

יסודות הסטטיסטיקה א



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



תוכן העניינים

1. סטטיסטיקה תיאורית- סיווג משתנים וסולמות מדידה 1
2. סטטיסטיקה תיאורית - טרנספורמציה על סולמות מדידה 5
3. סטטיסטיקה תיאורית- הצגה של נתונים 7
4. סטטיסטיקה תיאורית- סכימה 15
5. סטטיסטיקה תיאורית -מדדי מרכז (ללא ספר) 15
6. סטטיסטיקה תיאורית - מדדי פיזור - הטווח, השונות וסטיית התקן 19
7. סטטיסטיקה תיאורית - מדדי פיזור - טווח בין רבעוני 22
8. סטטיסטיקה תיאורית- מדדי מיקום יחסי-ציון תקן 26
9. סטטיסטיקה תיאורית-אחוזונים בטבלה בדידה 28
10. סטטיסטיקה תיאורית - טרנספורמציה לינארית 30
11. סטטיסטיקה תיאורית שאלות אמריקאיות 33
12. מדד הקשר של קרמר 40
13. מדד הקשר ספירמן 42
14. מדד הקשר פירסון 45
15. רגרסיה 51
16. מדדי קשר - בחירת מדד מתאים 54
17. התפלגות נורמלית 61
18. תרגול טענות על כל החומר 69
19. שאלות אמריקאיות על כל החומר (ללא ספר) 69

יסודות הסטטיסטיקה 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.
ג. בדיד.
- (3) א. רציף. ב. בדיד. ג. בדיד.
ד. רציף. ה. רציף. ו. רציף.

יסודות הסטטיסטיקה א

פרק 2 - סטטיסטיקה תיאורית - טרנספורמציה על סולמות מדידה

תוכן העניינים

1. כללי 5

סטטיסטיקה תיאורית – טרנספורמציה על סולמות מדידה:

רקע:

טרנספורמציה הינה מצב שבו עושים שינוי לערכים במשתנה הנחקר. להלן נפרט אילו טרנספורמציות מותרות על כל סולם מדידה:

סולם שמי (נומינאלי) –

הטרנספורמציה המותרת היא טרנספורמציה ששומרת על הזהות.

סולם סדר (אורדינאלי) –

הטרנספורמציה המותרת היא טרנספורמציה ששומרת על הסדר.

סולם רווחים (אינטרוואלי) –

הטרנספורמציה המותרת היא טרנספורמציה לינארית חיובית.

סולם מנה/יחס –

הטרנספורמציה המותרת היא הכפלה / חילוק במספר חיובי.

שאלות:

- (1) ציינו באילו סולמות מדידה מותרות הטרנספורמציות הבאות:
- הכפלה באפס.
 - הכפלה ב-2.
 - הכפלה במינוס 1.
 - הוספה של 3.
 - הפחתה של 3.
- (2) איזו טרנספורמציה שומרת על סולם המשתנה "הטמפרטורה בחדר הסגלגל"?
- טרנספורמציה שומרת סדר.
 - טרנספורמציה לינארית חיובית.
 - טרנספורמציה שומרת יחס.
 - תשובות ב' ו-ג' נכונות.
- (3) באיזה סולם/ות מדידה מותרת החסרה של קבוע מכל מספר?
- בסולם שמי בלבד.
 - בסולמות שמי וסדר בלבד.
 - בסולמות שמי, סדר ורווחים בלבד.
 - בכל ארבעת סולמות המדידה.
- (4) איזו טרנספורמציה שומרת על סולם מספרי האוטובוסים של "אגד"?
- טרנספורמציה שומרת סדר.
 - טרנספורמציה לינארית חיובית.
 - טרנספורמציה שומרת יחס.
 - כל התשובות נכונות.

תשובות סופיות:

- (1) א. אף סולם. ב. כל הסולמות. ג. רק על סולם שמי. ד. רווחים, סדר, שמי.
- (2) ד'.
- (3) ג'.
- (4) ד'.

יסודות הסטטיסטיקה א

פרק 3 - סטטיסטיקה תיאורית- הצגה של נתונים

תוכן העניינים

1. כללי 7

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, 5, 3, 4, 3.

טבלת שכיחויות בדידה:

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

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:

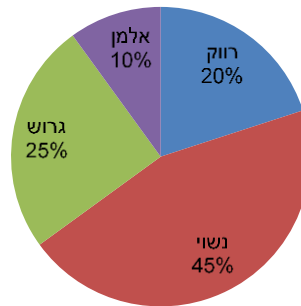
נתנו לקבוצת ילדים לבצע משימה, בדקו את התפלגות זמן הביצוע, בדקות. להלן ההתפלגות שהתקבלה:

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:

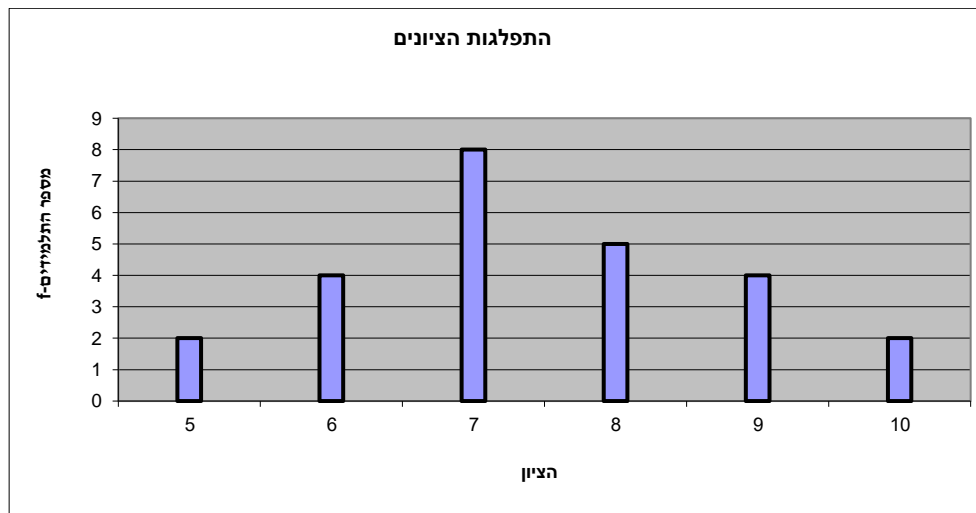
זהו התיאור הגרפי של משתנה איכותי. בדיאגרמת עוגה כל ערך במשתנה מקבל "נתח", שהוא פרופורציונלי לשכיחות היחסית של ערך המשתנה בנתונים.

התפלגות המצב המשפחתי



דיאגרמת מקלות:

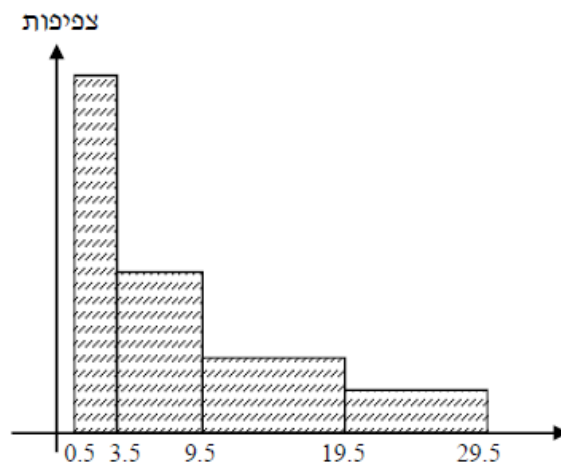
הציר האופקי הוא הציר של המשתנה והציר האנכי של השכיחות, כך שהגובה של המקל מעיד על השכיחות. רלבנטי למשתנה כמותי בדיד. לא נהוג להשתמש בתיאור למשתנה איכותי וכמו כן לא למשתנה כמותי רציף, וכן בסולמות מדידה עבור משתנה מסולם סדר.



