac{30}{100}$	30	180-190

- 6) א. עיין גרף מלא בסרטון הוידאו.
 ב. סימטרית.
- 7) א. המשתנה: גיל בשנים, משתנה רציף.
 ב. המטופלים של ד"ר שוורץ.
 ד. להלן טבלה:
 ה. 62.5%.

$f(x)$	x
8	10-20
40	20-30
16	30-50

מבוא לסטטיסטיקה א

פרק 3 - סטטיסטיקה תיאורית- סכימה

תוכן העניינים

1. כללי.....16

סטטיסטיקה תיאורית – סכימה:

רקע:

בסטטיסטיקה ישנה צורת רישום מקובלת לסכום 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

(הסבר מלא מופיע בסרטונים באתר).

שאלות:

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

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

א. $\sum_{i=1}^3 X_i$

ב. $\sum_{i=1}^5 Y_i$

ג. $\sum_{i=1}^4 X_i$

ד. $\left(\sum_{i=1}^4 X_i\right)^2$

ה. $\sum X_i$

ו. $\sum X_i Y_i$

ז. $\sum(X_i) \sum(Y_i)$

- (2) נתון לוח ערכי המשתנים X_i ו- Y_i , כאשר: $i = 1, 2, \dots, 6$, ונתונים הקבועים:
 $a = 2$, $b = 5$. חשבו את הנוסחאות הבאות:

i	1	2	3	4	5	6
X_i	3	2	4	-2	1	4
Y_i	2	0	0	1	-5	2

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) קבעו לכל זהות האם 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. $\left(\sum_{i=1}^n X_i \right)^2 = \sum_{i=1}^n X_i^2$

(4) נתון: $\sum_{i=1}^{10} X_i = 80$, $\sum_{i=1}^{10} X_i^2 = 1640$

חשבו: $\sum_{i=1}^{10} (X_i - 4)^2$

תשובות סופיות:

- | | | | |
|--------------|---------|-----------|--------------|
| ד. 121. | ג. 11. | ב. 9. | א. 7. (1 |
| | ז. 126. | ו. 27. | ה. 14. |
| | ג. 7. | ב. 12. | א. 3. (2 |
| | | ה. 14. | ד. 12. |
| ג. לא נכונה. | | ב. נכונה. | א. נכונה. (3 |
| | | | .1160 (4 |

מבוא לסטטיסטיקה א

פרק 4 - סטטיסטיקה תיאורית - מדדי מיקום מרכזי

תוכן העניינים

1. כללי 20

סטטיסטיקה תיאורית – מדדי מיקום מרכזי:

רקע:

המטרה במדדי המיקום המרכזי היא 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.

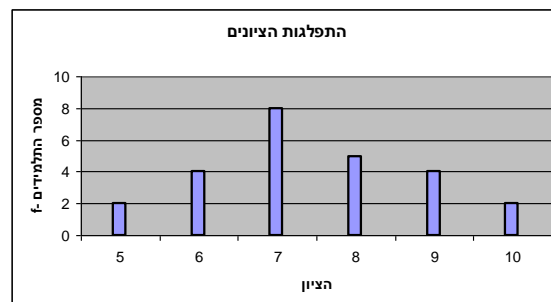
בטבלת שכיחויות בדידה

הערך שהשכיחות שלו היא הגבוהה ביותר.

$f(x)$	# תוכניות החיסכון
100	0
75	1
25	2
25	3
25	4

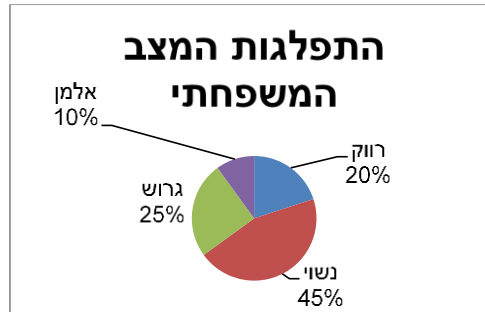
בדיאגרמת מקלות

שיעור ה- X של המקל הגבוה ביותר.



בעוגה

הערך של הפלח הגדול ביותר.



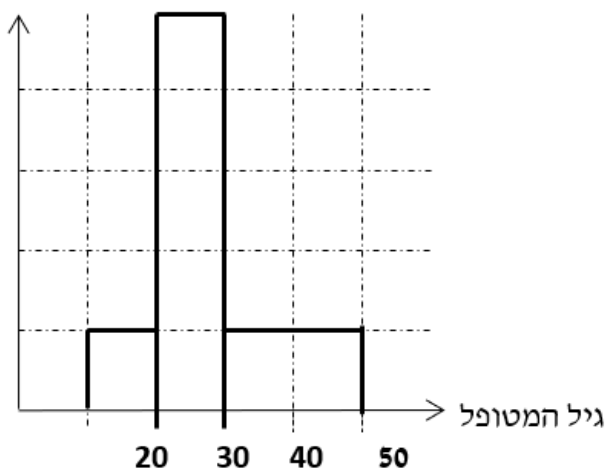
בטבלת שכיחויות במחלקות

אמצע המחלקה עם הצפיפות הגבוהה ביותר.
לדוגמה, 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.
לדוגמה, גיל המטופלים של ד"ר שוורץ בשנים :



= 8

כללי

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

אמצע תחום (טווח) – Midrange:

הממוצע בין התצפית הגבוהה ביותר לתצפית הנמוכה ביותר:

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

החציון – Median:

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 הוא ממוצע של האיבר ה- $\frac{n}{2}$,

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

וכשיש 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: נמצא 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 – Average :

הממוצע הוא 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.

