סטטיסטיקה



תוכן העניינים

1. סטטיסטיקה תיאורית- סיווג משתנים וסולמות מדידה 1
2. סטטיסטיקה תיאורית- הצגה של נתונים 5
3. סטטיסטיקה תיאורית- סכימה 13
4. סטטיסטיקה תיאורית- מדדי מרכז ..... (ללא ספר)
5. סטטיסטיקה תיאורית - מדדי פיזור - הטווח, השונות וסטיית התקן 17
6. סטטיסטיקה תיאורית - מדדי פיזור - טווח בין רבעוני 20
7. סטטיסטיקה תיאורית- מדדי מיקום יחסי-ציון תקן 24
8. סטטיסטיקה תיאורית-אחוזונים בטבלה בדידה 26
9. סטטיסטיקה תיאורית שאלות אמריקאיות 28
10. התפלגויות רציפות מיוחדות - התפלגות נורמלית 34
11. הסקה סטטיסטית - הקדמה 42
12. ניתוח שונות חד כיוונית 45
13. מקדם המתאם (מדד קשר) הלינארי ומובהקותו 54
14. רגרסיה ליניארית ..... (ללא ספר)
15. מבחנים אפרמטרים למדגמים מזווגים 81

סטטיסטיקה

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 (אינטרוולי) – 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.

© כל הזכויות שמורות לברק קנדל

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).

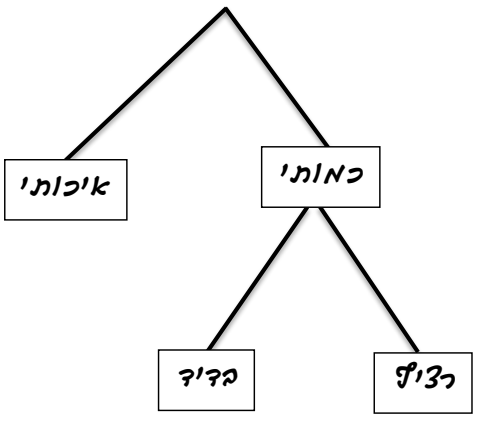
משתנה כמותי

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).



© כל הזכויות שמורות לברק קנדל

שאלות:

1) באיזה סולם מדידה המשתנים הבאים נחקרים (שמי/סדר/רווחים/מנה):

א. גובה (בס"מ).

ב. מספר ילדים למשפחה.

ג. מידת החרדה לפני מבחן.

ד. שביעות רצון משירות לקוחות בסקלה מ-1 עד 7 (1 - כלל לא מרוצה עד 7 - מרוצה מאד)

ה. השכלה.

ו. מספר אוטובוס.

ז. מקום מגורים.

ח. מין (1=גבר ; 2=אישה).

ט. 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.

א. מהו המשתנה הנחקר כאן?

ב. 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:

א. שכר ב-₪.

ב. ציון בחינת בגרות.

ג. תוצאה של הטלת קובייה.

ד. מהירות ריצה בתחרות.

ה. שיעור התמיכה בממשלה.

© כל הזכויות שמורות לברק קנדל

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.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה

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, 5, 3, 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	$\frac{f_3}{N}\cdot 100$
⋮	⋮	⋮
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.

© כל הזכויות שמורות לברק קנדל

דוגמה:

להלן התפלגות הציונים בכיתה מסוימת:

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

שכיחות מצטברת – צבירה של השכיחויות.

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

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

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

טבלת שכיחויות במחלקות:

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

דוגמה:

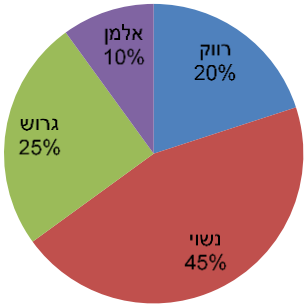
נתנו לקבוצת ילדים לבצע משימה, בדקו את התפלגות זמן הביצוע, בדקות.
להלן ההתפלגות שהתקבלה:

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

© כל הזכויות שמורות לברק קנדל

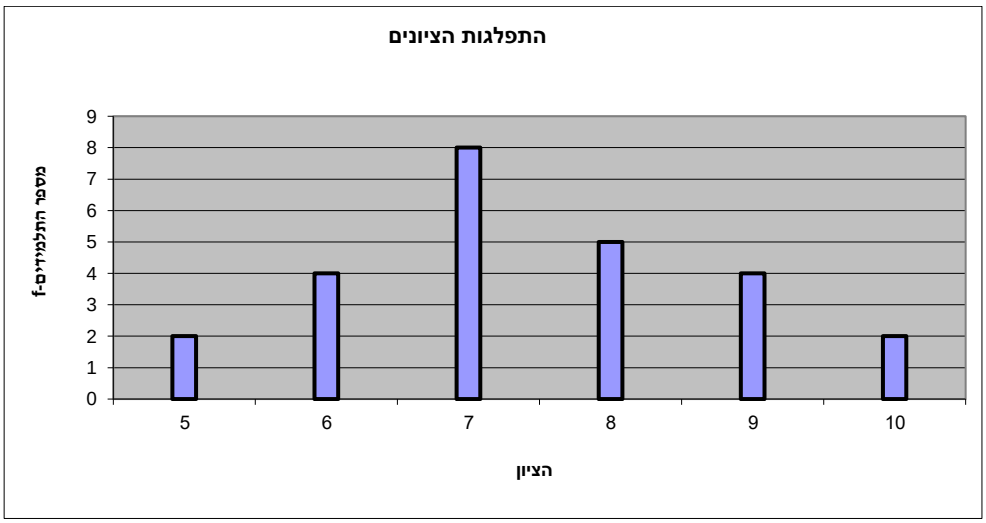
דיאגרמת עוגה:

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



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.

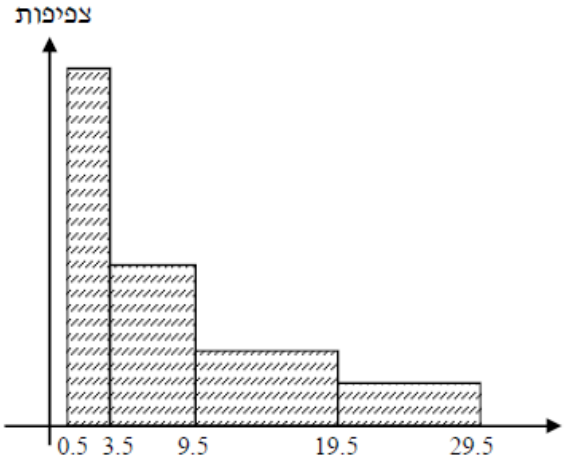


© כל הזכויות שמורות לברק קנדל

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

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



פוליגון – מצולעון:

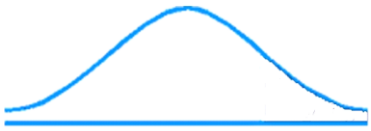
אם נחבר את אמצע קצה כל מלבן בקווים ישרים.
נותן מראה חזותי לצורה של התפלגות המשתנה.

© כל הזכויות שמורות לברק קנדל

צורות התפלגות נפוצות:

התפלגות סימטרית פעמונית

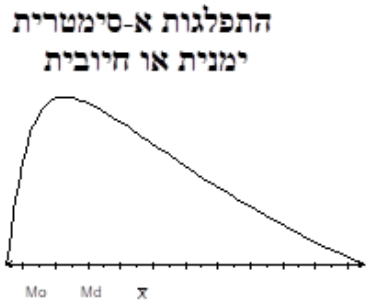
רוב התצפיות במרכז, וככל שנתרחק מהמרכז יהיו פחות תצפיות באופן סימטרי. לדוגמה, ציוני IQ.



ישנן התפלגויות סימטריות שאינן פעמוניות, כגון:

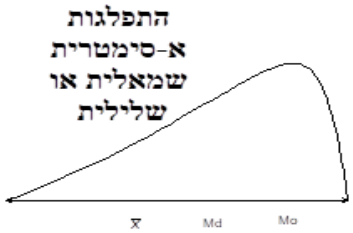
התפלגות אסימטרית ימנית (חיובית)

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



התפלגות אסימטרית שמאלית (שלילית)

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



© כל הזכויות שמורות לברק קנדל

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":

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

א. מהו המשתנה הנחקר?

ב. מהי פרופורציית התלמידים שמעדיפים תנ"ך?

3) להלן התפלגות ההשכלה במקום עבודה מסוים:

מספר העובדים	השכלה
60	נמוכה
120	תיכונית
20	אקדמאית

א. מהו המשתנה הנחקר?
מאיזה סולם הוא?

ב. תארו את הנתונים באופן גרפי.

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

© כל הזכויות שמורות לברק קנדל

תשובות סופיות:

1) א. להלן טבלה:

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

ב. עיין גרף מלא בסרטון הוידאו.

2) א. מקצוע מועדף.

ב. 19.6%.

3) א. משתנה נחקר: השכלה, סוג: סדר.

ב. עיין גרף מלא בסרטון הוידאו.

© כל הזכויות שמורות לברק קנדל

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

ב+ג. להלן טבלה:

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

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה

פרק 3 - סטטיסטיקה תיאורית- סכימה

תוכן העניינים

1. כללי 13

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).