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



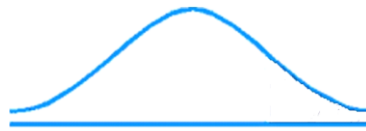
פוליגון – מצולעון:

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.

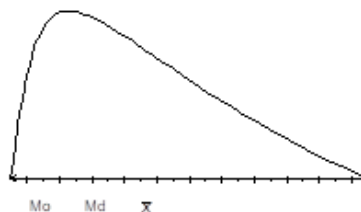


ישנן התפלגויות סימטריות שאינן פעמוניות, כגון:

התפלגות אסימטרית ימנית (חיובית)

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

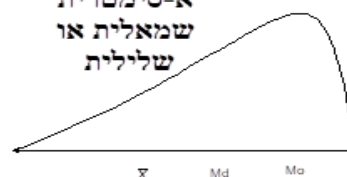
התפלגות א-סימטרית ימנית או חיובית



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.

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.

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

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

יסודות הסטטיסטיקה א

פרק 4 - סטטיסטיקה תיאורית- סכימה

תוכן העניינים

1. כללי.....15

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

$$\text{א. } \sum_{i=1}^4 y_i$$

$$\text{ב. } \sum_{i=1}^6 a$$

$$\text{ג. } \sum_{i=1}^6 x_i y_i$$

$$\text{ד. } \sum_{i=1}^6 (x_i + y_i)$$

$$\text{ה. } \sum_{i=1}^6 x_i + a$$

- (3) קבעו לכל זהות האם היא נכונה:

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

$$\text{ב. } \sum_{i=1}^n a = a \cdot n$$

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

תשובות סופיות:

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

יסודות הסטטיסטיקה א

פרק 5 - סטטיסטיקה תיאורית - מדדי מרכז

תוכן העניינים

1. כללי (ללא ספר)

יסודות הסטטיסטיקה א

פרק 6 - סטטיסטיקה תיאורית - מדדי פיזור - הטווח, השונות וסטיית התקן

תוכן העניינים

1. כללי 19

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.

$$S_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2$$

עבור סדרת נתונים:

דוגמאות:

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

$$S_x^2 = \frac{\sum (x - \bar{x})^2 f}{n} = \frac{\sum x^2 \cdot f}{n} - \bar{x}^2$$

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, 4, 6, 9, 10, 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

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.

יסודות הסטטיסטיקה א

פרק 7 - סטטיסטיקה תיאורית - מדדי פיזור - טווח בין רבעוני

תוכן העניינים

1. טווח בינרבעוני.....22

סטטיסטיקה תיאורית – מדדי פיזור – טווח בין רבעוני:

רקע:

הטווח הבין-רבעוני (יש הקוראים לו התחום הבין-רבעוני) נותן את הטווח בין הרבעונים בו נמצאים 50% מהתצפיות המרכזיות. הרעיון ליצור מדד פיזורי שלא רגיש לתצפיות החריגות ביותר. כדי לחשב את הטווח הבין-רבעוני יש למצוא את הרבעון התחתון והעליון של התפלגות התצפיות.

רבעון תחתון – ערך שמחלק את ההתפלגות לשניים.
25% מהמקרים נמוכים ממנו או שווים לו ו-75% מהמקרים גבוהים או שווים לו.
סימון: Q_1 .

רבעון עליון – ערך שמחלק את ההתפלגות לשניים.
75% מהמקרים נמוכים ממנו או שווים לו ו-25% מהמקרים גבוהים או שווים לו.
סימון: Q_3 .

הטווח הבין רבעוני הוא הפער בין שני הרבעונים: $IQR = Q_3 - Q_1$.

שלב ב: במציאת טווח בין-רבעוני בטבלת שכיחויות:

שלב א: נמצא את הרבעון התחתון: הוא הערך שהשכיחות היחסית המצטברת באחוזים עברה לראשונה את 25%.

שלב ב: נמצא את הרבעון העליון: הוא הערך שהשכיחות היחסית המצטברת באחוזים עברה לראשונה את 75%.

שלב ג: נמצא את הטווח הבין-רבעוני: נחסר את הרבעונים: $IQR = Q_3 - Q_1$.

דוגמה (פתרון בהקלטה):

בסניף בנק 250 לקוחות. ספרו לכל לקוח את מספר תוכניות החיסכון שלו. מהו הטווח הבין-רבעוני של מספר תוכניות החיסכון בסניף?

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

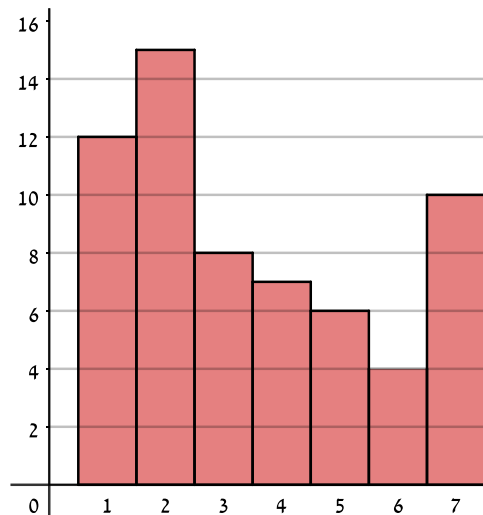
שאלות:

1) להלן התפלגות מספר המכוניות למשפחה בישוב "הגורן":

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

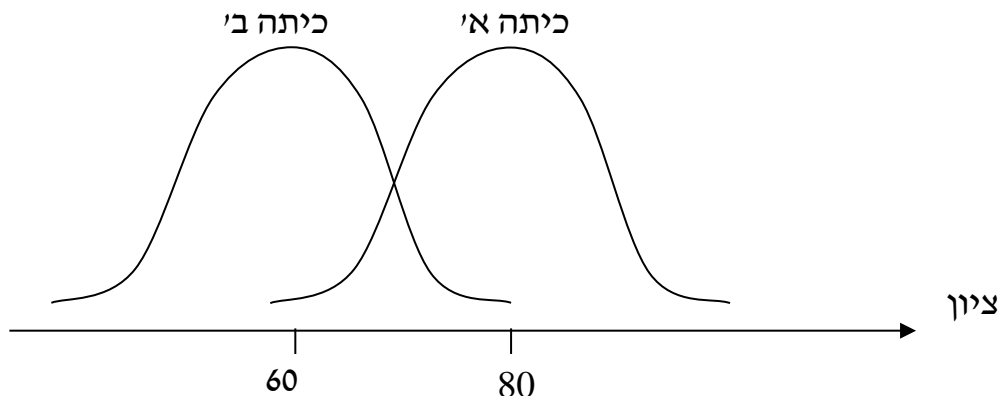
מהו הטווח הבין-רבעוני של מספר המכוניות למשפחה בישוב "הגורן"?

2) בסקר שנעשה בדקו את מספר ימי המחלה השנתיים של מורים בארץ.