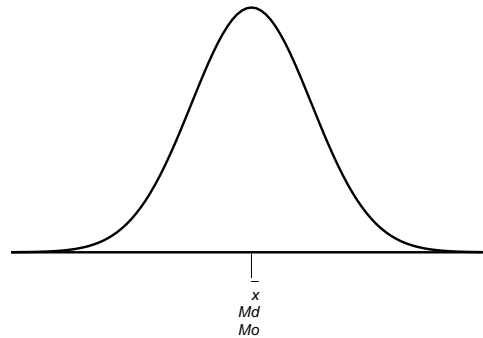
כללי

הממוצע רלבנטי רק למשתנה כמותי.

מדדי המיקום המרכזי בהתפלגויות המיוחדות:

בהתפלגות סימטרית פעמונית כל מדדי המרכז שווים זה לזה:

התפלגות סימטרית

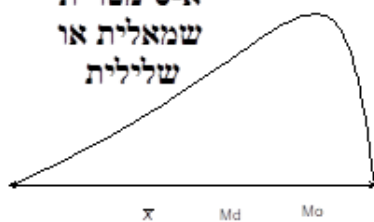


בהתפלגות סימטרית השכיח לא חייב להיות במרכז:

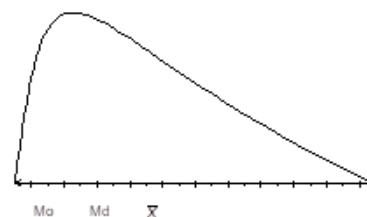
התפלגות U



התפלגות
א-סימטרית
שמאלית או
שלילית



התפלגות א-סימטרית
ימנית או חיובית



שאלות:

- (1) להלן רשימת הציונים של 20 תלמידים שנבחנו במבחן הבנת הנקרא:
 6, 5, 8, 7, 6, 6, 7, 8, 7, 6, 5, 4, 6, 10, 9, 8, 6, 7.
 חשבו 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 הטלוויזיה שנספרו עבור כל משפחה בישוב מסוים:

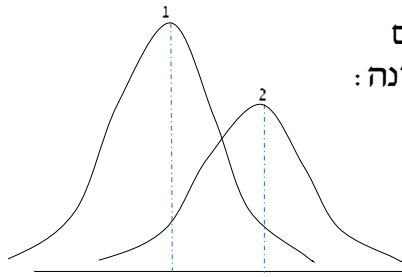
מספר משפחות	מספר מקלטים
22	0
28	1
18	2
22	3
10	4

- א. חשבו את הממוצע, החציון והשכיח של ההתפלגות.
 ב. הסבירו ללא חישוב כיצד כל מדד שחישבת בסעיף א' היה משתנה אם חלק מהמשפחות (לא כולן) שלא היה להם עד היום טלוויזיה היו רוכשים מקלט אחד.

- (4) להלן התפלגות מספר המכוניות למשפחה בישוב "הגורן":

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

- א. כמה משפחות יש בישוב?
 ב. מה אחוז המשפחות בישוב 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. בחרו בתשובה הנכונה:

- בכיתה 1 השכיח גבוה יותר מכיתה 2.
- בכיתה 2 השכיח גבוה יותר מכיתה 1.
- בשתי הכיתות אותו שכיח.
- 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	

- השלימו את הטבלה.
- מהו השכיח, אמצע טווח והחציון.
- חלק מהמשפחות להן הייתה טלוויזיה אחת בדיוק הוציאו את הטלוויזיה מביתם. כיצד כל מדד ישתנה (יגדל, יקטן או לא ישתנה). הסבירו ללא חישוב.

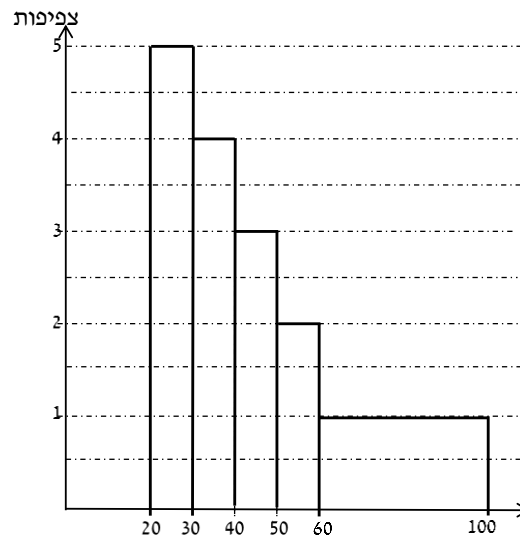
7) להלן התפלגות המשקל של קבוצה מסוימת בק"ג. מה הממוצע והחציון של ההתפלגות?

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

8) להלן התפלגות הגבהים בס"מ בקבוצה מסוימת. חשבו את הממוצע, החציון והשכיח של הגבהים בקבוצה זו.

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

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



- מצאו את השכיח בהתפלגות.
- מצאו את החציון בהתפלגות.
- הסבירו ללא חישוב האם הממוצע גדול/קטן/שווה לחציון.
- הסתבר שיש להוציא מספר תלמידים במחלקה בין 20-30 שקלים. כיצד הדבר ישפיע על הממוצע, החציון והשכיח? הסבירו ללא חישוב.

תשובות סופיות:

- (1) חציון: 7, שכיח: 6, ממוצע: 6.9.
- (2) א. 3. ב. שכיח: 3.4, חציון: 4.
- (3) א. ממוצע: 1.7, חציון: 1.5, שכיח: 1.
 ב. הממוצע יגדל ויתר המדדים לא ישתנו.
- (4) א. 630. ב. 34.13%. ג. שכיח וחציון: 3, ממוצע: 2.952.
- (5) ב'.
- (6) א. להלן טבלה: ב. חציון: 2, שכיח: 2, אמצע טווח: 1.5.

