מבוא לסטטיסטיקה א



תוכן העניינים

1. סטטיסטיקה תיאורית- סיווג משתנים וסולמות מדידה 1
2. סטטיסטיקה תיאורית- הצגה של נתונים 5
3. סטטיסטיקה תיאורית-גבולות מדומים ואמיתיים 16
4. סטטיסטיקה תיאורית- סכימה 18
5. סטטיסטיקה תיאורית - מדדי מיקום מרכזי 22
6. סטטיסטיקה תיאורית - מדדי פיזור - הטווח, השונות וסטיית התקן 31
7. סטטיסטיקה תיאורית - מדדי פיזור - טווח בין רבעוני 34
8. סטטיסטיקה תיאורית-מדדי מיקום יחסי-ציון תקן 36
9. סטטיסטיקה תיאורית- מדדי מיקום יחסי-אחוזונים במחלקות 38
10. סטטיסטיקה תיאורית- אחוזונים בטבלה בדידה 41
11. סטטיסטיקה תיאורית - טרנספורמציה לינארית 43
12. סטטיסטיקה תיאורית - שאלות מסכמות 45
13. התפלגויות רציפות מיוחדות - התפלגות נורמלית 52
14. מדדי קשר - מדד הקשר של קרמר 59
15. מדד הקשר ספירמן 61
16. מדד הקשר פירסון 64
17. רגרסיה 70
18. מדדי קשר-רגרסיה -שונות מוסברת ושונות לא מוסברת 73
19. מדדי קשר - בחירת מדד מתאים 76
20. שאלות אמריקאיות מסכמות על כל חומר הלימוד ..... (ללא ספר)

מבוא לסטטיסטיקה א

פרק 1 - סטטיסטיקה תיאורית- סיווג משתנים וסולמות מדידה

תוכן העניינים

1. כללי 1

סטטיסטיקה תיאורית – סיווג משתנים וסולמות מדידה:

רקע:

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

מיון משתנים לפי סולמות המדידה:

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

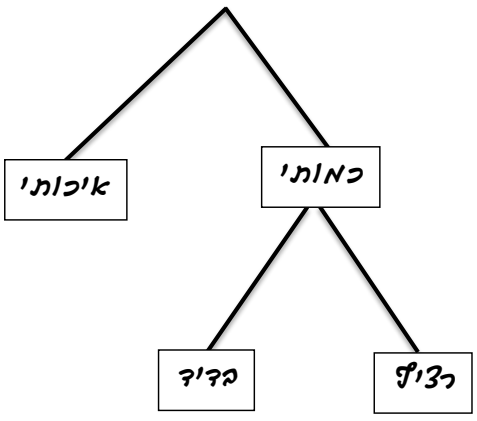
משתנה כמותי

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

מבוא לסטטיסטיקה א

פרק 2 - סטטיסטיקה תיאורית- הצגה של נתונים

תוכן העניינים

1. כללי 5

סטטיסטיקה תיאורית – הצגה של נתונים:

רקע:

דרכים להצגת נתונים שנאספו:

רשימה של תצפיות:

התצפית היא הערך שנצפה עבור ישות מסוימת בקבוצה. רושמים את התצפיות שהתקבלו כרשומה, יעיל שיש מספר מועט של תצפיות. ההצגה הזו רלבנטית לכל סוגי המשתנים. למשל, להלן מספר החדרים בבניין בן 5 דירות: 4, 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:

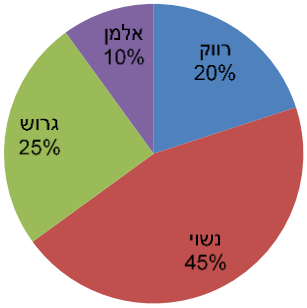
נתנו לקבוצת ילדים לבצע משימה, בדקו את התפלגות זמן הביצוע, בדקות.
להלן ההתפלגות שהתקבלה:

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

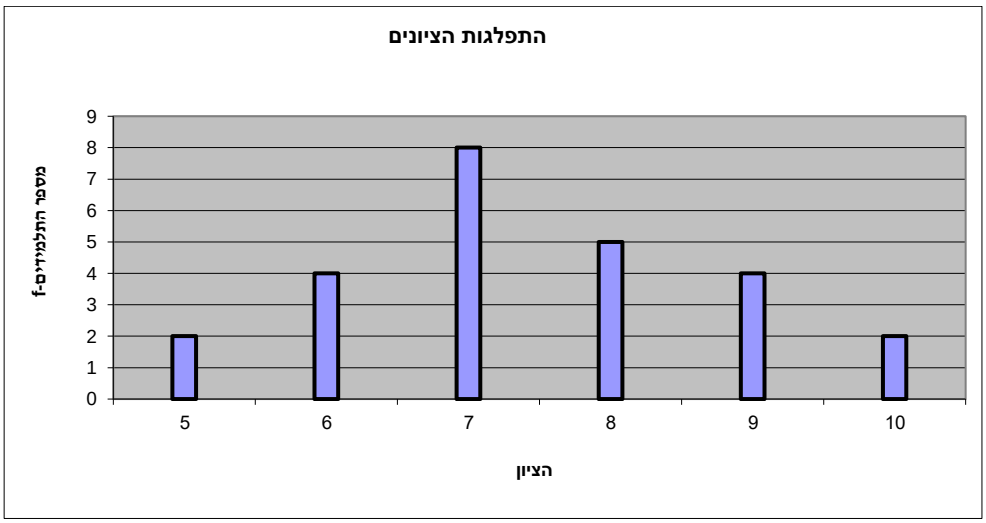
דיאגרמת עוגה:

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.