- א. מה מייצגים הערכים בציר האופקי?
- ב. מהו הטווח הבין-רבעוני של מספר ימי המחלה של המורים?
- ג. אם נוסף 25 מורים אשר הצהירו שמספר ימי המחלה השנתיים שלהם הוא 4 ימים, כיצד הדבר ישנה את הטווח הבין-רבעוני? הסבירו.
- ד. אם מסתבר שחלק מהמורים בסקר הצהירו שהם חלו 7 ימים בשנה אבל בפועל הם חלו 8 ימים, כיצד הדבר ישנה את הטווח הבין-רבעוני? הסבירו.

3) לפניך שתי עקומות המתארות את התפלגות הציונים בכל כיתה. באיזו כיתה הטווח הבין-רבעוני גדול יותר?



- א. כיתה א.
- ב. כיתה ב'.
- ג. לשתיהן אותו טווח בין-רבעוני.
- ד. לא ניתן לדעת, אין מספיק נתונים.

4) הוספת גודל קבוע לכל תצפיות סדרת נתונים:

- א. תגדיל את הטווח הבין-רבעוני.
- ב. תקטין את הטווח הבין-רבעוני.
- ג. לא תשנה הטווח הבין-רבעוני.
- ד. לא ניתן לדעת מה יקרה לטווח הבין-רבעוני.

5) חושב הטווח הבין-רבעוני עבור התפלגות מסוימת והתקבלה התוצאה אפס. לכן:

- א. לפחות 50% מהתצפיות זהות.
- ב. סטיית התקן היא אפס.
- ג. ההתפלגות היא סימטרית.
- ד. מצב זה כלל לא יתכן.

- 6) סניף מספר 543 של בנק "רווה" בדק ל-80 לקוחות את מספר הפעמים שכל לקוח נכנס לסניף הבנק במשך שבוע. התוצאות שהתקבלו הן:
- 50 אנשים נכנסו 0 פעמים לסניף.
 - 20 אנשים נכנסו פעם אחת לסניף.
 - 5 אנשים נכנסו פעמיים לסניף.
 - 5 אנשים נכנסו יותר מפעמיים.
- מהו הטווח הבין-רבעוני?
- א. 60.
 - ב. 2.
 - ג. 50.
 - ד. 1.

- 7) התפלגות הציונים במבחן ווקסלר היא סימטרית לכן:
- א. טווח הציונים הוא אפס.
 - ב. הטווח הבין רבעוני של הציונים אפס.
 - ג. סעיפים א ו-ב הם נכונים.
 - ד. אף סעיף אינו נכון.

תשובות סופיות:

- 1) 2.
- 2) א. מספר ימי המחלה השנתיים. ב. 3. ג. יקטן. ד. לא ישתנה.
- 3) ג.
- 4) ג.
- 5) א.
- 6) ד.
- 7) ד.

יסודות הסטטיסטיקה א

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

תוכן העניינים

1. כללי 26

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.

יסודות הסטטיסטיקה א

פרק 9 - סטטיסטיקה תיאורית-אחוזונים בטבלה בדידה

תוכן העניינים

1. כללי 28

סטטיסטיקה תיאורית – מדדי מיקום יחסי – אחוזונים בטבלה בדידה:

רקע:

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

חישוב האחוזון מתוך נתונים בטבלת שכיחויות בדידה:

האחוזון הוא הערך שבו בפעם הראשונה השכיחות היחסית המצטברת (באחוזים) גדולה או שווה ל- $p\%$.

דוגמה (פתרון בהקלטה):

בסניף בנק 250 לקוחות. ספרו לכל לקוח את מספר תוכניות החיסכון שלו:

שכיחות יחסית מצטברת	שכיחות מצטברת	$F(x)$	# תוכניות החיסכון
		100	0
		75	1
		25	2
		25	3
		25	4

א. מצאו את האחוזון ה-25.

ב. מצאו את הערך ש-20% מהמקרים מעליו.

שאלות:

(1) להלן התפלגות של משתנה כלשהו:

$F(x)$	X
10	0
40	1
30	2
15	3
5	4

מצאו להתפלגות את:

- א. האחוזון ה-60.
- ב. המאון ה-40.
- ג. העשירון העליון.
- ד. הטווח בין הרבעונים.

(2) להלן התפלגות מספר המכונניות למשפחה בישוב "הגורן":

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

חשבו את:

- א. העשירון התחתון.
- ב. האחוזון ה-30.
- ג. הערך ש-20% מהתצפית גדולות ממנו.
- ד. רבעון עליון.

תשובות סופיות:

- (1) א. 2 ב. 1 ג. 3 ד. 1
- (2) א. 1 ב. 2 ג. 4 ד. 4

יסודות הסטטיסטיקה א

פרק 10 - סטטיסטיקה תיאורית - טרנספורמציה לינארית

תוכן העניינים

1. כללי 30

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:

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

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

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

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

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) $a^2 \cdot S_x^2$

יסודות הסטטיסטיקה א

פרק 11 - סטטיסטיקה תיאורית שאלות אמריקאיות

תוכן העניינים

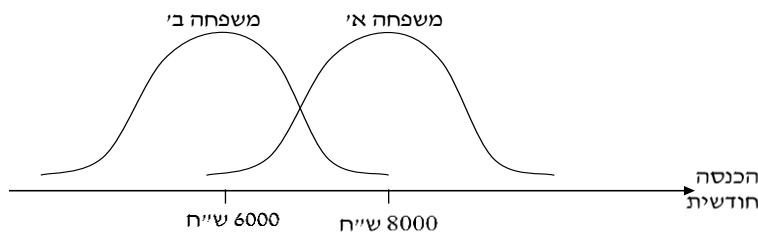
1. כללי 33

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.

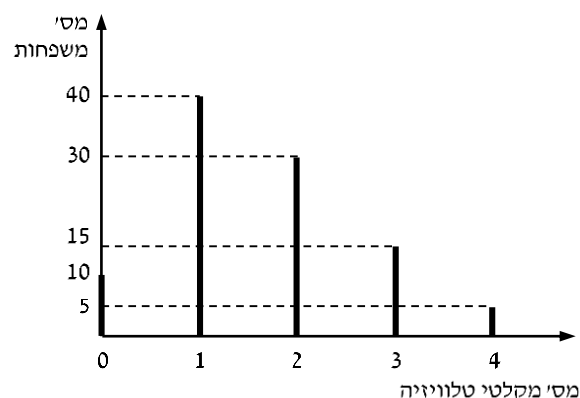
איזה מהטענות הבאות היא הנכונה והמלאה ביותר :

- 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'

יסודות הסטטיסטיקה א

פרק 12 - מדד הקשר של קרמר

תוכן העניינים

1. כללי 40

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	$f(y)$

יסודות הסטטיסטיקה א

פרק 13 - מדד הקשר ספירמן

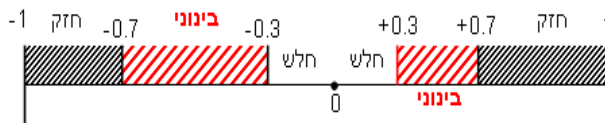
תוכן העניינים

42 1. כללי

מדדי קשר – מדד הקשר של ספירמן:

רקע:

מתי נשתמש במדד ספירמן?
 כאשר אחד המשתנים מסולם סדר והשני מסולם סדר ומעלה.
 הקשר שהמדד בודק הוא קשר דירוגי.
 מדד הקשר בודק את:



1. כיוון הקשר.

2. עצמת הקשר.

המדד מקבל ערכים בסקלה מ-(-1) ועד 1.