מספר משפחות	מספר טלויזיות
28	0
62	1
92	2
18	3

ג. שכיח: לא ישתנה, אמצע הטווח: לא ישתנה, חציון: לא ישתנה, ממוצע: יקטן.

- (7) חציון וממוצע: 55.
- (8) ממוצע: 172.6, חציון: 174.17, שכיח: 177.5.
- (9) א. 25. ב. 40. ג. גדול מהחציון.
 ד. שכיח: לא ישתנה, חציון: יגדל, ממוצע: יגדל.

מבוא לסטטיסטיקה א

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

תוכן העניינים

1. כללי 29

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 / תחום (RANGE):

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

שונות וסטיית תקן:

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

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

דוגמאות:

(1) נחשב את השונות של סדרת המספרים הבאה: 5, 4, 9.

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

(2) 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.

שאלות:

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

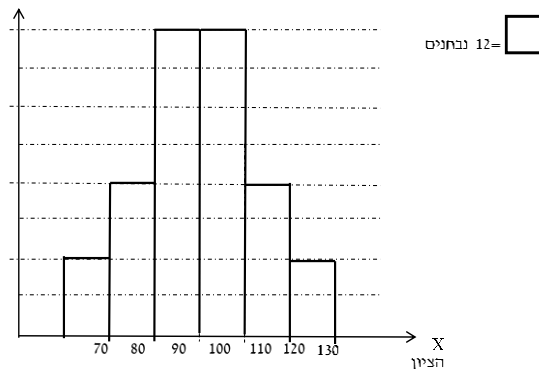
ב. חשבו את סטיית התקן של מספר החדרים לדירה.

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

6) להלן התפלגות המשקל של קבוצה מסוימת בק"ג: מהי סטיית התקן של התפלגות המשקל?

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

7) להלן התפלגות הציונים במבחן אינטליגנציה:



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

תשובות סופיות:

- 1) שונות: 2.19, סטיית תקן: 1.48, טווח: 6.
 2) א. סטיית תקן: 1.106. ב. טווח: 4.
 3) א. ממוצע לא ישתנה, סטיית התקן תקטן.
 ב. ממוצע לא ישתנה, סטיית התקן תגדל.
 4) 10.8
 5) א. 3.05. ב. 1.16. ג. ממוצע: יקטן, סטיית התקן: תגדל.
 6) 7.73
 7) א. 100. ב. 12.96. ג. ממוצע: לא ישתנה, סטיית תקן: תקטן.

מבוא לסטטיסטיקה א

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

תוכן העניינים

1. כללי 32

סטטיסטיקה תיאורית - מדדי פיזור - טווח בין רבעוני:

רקע:

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

שלבם במציאת טווח בין-רבעוני במחלקות:

F	f מספר עובדים (שכחות)	רוחב $L_1 - L_0$	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:

נמצא את הרבעון התחתון (אחוזון 25) והרבעון העליון (האחוזון ה-75).

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

נוסחאות הרבעונים יהיו:

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

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

נציב:

$$Q_1 = 0.5 + \frac{\frac{200}{4} - 0}{56} \cdot 4 = 4.07 \text{ שניות}$$

$$Q_3 = 9.5 + \frac{\frac{3 \cdot 200}{4} - 106}{48} \cdot 2 = 11.33 \text{ שניות}$$

שלב ב:

$$IQR = Q_3 - Q_1 = 11.33 - 4.07 = 7.26 \text{ שניות}$$

נחסר את הרבעונים:

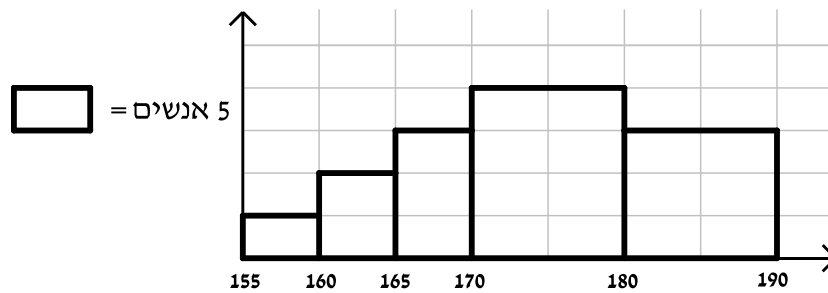
שאלות:

(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

מצאו את הטווח הבין-רבעוני.

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



מצאו את הטווח הבין-רבעוני.

תשובות סופיות:

(1) 13.75 ק"ג.

(2) 13.33 ק"ג.

מבוא לסטטיסטיקה א

פרק 7 - סטטיסטיקה תיאורית-ממוצע משוקלל ושונות מצורפת

תוכן העניינים

1. כללי 34

סטטיסטיקה תיאורית – ממוצע משוקלל ושונות מצורפת:

רקע:

מדובר על מצב שבו ישנן כמה קבוצות שנרצה לאחד לקבוצה אחת. מתעניינים בממוצע והשונות של הקבוצה הגדולה המתקבלת מאיחוד הקבוצות הקטנות.

n_j - מס' התצפיות בקבוצת ה- j .

j - אינדקס של הקבוצה.

N - מס' התצפיות בכל הקבוצות יחד (סכום כל ה- n_j).

\bar{x}_j - הממוצע בקבוצה ה- j .

S_j^2 - השונות בקבוצה ה- j .

הנוסחאות לממוצע משוקלל ושונות מצורפת:

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

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?

שאלות:

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. חשבו את הממוצע המשוקלל בשכבה.

ב. חשבו 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. מצאו את הממוצע של התצפיות לאחר שאוחדו שתי הקבוצות לקבוצה אחת.