© כל הזכויות שמורות לברק קנדל

שאלות:

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

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

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

© כל הזכויות שמורות לברק קנדל

2) נתון לוח ערכי המשתנים X_i ו- Y_i, כאשר: $i=1,2,...,6$, 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. $\left(\sum_{i=1}^{n} X_i \right)^2 = \sum_{i=1}^{n} X_i^{\,2}$.

© כל הזכויות שמורות לברק קנדל

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.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה

פרק 4 - סטטיסטיקה תיאורית- מדדי מרכז

תוכן העניינים

1. כללי ..... (ללא ספר)

סטטיסטיקה

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

תוכן העניינים

1. כללי 17

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.

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

דוגמאות:

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

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

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**

$$S_x^2 = \frac{\sum_x x^2 f(x)}{n} - \bar{x}^2 = \frac{1430}{25} - 7.44^2 = 1.8464$$
$$S = \sqrt{S_x^2} = \sqrt{1.8464} = 1.3588$$

כשיש מחלקות נעזר באמצע המחלקה כדי לחשב את השונות.

© כל הזכויות שמורות לברק קנדל

שאלות:

1) להלן רשימת הציונים של 20 תלמידים שנבחנו במבחן הבנת הנקרא:
6, 5, 8, 7, 6, 9, 8, 6, 7, 6, 7, 8, 5, 4, 6, 10, 9, 8, 6, 7.
חשבו את השונות, סטיית התקן והטווח של הציונים.

2) להלן התפלגות מספר המכוניות למשפחה בישוב "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

© כל הזכויות שמורות לברק קנדל

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.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה

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

תוכן העניינים

1. טווח בינרבעוני 20

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 ו75% מהמקרים גבוהים או שווים לו.
סימון: Q_1.

רבעון עליון – ערך שמחלק את ההתפלגות לשניים.
75% מהמקרים נמוכים ממנו או שווים לו ו25% מהמקרים גבוהים או שווים לו.
סימון: Q_3.

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

שלבים 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?

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

© כל הזכויות שמורות לברק קנדל

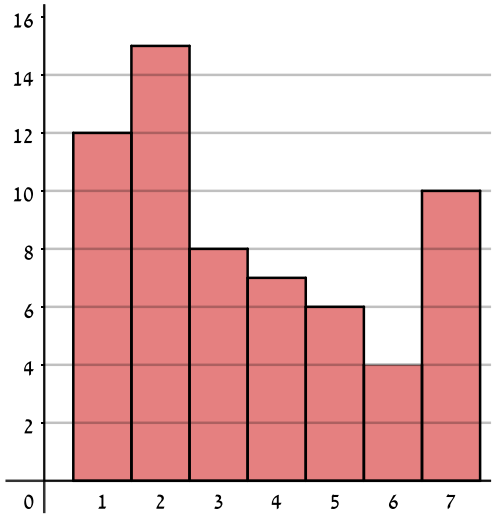
שאלות:

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?

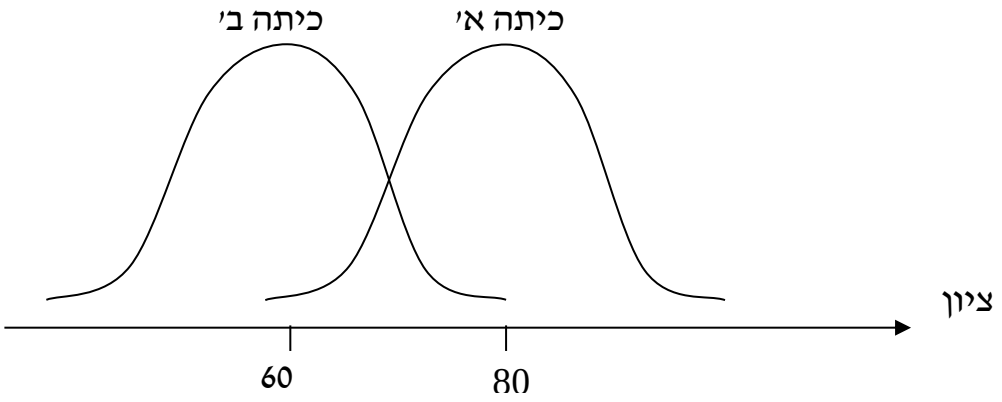
ב. מהו הטווח הבין-רבעוני של מספר ימי המחלה של המורים

ג. אם נוסיף 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.

© כל הזכויות שמורות לברק קנדל

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.

© כל הזכויות שמורות לברק קנדל

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'.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה

פרק 7 - סטטיסטיקה תיאורית- מדדי מיקום יחסי-ציון תקן

תוכן העניינים

1. כללי 24

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: $Z = \frac{X - \overline{X}}{S}$.

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?

© כל הזכויות שמורות לברק קנדל

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:

	תפוקה	מספר פועלים
ממוצע	48	15
סטיית תקן	10	2

באחד הימים מתוך כלל הימים שנבדקו התפוקה הייתה 50 מאות מצברים ובאותו היום עבדו 13 פועלים.
מה יותר חריג באותו היום, יחסית לשאר הימים שנבדקו: נתוני התפוקה או כמות הפועלים?

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.
ערן התגייס כשגובהו 180 ס"מ ומשקלו 59 ק"ג.

א. במה ערן חריג יותר ביחס לשאר המתגייסים, גובהו או משקלו?

ב. כמה ערן אמור לשקול כדי שמשקלו יהיה שקול לגובהו?

תשובות סופיות:

1) א. לשון. ב. 72.

2) ב'.

3) א. משקל. ב. 70.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה

פרק 8 - סטטיסטיקה תיאורית-אחוזונים בטבלה בדידה

תוכן העניינים

1. כללי 26

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:

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

א. מצאו את האחוזון ה-25.

ב. מצאו את הערך ש-20% מהמקרים מעליו.

© כל הזכויות שמורות לברק קנדל

שאלות:

1) להלן התפלגות של משתנה כלשהו:

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

מצאו להתפלגות את:

א. האחוזון ה-60.

ב. המאון ה-40.

ג. העשירון העליון.

ד. הטווח בין הרבעונים.

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

מספר מכוניות 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

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה

פרק 9 - סטטיסטיקה תיאורית שאלות אמריקאיות

תוכן העניינים

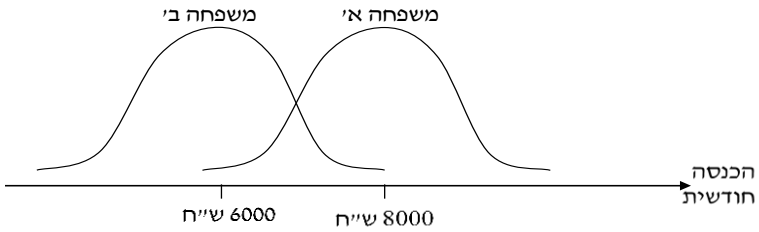
1. כללי 28

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?
א. משפחה א'.
ב. משפחה ב'.
ג. לשתיהן אותה הכנסה שכיחה.
ד. לא ניתן לדעת – אין מספיק נתונים.

2) באיזו משפחה ההכנסה החציונית שווה להכנסה הממוצעת?
א. משפחה א'.
ב. משפחה ב'.
ג. בשתיהן ההכנסה החציונית שווה להכנסה הממוצעת.
ד. לא ניתן לדעת – אין מספיק נתונים.

3) באיזו משפחה סטית התקן של ההכנסה החודשית גבוהה יותר?
א. משפחה א'.
ב. משפחה ב'.
ג. לשתיהן אותה סטית תקן.
ד. לא ניתן לדעת – אין מספיק נתונים.

© כל הזכויות שמורות לברק קנדל

הנתונים הבאים מתייחסים לשאלות 4-6:

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

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

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.

© כל הזכויות שמורות לברק קנדל

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.

9) 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 10-11:

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

10) התלמיד הטוב ביותר ביחס לנבחנים באותו מועד בו נבחן הוא:

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

11) 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.

© כל הזכויות שמורות לברק קנדל

הנתונים הבאים מתייחסים לשאלות 12-15:

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

12) מהו החציון של מספר הפגומים בקופסא:

א. 1.
ב. 2.
ג. 4.
ד. לא ניתן לדעת.

13) מהו הרבעון התחתון של מספר הפגומים בקופסא?

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

14) מספר הפגומים בקופסא הוא משתנה:

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

15) השכיח של מספר הפגומים בקופסא:

א. 63.
ב. 1.
ג. 200.
ד. לא ניתן לדעת.

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

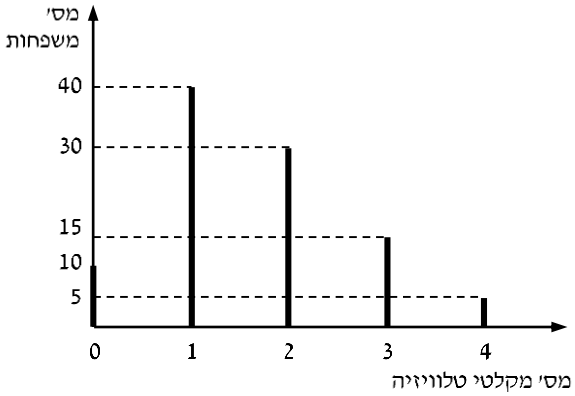
א. בערכים הגבוהים.
ב. בחלוקה זהה בין הערכים הגבוהים והנמוכים.
ג. בערכים הנמוכים.
ד. לא ניתן לדעת.
ה. אף לא תשובה מהנ"ל נכונה.