קשר דירוגי חיובי מלא:

מדד הקשר של ספירמן יוצא 1.
 ככל שמשנתנה אחד עולה, השני עולה ללא יוצא מן הכלל.

קשר דירוגי חיובי חלקי:

מקדם המתאם בין 0 ל-1.
 ככל שמשנתנה אחד עולה, לשני יש נטייה לעלות אך לא באופן מוחלט.

קשר דירוגי שלילי מלא:

מדד הקשר של ספירמן יוצא -1.
 ככל שמשנתנה אחד עולה השני יורד ללא יוצא מן הכלל.

קשר דירוגי שלילי חלקי:

מקדם המתאם הוא בין 0 ל-(-1).
 ככל שמשנתנה אחד עולה לשני יש נטייה לרדת אך לא באופן מוחלט.
 על מנת לחשב את הקשר יש לבצע פעולת דירוג (RANK).
 כאשר מדרגים, אם יש כמה תצפיות שתופסות את אותו הערך אז הדירוג שלהם הוא הממוצע של המקומות שהן תופסות.

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}$$

הנוסחה של מדד הקשר:

דוגמה (פתרון בהקלטה):

בתחרות רוקדים עם כוכבים השתתפו 7 זוגות, 2 שופטים נתנו את ציוניהם לריקוד של כל זוג. להלן התוצאות שהתקבלו:

מספר הזוג	ציון שופט א'	R_x	ציון שופט ב'	R_y	$d = r_x - r_y$	d^2
1	4		5			
2	5		5			
3	6		7			
4	5		7			
5	8		9			
6	7		9			
7	3		7			

מהי מידת ההתאמה בין ציוני השופטים?

X - ציון שופט א' (סולם סדר).

Y - ציון שופט ב' (סולם סדר).

שאלות:

(1) בתחרות יופי חילקו שני שופטים ציונים למועמדות:

מספר מועמדת	1	2	3	4	5	6	7
ציון שופט א'	7	8	6	8	9	5	6
ציון שופט ב'	8	8	7	8	9	5	7

האם קיים קשר בין שתי הערכות השופטים? נמקו והסבירו!

(2) משרד רצה לבחון האם קיים קשר בין מידת המוטיבציה של העובדים שלו לבין מספר החיסורים של העובדים בחודש עבודה. להלן התוצאות שהתקבלו:

מספר חיסורים	מידת מוטיבציה
0	גבוהה
4	נמוכה
2	בינונית
5	נמוכה
1	גבוהה

האם קיים קשר בין רמת המוטיבציה של העובד ומספר החיסורים שלו? חשבו באמצעות מדד הקשר המתאים והסבירו.

(3) אם: $r_s = 1$, הדבר אומר שערכי X תמיד שווים לערכי Y . האם הטענה נכונה? הסבר.

תשובות סופיות:

- (1) קיים קשר דירוג חיובי חזק בין הערכת שופט א' להערכת שופט ב'.
מדד הקשר: 0.973.
- (2) קיים קשר שלילי בעוצמה חזקה בין רמת המוטיבציה של העובד למס' החיסורים שלו.
מדד הקשר: -0.85.
- (3) לא נכון.

יסודות הסטטיסטיקה א

פרק 14 - מדד הקשר פירסון

תוכן העניינים

1. כללי 45

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.

דוגמה:

בבניין של 5 דירות בדקו את הנתונים הבאים:

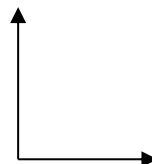
X - מס' חדרים בדירה.

Y - מס' נפשות הגרות בדירה.

להלן התוצאות שהתקבלו:

מס' דירה	X	Y
1	3	2
2	2	2
3	4	3
4	3	3
5	5	4

נשרטט מנתונים הללו דיאגרמת פיזור:



בשלב השני, מחשבים את מקדם המתאם (מדד הקשר) שבודק עד כמה קיים קשר לינארי בין שני המשתנים.

המדד (נקרא 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: $y = bx + a$.

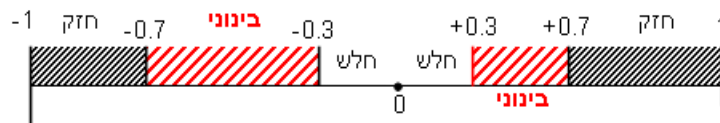
מתאם חיובי מלא (מקדם מתאם 1) אומר שקיים קשר לינארי מלא בו השיפוע b יהיה חיובי ואילו מתאם שלילי מלא אומר שקיים קשר לינארי מלא בו השיפוע b שלילי (מקדם מתאם -1).

מתאם חיובי חלקי אומר שככל שמשנתה אחד עולה לשני יש נטייה לעלות בערכו אבל לא קיימת נוסחה לינארית שמקשרת את X ל- Y באופן מוחלט.

מתאם שלילי חלקי אומר שככל שמשנתה 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.

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

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

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

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

שאלות:

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

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

- 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 הללו עבור חמישה חולים. להלן התוצאות שהתקבלו:

X	Y
10	12
14	15
15	15
18	17
20	21

- א. מה הממוצע של כל רמת הורמון?
- ב. מהו מקדם המתאם בין ההורמונים? ומה משמעות התוצאה?

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

$$\sum_{i=1}^{20} (X_i - \bar{X})^2 = 76, \quad \sum_{i=1}^{20} (Y_i - \bar{Y})^2 = 76, \quad \sum_{i=1}^{20} X_i = 240, \quad \sum_{i=1}^{20} Y_i = 200$$

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

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:

$$\sum_{i=1}^{20} X_i = 240, \quad \sum_{i=1}^{20} Y_i = 200, \quad \sum_{i=1}^{20} X_i^2 = 2960, \quad \sum_{i=1}^{20} Y_i^2 = 2080, \quad \sum_{i=1}^{20} X_i Y_i = 2464$$

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?

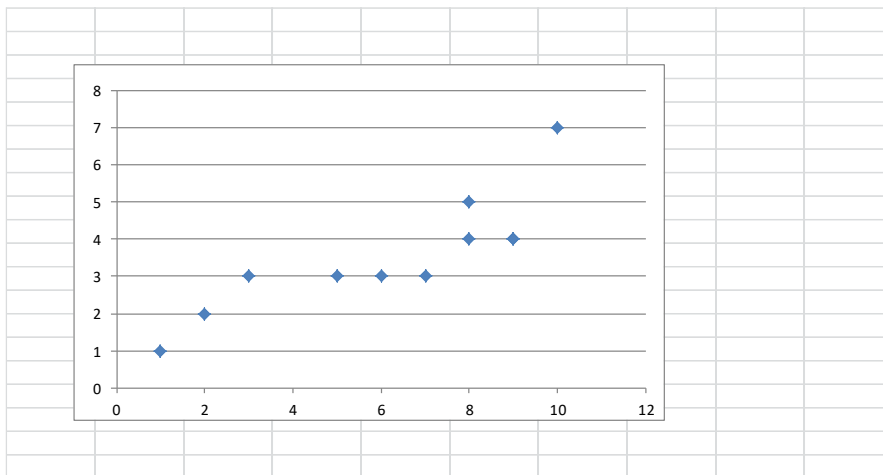
(6) להלן רשימת טענות. לגבי כל טענה קבע נכון/לא נכון ונמק!
 א. מתווך דירות המיר מחירי דירות מדולר לשקל. נניח שדולר אחד הוא 3.5 ₪. אם מתווך הדירות יחשב את מדד הקשר של פירסון בין מחיר הדירה בשקלים למחיר הדירה בדולרים הוא יקבל 1.
 ב. לסדרה של נתונים התקבל: $\bar{X} = \bar{Y} = 6, S_x = S_y = 1$ לכן מדד הקשר של פירסון יהיה 1.
 ג. אם השונות המשותפת של X ושל Y הינה 0 אז בהכרח גם מקדם המתאם של פירסון יהיה 0.