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

תשובות סופיות:

1) א. 76.22 ב. 173.5

2) א. 70 ב. 14.14

מבוא לסטטיסטיקה א

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

תוכן העניינים

1. כללי 36

סטטיסטיקה תיאורית – מדדי מיקום יחסי – ציון תקן:

רקע:

המטרה למדוד איך תצפית ממוקמת ביחס לשאר התצפיות בהתפלגות.

ציון תקן:

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

- 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 הוא 8 אלף ₪, עם סטית תקן של אלפיים ₪. באותו מקום עבודה ההשכלה הממוצעת של העובדים הנה 14 שנים, עם סטית תקן של 1.5 שנים. ערן מרוויח במקום עבודה זה 11 אלף ₪ והשכלתו 16 שנים. מה ערן יותר, באופן יחסי, משכיל או משתכר?

שאלות:

- 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 ו-70 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 המידע שנאסף על שני המשתנים:

מספר פועלים	תפוקה	ממוצע
15	48	
2	10	סטיית תקן

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

תשובות סופיות:

- 1) א. לשון. ב. 72.
 2) ב'.
 3) א. משקל. ב. 70.

מבוא לסטטיסטיקה א

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

תוכן העניינים

1. כללי 38

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 שרבע מהתצפיות קטנות ממנו והשאר גבוהות ממנו. מסומן: $X_{0.25}$.

מציאת מאון במחלקות:

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

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

$F(x_{m-1})$ - 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 נרצה לחלץ את אחוז התצפיות שמתחת לערך מסוים נשתמש בנוסחה

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

הבאה:

דוגמה (פתרון בהקלטה):

להלן התפלגות השכר של עובדים בחברה מסוימת:

שכר ב-₪	
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 שמשתכרים מתחת ל-5,000 ₪?

שאלות:

(1) להלן התפלגות השכר (באלפי שקלים) בחברה:

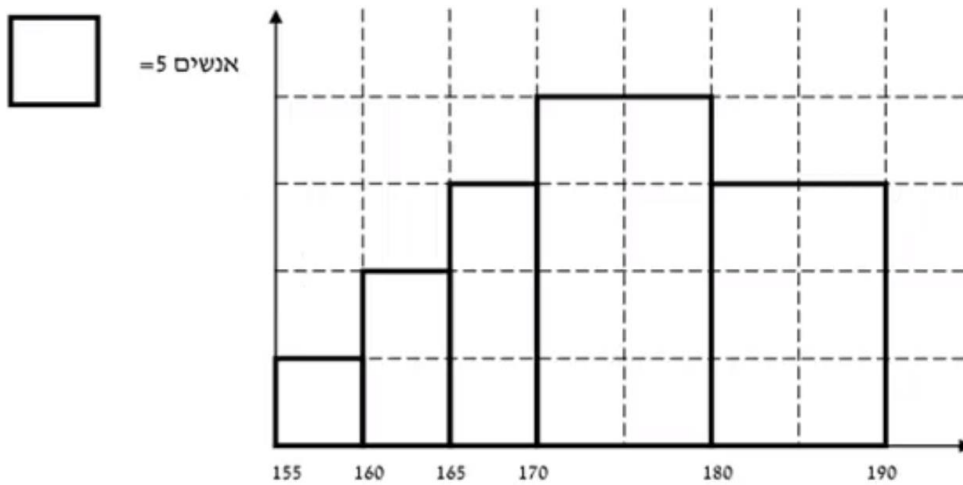
שכחות מצטברת	שכר - X
48	6-10
100	10-15
120	15-20
132	20-30
136	30-60

- 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 שמשכרם מתחת ל-7,000 ₪?
 ה. איזה אחוז מהעובדים משכרם מעל ל-25,000 ₪?
 ו. איזה אחוז מהעובדים משכרם בין 7,000 ₪ 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 השכיחויות 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:



חשבו:

- העשירון התחתון.
- האחוזון ה-30.
- הגובה ש-20% מהתצפית גדולות ממנו.
- את אחוז התצפיות מתחת לגובה 158 ס"מ.
- את אחוז התצפיות מעל לגובה 185 ס"מ.
- את אחוז התצפיות בין גובה 170 ס"מ ל-185 ס"מ.

תשובות סופיות:

- 1) א. 13.23 ב. 22 ג. 17.2 ד. 8.82% ה. 7.36%
- ו. 83.82%
- 2) להלן טבלה:

ציון- X	$f(x)$
50-60	40
60-70	60
70-80	200
80-90	60
90-100	40

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

מבוא לסטטיסטיקה א

פרק 10 - סטטיסטיקה תיאורית - טרנספורמציה לינארית

תוכן העניינים

1. כללי 41

סטטיסטיקה תיאורית – טרנספורמציה לינארית:

רקע:

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

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

$$MO_y = a \cdot MO + b$$

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

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

מדדי המרכז:

$$R_y = |a| R_x$$

$$S_y = |a| S_x$$

$$S_y^2 = a^2 S_x^2$$

מדדי הפיזור:

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

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

מדדי המיקום היחסי:

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 הינו 9000 ₪ וטווח 6000 ₪. חשבו את המדדים הללו לאחר שהעלו את כל המשכורות ב-10% ואחר כך קנסו אותם ב-100 ₪.