© כל הזכויות שמורות לברק קנדל

17) 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 18 עד 22:

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



18) המשתנה הנחקר כאן הוא:
א. משתנה שמי.
ב. משתנה מסולם סדר.
ג. משתנה כמותי בדיד.
ד. משתנה כמותי רציף.

19) הטווח של ההתפלגות הוא:
א. 35.
ב. 4.
ג. 3.
ד. 2.

© כל הזכויות שמורות לברק קנדל

20) ממוצע מספר מקלטי הטלוויזיה למשפחה הוא:

א. 1.65.
ב. 1.5.
ג. 1.
ד. 2.

21) השכיח של התפלגות זו היא:

א. 40.
ב. 1.5.
ג. 1.
ד. 2.

22) 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:

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

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה

פרק 10 - התפלגויות רציפות מיוחדות - התפלגות נורמלית

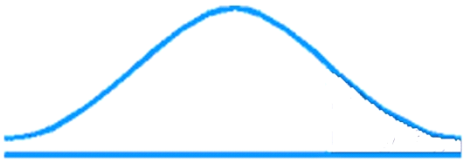
תוכן העניינים

1. כללי 34

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: $X \sim N\left(\mu,\sigma^2\right)$.

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

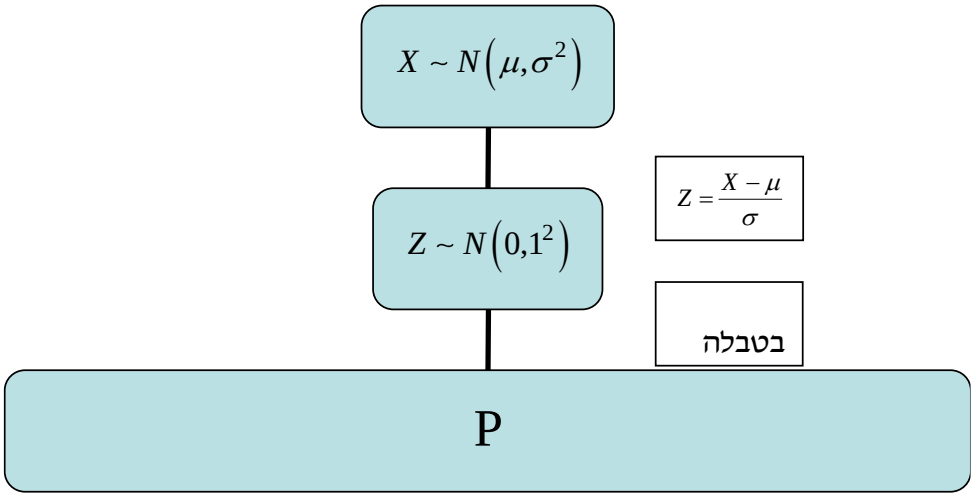
כדי לחשב הסתברויות בהתפלגות נורמלית יש לחשב את השטחים הרלוונטיים שמתחת לעקומה. כדי לחשב שטחים אלה נמיר כל התפלגות נורמלית להתפלגות נורמלית סטנדרטית על ידי תהליך הנקרא תקנון. התפלגות נורמלית סטנדרטית היא התפלגות נורמלית שהממוצע שלה הוא אפס וסטיית התקן היא אחת, והיא תסומן באות Z: $Z \sim N\left(0,1^2\right)$.

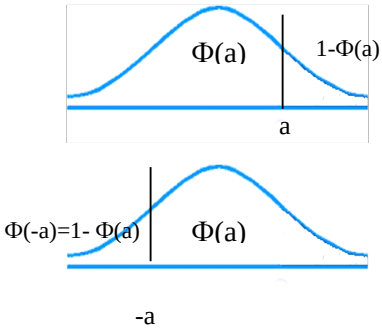
תהליך התקנון מבוצע על ידי הנוסחה הבאה: $Z=\frac{X-\mu}{\sigma}$.

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:

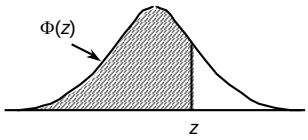
© כל הזכויות שמורות לברק קנדל





© כל הזכויות שמורות לברק קנדל

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
Φ(z)	0.90	0.95	0.975	0.99	0.995	0.999	0.9995	0.99995	0.999995

© כל הזכויות שמורות לברק קנדל

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 השוקלות מעל 110 גרם?

3) מה אחוז חפיסות השוקולד השוקלות מתחת ל 92 גרם?

4) מהו המשקל ש-90% מהחפיסות בקו הייצור שוקלים פחות מהם?

© כל הזכויות שמורות לברק קנדל

שאלות:

1) הגובה של אנשים באוכלוסייה מסוימת מתפלג נורמלית עם ממוצע של 170 ס״מ וסטית תקן של 10 ס״מ.

א. מה אחוז האנשים שגובהם מתחת ל-182.4 ס״מ?

ב. מה אחוז האנשים שגובהם מעל 190 ס״מ?

ג. מה אחוז האנשים שגובהם בדיוק 173.6 ס״מ?

ד. מה אחוז האנשים שגובהם מתחת ל-170 ס״מ?

ה. 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.

א. מהי פרופורציית המקרים בהן התרופה תעזור אחרי יותר משעה?

ב. מה אחוז מהמקרים שבהן התרופה תעזור בין 35 ל-37 דקות?

ג. מה הסיכוי שהתרופה תעזור בדיוק תוך 36 דקות?

ד. 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 וסטיית תקן של 8 ק״ג.

א. מה אחוז האנשים שמשקלם נמוך מ-55 ק״ג?

ב. מהי פרופורציית האנשים באוכלוסייה שמשקלם לפחות 50 ק״ג?

ג. מהי השכיחות היחסית של האנשים באוכלוסייה שמשקלם בין 60 ל-70 ק״ג?

ד. לאיזה חלק מהאוכלוסייה משקל הסוטה מהמשקל הממוצע בלא יותר מ-4 ק״ג?

ה. מה הסיכוי שאדם אקראי ישקול מתחת ל-140 ק״ג?

4) משקל תינוקות ביום היוולדם מתפלג נורמלית עם ממוצע של 3300 גרם וסטיית תקן 400 גרם.

א. מצאו את העשירון העליון.

ב. מצאו את האחוזון ה-95.

ג. מצאו את העשירון התחתון.

© כל הזכויות שמורות לברק קנדל

5) ציוני מבחן אינטליגנציה מתפלגים נורמלית עם ממוצע 100 ושונות 225.

א. מה העשירון העליון של הציונים במבחן האינטיליגנציה?

ב. מה העשירון התחתון של ההתפלגות?

ג. מהו הציון ש-20% מהנבחנים מקבלים מעליו?

ד. מהו האחוזון ה-20?

ה. 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.

א. מה ממוצע נפח משקה בבקבוק?

ב. 5% מהבקבוקים המיוצרים עם הנפח הגבוה ביותר נשלחים לבדיקה, החל מאיזה נפח שולחים בקבוק לבדיקה?

ג. 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.

א. מהו ממוצע אורך חיי מכשיר?

ב. מהי סטית בתקן של אורך חיי מכשיר?

ג. מה הסיכוי שמכשיר אקראי יחיה פחות מ-460 שעות?

ד. מהו המאיון העליון של אורח חיי מכשיר?

ה. 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.

א. לאיזו התפלגות הממוצע הגבוה ביותר?

ב. במה מבין המדדים הבאים התפלגות 1 ו-2 זהות?

i. בעשירון העליון.

ii. בממוצע.

iii. בשונות.

ג. לאיזו התפלגות סטיית התקן הקטנה ביותר?

i. 1.

ii. 2.

iii. 3.

iv. אין לדעת.



© כל הזכויות שמורות לברק קנדל

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 וסטיית 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?

© כל הזכויות שמורות לברק קנדל

תשובות סופיות:

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

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה

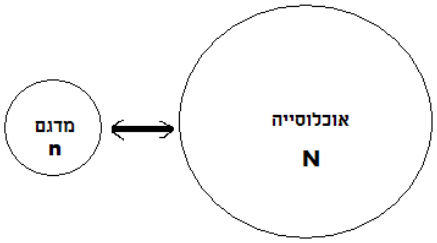
פרק 11 - הסקה סטטיסטית - הקדמה

תוכן העניינים

1. כללי 42

הסקה סטטיסטית – הקדמה:

רקע:



אוכלוסייה:
קבוצה שאליה מפנים שאלה מחקרית.
למשל, חברת תרופות שמעוניינת לפתח תרופה למחלת הסוכרת מתעניינת באוכלוסיית חולי הסוכרת בעולם.

מדגם:

חלק מתוך האוכלוסייה.
למשל, אם נדגום באקראי 10 אנשים מתוך חולי הסוכרת אז זהו מדגם מתוך אוכלוסיית חולי הסוכרת.

במקרים רבים אין אפשרות לחקור את כל האוכלוסייה כיוון שאין גישה לכולה, היא גדולה מידי, אנו מוגבלים בזמן ובאמצעים טכניים ולכן מבצעים מדגם במטרה לבצע הסקה סטטיסטית מהמדגם לאוכלוסייה.
הדגימה בקורס תהיה דגימה מקרית - הכוונה לדגימה שבה לכל תצפית באוכלוסייה יש את אותו סיכוי להיכלל במדגם.