(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 בשי"ח (באותו היום ערך הדולר היה 4.2 ₪). מהו מקדם המתאם בין המחיר בדולר למחיר ב-₪?

- 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.

תשובות סופיות:

- (1) ב. -0.9325
- (2) א. $\bar{x} = 15.4$, $\bar{y} = 16$ ב. $r_{xy} = 0.96$
- (3) א. 0.8
- (4) 0.8
- (5) 1
- (6) א. נכון. ב. לא נכון. ג. נכון.
- (7) ג'.
- (8) א'.
- (9) ב'.

יסודות הסטטיסטיקה א

פרק 15 - רגרסיה

תוכן העניינים

1. כללי 51

מדדי קשר – רגרסיה ליניארית:

רקע:

במידה וקיים קשר חזק בין שני המשתנים הכמותיים נהוג לבצע ניבוי. 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 - נותן את ערך Y כאשר X הנו אפס על גבי קו הניבויים. הוא נקרא 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: $Y = bX + a$, $b = r \frac{S_y}{S_x}$.

לצורך בניית קו ניבויים לניבוי X על סמך Y נצטרך לעדכן את הנוסחאות בהתאם.

שאלות:

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

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

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

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

- א. חשבו את מדד הקשר הלינארי בין X ל- Y . מיהו המשתנה התלוי?
 ב. מצאו את קו הרגרסיה לניבוי ההוצאה של משפחה על סמך הכנסה שלה. הסבירו 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 לימוד. נסמן ב- Y את הכנסתו באלפי ₪. במחקר התקבלו התוצאות הבאות:

$$S_x = 2, \quad S_y = 5, \quad \bar{X} = 14, \quad \bar{Y} = 8, \quad \text{COV}(X, Y) = 7.5$$

- א. חשבו 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-?
 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 X ו- Y . כמו כן נתון: $\bar{X} = 1.5, S_x = S_y = 4$,
 וכן שקו הרגרסיה של Y על בסיס X הינו: $Y = -0.2X + 0.5$.
 חשבו מהו מקדם המתאם בין X ל- Y .

תשובות סופיות:

- | | | |
|----------------------|-----------------------|-------------|
| ג. 12.4 אלפי ₪. | ב. $Y = 0.8X + 0.4$. | א. 0.8 (1) |
| ג. 14.6 שנים. | ב. 4.25 אלפי ₪. | א. 0.75 (2) |
| ג. $Y = 1.2X + 29$. | ב. 29. | א. 1.2 (3) |
| | | א. -0.2 (4) |

יסודות הסטטיסטיקה א

פרק 16 - מדדי קשר - בחירת מדד מתאים

תוכן העניינים

1. בחירת מדד מתאים 54

מדדי קשר – בחירת מדד מתאים:

רקע:

בפרק זה נתרגל את התהליך של בחירת מדד הקשר (מקדם המתאם) המתאים. נתרכז בשלושת מדדי הקשר הנפוצים ביותר:

- מדד הקשר של קרמר.
- מדד הקשר של ספירמן.
- מדד הקשר של פירסון (מדד הקשר הלינארי).

בחירת מדד הקשר נעשה לפי סולמות המדידה של שני המשתנים שאנחנו רוצים לבדוק את הקשר בינם. הנושא של סולמות מדידה נלמד כבר בפרק אחר, כמו כן כל מדד קשר נלמד בפרק נפרד. אנו מתרכזים ב 3 סולמות מדידה:

- סולם שמי/ זהות (nominal).
- סולם סדר (ordinal).
- סולם כמותי (scale): לכאן אנו מאחדים את סולם רווחים ומנה יחד.

שלושת מדדי הקשר שלעיל דנים בקשר בין שני משתנים. מדדי הקשר הם סימטריים, כלומר אין זה משנה איזה משתנה נגדיר בתור משתנה X ואיזה יוגדר בתור משתנה Y .

להלן טבלה שמסכמת את בחירת המדד המתאים:

X / Y	שמי	סדר	כמותי
שמי	קרמר	קרמר	קרמר
סדר	קרמר	ספירמן	ספירמן
כמותי	קרמר	ספירמן	פירסון

דוגמה (פתרון בהקלטה):

במפעל לייצור מצברים לרכב בדקו במשך 40 ימים את התפוקה היומית (מספר מצברים במאות) ואת מספר הפועלים שעבדו באותו היום. איזה מדד קשר מתאים כדי לבדוק האם קיים קשר בין התפוקה היומית לכמות עובדים באותו היום במפעל?

- א. פירסון.
- ב. ספירמן.
- ג. קרמר.
- ד. חסרים נתונים כדי לדעת זאת.

שאלות:

- (1) בקרב תלמידי כיתות א' בבית הספר גבריאלי אשר בתל אביב בדקו לכל תלמיד את גובהו בס"מ ואת משקלו בק"ג. מהו מדד הקשר המתאים כדי לבדוק האם קיים קשר בין גובה התלמיד למשקלו?
 א. פירסון.
 ב. ספירמן.
 ג. קרמר.
 ד. אין מספיק נתונים כדי לדעת.
- (2) בסקר שנעשה על אזרחים במדינה בדקו לכל אזרח את השכלתו ואת שכרו. מהו מדד הקשר המתאים כדי לבדוק האם קיים קשר בין השכלה לשכר?
 א. פירסון.
 ב. ספירמן.
 ג. קרמר.
 ד. אין מספיק נתונים כדי לדעת.
- (3) בגן הילדים של שולה אספו נתונים על 25 הילדים שבגן. על כל ילד בדקו את רמת הביטחון העצמי שלו ($X =$ מדד שמקבל ערכים בין 1 - נמוך ועד 5 - גבוה), ואת אוצר המילים שלו ($Y =$ לפי מבחן שנעשה לכל ילד בו ספרו את מספר המילים שידע מתוך רשימה של 20 מילים). איסוף הנתונים נעשה על ידי איש מקצוע שצפה בילדים ובחן אותן.
 מהו מקדם המתאם המתאים לבדיקת התלות בין X לבין Y ?
 א. פירסון.
 ב. ספירמן.
 ג. קרמר.
 ד. אין מספיק נתונים כדי לדעת.
- (4) הועלתה השערה בבית הספר למדעי ההתנהגות שיש קשר בין המרצה להצלחת הסטודנט. לצורך בדיקת הטענה בדקו לגבי כל סטודנט שלמד סטטיסטיקה אצל איזה מרצה הוא למד (היו 3 מרצים שונים) והאם הוא עבר את הבחינה. מהו מדד הקשר המתאים במקרה זה?
 א. פירסון.
 ב. ספירמן.
 ג. קרמר.
 ד. אין מספיק נתונים כדי לדעת.

5) בבורסה בתל אביב רצו לבדוק את הקשר בין גובה הריבית במשק בסוף החודש (באחוזים), לבין תשואת מניית אקטר (באחוזים) בסוף החודש. מהו מדד הקשר המתאים?

- א. פירסון.
- ב. ספירמן.
- ג. קרמר.
- ד. אין מספיק נתונים כדי לדעת.