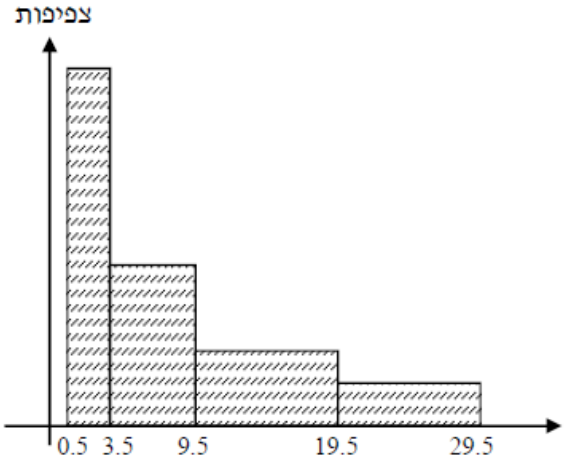


© כל הזכויות שמורות לברק קנדל

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:

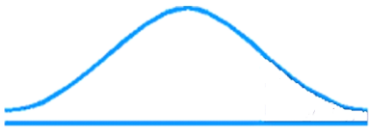
אם נחבר את אמצע קצה כל מלבן בקווים ישרים.
נותן מראה חזותי לצורה של התפלגות המשתנה.

© כל הזכויות שמורות לברק קנדל

צורות התפלגות נפוצות:

התפלגות סימטרית פעמונית

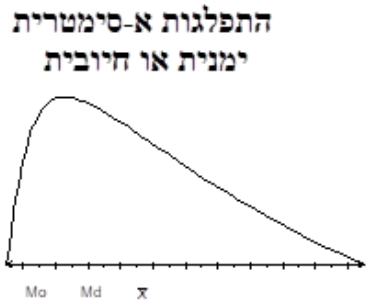
רוב התצפיות במרכז, וככל שנתרחק מהמרכז יהיו פחות תצפיות באופן סימטרי. לדוגמה, ציוני IQ.



ישנן התפלגויות סימטריות שאינן פעמוניות, כגון:

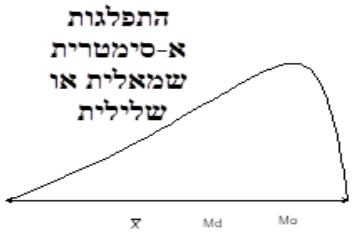
התפלגות אסימטרית ימנית (חיובית)

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?

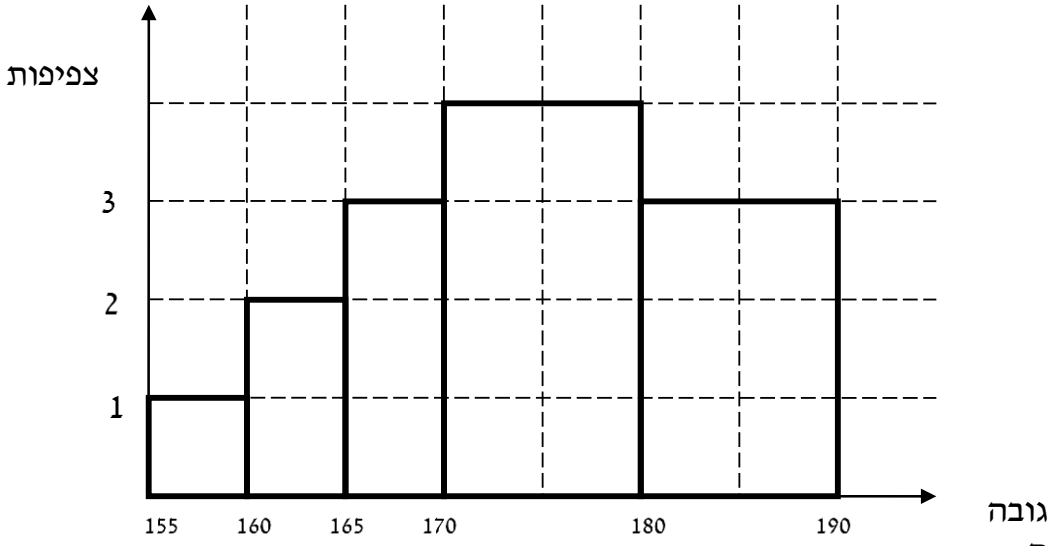
ב. תארו את הרשימה בטבלת שכיחויות.

ג. הוסיפו שכיחויות יחסיות לטבלה.

ד. תארו את הנתונים באופן גרפי.

© כל הזכויות שמורות לברק קנדל

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



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

6) להלן התפלגות המשקל של קבוצה מסוימת בק״ג:

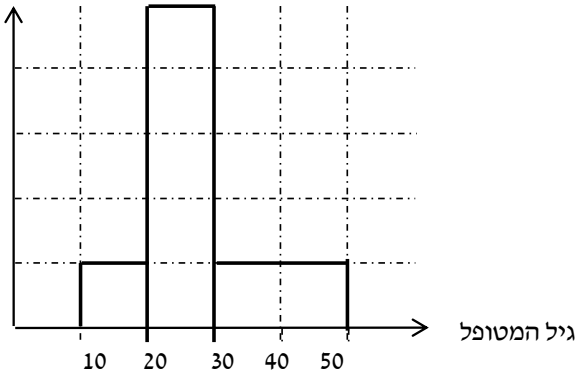
משקל	מספר מקרים
40-45	10
45-50	20
50-60	30
60-65	20
65-70	10

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

© כל הזכויות שמורות לברק קנדל

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





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

© 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%

5) א. גובה בס"מ, רציף.

ב+ג+ד. להלן טבלה:

ה. אסימטרית.

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

© כל הזכויות שמורות לברק קנדל

6) א. עיין גרף מלא בסרטון הוידאו.
ב. סימטרית.

7) א. המשתנה: גיל בשנים, משתנה רציף.
ב. המטופלים של ד"ר שוורץ.
ד. להלן טבלה:
ה. 62.5%.

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

© כל הזכויות שמורות לברק קנדל

מבוא לסטטיסטיקה א

פרק 3 - סטטיסטיקה תיאורית-גבולות מדומים ואמיתיים

תוכן העניינים

1. כללי 16

סטטיסטיקה תיאורית – גבולות מדומים וגבולות אמיתיים:

רקע:

עבור משתנה רציף נהוג לתאר את הנתונים בטבלת שכיחויות במחלקות.
הנתונים שנאספים הם ברמת דיוק מסוימת.
לדוגמה: משקל של בני אדם ומשקל של יהלומים ישקלו ברמת דיוק שונה.