סטטיסטי:

גודל המחושב על המדגם.

פרמטר:

גודל המתאר את האוכלוסייה.

© כל הזכויות שמורות לברק קנדל

הסימונים לפרמטר וסטטיסטי הם שונים:

	סטטיסטי (מדגם)	פרמטר (אוכלוסייה)
ממוצע	$\overline{X}$	μ
פרופורציה (שכיחות יחסית)	$\hat{p}$	P

פרמטר הוא גודל קבוע גם אם אנו לא יודעים אותו סטטיסטי הוא משתנה ממדגם למדגם ולכן יש לו התפלגות הנקראת התפלגות הדגימה.

דוגמה (פתרון בהקלטה):

25% מאזרחי המדינה תומכים בהצעת החוק של חבר כנסת מסוים.
הוחלט לדגום 200 אזרחים ומתוכם לבדוק מהו אחוז התומכים בהצעת החוק.

א. מי האוכלוסייה?

ב. מה המשתנה?

ג. מה הפרמטרים?

ד. מהו גודל המדגם?

ה. מהו הסטטיסטי שמתכננים להוציא מהמדגם?

ו. האם הפרמטר או הסטטיסטי הוא משתנה מקרי?

© כל הזכויות שמורות לברק קנדל

שאלות:

1) מתוך כלל הסטודנטים במכללה שסיימו סטטיסטיקה א נדגמו שני סטודנטים. נתון שממוצע הציונים של כלל הסטודנטים היה 78 עם סטיית תקן של 15.

א. מי האוכלוסייה?

ב. מה המשתנה?

ג. מהם הפרמטרים?

ד. מהו גודל המדגם?

2) להלן התפלגות מספר מקלטי הטלוויזיה למשפחה בישוב "העוגן".
נגדיר את X להיות מספר המקלטים של משפחה אקראית. מתכננים לדגום מאוכלוסייה זו 4 משפחות ולהתבונן בממוצע מספר מקלטי הטלוויזיה במדגם.

א. מיהי האוכלוסייה ומהו המשתנה הנחקר?

ב. מהו הסטטיסטי שיילקח מהמדגם ומה סימונו?

מספר המשפחות	מספר מקלטים
50	0
250	1
350	2
300	3
50	4
סך הכול $N=1000$	

3) נתון כי 20% מהשכירים במדינה הם אקדמאיים. נבחרו באקראי 10 שכירים באותה אוכלוסייה ומתכננים לפרסם את מספר האקדמאיים שנדגמו.

א. מהי האוכלוסייה?

ב. מה המשתנה באוכלוסייה?

ג. מהם הפרמטרים?

ד. מהו הסטטיסטי?

תשובות סופיות:

1) א. כלל הסטודנטים במכללה שסיימו סטטיסטיקה א. ב. ציון.
ג. ממוצע: 78, סטיית תקן: 15. ד. 2.

2) א. האוכלוסייה: 1000 משפחות בישוב העוגן, המשתנה הנחקר: מס' מקלטים.
ב. $\bar{X}$ = ממוצע מדגם.

3) א. השכירים במדינה. ב. השכלה: אקדמאי, לא אקדמאי.
ג. שיעור ההצלחות באוכלוסייה: 0.2. ג. מס' האקדמאים במדגם.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה

פרק 12 - ניתוח שונות חד כיוונית

תוכן העניינים

1. כללי 45

ניתוח שונות חד כיוונית

רקע תיאורטי

ניתוח שונות (חד כיווני) הוא מבחן להשוואת תוחלות $(\mu_{1..},\mu_k)$ של k אוכלוסיות שונות.
לכן, בניתוח שונות, <u>השערות</u> המחקר הן:

$H_0: \mu_1 = \mu_2 = \ldots = \mu_k$ (התוחלות של כל האוכלוסיות שוות)

$H_1:$ אחרת (לפחות שתיים מהתוחלות שונות)

ההנחות הדרושות לביצוע התהליך:

2) בכל אוכלוסייה מתוך k האוכלוסיות ההתפלגות נורמלית.
3) כל האוכלוסיות הן עם אותה שונות σ^2.
4) המדגמים בלתי תלויים זה בזה.

ישנו משתנה המבדיל בין הקבוצות השונות, הוא המשתנה הבלתי תלוי הנקרא גורם (factor). משתנה זה הוא קטגוריאלי עם k רמות (levels).
כדי לבצע את התהליך יש לבצע מדגם מכל אוכלוסייה:
נסמן ב- n_i את גודל המדגם בקבוצה i.

$n = \sum_{1}^{k} n_i$ - מספר התצפיות סך הכול (בכל המדגמים).

$\overline{X}_1$ - ממוצע המדגם הראשון,....., $\overline{X}_k$ -ממוצע המדגם ה- k -י.

$\overline{X}$ - ממוצע כללי (של כל המדגמים).

סכום ריבועים בין הקבוצות: $SS_B = \sum_{i=1}^{k} n_i \left[\overline{X}_i - \overline{X}\right]^2$

סכום ריבועים בתוך הקבוצות: $SS_W = \sum_{i=1}^{k} n_i \left[n_i - 1\right] \bullet \hat{S}_i^2$

סכום ריבועים כללי: $SS_T = \sum_{i=1}^{k}\sum_{j=1}^{n_j} [X_{ij} - \overline{X}]^2$

$$SST = SSB + SSW$$

יש למלא את טבלת ניתוח השונות הבאה:

© כל הזכויות שמורות לברק קנדל

מקור השונות	סכום הריבועים SS	דרגות חופש df	ממוצע הריבועים MS	F
בין הקבוצות - B	SSB	$k-1$	$\frac{SSB}{k-1}$	$\frac{MSB}{MSW}$
בתוך הקבוצות - W	SSW	$n-k$	$\frac{SSW}{n-k}$	
סה"כ - T	SST	$n-1$		

$$F=\frac{\frac{SSB}{k-1}}{\frac{SSW}{n-k}}\sim F\left(k-1,n-k\right)$$

אזור דחיית H_0 : $1-\alpha$: $F>F_{(k-1,n-k)}$

© כל הזכויות שמורות לברק קנדל

שאלות

1) מחקר מעוניין להשוות בין שלוש תרופות לשיכוך כאבים במטרה לבדוק האם קיים הבדל בין התרופות מבחינת הזמן בדקות שלוקח עד שהתרופה משפיעה. לצורך הבדיקה נלקחו 15 אנשים שסובלים מכאבי ראש. אנשים אלה חולקו באקראי לשלוש: קבוצה 1קיבלה "אקמול" קבוצה 2 קיבלה "אופטלגין" קבוצה 3 קיבלה "נורופן".
כל אדם במחקר מסר את מספר הדקות עד שהתרופה השפיעה עליו.

א. מהו המשתנה התלוי ומהו המשתנה הבלתי תלוי במחקר?
מהו ה"גורם" וכמה רמות יש לו?

ב. מהו המבחן הסטטיסטי המתאים כאן? רשמו את ההשערות.

ג. מה הן ההנחות הדרושות כדי לבצע את המבחן הסטטיסטי שהצעת בסעיף הקודם?

2) בעיר מסוימת שלושה בתי ספר תיכון. ראש העיר התעניין לבדוק האם קיים הבדל בהצלחה של בתי הספר במקצוע מתמטיקה. לצורך כך הוא דגם מספר תלמידים שנבחנו במבחן הבגרות במתמטיקה ברמה של 3 יחידות בעירו ובדק עבור כל תלמיד מה ציון הבגרות שלו במתמטיקה.
להלן הציונים שהתקבלו:

"המתמיד"	"רבין"	"הס"
78	98	85
65	62	83
70	55	74
90	80	85
56		75

א. מהו המבחן הסטטיסטי המתאים?
רשמו את ההשערות ואת ההנחות של המבחן.

ב. מהו גודל המדגם? מהו המשתנה הבלתי תלוי (factor) כמה רמות יש לו?

ג. חשבו את הממוצע ואת סטיית התקן של הציונים בכל אחד מהמדגמים.

ד. מלאו את טבלת ANOVA.

ה. רשמו את כלל ההכרעה למבחן שהוצע בסעיף א ברמת מובהקות של 5%.

ו. האם קיים הבדל בין בתי הספר בעיר מבחינת רמת הצלחת התלמידים במקצוע המתמטיקה? ענה על סמך הסעיפים הקודמים.

3) מעוניינים לבדוק האם יש הבדל בהשפעה של שיטות טפול שונות על לחץ הדם הסיסטולי (SBP) באוכלוסייה של קשישים. נבדקו 4 שיטות שונות.
בטבלה המצורפת מרוכזים ממצאי המחקר.

השיטה	A	B	C	D
גודל המדגם	12	14	8	12
הממוצע	178	172	180	182
סטיית התקן	4	8	5	3

א. רשמו את השערות המחקר וההנחות הדרושות כדי לבצע את המבחן המתאים.

ב. מה מסקנת המחקר ברמת מובהקות של 5%?

ג. האם יש צורך לבצע השוואות מרובות?