6) בעל מסעדה ביצע סקר על לקוחותיו, בין השאלות שנשאלו בסקר:

- מה מידת שביעות הרצון של הלקוח מאדיבות השירות של המלצר בסקלה של 1 עד 5.
- מה גילו של הלקוח בשנים.
- מה גובה התשר (טיפ) ב-ש אשר נתן הלקוח למלצר בלכתו מהמסעדה.

מהו המדד המתאים כדי לבדוק האם קיים מתאם חיובי בין מידת שביעות הרצון של הלקוח מאדיבות השירות לבין גובה התשר שהוא נתן למלצר?

- א. פירסון.
- ב. ספירמן.
- ג. קרמר.
- ד. אין מספיק נתונים כדי לדעת.

7) בעל מסעדה ביצע סקר על לקוחותיו, בין השאלות שנשאלו בסקר:

- מה מידת שביעות הרצון של הלקוח מאדיבות השירות של המלצר בסקלה של 1 עד 5.
- מה גילו של הלקוח בשנים.
- מה גובה התשר (טיפ) ב-ש אשר נתן הלקוח למלצר בלכתו מהמסעדה.

מהו המדד המתאים כדי לבדוק האם קיים מתאם בין גיל הלקוח לגובה התשר שהעניק לשירות?

- א. פירסון.
- ב. ספירמן.
- ג. קרמר.
- ד. אין מספיק נתונים כדי לדעת.

הנתונים הבאים מתאימים ל-3 השאלות הבאות:

חוקרים ערכו מדגם של ילדים מכיתות ב' ו-ג' מ-4 בתי ספר שונים. הועבר לילדים שאלון בו תואר מצב מסוים והילדים התבקשו לציין את רמת החרדה שלהם באשר לאותו מצב. המשתנים שלגביהם נאספו נתונים:

- מגדר (1 - בן, 2 - בת).
- כיתה (0 - ג', 1 - ב').
- בית ספר (A, B, C, D).
- רמת חרדה (ציון שהילד היה צריך לתת בסקלה של 1 עד 10).
- גיל התלמיד בחודשים.

8) מהו מדד הקשר המתאים כדי לבדוק את הקשר בין גיל התלמיד לבין רמת החרדה שלו מהמצב?

- א. פירסון.
- ב. ספירמן.
- ג. קרמר.
- ד. אין מספיק נתונים כדי לדעת.

9) מהו מדד הקשר המתאים כדי לבדוק את הקשר בין המגדר לבין רמת החרדה שלו מהמצב?

- א. פירסון.
- ב. ספירמן.
- ג. קרמר.
- ד. אין מספיק נתונים כדי לדעת.

10) מהו מדד הקשר המתאים כדי לבדוק את הקשר בין המגדר לבין בית הספר?

- א. פירסון.
- ב. ספירמן.
- ג. קרמר.
- ד. אין מספיק נתונים כדי לדעת.

11) בטבלה שלהלן נתונות שכיחויות ההצלחה והכישלון של 150 חולים :

C	B	A	תוצאה/ התרופה
45	13	35	נרפא
5	37	15	לא נרפא

החולים קיבלו 3 תרופות שונות ובדקו עבור כל חולה אם התרופה הצליחה בריפוי. מהו מדד הקשר המתאים?

- א. פירסון.
- ב. ספירמן.
- ג. קרמר.
- ד. אין מספיק נתונים כדי לדעת.

12) שני מוסיקאים מפורסמים נתנו ציון בסולם של 1-10 לקולם של 8 מתמודדים בתוכנית ריאליטי ידועה. ציון 10 ניתן לקול שמצא חן ביותר בעיני המוסיקאי. מפיק התוכנית רצה לבדוק האם יש קורלציה בין המוסיקאים מבחינת הטעם. בטבלה הבאה נתונים הציונים של כל אחד מהמוסיקאים את שמונת המתמודדים :

8	7	6	5	4	3	2	1	
4	1	1	3	4	7	5	6	מוסיקאי א'
7	2	3	3	2	5	7	5	מוסיקאי ב'

מהו מדד הקשר המתאים?

- א. פירסון.
- ב. ספירמן.
- ג. קרמר.
- ד. אין מספיק נתונים כדי לדעת.

13) להלן טבלה המסכמת את השכר באלפי ₪ של עובדים בחברה ואת רמת המוטיבציה שלהם מ-1 עד 5 :

30	15	20	18	12	שכר
5	3	5	4	4	מוטיבציה

מהו מקדם המתאם המתאים לבדיקת רמת ההתאמה בין המוטיבציה לשכר של העובד?

- א. פירסון.
- ב. ספירמן.
- ג. קרמר.
- ד. אין מספיק נתונים כדי לדעת.

14) להלן טבלה על נתונים שנאספו על מספר תצפיות:

5	4	3	2	1	X
20	17	17	14	12	Y

אם מעוניינים לבדוק עד כמה קיים קשר לנארי בין שני המשתנים.
מהו המדד המתאים?

- א. פירסון.
- ב. ספירמן.
- ג. קרמר.
- ד. אין מספיק נתונים כדי לדעת.

תשובות סופיות:

(1) א'	(2) ד'	(3) ב'	(4) ג'	(5) א'
(6) ב'	(7) א'	(8) ב'	(9) ג'	(10) ג'
(11) ג'	(12) ב'	(13) ב'	(14) א'	

יסודות הסטטיסטיקה א

פרק 17 - התפלגות נורמלית

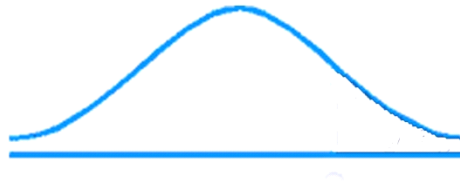
תוכן העניינים

1. כללי 61

התפלגויות רציפות מיוחדות – התפלגות נורמלית:

רקע:

התפלגות נורמלית הינה התפלגות של משתנה רציף. ישנם משתנים רציפים מסוימים שנהוג להתייחס אליהם כנורמליים כגון: זמן ייצור, משקל תינוק ביום היוולדו ועוד. פונקציית הצפיפות של ההתפלגות הנורמלית נראית כמו פעמון:



לעקומה זו קוראים גם עקומת גאוס ועקומה אחת נבדלת מהשנייה באמצעות הממוצע וסטיית התקן שלה.

אלה הם הפרמטרים שמאפיינים את ההתפלגות: $X \sim N(\mu, \sigma^2)$.