גבולות מדומים:

כאשר גבול עליון של מחלקה אחת שונה מגבול תחתון של המחלקה הבאה אז הגבולות הם גבולות מדומים. כשהגבולות מדומים, ההפרש בין גבול תחתון של מחלקה לבין גבול עליון של המחלקה הקודמת יהיה רמת הדיוק.

רמת הדיוק חייבת להיות קבועה - אין אפשרות שחלק מהאנשים נדייק ברמה אחת ואת השאר ברמה אחרת. בגלל שהמשתנה הוא משתנה רציף, כשננתח את הנתונים נעבור מגבולות מדומים לגבולות אמתיים. אם הנתונים יינתנו בגבולות מדומים נהפוך אותם תמיד לגבולות אמיתיים.

כיצד עוברים מגבולות מדומים לגבולות אמיתיים?

לוקחים את רמת הדיוק ומחלקים אותה ב-2, ואת התוצאה המתקבלת מוסיפים לגבולות העליונים ומפחיתים מהגבולות התחתונים. אם יתנו נתונים בגבולות מדומים אנחנו מוכרחים לעבור לגבולות אמיתיים על מנת להמשיך ולנתח, אך אם הנתונים כבר יינתנו בגבולות אמיתיים נשאיר אותם כמו שהם.

דוגמה (פתרון בהקלטה):

להלן התפלגות הגבהים בס״מ של תלמידי כיתה ח׳:
יש להעביר את הנתונים לגבולות אמתיים.

$f(x)$	X
20	130-139
25	140-149
30	150-159
20	160-169
10	170-189

© כל הזכויות שמורות לברק קנדל

שאלות:

1) להלן התפלגות של משתנה בהצגה של מחלקות.
יש להעביר את הנתונים לגבולות אמתיים:

$f(x)$	X
542	500-590
32	600-690
154	700-790
254	800-890

2) להלן התפלגות המשקלים בק"ג של קבוצת אנשים מסוימת.
יש לרשום את הנתונים בגבולות אמיתיים:

משקל בק"ג	מספר אנשים
60-64	18
65-69	24
70-79	52
80-89	19

תשובות סופיות:

1) להלן טבלה:

$f(x)$	x
542	495-595
32	595-695
154	695-795
254	795-895

2) להלן טבלה:

$f(x)$	x
18	59.5-64.5
24	64.5-69.5
52	69.5-79.5
19	79.5-89.5

© כל הזכויות שמורות לברק קנדל

מבוא לסטטיסטיקה א

פרק 4 - סטטיסטיקה תיאורית- סכימה

תוכן העניינים

1. כללי 18

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

מבוא לסטטיסטיקה א

פרק 5 - סטטיסטיקה תיאורית - מדדי מיקום מרכזי

תוכן העניינים

1. כללי 22

סטטיסטיקה תיאורית – מדדי מיקום מרכזי:

רקע:

המטרה במדדי המיקום המרכזי היא למדוד את מרכז ההתפלגות של התצפיות.

השכיח – Mode:

השכיח הוא הערך הנפוץ ביותר בהתפלגות.

ברשימה

הערך החוזר על עצמו הכי הרבה פעמים: 6, 10, 4, 8, 4, 9, 7.

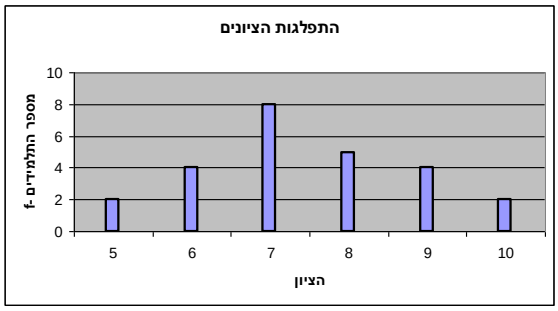
בטבלת שכיחויות בדידה

הערך שהשכיחות שלו היא הגבוהה ביותר.

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

בדיאגרמת מקלות

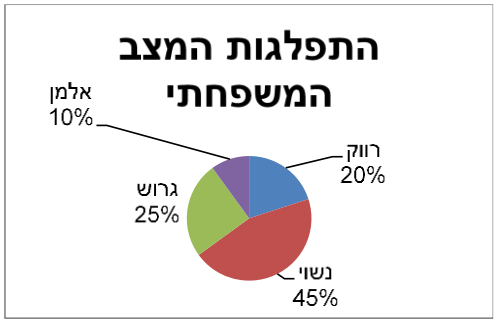
שיעור ה- X של המקל הגבוה ביותר.



© כל הזכויות שמורות לברק קנדל

בעוגה

הערך של הפלח הגדול ביותר.



בטבלת שכיחויות במחלקות

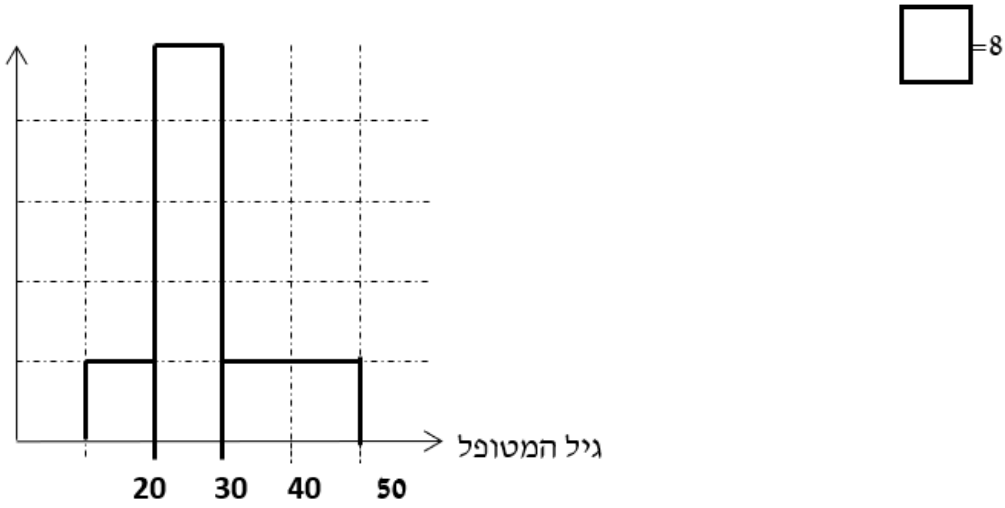
אמצע המחלקה עם הצפיפות הגבוהה ביותר.
לדוגמה, התפלגות הציונים בכיתה:

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

בהיסטוגרמה

שיעור ה- X של אמצע המחלקה הגבוהה ביותר.