שאלות:

- (1) עבור סדרת נתונים התקבל: $\bar{x} = 80, S = 15, MO = 70$.
 הוחלט להכפיל את כל התצפיות ב-4 ולהחסיר מהתוצאה 5.
 חשבו את המדדים הללו לאחר השינוי.
- (2) בחברה מסוימת השכר הממוצע הוא 40 ש"ח לשעה עם סטיית תקן של 5 ש"ח לשעה.
 הוחלט להעלות את כל המשכורות ב-10%, אך זה לא סיפק את העובדים ולכן
 הם קיבלו לאחר מכן תוספת של 2 ש"ח לשעה.
 מה הממוצע ומהי השונות של השכר לשעה לאחר כל השינויים.
- (3) במבחן מסוים 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. התקבל שבממוצע יש 3 גפרורים
 פגומים בקופסא, עם סטיית תקן של 1.5 גפרורים.
 מה יהיה הממוצע ומה תהיה סטיית התקן של מספר התקינים בקופסא?
- (5) חברת בזק הציעה את ההצעה הבאה: שלושים שקלים דמי מנוי חודשיים
 קבועים וכן 10 אגורות לכל דקה של שיחה יוצאת. אדם בדק במשך שנה את
 דקות השיחות היוצאות שלו, וקיבל שבממוצע חודשי יש לו 600 דקות שיחות
 יוצאות עם שונות של 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) הוכיחו שאם כל התצפיות בהתפלגות עברו טרנספורמציה לינארית: $Y_i = a \cdot X_i + b$,
 אזי הממוצע והשונות של כלל התצפיות לאחר הטרנספורמציה יהיו בהתאמה:

$$\bar{y} = a \cdot \bar{x} + b, S_y^2 = a^2 S_x^2$$

תשובות סופיות:

- (1) ממוצע: 315, סטיית תקן: 60, שכיח: 275.
- (2) ממוצע: 46, שונות: 30.25.
- (3) טווח: 40, חציון: 77, עשירון עליון: 91.
- (4) ממוצע: 37, סטיית תקן: 1.5.
- (5) ממוצע: 90, שונות: 25, ציון תקן: 2.
- (6) $a^2 \cdot S_x^2$

מבוא לסטטיסטיקה א

פרק 11 - סטטיסטיקה תיאורית- ניתוח פלטים

תוכן העניינים

1. כללי 44

סטטיסטיקה תיאורית – ניתוח פלטים:

שאלות:

1) להלן פלט על התפלגות הגילאים באוכלוסייה מסוימת:

		Statistic
Age of Respondent	Mean	45.63
	Median	41.00
	Variance	317.140
	Std. Deviation	a
	Minimum	18
	Maximum	b
	Range	71
	Interquartile Range	28

- א. מצאו את הערכים בטבלה המסומנים ב- a ו- b .
 ב. נתון שההתפלגות היא אסימטרית, האם היא נוטה ימינה או שמאלה?

2) להלן התפלגות ההשכלה של העובדים בחברת "מתאר":

		Statistic
years of education	Mean	?
	Median	12.0000
	Variance	?
	Std. Deviation	2.54786
	Minimum	?
	Maximum	?
	Range	?
	Interquartile Range	?

מלאו את הערכים המסומנים בסימני שאלה.

תשובות סופיות:

- (1) א. $b = 89$, $a = 17.81$. ב. אסימטרית ימנית.
- (2) ממוצע: 11.909, שונות: 6.492, טווח: 10, טב"ר: 3.

מבוא לסטטיסטיקה א

פרק 12 - סטטיסטיקה תיאורית - שאלות מסכמות

תוכן העניינים

1. כללי 46

סטטיסטיקה תיאורית – שאלות מסכמות:

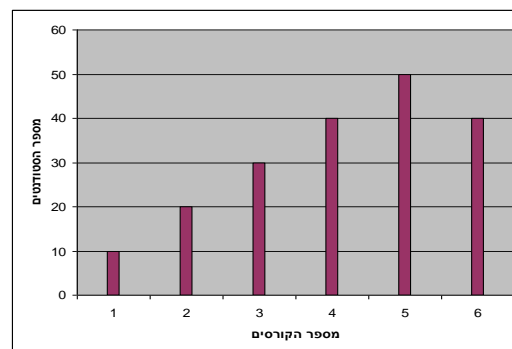
שאלות:

1) בדקו עבור 5 תלמידים את המשקל שלהם:

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

- מהו המשתנה הנחקר? בדיד או רציף?
- מה המשקל החציוני, הממוצע והשכיח?
- מה הטווח וסטיית התקן של המשקל?
- 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?
- הוסיפו עוד תלמיד השוקל 52 ק"ג בדיוק. הסבירו ללא חישוב כיצד הדבר ישפיע על הממוצע וסטיית התקן (יגדילו, יקטין או לא ישנה).

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



- מה המשתנה הנחקר? בדיד או רציף?
- מהי צורת ההתפלגות?
- תארו את הנתונים בטבלת שכיחויות.
- חשבו את השכיח, החציון והטווח.

3) להלן התפלגות הציונים בבחינה בלשון שנעשתה עבור תלמידי כיתות ד'.

במחקר השתתפו 150 תלמידים. ממוצע הציונים שהתקבל: $\bar{X} = 7\frac{1}{15}$.

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

- השלימו את השכיחויות החסרות בטבלה.
- חשבו את הציון החציוני, השכיח.
- 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 ו-5 - שביעות רצון גבוהה). להלן ההתפלגות שהתקבלה:

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

- מה אחוז האנשים עם רמת שביעות רצון נמוכה?
- מה המשתנה הנחקר ומאיזה סוג הוא?
- מהי הדרך הגרפית המתאימה ביותר לתיאור הנתונים?
 - היסטוגרמה.
 - דיאגרמת מקלות.
 - דיאגרמת עוגה.
- חשבו את המדדים הבאים:
 - טווח.
 - שכיח.
 - חציון.

5) להלן התפלגות מספר שעות העבודה לשבוע של העובדים (כ-200) בחברת "סטאר":

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

- א. השלימו את הטבלה.
- ב. חשבו את החציון, השכיח והממוצע של התפלגות מס' שעות העבודה בחברה.
- ג. מה סטיית התקן של מספר 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? הסבירו.