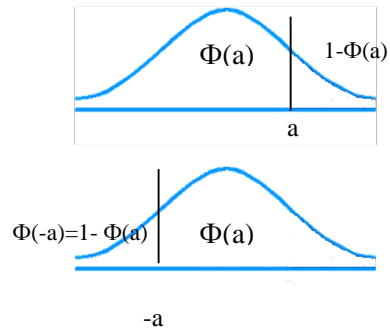
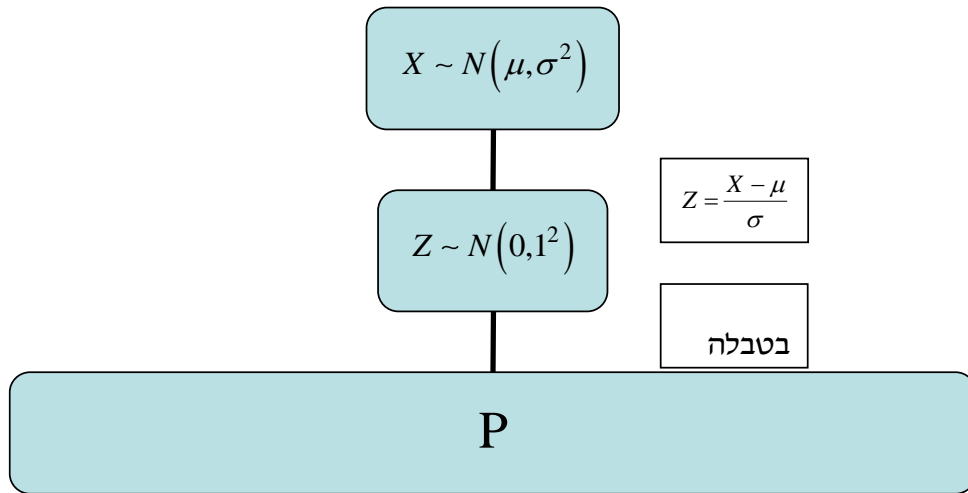
$$\text{נוסחת פונקציית הצפיפות: } f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

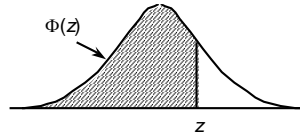
כדי לחשב הסתברויות בהתפלגות נורמלית יש לחשב את השטחים הרלוונטיים שמתחת לעקומה. כדי לחשב שטחים אלה נמיר כל התפלגות נורמלית להתפלגות נורמלית סטנדרטית על ידי תהליך הנקרא תקנון. התפלגות נורמלית סטנדרטית היא התפלגות נורמלית שהממוצע שלה הוא אפס וסטיית התקן היא אחת, והיא תסומן באות Z : $Z \sim N(0, 1^2)$.

$$\text{תהליך התקנון מבוצע על ידי הנוסחה הבאה: } Z = \frac{X - \mu}{\sigma}$$

אחרי תקנון מקבלים ערך הנקרא ציון תקן. ציון התקן משמעו בכמה סטיות תקן הערך סוטה מהממוצע.

לאחר חישוב ציון התקן של ערך מסוים נעזרים בטבלה של ההתפלגות הנורמלית הסטנדרטית לחישוב השטח הרצוי, ובאופן כללי נתאר את הסכמה הבאה:



טבלת ההתפלגות המצטברת הנורמלית סטנדרטית – ערכי $\Phi(z)$:


z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

z	1.282	1.645	1.960	2.326	2.576	3.090	3.291	3.891	4.417
$\Phi(z)$	0.90	0.95	0.975	0.99	0.995	0.999	0.9995	0.99995	0.999995

דוגמה (הפתרון בהקלטה):

משקל חפיסות שוקולד המיוצרות בחברה מתפלג נורמלית עם ממוצע 100 גרם בסטיית תקן של 8 גרם.

- (1) מה אחוז חפיסות השוקולד ששוקלות מתחת ל-110 גרם?
- (2) מה אחוז חפיסות השוקולד ששוקלות מעל 110 גרם?
- (3) מה אחוז חפיסות השוקולד ששוקלות מתחת ל-92 גרם?
- (4) מהו המשקל ש-90% מהחפיסות בקו הייצור שוקלים פחות מהם?

שאלות:

- (1) הגובה של אנשים באוכלוסייה מסוימת מתפלג נורמלית עם ממוצע של 170 ס"מ וסטית תקן של 10 ס"מ.
- מה אחוז האנשים שגובהם מתחת ל-182.4 ס"מ?
 - מה אחוז האנשים שגובהם מעל 190 ס"מ?
 - מה אחוז האנשים שגובהם בדיוק 173.6 ס"מ?
 - מה אחוז האנשים שגובהם מתחת ל-170 ס"מ?
 - מה אחוז האנשים שגובהם לכל היותר 170 ס"מ?
- (2) נתון שהזמן שלוקח לתרופה מסוימת להשפיע מתפלג נורמלית עם ממוצע של 30 דקות ושונות של 9 דקות רבועות.
- מהי פרופורציית המקרים בהן התרופה תעזור אחרי יותר משעה?
 - מה אחוז מהמקרים שבהן התרופה תעזור בין 35 ל-37 דקות?
 - מה הסיכוי שהתרופה תעזור בדיוק תוך 36 דקות?
 - מה שיעור המקרים שבהן ההשפעה של התרופה תסטה מ-30 דקות בפחות מ-3 דקות?
- (3) המשקל של אנשים באוכלוסייה מסוימת מתפלג נורמלית עם ממוצע של 60 ק"ג וסטית תקן של 8 ק"ג.
- מה אחוז האנשים שמשקלם נמוך מ-55 ק"ג?
 - מהי פרופורציית האנשים באוכלוסייה שמשקלם לפחות 50 ק"ג?
 - מהי השכיחות היחסית של האנשים באוכלוסייה שמשקלם בין 60 ל-70 ק"ג?
 - לאיזה חלק מהאוכלוסייה משקל הסוטה מהמשקל הממוצע בלא יותר מ-4 ק"ג?
 - מה הסיכוי שאדם אקראי ישקול מתחת ל-140 ק"ג?
- (4) משקל תינוקות ביום היוולדם מתפלג נורמלית עם ממוצע של 3300 גרם וסטית תקן 400 גרם.
- מצאו את העשירון העליון.
 - מצאו את האחוזון ה-95.
 - מצאו את העשירון התחתון.

- 5) ציוני מבחן אינטליגנציה מתפלגים נורמלית עם ממוצע 100 ושונויות 225.
- מה העשירון העליון של הציונים במבחן האינטליגנציה?
 - מה העשירון התחתון של ההתפלגות?
 - מהו הציון ש-20% מהנבחנים מקבלים מעליו?
 - מהו האחוזון ה-20?
 - מהו הציון ש-5% מהנבחנים מקבלים מתחתיו?
- 6) נפח משקה בבקבוק מתפלג נורמלית עם סטיית תקן של 20 מ"ל, ונתון ש-33% מהבקבוקים בעלי נפח שעולה על 508.8 מ"ל.
- מה ממוצע נפח משקה בבקבוק?
 - 5% מהבקבוקים המיוצרים עם הנפח הגבוה ביותר נשלחים לבדיקה, החל מאיזה נפח שולחים בקבוק לבדיקה?
 - 1% מהבקבוקים עם הנפח הקטן ביותר נתרמים לצדקה, מהו הנפח המקסימלי לצדקה?
- 7) אורך חיים של מכשיר מתפלג נורמלית. ידוע שמחצית מהמכשירים חיים פחות מ-500 שעות, כמו כן ידוע ש-67% מהמכשירים חיים פחות מ-544 שעות.
- מהו ממוצע אורך חיי מכשיר?
 - מהי סטיית בתקן של אורך חיי מכשיר?
 - מה הסיכוי שמכשיר אקראי יחיה פחות מ-460 שעות?
 - מהו המאיון העליון של אורח חיי מכשיר?
 - 1% מהמכשירים בעלי אורך החיים הקצר ביותר נשלח למעבדה לבדיקה מעמיקה. מהו אורך החיים המקסימלי לשליחת מכשיר למעבדה?
- 8) להלן שלוש התפלגויות נורמליות של שלוש קבוצות שונות ששורטטו באותה מערכת צירים. ההתפלגויות מוספרו כדי להבדיל ביניהן.
- לאיזו התפלגות הממוצע הגבוה ביותר?
 - במה מבין המדדים הבאים התפלגות 1 ו-2 זהות?
 - בעשירון העליון.
 - בממוצע.
 - בשונויות.
 - לאיזו התפלגות סטיית התקן הקטנה ביותר?
 - 1.
 - 2.
 - 3.
 - אין לדעת.