לדוגמה, גיל המטופלים של ד"ר שוורץ בשנים:



© כל הזכויות שמורות לברק קנדל

כללי

יתכן שלהתפלגות יותר משכיח אחד.

השכיח הוא מדד הרלבנטי לכל סוגי המשתנים.

אמצע תחום (טווח) – Midrange:

הממוצע בין התצפית הגבוהה ביותר לתצפית הנמוכה ביותר:

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

החציון – Median:

החציון הוא ערך שמחצית מהתצפיות קטנות או שוות לו ומחצית מהתצפיות גדולות או שוות לו.

ברשימה

נסדר את התצפיות בסדר עולה.

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

אם יש מספר זוגי של איברים – החציון הוא ממוצע של האיבר ה- $\frac{n}{2}$,

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

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

בטבלת שכיחויות בדידה

נעשה תהליך דומה אך נעזר בשכיחות המצטברת.

דיאגרמת מקלות

נמיר לטבלת שכיחויות בדידה במטרה למצוא את החציון.

© כל הזכויות שמורות לברק קנדל

בטבלת שכיחויות במחלקות

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

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

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

$f\left(x_m\right)$ - השכיחות של המחלקה החציונית.

L_0 - גבול התחתון של המחלקה.

L_1 - גבול העליון של המחלקה.

היסטוגרמה

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

כללי

החציון אינו רלבנטי למשתנה מסולם שמי ולא רלבנטי למשתנה איכותי.

הממוצע – Average:

הממוצע הוא מרכז הכובד של ההתפלגות.

ברשימה

$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$.

בטבלת שכיחויות

$\overline{x} = \frac{\sum x \cdot f}{n}$.

© כל הזכויות שמורות לברק קנדל

במחלקות

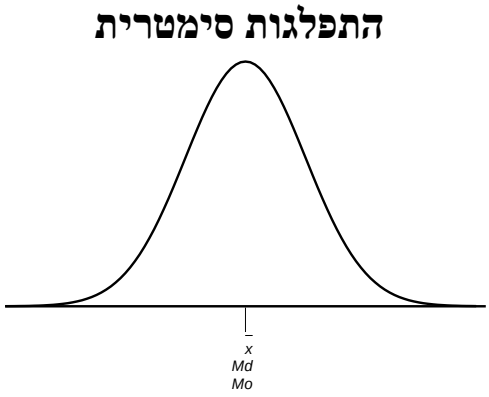
נשתמש באותה נוסחה רק נתייחס לאמצע המחלקה בתור ה- X .
הממוצע הזה יהיה ממוצע מקורב.

כללי

הממוצע רלבנטי רק למשתנה כמותי.

מדדי המיקום המרכזי בהתפלגויות המיוחדות:

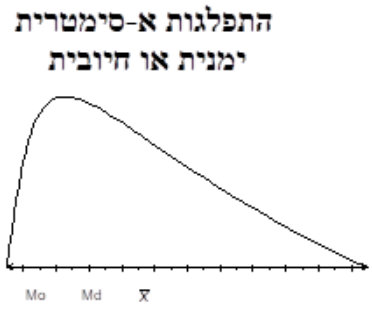
בהתפלגות סימטרית פעמונית כל מדדי המרכז שווים זה לזה:



בהתפלגות סימטרית השכיח לא חייב להיות במרכז:







© 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 החציון, השכיח, והממוצע של הציונים.

2) בדקו את מספר החדרים לדירה בבניין בן 5 דירות והתקבל ממוצע 3.8.
לגבי 4 דירות נמצא מספר חדרים: 5, 4, 3, 4.

א. כמה חדרים יש בדירה החמישית?

ב. מהו השכיח ומהו החציון?

3) להלן התפלגות מספר מקלטי הטלוויזיה שנספרו עבור כל משפחה בישוב מסוים:

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

א. חשבו את הממוצע, החציון והשכיח של ההתפלגות.

ב. הסבירו ללא חישוב כיצד כל מדד שחישבת בסעיף א' היה משתנה אם חלק מהמשפחות (לא כולן) שלא היה להם עד היום טלוויזיה היו רוכשים מקלט אחד.

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

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

א. כמה משפחות יש בישוב?

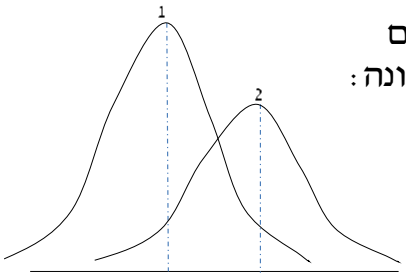
ב. מה אחוז המשפחות בישוב עם לכל היותר 2 מכוניות?

ג. חשבו את הממוצע, החציון והשכיח.

הקפידו להסביר לגבי כל סעיף מה משמעות התוצאה שקיבלתם!

© כל הזכויות שמורות לברק קנדל

5) מורה לימד 2 כיתות, הוא תיאר באותה מערכת צירים את התפלגות הציונים בכל כיתה. בחרו בתשובה הנכונה:



א. בכיתה 1 השכיח גבוה יותר מכיתה 2.

ב. בכיתה 2 השכיח גבוה יותר מכיתה 1.

ג. בשתי הכיתות אותו שכיח.

ד. לא ניתן לדעת באיזו כיתה השכיח גדול יותר.