© כל הזכויות שמורות לברק קנדל

4) שלושה אופים נתבקשו להכין עוגת שוקולד. לכל אופה בדקו את משך זמן ההכנה בדקות. כל אופה נדרש לאפות בכל יום 4 עוגות.
האם קיים הבדל בין האופים מבחינת תוחלת זמני ההכנה של העוגות?
בדקו ברמת מובהקות של 5%.

האופה	ניר	מוזס	שלום
סכום הזמנים	206	212	182
סכום ריבועי הזמנים	10644	11250	8982

5) להלן טבלת ניתוח שונות חד כיוונית. במחקר בחנו 4 סוגי סוללות. רצו לבדוק האם לסוג הסוללה השפעה על תוחלת אורך החיים שלה. הפעילו את כל הסוללות על אותו מכשיר ובדקו את אורך החיים של כל סוללה בשעות.
מה המסקנה ברמת מובהקות של 10%? רשמו את ההשערות וההנחות הדרושות.

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	10.317	3	3.439	1.361	.279
Within Groups	60.648	24	2.527		
Total	70.964	27			

6) להלן טבלת ANOVA בטבלה הושמטו חלקים.
השלימו את החלקים בטבלה שהושמטו ומסומנים באותיות.

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	357.450	ב	ג	ה	.000
Within Groups	א	17	ד		
Total	522.950	19			

© כל הזכויות שמורות לברק קנדל

7) חברת תרופות לקחה 15 אנשים ברמת בריאות דומה. החברה חילקה את האנשים ל שלוש קבוצות שוות בגודלן. לכל קבוצה ניתנה אותה תרופה במינון שונה (dosage). המינונים שניתנו הם : 10 מ״ג, 20 מ״ג ו-30 מ״ג. לאחר שעה מזמן לקיחת התרופה ניבדק קצב פעימות הלב של כל אדם (pulse). הנתונים הוזנו לתוכנה סטטיסטית והתקבלו התוצאות הבאות:

ANOVA

pulse

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	414.400	2	207.200	19.733	.000
Within Groups	126.000	12	10.500		
Total	540.400	14			

pulse

Tukey HSD[a]

dosage	N	Subset for alpha = 0.05	
		1	2
30.00	5	71.0000	
20.00	5		80.2000
10.00	5		83.4000
Sig.		1.000	.299

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 5.000.

Post Hoc Tests

Multiple Comparisons

pulse
Tukey HSD

(I) dosage	(J) dosage	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
10.00	20.00	3.20000	2.04939	.299	-2.2675	8.6675
	30.00	12.40000*	2.04939	.000	6.9325	17.8675
20.00	10.00	-3.20000	2.04939	.299	-8.6675	2.2675
	30.00	9.20000*	2.04939	.002	3.7325	14.6675
30.00	10.00	-12.40000*	2.04939	.000	-17.8675	-6.9325
	20.00	-9.20000*	2.04939	.002	-14.6675	-3.7325

*. The mean difference is significant at the 0.05 level.

א. בדקו ברמת מובהקות של 5% האם קיים הבדל בין המינונים השונים מבחינת תוחלת הדופק של האנשים? רשמו את ההשערות וההנחות הדרושות לצורך פתרון.

ב. הסבירו ללא חישוב כיצד הייתה משתנה התשובה לסעיף הקודם אם הינו מעלים את הדופק של כל התצפיות במחקר ב-2.

ג. האם יש צורך במחקר בהשוואת מרובות. נמקו!

ד. לטבלת הANOVA צורפו טבלאות של השוואות מרובות בשיטה הנקראת ״טוקי״. ברמת בטחון של 95% מה הם הממצאים לפי שיטה זו?

© כל הזכויות שמורות לברק קנדל

8) בעיר מסוימת רצו לבדוק האם קיים הבדל ברמה של התלמידים בין בתי הספר השונים בעיר. ביצעו מדגם מכל בית ספר ונתנו מבחן זהה לכל הנדגמים. לאחר מכן ריכזו את הנתונים בתוכנה סטטיסטית והפעילו ניתוח שונות. מצורפים הפלטים שהתקבלו. ענו על הסעיפים הבאים:

א. כמה בתי ספר יש בעיר?

ב. כמה תלמידים השתתפו בסך הכול במחקר?

ג. האם קיים הבדל בין בתי הספר בעיר מבחינה רמת הציונים? בדקו ברמת מובהקות של 1%

ד. בביטחון של 95% אילו בתי ספר שונים זה מזה ברמת התלמידים? נמקו והסבירו.

Oneway

ANOVA

grade

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	7799.600	4	1949.900	13.586	.000
Within Groups	2870.400	20	143.520		
Total	10670.000	24			

© כל הזכויות שמורות לברק קנדל

Post Hoc Tests

Multiple Comparisons

grade

Scheffe

(I) school	(J) school	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1.00	2.00	5.40000	7.57681	.971	-20.2543	31.0543
	3.00	36.80000*	7.57681	.003	11.1457	62.4543
	4.00	36.40000*	7.57681	.003	10.7457	62.0543
	5.00	-2.60000	7.57681	.998	-28.2543	23.0543
2.00	1.00	-5.40000	7.57681	.971	-31.0543	20.2543
	3.00	31.40000*	7.57681	.011	5.7457	57.0543
	4.00	31.00000*	7.57681	.013	5.3457	56.6543
	5.00	-8.00000	7.57681	.888	-33.6543	17.6543
3.00	1.00	-36.80000*	7.57681	.003	-62.4543	-11.1457
	2.00	-31.40000*	7.57681	.011	-57.0543	-5.7457
	4.00	-.40000	7.57681	1.000	-26.0543	25.2543
	5.00	-39.40000*	7.57681	.001	-65.0543	-13.7457
4.00	1.00	-36.40000*	7.57681	.003	-62.0543	-10.7457
	2.00	-31.00000*	7.57681	.013	-56.6543	-5.3457
	3.00	.40000	7.57681	1.000	-25.2543	26.0543
	5.00	-39.00000*	7.57681	.001	-64.6543	-13.3457
5.00	1.00	2.60000	7.57681	.998	-23.0543	28.2543
	2.00	8.00000	7.57681	.888	-17.6543	33.6543
	3.00	39.40000*	7.57681	.001	13.7457	65.0543
	4.00	39.00000*	7.57681	.001	13.3457	64.6543

*. The mean difference is significant at the 0.05 level.

Homogeneous Subsets

grade

Scheffe[a]

school	N	Subset for alpha = 0.05	
		1	2
3.00	5	45.0000	
4.00	5	45.4000	
2.00	5		76.4000
1.00	5		81.8000
5.00	5		84.4000
Sig.		1.000	.888

Means for groups in homogeneous subsets are displayed.
a. Uses Harmonic Mean Sample Size = 5.000.

© כל הזכויות שמורות לברק קנדל

תשובות סופיות

1) א. משתנה בלתי תלוי: סוג התרופה.
משתנה תלוי: הזמן עד להשפעת התרופה בדקות.

ב. ניתוח שונות חד כיווני
$H_0: \mu_1 = \mu_2 = \mu_3$
$H_1: otherwise$

ג. 1. מדגמים בלתי תלויים.
2. שוויון שונויות.
3. משתנים מתפלגים נורמלית.

2) א. המבחן לניתוח שונות חד כיוונית.
$H_0: \mu_1 = \mu_2 = \mu_3$
$H_1: otherwise$
הנחות:
1. מדגמים בלתי תלויים.
2. משתנים מתפלגים נורמלית.
3. שוויון שונות.

ב. גודל המדגם: 14. משתנה ב"ת: בית הספר, בעל 3 רמות.

ג. $\bar{X} = 71.8, \hat{S} = 12.93$, $\bar{X} = 73.75, \hat{S} = 19.29$, $\bar{X} = 80.4, \hat{S} = 5.46$.

ד. להלן טבלה:

מקור השונות	SS	df	MS	F
B	200.6	2	100.3	
W	1904.75	11	173.2	
סה"כ	2105.35	13		0.58

ה. $F > 3.98$.

ו. נקבל את H_0 .

3) א. $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$
$H_1: otherwise$

ב. נדחה את H_0 .

ג. כן.

הנחות:
1. מדגמים בלתי תלויים.
2. שוויון שונויות.
3. משתנים מתפלגים נורמלית.

© כל הזכויות שמורות לברק קנדל

4) נקבל את H_0 : נכריע שאין הבדל מובהק בין האופים מבחינת תוחלת זמן הכנה.

5) $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$
$H_1 : otherwise$

הנחות:
1. מדגמים בלתי תלויים.
2. שוויון שונויות.
3. משתנים מתפלגים נורמלית.

נקבל את H_0 : לסוג סוללה אין השפעה של תוחלת החיים ברמת ביטחון של 10%.

6) להלן טבלה:

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	357.450 ב 2		178.725 ג	18.36 ה	.000
Within Groups	165.5 א	17	9.735 ד		
Total	522.950	19			

7) א. $H_0 : \mu_1 = \mu_2 = \mu_3$
$H_1 : otherwise$

הנחות:
1. מדגמים בלתי תלויים.
2. משתנים מתפלגים נורמלית.
3. שוויון שונות.

נדחה את H_0 : ברמת ביטחון של 5% קיים הבדל במינונים השונים מבחינת תוחלת הדופק.

ב. ראה וידאו. ג. כן. ד. $\mu_{20} = \mu_{10} > \mu_{30}$.