9 הזמן שלוקח לאדם להגיע לעבודתו מתפלג נורמלית עם ממוצע של 40 דקות וסטית תקן של 5 דקות.

- א. מה ההסתברות שמשך הנסיעה של האדם לעבודתו יהיה לפחות שלושת רבעי השעה?
- ב. אדם יצא לעבודתו בשעה 08:10 מביתו. הוא צריך להגיע לעבודתו בשעה 09:00. מה הסיכוי שיאחר לעבודתו?
- ג. אם ידוע שזמן נסיעתו לעבודה היה יותר משלושת רבעי השעה. מה ההסתברות שזמן הנסיעה הכולל יהיה פחות מ-50 דקות?
- ד. מה הסיכוי שבשבוע (חמישה ימי עבודה) בדיוק פעם אחת יהיה זמן הנסיעה לפחות שלושת רבעי השעה?

10 אורך שיר אקראי המשודר ברדיו מתפלג נורמלית עם תוחלת של 3.5 דקות וסטית תקן של שלושים שניות.

- א. מה ההסתברות שאורך של שיר אקראי המנוגן ברדיו יהיה בין 3 ל-2.5 דקות?
- ב. מהו הטווח הבין רבעוני של אורך שיר המשודר ברדיו?
- ג. ביום מסוים מנוגנים 200 שירים ברדיו. כמה שירים מתוכם תצפה שיהיו באורך הנמוך מ-3.5 דקות?
- ד. בשעה מסוימת שודרו 8 שירים. מה ההסתברות שרבע מהם בדיוק היו ארוכים מ-4 דקות והיתר לא?

תשובות סופיות:

ה. 50%	ד. 50%	ג. 0	ב. 2.28%	א. 89.25%	(1
	ד. 68.26%	ג. 0%	ב. 3.76%	א. 0%	(2
	ד. 0.383	ג. 39.44%	ב. 89.44%	א. 26.43%	(3
				ה. 100%	
		ג. 2787.2	ב. 3958	א. 3812.8	(4
	ד. 87.4	ג. 112.6	ב. 80.8	א. 119.2	(5
		ג. 453.48	ב. 532.9	א. 500	(6
	ד. 733	ג. 0.3446	ב. 100	א. 500	(7
				ה. 267	
		ג. 1	ב. בממוצע.	א. 3	(8
	ד. 0.3975	ג. 0.8563	ב. 0.0228	א. 0.1587	(9
	ד. 0.25	ג. 100	ב. 0.675	א. 0.1359	(10

יסודות הסטטיסטיקה א

פרק 18 - תרגול טענות על כל החומר

תוכן העניינים

1. כללי.....69

תרגול טענות:

שאלות:

להלן מספר טענות.

ציינו לגבי כל טענה נכון/לא נכון ונמקו (תשובה ללא נימוק לא תתקבל).

- (1) בסדרה שבה כל התצפיות שוות זו לזו, השונות הינה 0.
- (2) ציון התקן של החציון תמיד יהיה 0.
- (3) ציון התקן של האחוזון ה-70 בהתפלגות אסימטרית ימנית (חיובית) תמיד יהיה חיובי.
- (4) אם נוסיף תצפיות לסדרה של תצפיות, הדבר בהכרח יגדיל את הממוצע של הסדרה.
- (5) בסדרה החציון הינו 80. הוספו שתי תצפיות, אחת 79 ואחת 100, לכן החציון יגדל.
- (6) אם נוסיף את הערך 4 לכל התצפיות אז סטיית התקן לא תשתנה.
- (7) אם נחלק את כל התצפיות בהתפלגות ב-2 אז השונות תקטן פי 2.
- (8) אם נגדיל את ממוצע המשכורות של עובדים בחברה אז גם השונות תגדל.
- (9) מתווך דירות המיר מחירי דירות מדולר לשקל. נניח שדולר אחד הוא 3.5 ש"ח. אם מתווך הדירות יחשב את מדד הקשר של פירסון בין מחיר הדירה בשקלים למחיר הדירה בדולרים הוא יקבל 1.
- (10) לסדרה של נתונים התקבל: $\bar{X} = \bar{Y} = 6$, $S_x = S_y = 1$, לכן מדד הקשר של פירסון יהיה 1.
- (11) בסדרה המונה 13 תצפיות, ידוע כי הממוצע הוא 40 והשונות היא 100. מוסיפים שתי תצפיות חדשות, שהן 35 ו-45. כתוצאה מכך, הממוצע בסדרה החדשה (הכוללת 15 תצפיות) יקטן והשונות תקטן.

- 12** לסדרה סטטיסטית בת 61 תצפיות הממוצע 120 והחציון 110.
 לסדרה זו הוסיפו עוד שתי תצפיות: 100, 140.
 בעקבות כך, הממוצע והחציון של הסדרה בת 63 התצפיות אינם משתנים.
- 13** לסדרה סטטיסטית בת 100 תצפיות הממוצע 75 וסטיית התקן 10.
 נוספו לסדרה זו עוד 2 תצפיות: 75; 75. כתוצאה מכך, הממוצע החדש
 (של 103 התצפיות) לא ישתנה, אך סטיית התקן תקטן.
- 14** לסדרת נתונים המונה 10 תצפיות ממוצע 25 וסטיית תקן 2.
 נתון כי הסדרה סימטרית סביב הממוצע. בשלב מאוחר יותר נוספו שלוש
 תצפיות לסדרה: 23, 25 ו-27. לכן סטיית התקן של 13 התצפיות לא תשתנה.
- 15** בהתפלגות אסימטרית חיובית, הערך המתאים למאון ה-30, ציון התקן שלו
 בהכרח שלילי.
- 16** סטיית התקן של סדרת נתונים תמיד תגדל אם נוסיף גודל קבוע לכל נתוני
 הסדרה.
- 17** בוצע מחקר על מספר העובדים בחברות מזון לעומת חברות תקשורת.
 החציון והממוצע בשתייהן שווה 8. לכן גם השכיח שווה בין שתי החברות.
- 18** הציון הממוצע של 5 סטודנטים הוא 78. 4 סטודנטים מתוכם קיבלו את
 הציונים הבאים: 74, 72, 86, 70. הציון של הסטודנט החמישי הוא: 76.
- 19** חושב מקדם המתאם של ספירמן בין שני משתנים והתקבל 1 לכן אם יחושב
 מדד הקשר של פירסון יתקבל גם כן 1.
- 20** חושב מקדם המתאם של פירסון בין שני משתנים והתקבל 1 אם יחושב מדד
 הקשר של ספירמן יתקבל גם 1.

תשובות סופיות:

- | | | | | |
|--------------|--------------|--------------|--------------|--------------|
| 1) נכון. | 2) לא נכון. | 3) לא נכון. | 4) לא נכון. | 5) לא נכון. |
| 6) נכון. | 7) לא נכון. | 8) לא נכון. | 9) נכון. | 10) לא נכון. |
| 11) לא נכון. | 12) נכון. | 13) נכון. | 14) לא נכון. | 15) נכון. |
| 16) לא נכון. | 17) לא נכון. | 18) לא נכון. | 19) לא נכון. | 20) נכון. |

יסודות הסטטיסטיקה א

פרק 19 - שאלות אמריקאיות על כל החומר

תוכן העניינים

1. כללי (ללא ספר)