6) ביישוב מסוים בדקו לכל משפחה את מספר הטלוויזיות שיש לה בבית. ביישוב גרות 200 משפחות. בממוצע יש למשפחה 1.5 טלוויזיות.

מספר טלוויזיות	מספר משפחות
0	28
1	62
2	
3	

א. השלימו את הטבלה.

ב. מהו השכיח, אמצע טווח והחציון.

ג. חלק מהמשפחות להן הייתה טלוויזיה אחת בדיוק הוציאו את הטלוויזיה מביתם. כיצד כל מדד ישתנה (יגדל, יקטן או לא ישתנה). הסבירו ללא חישוב.

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

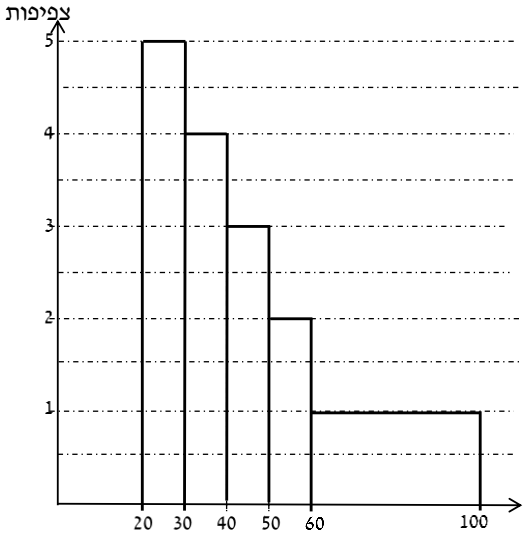
משקל	מספר מקרים
40-45	10
45-50	20
50-60	30
60-65	20
65-70	10

© כל הזכויות שמורות לברק קנדל

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

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

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



א. מצאו את השכיח בהתפלגות.

ב. מצאו את החציון בהתפלגות.

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

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

© כל הזכויות שמורות לברק קנדל

תשובות סופיות:

1) חציון: 7, שכיח: 6, ממוצע: 6.9.

2) א. 3. ב. שכיח: 3.4, חציון: 4.

3) א. ממוצע: 1.7, חציון: 1.5, שכיח: 1.
ב. הממוצע יגדל ויתר המדדים לא ישתנו.

4) א. 630. ב. 34.13%. ג. שכיח וחציון: 3, ממוצע: 2.952.

5) ב'.

6) א. להלן טבלה: ב. חציון: 2, שכיח: 2, אמצע טווח: 1.5.

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

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

7) חציון וממוצע: 55.

8) ממוצע: 172.6, חציון: 174.17, שכיח: 177.5.

9) א. 25. ב. 40. ג. גדול מהחציון.
ד. שכיח: לא ישתנה, חציון: יגדל, ממוצע: יגדל.

© כל הזכויות שמורות לברק קנדל

מבוא לסטטיסטיקה א

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

תוכן העניינים

1. כללי 31

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.

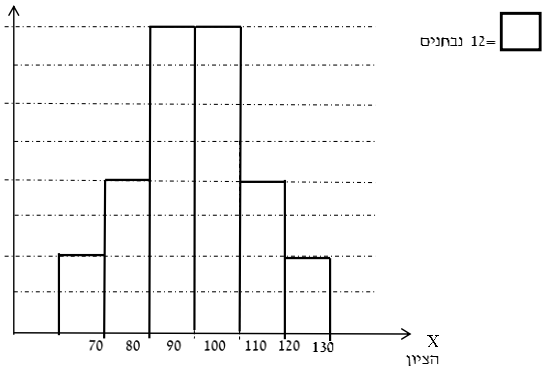
ג. חלק מבעלי הדירות בנות 2 החדרים הפכו את דירתם לדירת חדר. כיצד הדבר ישפיע (יקטין, יגדל, לא ישנה) על כל מדד שחישבתם בסעיפים הקודמים.

© כל הזכויות שמורות לברק קנדל

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

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

7) להלן התפלגות הציונים במבחן אינטליגנציה:



א. מה הממוצע ומה החציון של ההתפלגות?

ב. חשבו את סטיית התקן של הציונים.

ג. מסתבר שיש להוסיף 20 תצפיות לכל אחת משתי המחלקות 90-100 ו- 100-110. כיצד הדבר ישתנה את כל אחד מהמדדים של הסעיפים הקודמים?

תשובות סופיות:

1) שונות: 2.19, סטיית תקן: 1.48, טווח: 6.

2) א. סטיית תקן: 1.106. ב. טווח: 4.

3) א. ממוצע לא ישתנה, סטיית התקן תקטן.

ב. ממוצע לא ישתנה, סטיית התקן תגדל.

4) 10.8.

5) א. 3.05. ב. 1.16. ג. ממוצע: יקטן, סטיית התקן: תגדל.

6) 7.73.

7) א. 100. ב. 12.96. ג. ממוצע: לא ישתנה, סטיית תקן: תקטן.

© כל הזכויות שמורות לברק קנדל

מבוא לסטטיסטיקה א

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

תוכן העניינים

1. כללי 34

סטטיסטיקה תיאורית – מדדי פיזור – טווח בין רבעוני:

רקע:

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

שלבים במציאת טווח בין-רבעוני במחלקות:

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

שלב א:

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

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

נוסחאות הרבעונים יהיו:

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

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

נציב:

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

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

שלב ב:

נחסר את הרבעונים: $IQR = Q_3 - Q_1 = 11.33 - 4.07 = 7.26$ שניות

© כל הזכויות שמורות לברק קנדל

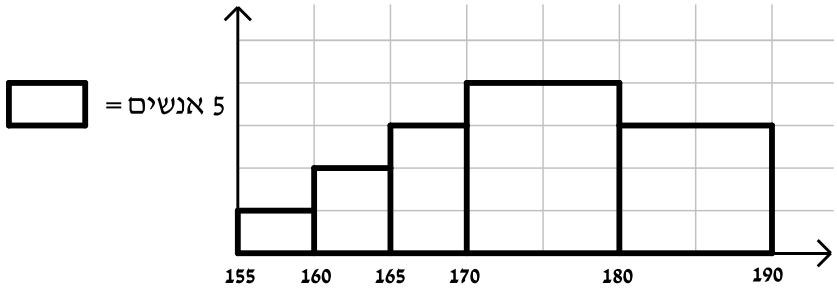
שאלות:

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

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

מצאו את הטווח הבין-רבעוני.