8) א. 5 ב. 25

ג. נדחה את H_0 : יש לפחות שני בתי ספר בעיר עם תוחלת רמת ציונים שונה.

ד. $(\mu_3 = \mu_4) < (\mu_1 = \mu_2 = \mu_3)$.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה

פרק 13 - מקדם המתאם (מדד קשר) הלינארי ומובהקותו

תוכן העניינים

1. מקדם המתאם הלנארי (פירסון) 54
2. חישוב מקדם המתאם הלנארי (פירסון) 65
3. בדיקת השערות על מקדם המתאם הלנארי 70
4. ניתוח פלטים על מקדם המתאם הלנארי 74

מקדם המתאם (מדד קשר) הלינארי ומובהקותו

מדד הקשר הלינארי (פירסון) – מבוא

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

דוגמה (פתרון בהקלטה) :

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

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

1) כמה תצפיות ישנן בדוגמה?

2) כמה משתנים ישנם בדוגמה, מי הם?

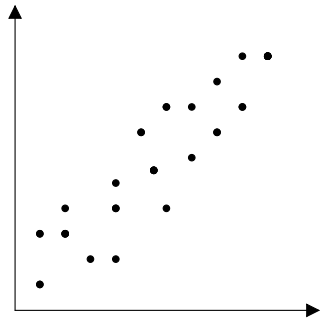
3) שרטטו לנתונים דיאגרמת פיזור.

4) מי המשתנה התלוי ומיהו המשתנה הבלתי תלוי?

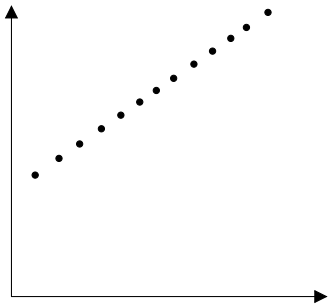
© כל הזכויות שמורות לברק קנדל

דיאגרמות פיזור לקשר בין משתנים וניתוחם

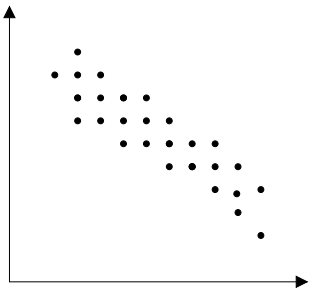
קשר לנארי חיובי חלקי



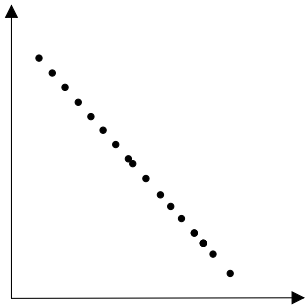
קשר לנארי חיובי מלא



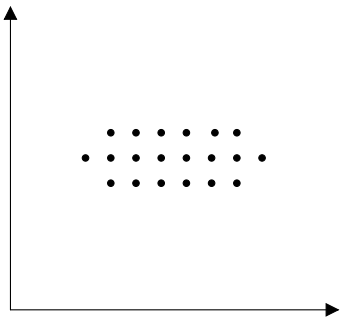
קשר לינארי שלילי חלקי



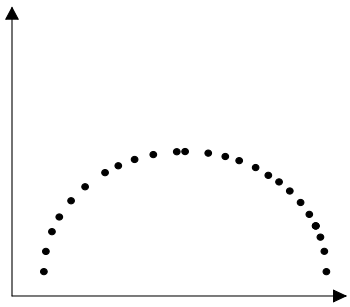
קשר לינארי שלילי מלא



אין קשר



אין קשר לינארי



משמעות מקדם המתאם:

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

מקדם מתאם זה מקבל ערכים בין 1 ל-1.

-1 0 1

© כל הזכויות שמורות לברק קנדל

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

מתאם חיובי מלא (מקדם מתאם 1):

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

מתאם חיובי חלקי:

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

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

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

r - מדד הקשר הלינארי במדגם שמהווה אומד לפרמטר ρ.

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

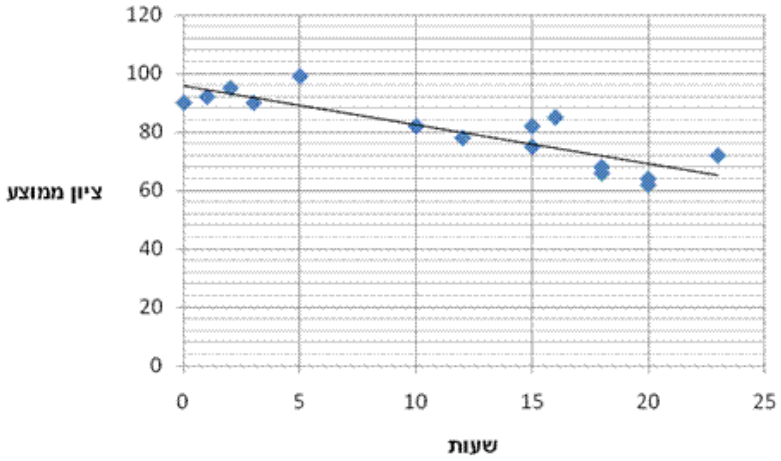
דוגמה (פתרון בהקלטה):

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

© כל הזכויות שמורות לברק קנדל

שאלות

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



א. מיהו המשתנה הבלתי תלוי?

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

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

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

א. השלימו את מקדמי המתאם שמסומנים בסימן שאלה בטבלה.

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

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

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

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

ב. שרטטו דיאגרמת פיזור לנתונים.

ג. מה ניתן לומר על הקשר בין המשתנים במדגם?

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

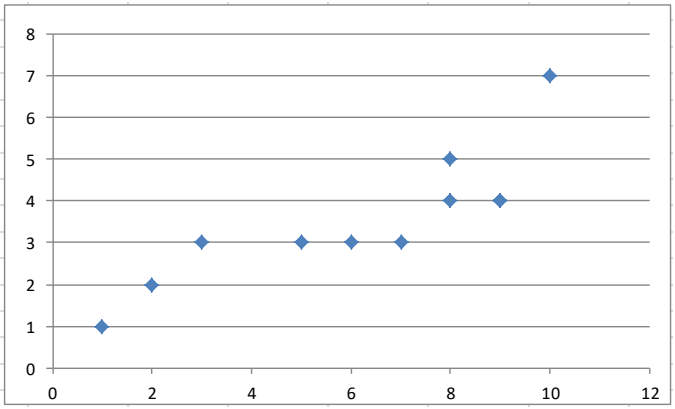
© כל הזכויות שמורות לברק קנדל

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

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

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

5) להלן דיאגרמת פיזור המראה קשר בין שני משנים:



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

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

שאלות רב ברירה (יש לבחור את התשובה הנכונה):

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

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

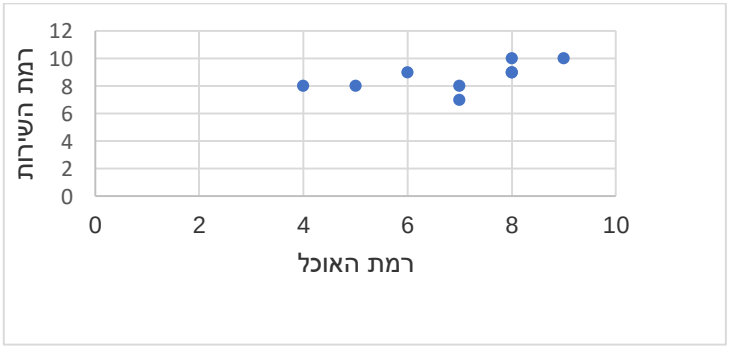
ב. במדגם, רוב הטמפרטורות בטורונטו היו שליליות.

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

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

© כל הזכויות שמורות לברק קנדל

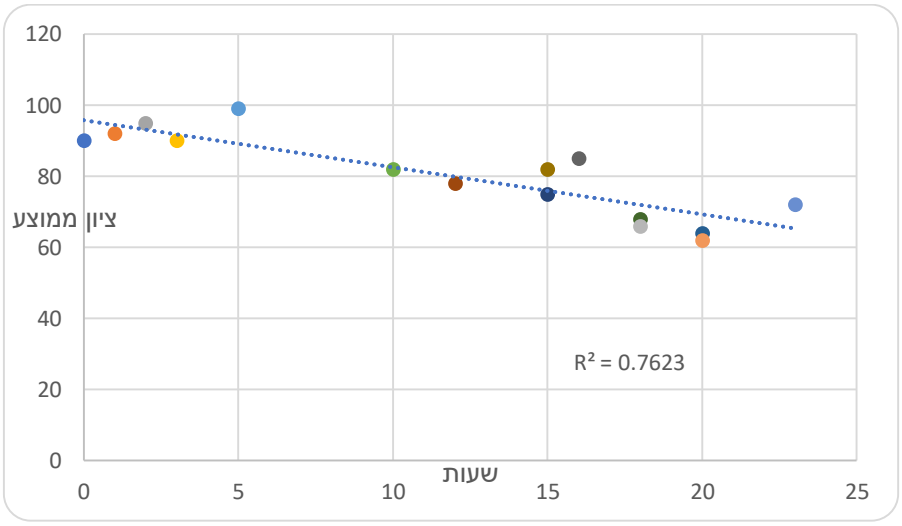
7) בסקר שביעות רצון שנערך בבית הקפה "פת לחם" התבקשו הלקוחות לדרג את מידת שביעות הרצון שלהם (בסולם 1-10) בשני נושאים: רמת האוכל ורמת השירות.