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 שמשלמים במספר הרב ביותר של מסרונים בחודש. החל מאיזה כמות של מסרונים תחולק המתנה?
- ד. ציינו איזה מדד ניתן לחשב ואיזה לא ניתן. אם ניתן, חשבו:
 - i. ממוצע.
 - ii. שכיח.
 - iii. חציון.
 - iv. שונות.

7) 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

- שרטטו היסטוגרמה לתיאור התפלגות זמן ביצוע המשימה.
- מתוך ההיסטוגרמה שבנית בסעיף א', מהי צורת ההתפלגות?
- חשבו את השכיח והחציון של ההתפלגות.
- הסבירו, ללא חישוב, האם הזמן הממוצע לביצוע המשימה, קטן או גדול או שווה ביחס לשכיח ולחציון.

8) התפלגות ציוני מבחן אינטליגנציה היא סימטרית. נתון שהעשירון העליון הוא 130, הרבעון התחתון הוא 90, ושלמבחן נגשו 500 מועמדים.

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

- השלימו את הטבלה.
- מהו הממוצע והחציון של ההתפלגות?
- מהו הציון ש-40% מהתלמידים קיבלו מעליו? באיזה אחוזון מדובר?
- הוחלט להעלות את כל הציונים ב-10 נקודות. כיצד הדבר ישפיע על הממוצע וסטיית התקן של הציונים?

- 9) להלן מספר טענות, עבור כל טענה ציינו 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.

תשובות סופיות:

- (1) א. המשתנה הנחקר: משקל תלמיד בק"ג, משתנה כמותי רציף.
 ב. $\bar{X} = 52$, $Md = X_{\frac{n+1}{2}} = X_3 = 58$, שכיח: 58.
 ג. $R = 28$, $S = 10.12$.
 ד. הוא חריג יותר בגובה כי שם ציון התקן בערך מוחלט יותר גבוה.
 ה. הממוצע לא ישתנה אך סטיית התקן תקטן.
- (2) א. מספר הקורסים, בדיד. ב. התפלגות אסימטרית שמאלית.
 ג. להלן טבלה: ד. שכיח: 5, טווח: 5.

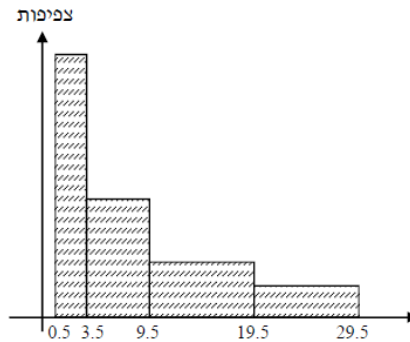
$f(x)$	x
10	1
20	2
30	3
40	4
50	5
40	6
190	סה"כ

- (3) א. 20 תלמידים קיבלו ציון 6 ו-40 תלמידים קיבלו ציון 8.
 ב. חציון: 7, שכיח: 8. ג. שונות: 2.533, סטיית תקן: 1.592.
- (4) א. 20%. ב. שביעות רצון (סדר).
 ג. ii. ד. טווח: 4, שכיח: 2, חציון: 2.
- (5) א. להלן טבלה: ב. חציון: 35, שכיח: 35, ממוצע: 35.

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

- ג. סטיית תקן: 12.65. ד. 53.333.
 ה. 25%. ו. -0.395.
 ז. חציון לא ישתנה, ממוצע יגדל, סטיית תקן תקטן.
- (6) א. 38%. ב. 40%. ג. 150. ד. חציון: 100.

7) א. שרטוט: ב. ההתפלגות היא א-סימטרית ימנית.



ג. שכיח: 2, חציון: 6.83.

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

8) א. ראו טבלה:

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

ב. 100. ג. 104.

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

9) א. נכון. ב. לא נכון. ג. לא נכון. ד. לא נכון. ה. לא נכון.
ו. נכון. ז. לא נכון. ח. לא נכון.

מבוא לסטטיסטיקה א

פרק 13 - סטטיסטיקה תיאורית שאלות אמריקאיות

תוכן העניינים

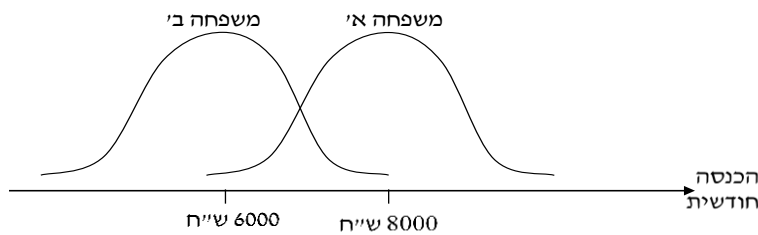
1. כללי 53

סטטיסטיקה תיאורית – שאלות אמריקאיות:

שאלות:

שאלות 1-3 מתייחסות לקטע הבא:

להלן שתי עקומות המתארות את התפלגות ההכנסות החודשיות של שתי משפחות שנבחרו באקראי:



- (1) לאיזו משפחה הכנסה שכיחה גבוהה יותר?
- משפחה א'.
 - משפחה ב'.
 - לשתיהן אותה הכנסה שכיחה.
 - לא ניתן לדעת – אין מספיק נתונים.
- (2) באיזו משפחה ההכנסה החציונית שווה להכנסה הממוצעת?
- משפחה א'.
 - משפחה ב'.
 - בשתיהן ההכנסה החציונית שווה להכנסה הממוצעת.
 - לא ניתן לדעת – אין מספיק נתונים.
- (3) באיזו משפחה סטית התקן של ההכנסה החודשית גבוהה יותר?
- משפחה א'.
 - משפחה ב'.
 - לשתיהן אותה סטית תקן.
 - לא ניתן לדעת – אין מספיק נתונים.