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



מצאו את הטווח הבין-רבעוני.

תשובות סופיות:

1) 13.75 ק"ג.

2) 13.33 ק"ג.

© כל הזכויות שמורות לברק קנדל

מבוא לסטטיסטיקה א

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

תוכן העניינים

1. כללי 36

סטטיסטיקה תיאורית – מדדי מיקום יחסי – ציון תקן:

רקע:

המטרה למדוד איך תצפית ממוקמת ביחס לשאר התצפיות בהתפלגות.

ציון תקן:

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

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

מבוא לסטטיסטיקה א

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

תוכן העניינים

1. כללי 38

סטטיסטיקה תיאורית – מדדי מיקום יחסי – אחוזונים במחלקות:

רקע:

האחוזון (המאון) ה- p הוא הערך בנתונים המחלק את הנתונים בצורה כזאת שעד אליו יש p % מהנתונים. מסמנים את האחוזון ה- p ב- X_p.
למשל, המאון ה-25 הוא האחוזון ה-25 או הרבעון התחתון:
ערך שרבע מהתצפיות קטנות ממנו והשאר גבוהות ממנו. מסומן: $X_{0.25}$.

מציאת מאון במחלקות:

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

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

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

$f\left(x_m\right)$ - השכיחות של המחלקה הרלבנטית.

L_0 - גבול התחתון של המחלקה.

L_1 - גבול העליון של המחלקה.

אם נרצה לחלץ את אחוז התצפיות שמתחת לערך מסוים נשתמש בנוסחה הבאה: $P_x = \left[\frac{\left(x - L_0\right)}{\left(L_1 - L_0\right)} \cdot f\left(x_m\right) + F\left(x_{m-1}\right)\right] \cdot \frac{100}{n}$.

דוגמה (פתרון בהקלטה):

להלן התפלגות השכר של עובדים בחברה מסוימת:

שכר ב-₪	
4000-6000	140
6000-10000	128
10000-15000	60
15000-20000	54
20000-40000	18

א. מצאו את המאון ה-40.

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

© כל הזכויות שמורות לברק קנדל

שאלות:

1) להלן התפלגות השכר (באלפי שקלים) בחברה:

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

א. חשבו את המאון ה-60.

ב. מהו העשירון העליון?

ג. 20% מהמשכורות הגבוהות ביותר הן משכורות של הבכירים, מהי המשכורת המינימאלית לבכיר?

ד. מה אחוז האנשים שמשתכרים מתחת ל-7,000 ₪?

ה. איזה אחוז מהעובדים משתכרים מעל ל-25,000 ₪?

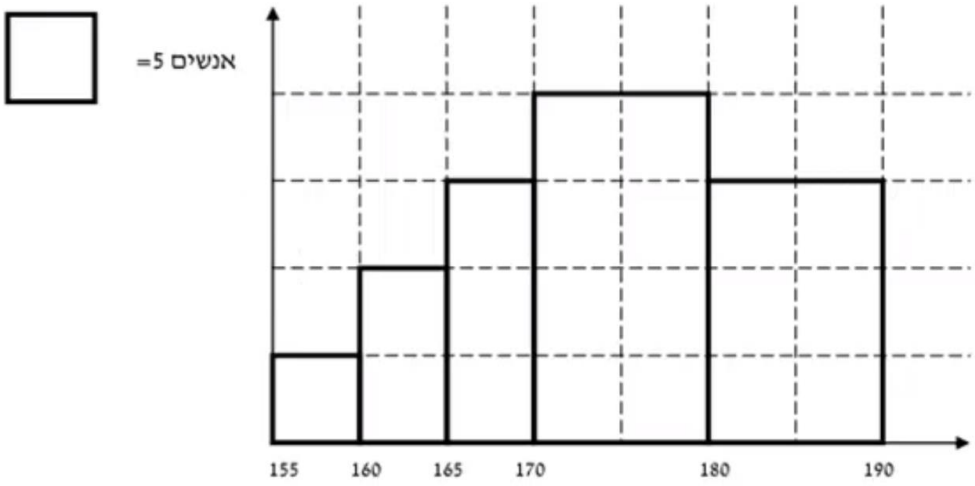
ו. איזה אחוז מהעובדים משתכרים בין 7,000 ₪ ל-25,000 ₪?

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

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

© כל הזכויות שמורות לברק קנדל

3) להלן היסטוגרמה המתארת את התפלגות הגבהים בס״מ של קבוצה מסוימת:



חשבו:

א. העשירון התחתון.

ב. האחוזון ה-30.

ג. הגובה ש-20% מהתצפית גדולות ממנו.

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

ה. את אחוז התצפיות מעל לגובה 185 ס״מ.

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

תשובות סופיות:

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

2) להלן טבלה:

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

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

© כל הזכויות שמורות לברק קנדל

מבוא לסטטיסטיקה א

פרק 10 - סטטיסטיקה תיאורית- אחוזונים בטבלה בדידה

תוכן העניינים

1. כללי 41

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

מבוא לסטטיסטיקה א

פרק 11 - סטטיסטיקה תיאורית - טרנספורמציה לינארית

תוכן העניינים

1. כללי 43

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:

מדדי המרכז:

$$MR_y = a \cdot MR + b$$
$$MO_y = a \cdot MO + b$$
$$\bar{y} = a \cdot \bar{x} + b$$
$$Md_y = a \cdot \leq Md_x + b$$

מדדי הפיזור:

$$R_y = |a| R_x$$
$$S_y = |a| S_x$$
$$S_y^{\,2} = a^2 S_x^{\,2}$$

מדדי המיקום היחסי:

$$Y_p = a \cdot X_p + b$$
$$Z_Y = \frac{a}{|a|} Z_x$$

שלבי העבודה:

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

דוגמה (פתרון בהקלטה):

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

© כל הזכויות שמורות לברק קנדל

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

3) במבחן 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.