מה יהיה ערכו של מקדם המתאם (r)?

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

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

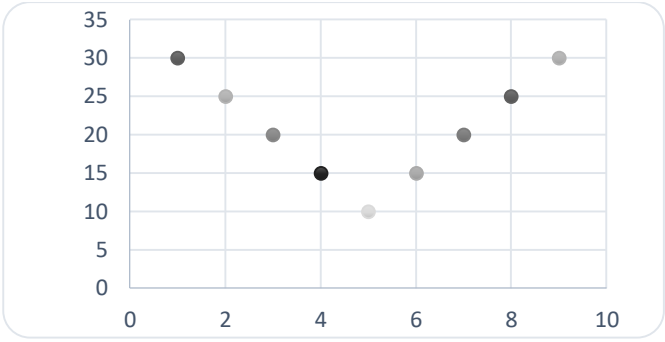


מה ניתן לומר על כיוון הקשר במדגם בין מספר שעות הבילוי השבועיות לבין הציון הממוצע של הסמסטר?

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

© כל הזכויות שמורות לברק קנדל

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



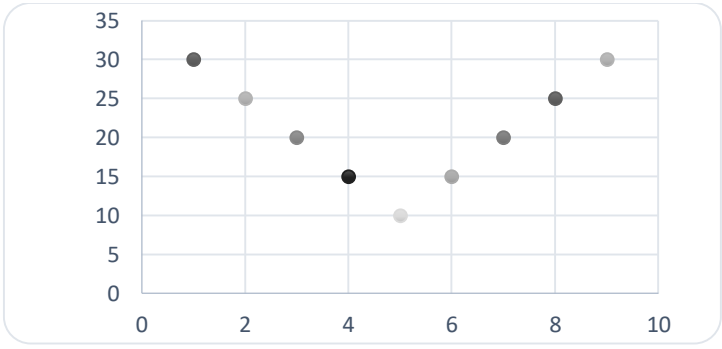
א. $r = 1$ היות ושני המשתנים יוצרים קוים ישרים.

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

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

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

10) התרשים הבא מתאר דיאגרמת פיזור.



איזו טענה נכונה?

א. בתרשים מוצג הקשר בין שני משתנים.

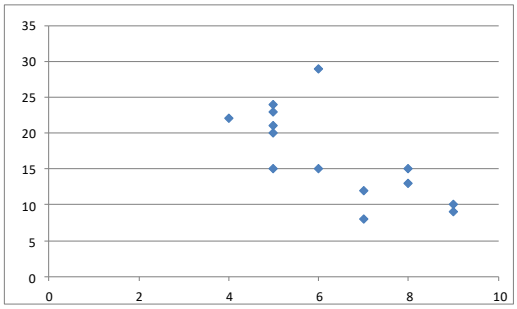
ב. בתרשים מוצג הקשר בין 9 משתנים.

ג. בתרשים מוצג הקשר בין 10 משתנים.

ד. אין לדעת כמה משתנים מוצגים בתרשים.

© כל הזכויות שמורות לברק קנדל

בגרף הבא מתוארת דיאגרמת פיזור של שני משתנים:



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

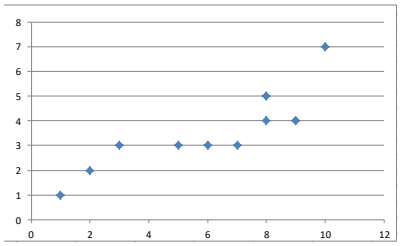
11) לאור הנתונים המופיעים בדיאגרמה, איזה מבין הערכים הבאים מתאים להיות התוצאה של r?

א. -0.52
ב. 0.72
ג. -0.72
ד. 0.52

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

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

13) להלן דיאגרמת פיזור:



מה יהיה מקדם המתאם בין שני המשתנים?

א. 1
ב. 0.85
ג. 0.15
ד. 0

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

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

© כל הזכויות שמורות לברק קנדל

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

א. חיובי

ב. שלילי

ג. אפסי

ד. לא ניתן לדעת.

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

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

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

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

א. -1

ב. 1

ג. -0.5

ד. 0.5

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

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

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

ג. לא קיים קשר בין שטח הדירה למחיר הדירה.

ד. מצב כזה שמתואר הנתונים לא אפשרי.

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

א. 1

ב. 0

ג. -1

ד. לא ניתן לדעת.

© כל הזכויות שמורות לברק קנדל

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

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

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

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

ד. אף אחד מהטענות שהוצגו אינה בהכרח נכונה.

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

א. 1

ב. -1

ג. 0

ד. אף אחת מהתשובות.

© כל הזכויות שמורות לברק קנדל

תשובות סופיות

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

2) א. להלן טבלה :

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

ב. ספורט ולשון.

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

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

5) א. חלקי, חיובי. ב. קטן.

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

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

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

© כל הזכויות שמורות לברק קנדל

מדדי קשר – מדד הקשר הלינארי (פירסון) – רקע

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

דוגמה:

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

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

דוגמה:

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

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

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

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

מתאם חיובי מלא (מקדם מתאם 1):

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

© כל הזכויות שמורות לברק קנדל

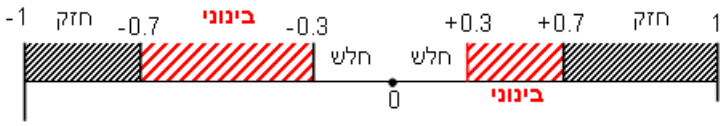
מתאם חיובי חלקי:

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

מתאם שלילי חלקי:

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

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



מקדם המתאם יסומן באות r.

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

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

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

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

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

© כל הזכויות שמורות לברק קנדל

שאלות

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

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

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

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

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

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

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

א. מה הממוצע של כל רמת הורמון?

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

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

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

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

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

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

ב. מה המשמעות של התוצאה שקיבלת בסעיף א?

© כל הזכויות שמורות לברק קנדל

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

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

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

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

חשבו את מדד הקשר הלינארי בין X ל- Y.

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

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

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

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

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

שאלות רב-ברירה:

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

א. הדבר מעיד שהציונים בכתה היו שליליים.

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

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

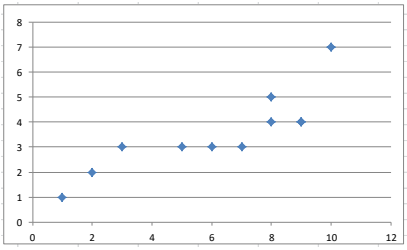
ד. אף אחת מהתשובות לא נכונה.

© כל הזכויות שמורות לברק קנדל

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

א. 1
ב. 0
ג. 4.2
ד. לא ניתן לדעת.

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



א. 1
ב. 0.85
ג. 0.15
ד. 0

תשובות סופיות

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

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

3) א. 0.8.

4) 0.8.

5) 1.

6) א. נכון. ב. לא נכון. ג. נכון.

7) ג'.

8) א'.

9) ב'.

© כל הזכויות שמורות לברק קנדל

בדיקת השערות על מקדם המתאם הלינארי – רקע

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

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

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

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

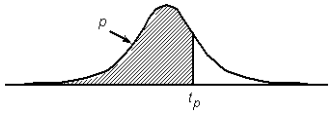
סטטיסטי זה מתפלג t עם $n-2$ דרגות חופש.

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

© כל הזכויות שמורות לברק קנדל

טבלת ערכים קריטיים של t - נספח: טבלת התפלגות T

P



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

© כל הזכויות שמורות לברק קנדל

שאלות

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

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

מקדם המתאם חושב והתקבל: 0.31-.

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

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

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

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

במדגם חושב מדד הקשר של פירסון להיות 0.7.

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

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

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

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

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

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

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

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

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

© כל הזכויות שמורות לברק קנדל

תשובות סופיות

1) א. לא נדחה את H_0 .
ב. לא תשתנה.

2) א. לא
ב. לא נדחה את H_0 .

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

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

© כל הזכויות שמורות לברק קנדל

מדד הקשר הלינארי – ניתוח פלטים – רקע

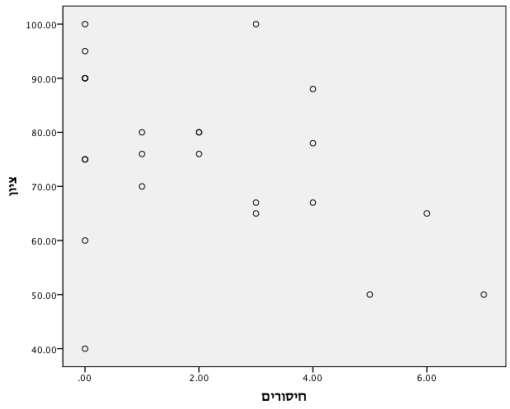
מדד הקשר הלינארי באוכלוסייה, שנקרא גם מקדם המתאם של פירסון או מדד הקשר של פירסון באוכלוסייה מסומן ב: ρ - פרמטר המאפיין את עוצמת הקשר הלינארי באוכלוסייה וכיוונו בין שני המשתנים הנחקרים. כאשר:
r - מדד הקשר הלינארי במדגם שמהווה אומד לפרמטר ρ.