הנתונים הבאים מתייחסים לשאלות 4-6:

להלן נתונים חלקיים של טבלת שכיחויות:
כמו כן, נתון כי הממוצע הוא 1.66.

$F(x)$	x
?	0
10	1
6	2
15	3
?	4
50	סה"כ

(4) השכיח של הנתונים הוא:

- א. 0.
- ב. 15.
- ג. ישנם שני שכיחים: 0 ו-3.
- ד. על סמך הנתונים החלקיים אי אפשר לקבוע מה יהיה ערכו של השכיח.

(5) חציון הנתונים הוא:

- א. 2.
- ב. 1.5.
- ג. 25.5.
- ד. על סמך הנתונים החלקיים אי אפשר לקבוע מה יהיה ערכו של החציון.

(6) הטווח של הנתונים:

- א. 11.
- ב. 3.
- ג. 4.
- ד. על סמך הנתונים החלקיים אי אפשר לקבוע מה יהיה ערכו של החציון.

(7) בהתפלגות אסימטרית ימנית של משתנה כמותי רציף, הערך המתאים למאון

ה-30, ציון התקן שלו הוא בהכרח:

- א. שלילי.
- ב. חיובי.
- ג. אפס.
- ד. לא ניתן לדעת ללא הנתונים.

- 8) סדרת נתונים סטטיסטיים מונה 10 תצפיות. נתון כי סדרת הנתונים סימטרית סביב הממוצע. ממוצע הסדרה-40 ושונות הסדרה-100. בשלב מאוחר יותר נוספו שתי תצפיות נוספות לסדרה : 50 ו-30. השונות של 12 התצפיות :
- א. תקטן.
 - ב. תגדל.
 - ג. לא תשתנה.
 - ד. לא ניתן לחשב את השונות ללא ידיעת התצפיות.

הנתונים הבאים מתייחסים לשאלות 9-10:

בחברת "טיק" המשכורת הממוצעת היא 4,600 ₪ וסטיית התקן של משכורת זו הינה 200 ₪. לאחר מו"מ עם ועד עובדי ההנהלה סוכם כי המשכורת תוכפל פי 1.5.

- 9) מהי המשכורת הממוצעת החדשה (ב-₪)?

- א. 2,300.
- ב. 6,900.
- ג. 4,650.
- ד. 4,600.
- ה. חסרים נתונים כדי לדעת.

- 10) מהי סטיית התקן של המשכורת לאחר יישום המו"מ לגבי השכר (ב-₪)?

- א. 200.
- ב. 300.
- ג. 675.
- ד. לא ניתן לדעת.

- 11) הוספת גודל קבוע לכל תצפיות סדרת נתונים :

- א. תגדיל את סטיית התקן.
- ב. תקטין את סטיית התקן.
- ג. לא תשנה את סטיית התקן.
- ד. לא ניתן לדעת.

הנתונים הבאים מתייחסים לשאלות 12-13:

להלן נתונים על ציוני תלמידים שנבחנו במועדים שונים בסטטיסטיקה:

שם התלמיד	ציון	ממוצע הציונים במועד בו נבחן	סטיית התקן של הציונים במועד בו נבחן
צבי	50	50	12
סטף	82	80	5
שרית	65	60	15
לובה	60	63	1.5
מיטב	70	70	10

12 התלמיד הטוב ביותר ביחס לנבחנים באותו מועד בו נבחן הוא:

- א. מיטב.
- ב. צבי.
- ג. לובה.
- ד. שרית.
- ה. סטף.

13 פנינה נבחנה עם סטף וציון התקן שלה שווה לציון התקן של שרית לכן ציונה הוא:

- א. 80.55.
- ב. 65.
- ג. 80.
- ד. 81.66.

הנתונים הבאים מתייחסים לשאלות 14-17:

בבדיקת פתע של משרד הבריאות במפעל שוקולד, נמצא ש:

שוקולד פגום	0	1	2	3	4	5	6	7
מס' קופסאות	35	63	48	12	13	11	10	8

14 מהו החציון של מספר הפגומים בקופסא:

- א. 1.
- ב. 2.
- ג. 4.
- ד. לא ניתן לדעת.

15 מהו הרבעון התחתון של מספר הפגומים בקופסא?

- א. 1.
- ב. 2.
- ג. 3.
- ד. 4.
- ה. לא ניתן לדעת.

16) מספר הפגומים בקופסא הוא משתנה :

- א. סדר.
- ב. שמי.
- ג. כמותי בדיד.
- ד. כמותי רציף.

17) השכיח של מספר הפגומים בקופסא :

- א. 63.
- ב. 1.
- ג. 200.
- ד. לא ניתן לדעת.

18) ביחס לציר המספרים, רוב הערכים בהתפלגות א-סימטרית ימנית נמצאים :

- א. בערכים הגבוהים.
- ב. בחלוקה זהה בין הערכים הגבוהים והנמוכים.
- ג. בערכים הנמוכים.
- ד. לא ניתן לדעת.
- ה. אף לא תשובה מהני"ל נכונה.

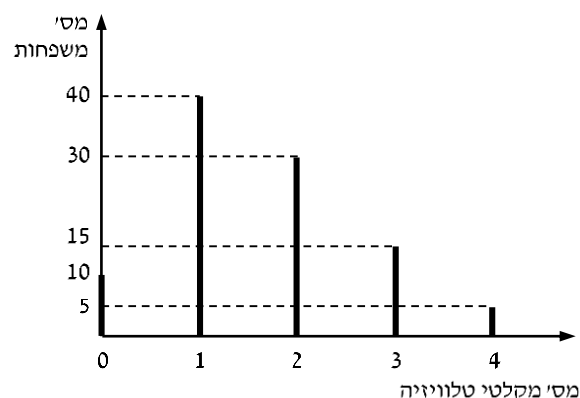
19) בוצע מחקר על מספר העובדים בחברות מזון לעומת חברות תקשורת. החציון והממוצע בשתיהן שווה 8.