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.

© כל הזכויות שמורות לברק קנדל

מבוא לסטטיסטיקה א

פרק 12 - סטטיסטיקה תיאורית - שאלות מסכמות

תוכן העניינים

1. כללי 45

סטטיסטיקה תיאורית – שאלות מסכמות:

שאלות:

1) בדקו עבור 5 תלמידים את המשקל שלהם:

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

א. מהו המשתנה הנחקר? בדיד או רציף?

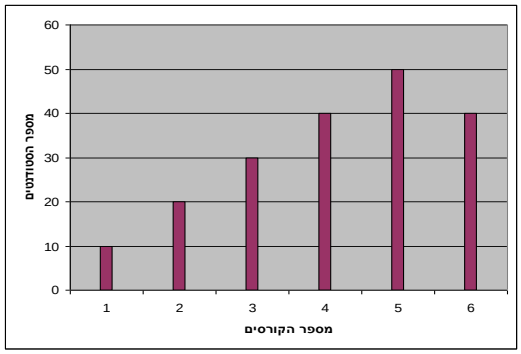
ב. מה המשקל החציוני, הממוצע והשכיח?

ג. מה הטווח וסטיית התקן של המשקל?

ד. לאותם תלמידים חישבו גם את הגובה בס"מ וקיבלו גובה ממוצע של 168 וסטיית תקן 6. במה תלמיד מספר 3, שגובהו 162, יותר חריג – במשקל או בגובה?

ה. הוסיפו עוד תלמיד השוקל 52 ק"ג בדיוק. הסבירו ללא חישוב כיצד הדבר ישפיע על הממוצע וסטיית התקן (יגדיל, יקטין או לא ישנה).

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



א. מה המשתנה הנחקר? בדיד או רציף?

ב. מהי צורת ההתפלגות?

ג. תארו את הנתונים בטבלת שכיחויות.

ד. חשבו את השכיח, החציון והטווח.

© כל הזכויות שמורות לברק קנדל

3) להלן התפלגות הציונים בבחינה בלשון שנעשתה עבור תלמידי כיתות ד'.
במחקר השתתפו 150 תלמידים. ממוצע הציונים שהתקבל: $\overline{X} = 7\frac{1}{15}$.

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

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

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

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

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

© כל הזכויות שמורות לברק קנדל

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

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

א. השלימו את הטבלה.

ב. חשבו את החציון, השכיח והממוצע של התפלגות מס' שעות העבודה בחברה.

ג. מה סטיית התקן של מספר שעות העבודה?

ד. מה העשירון העליון של ההתפלגות?

ה. איזה אחוז מהעובדים עובדים מעל 45 שעות בשבוע?

ו. מה ציון התקן של רינה, שעובדת 30 שעות בשבוע?

ז. כיצד ישתנה החציון, הממוצע וסטיית התקן אם מספר שעות העבודה המינימאלי אינו 10 אלא 15? הסבירו.

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

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

א. מה אחוז האנשים ששלחו פחות מ-80 מסרונים בחודש?

ב. מה אחוז האנשים ששלחו בין 50 ל-120 מסרונים?

ג. הוחלט להעניק מתנה עבור $\frac{1}{4}$ מהלקוחות שמשתמשים במספר הרב ביותר של מסרונים בחודש. החל מאיזה כמות של מסרונים תחולק המתנה?

ד. ציינו איזה מדד ניתן לחשב ואיזה לא ניתן. אם ניתן, חשבו:

i. ממוצע.

ii. שכיח.

iii. חציון.

iv. שונות.

© כל הזכויות שמורות לברק קנדל

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

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

א. שרטטו היסטוגרמה לתיאור התפלגות זמן ביצוע המשימה.

ב. מתוך ההיסטוגרמה שבנית בסעיף א', מהי צורת ההתפלגות?

ג. חשבו את השכיח והחציון של ההתפלגות.

ד. הסבירו, ללא חישוב, האם הזמן הממוצע לביצוע המשימה, קטן או גדול או שווה ביחס לשכיח ולחציון.

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

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

א. השלימו את הטבלה.

ב. מהו הממוצע והחציון של ההתפלגות?

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

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

© כל הזכויות שמורות לברק קנדל

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

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

ב. ציון התקן של החציון תמיד יהיה 0.

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

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

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

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

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

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

© כל הזכויות שמורות לברק קנדל

תשובות סופיות:

1) א. המשתנה הנחקר: משקל תלמיד בק"ג, משתנה כמותי רציף.

ב. $\bar{X}=52$, $Md=X_{\frac{n+1}{2}}=X_3=58$, שכיח: 58.

ג. $S=10.12$, $R=28$.

ד. הוא חריג יותר בגובה כי שם ציון התקן בערך מוחלט יותר גבוה.

ה. הממוצע לא ישתנה אך סטיית התקן תקטן.

2) א. מספר הקורסים, בדיד. ב. התפלגות אסימטרית שמאלית.

ג. להלן טבלה: ד. שכיח: 5, טווח: 5.

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

3) א. 20 תלמידים קיבלו ציון 6 ו-40 תלמידים קיבלו ציון 8.

ב. חציון: 7, שכיח: 8. ג. שונות: 2.533 , סטיית תקן: 1.592.

4) א. 20%.

ב. שביעות רצון (סדר).

ג. ii.

ד. טווח: 4, שכיח:2, חציון: 2.

5) א. להלן טבלה:

ב. חציון: 35, שכיח: 35, ממוצע: 35.

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

ג. סטיית תקן: 12.65.

ד. 53.333.

ה. 25%.

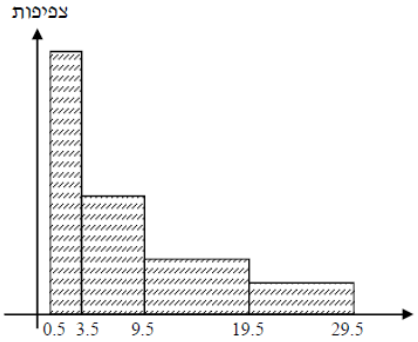
ו. -0.395.