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

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

דוגמה (פתרון בהקלטה):

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



א. מיהו המשתנה התלוי ומיהו המשתנה הבלתי תלוי במחקר?

ב. מה ניתן לראות לגבי הקשר הלינארי בין המשתנים שהתקבל במדגם?

Correlations

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

ג. מהו מקדם המתאם שהתקבל במדגם? מה המשמעות שלו?

ד. האם ניתן להגיד ברמת מובהקות של 5% שקיים מתאם לינארי שלילי בין מספר החיסורים של הסטודנטים מהקורס לבין הציון של הסטודנטים בקורס?

© כל הזכויות שמורות לברק קנדל

שאלות

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

Correlations

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

א. מהי אוכלוסיית המחקר?
ב. מהן השערות המחקר?
ג. מהו המשתנה הבלתי תלוי ומהו המשתנה התלוי במחקר?
ד. מהי מסקנת המחקר ברמת מובהקות של 5%?

© כל הזכויות שמורות לברק קנדל

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

Correlations

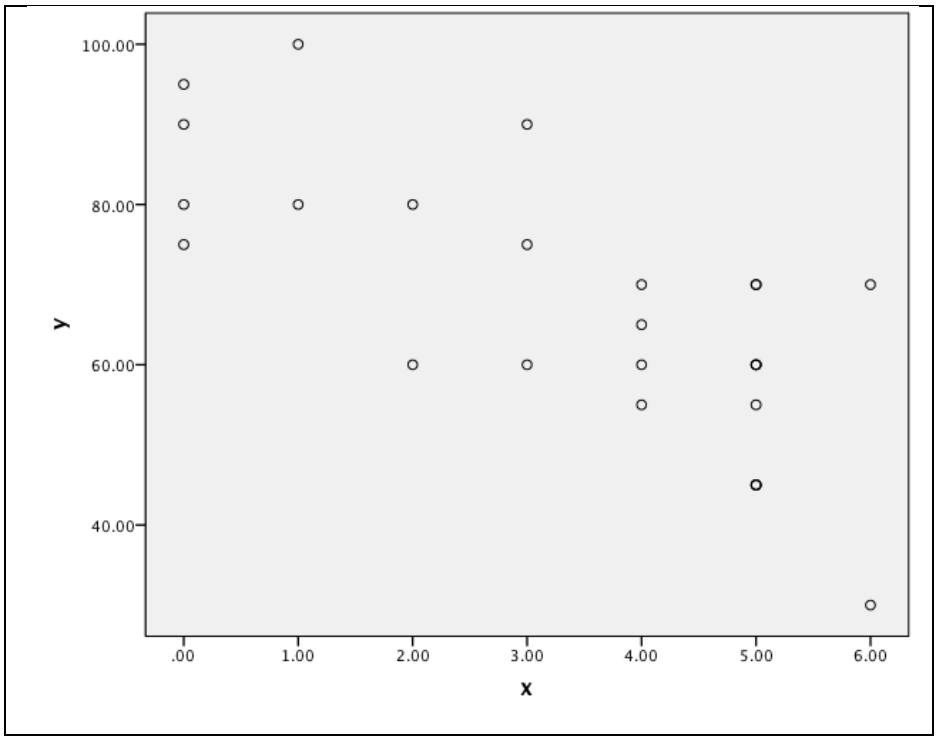
		x	y	z	w
x	Pearson Correlation	???	-.682	.134	.176
	Sig. (2-tailed)		.005	.634	.530
	N	15	15	15	15
y	Pearson Correlation	-.682	1	???	-.555
	Sig. (2-tailed)	.005		.544	.032
	N	15	15	15	15
z	Pearson Correlation	.134	.170	1	-.247
	Sig. (2-tailed)	???	.544		.374
	N	15	15	15	15
w	Pearson Correlation	.176	-.555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

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

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

ג. השלימו את המספרים המסומנים בפלט בסימני שאלה.

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



Correlations

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

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

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

© כל הזכויות שמורות לברק קנדל

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

Correlations

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

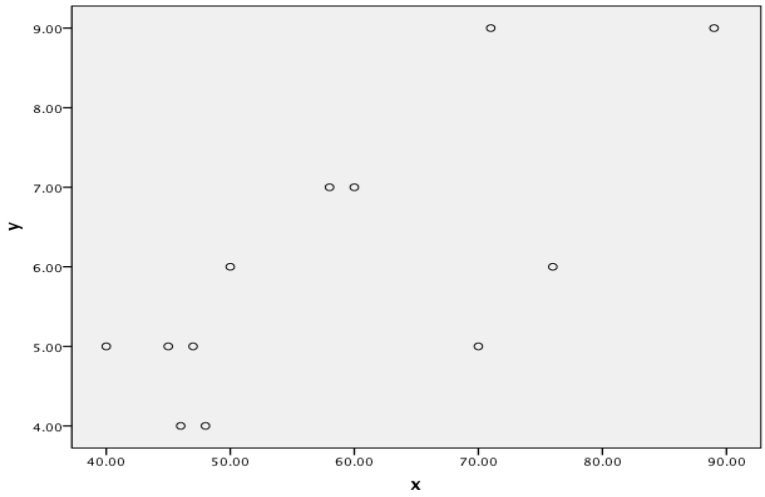
א. בדקו ברמת מובהקות של 5% האם קיים קשר לנארי חיובי בין המשקל והגובה.

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

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

© כל הזכויות שמורות לברק קנדל

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



Correlations

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

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

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

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

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

© כל הזכויות שמורות לברק קנדל

תשובות סופיות

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

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

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

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

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

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

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

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

ג. מקדם המתאם קטן ומובהקות התוצאה גדלה.

© כל הזכויות שמורות לברק קנדל

סטטיסטיקה

פרק 14 - רגרסיה ליניארית

תוכן העניינים

1. כללי ..... (ללא ספר)

סטטיסטיקה

פרק 15 - מבחנים אפרמטרים למדגמים מזווגים

תוכן העניינים

1. מבחן מקנמר 81

מבחן מקנמר – רקע

מבחן סטטיסטי זה נכנס לקטגוריית המבחנים האפרמטריים.
המבחן רלבנטי כשהמדגם מזווג וכשהמשתנה התלוי (הנחקר) הוא דיכוטומי, כלומר מקבל שני ערכים בלבד.
מספר התצפיות שחל בהן שינוי צריך להיות לפחות 20; אם תנאי זה לא מתקיים אפשר לבצע את מבחן הסימן במקום.

דוגמה:

50 איש נשאלו האם הם נוהגים לשלוח מסרונים בזמן הנהיגה. הם נשאלו בשני מצבים: פעם ראשונה לפני שצפו בסרטון ופעם שניה שבוע אחרי שצפו בסרטון.
התוצאות שהתקבלו היו:
13 אנשים טענו שהם נוהגים לשלוח מסרונים בזמן הנהיגה גם לפני הצפייה בסרטון וגם אחרי הצפייה בסרטון.
9 אנשים טענו שלפני הצפייה בסרטון הם לא שלחו מסרונים אבל אחרי הצפייה בסרטון הם כן שלחו מסרונים בזמן הנהיגה.
20 איש טענו שהם נהגו לשלוח מסרונים לפני הסרטון אך אחרי צפייה בסרטון הם הפסיקו לשלוח מסרונים בזמן הנהיגה.
ו-8 אנשים לא שלחו מסרונים בזמן הנהיגה לא לפני ולא אחרי צפייה בסרטון.
מהן השערות המחקר ומהו המבחן הסטטיסטי המתאים?

מבנה המבחן:

סטטיסטי המבחן הוא: $Z=\frac{B-C}{\sqrt{B+C}}$ כאשר B,C – שכיחויות שבהן חל שינוי ומתעלמים מהשכיחויות שלא חל בהן שינוי. הסטטיסטי מתפלג $Z\sim N(0,1)$.

דוגמה:

בדקו ברמת מובהקות של 5% האם לסרטון השפעה על שליחת המסרונים בזמן הנהיגה?

© כל הזכויות שמורות לברק קנדל

שאלות

1) פוליטיקאי הופיע אמש בתכנית טלוויזיה והוא מעוניין לבדוק האם התוכנית שיפרה את אמון הציבור בו. לצורך כך בוצע סקר שבו נשאל הצופה האם בעניו הפוליטיקאי נתפס כאמין לפני התוכנית והאם הוא נתפס כאמין לאחר התוכנית. להלן התוצאות שהתקבלו (המספרים מייצגים מספר צופים):
מה המסקנה ברמת מובהקות של 5%?

		לפני	
		אמין	לא אמין
אחרי	אמין	12	21
	לא אמין	7	17

2) חברת משקאות יצאה בקמפיין שנוי במחלוקת. החברה מעוניינת לבדוק האם הקמפיין השפיע על הרגלי הצריכה. במחקר שבו השתתפו 50 נשאלים 30 טענו שלא שינו את הרגלי הצריכה. 15 טענו שהחלו לרכוש את המשקה בעקבות הקמפיין ו-5 טענו שהפסיקו לרכוש את המשקה בעקבות הקמפיין.

א. מהי מובהקות התוצאה?

ב. מה מסקנתכם ברמת מובהקות של 2.5%?

תשובות סופיות

1) לא נדחה H_0.

2) א. 0.025 ב.נדחה H_0.

© כל הזכויות שמורות לברק קנדל