איזה מהטענות הבאות היא הנכונה והמלאה ביותר :

- א. השכיחות ב-2 החברות זהה אך שונה מ-8.
- ב. השכיח ב-2 החברות זהה אך לא ניתן לדעת מהו.
- ג. השכיח בשתי חברות הינו בהכרח 8.
- ד. שכיח בחברה אחת שונה מ-8 ובשנייה הוא 8.
- ה. אף תשובה אינה נכונה.

הנתונים הבאים מתייחסים לשאלות 20 עד 24 :

נערך סקר על מספר מקלטי הטלוויזיה הנמצאים בבית. תוצאות הסקר נתונות בדיאגרמת מקלות הבאה :



(20) המשתנה הנחקר כאן הוא :

- א. משתנה שמי.
- ב. משתנה מסולם סדר.
- ג. משתנה כמותי בדיד.
- ד. משתנה כמותי רציף.

(21) הטווח של ההתפלגות הוא :

- א. 35.
- ב. 4.
- ג. 3.
- ד. 2.

(22) ממוצע מספר מקלטי הטלוויזיה למשפחה הוא :

- א. 1.65.
- ב. 1.5.
- ג. 1.
- ד. 2.

(23) השכיח של התפלגות זו היא :

- א. 40.
- ב. 1.5.
- ג. 1.
- ד. 2.

(24) מסתבר שיש בין 2 ל-5 משפחות נוספות שאין להם מקלטי טלוויזיה ויש לצרף את המשפחות הללו להתפלגות. כיצד הנתון זה ישפיע על סטיית התקן?

- א. יקטין אותו.
- ב. יגדיל אותו.
- ג. לא ישנה אותו.
- ד. אין לדעת.

תשובות סופיות:

(5 ב'	(4 ג'	(3 ג'	(2 ג'	(1 א'
(10 ב'	(9 ב'	(8 ג'	(7 א'	(6 ג'
(15 א'	(14 ב'	(13 ד'	(12 ה'	(11 ג'
(20 ג'	(19 ה'	(18 ג'	(17 ב'	(16 ג'
	(24 ב'	(23 ג'	(22 א'	(21 ב'

מבוא לסטטיסטיקה א

פרק 14 - מדדי קשר - מדד הקשר של קרמר

תוכן העניינים

1. כללי 60

מדדי קשר – מדד הקשר של קרמר:

רקע:

משתמשים במדד זה כאשר אחד המשתתפים הוא מסולם שמי והשני מכל סולם אפשרי. מדד הקשר מקבל ערכים בין 0 ל-1. ככל שהמדד יותר קרוב לאחד קיים קשר בעוצמה יותר חזקה בין המשתתפים.

דוגמה (פתרון בהקלטה):

במחקר רוצים לבדוק את הקשר בין מין לדעה בנושא מסוים. שאלו 100 גברים ו-100 נשים על דעתם באיזשהו נושא. להלן טבלת השכיחויות המשותפת שהתקבלה:

$F(x)$	נמנע	נגד	בעד	Y / X
100	10	40	50	גבר
100	10	60	30	אישה
$n = 200$	20	100	80	$F(y)$

הטבלה נקראת טבלת O (observe):

X - מין (גבר/אישה) – סולם שמי.

Y - דעה (בעד/נמנע/נגד) – סולם שמי/סדר.

שלבים בחישוב r_c :

שלב א': נבנה את טבלת E (Expected).

נעתיק את המסגרת של טבלת O ואז כל: $E_i = (F(x) \cdot F(y)) / n$.

$f(x)$	נמנע	נגד	בעד	$\frac{Y}{X}$
100				גבר
100				אישה
$n = 200$	20	100	80	$f(y)$

שלב ב': נחשב $\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$.

שלב ג': נחשב: $r_c = \sqrt{\frac{1}{n(L-1)} \chi^2}$,

כאשר L מבטא את המספר הקטן מבין מספר השורות או העמודות.

שאלות:

- (1) להלן תוצאות מחקר שבדק את הקשר בין מין להשכלה. לגבי כל נחקר נבדק המין שלו והשכלתו. האם קיים קשר בין מין להשכלה? נמקו!

מין/ השכלה	נמוכה	תיכונית	גבוהה
גבר	120	40	20
אישה	20	20	80

- (2) נלקחו 200 אנשים שמתוכם 60 הצהירו שהם עוסקים בפעילות גופנית סדירה. מתוך אלו שעוסקים בפעילות גופנית סדירה 50 נמצאו במצב בריאותי תקין. מתוך אלו שלא עוסקים בפעילות גופנית סדירה 90 נמצאו במצב בריאותי תקין.
- א. בנו טבלת שכיחות משותפת לנתונים שהוצגו בשאלה.
- ב. האם קיים קשר בין פעילות גופנית למצב בריאותי? חשבו לפי מדד הקשר של קרמר.

תשובות סופיות:

- (1) ישנו קשר בעוצמה בינונית, מקדם המתאם של קרמר: 0.595.
- (2) א. להלן טבלה: ב. 0.19.

$f(x)$	לא תקין	תקין	y/x
60	10	50	כן
140	50	90	לא
$n = 200$	60	140	$f(y)$

מבוא לסטטיסטיקה א

פרק 15 - התפלגויות רציפות מיוחדות - התפלגות נורמלית

תוכן העניינים

1. התפלגות נורמלית (טבלת z כוללת ערכים שליליים) (ללא ספר)