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

6) א. 38% ב. 40% ג. 150 ד. חציון: 100.

© כל הזכויות שמורות לברק קנדל

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



ג. שכיח: 2, חציון: 6.83.

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

8) א. ראו טבלה:

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

ב. 100. ג. 104.

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

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

© כל הזכויות שמורות לברק קנדל

מבוא לסטטיסטיקה א

פרק 13 - התפלגויות רציפות מיוחדות - התפלגות נורמלית

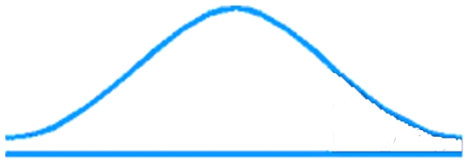
תוכן העניינים

1. כללי 52

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

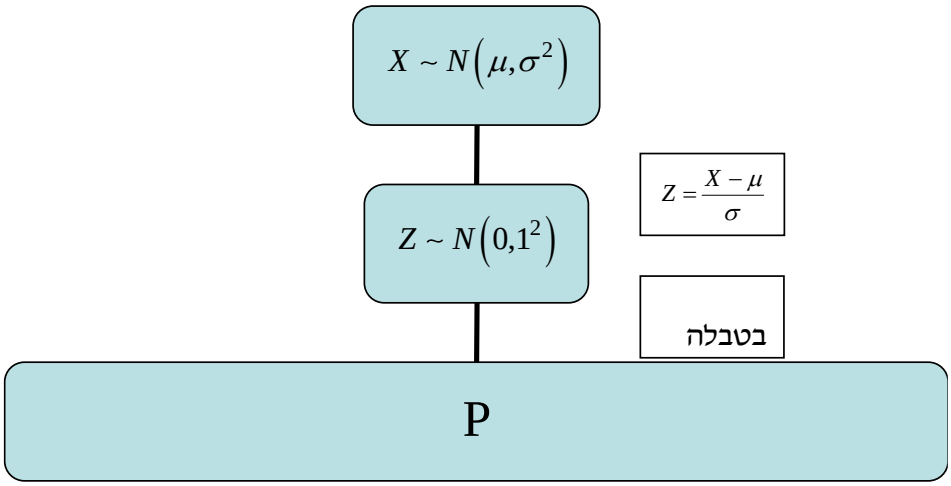
כדי לחשב הסתברויות בהתפלגות נורמלית יש לחשב את השטחים הרלוונטיים שמתחת לעקומה. כדי לחשב שטחים אלה נמיר כל התפלגות נורמלית להתפלגות נורמלית סטנדרטית על ידי תהליך הנקרא תקנון. התפלגות נורמלית סטנדרטית היא התפלגות נורמלית שהממוצע שלה הוא אפס וסטיית התקן היא אחת, והיא תסומן באות Z: $Z \sim N\left(0,1^2\right)$.

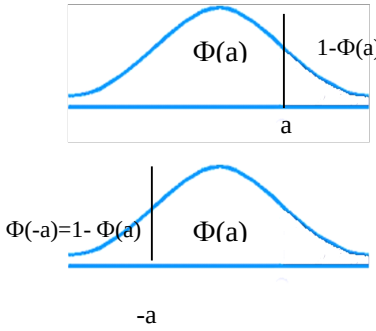
תהליך התקנון מבוצע על ידי הנוסחה הבאה: $Z=\frac{X-\mu}{\sigma}$.

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:

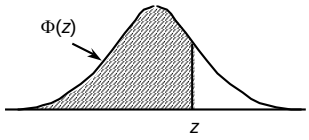
© כל הזכויות שמורות לברק קנדל





© 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 האינטליגנציה?

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?

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 ה. 75.3

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

© כל הזכויות שמורות לברק קנדל

מבוא לסטטיסטיקה א

פרק 14 - מדדי קשר - מדד הקשר של קרמר

תוכן העניינים

1. כללי 59

מדדי קשר – מדד הקשר של קרמר:

רקע:

משתמשים במדד זה כאשר אחד המשתנים 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 המשתנים.

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:

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

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: $E_i=\left(F(x)\cdot F(y)\right)/n$.

Y/X	בעד	נגד	נמנע	$f(x)$
גבר				100
אישה				100
$f(y)$	80	100	20	$n=200$

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.

© כל הזכויות שמורות לברק קנדל

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.

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

© כל הזכויות שמורות לברק קנדל

מבוא לסטטיסטיקה א

פרק 15 - מדד הקשר ספירמן

תוכן העניינים

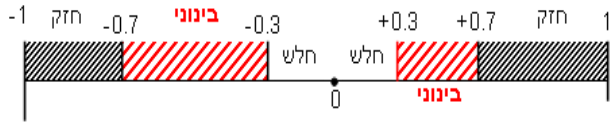
1. כללי 61

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 שמשתנה אחד עולה, השני עולה ללא יוצא מן הכלל.

קשר דירוגי חיובי חלקי:
מקדם המתאם בין 0 ל-1.
ככל שמשתנה אחד עולה, לשני יש נטייה לעלות אך לא באופן מוחלט.

קשר דירוגי שלילי מלא:
מדד הקשר של ספירמן יוצא 1-.
ככל שמשתנה אחד עולה השני יורד ללא יוצא מן הכלל.

קשר דירוגי שלילי חלקי:
מקדם המתאים הוא בין 0 ל-(1-).
ככל שמשתנה 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.

הנוסחה של מדד הקשר: $r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n\left(n^2-1\right)}$.

© כל הזכויות שמורות לברק קנדל

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).

© כל הזכויות שמורות לברק קנדל

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.

© כל הזכויות שמורות לברק קנדל

מבוא לסטטיסטיקה א

פרק 16 - מדד הקשר פירסון

תוכן העניינים

1. כללי 64

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:

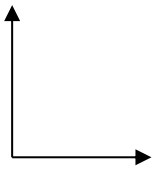
X - מס' חדרים בדירה.

Y - מס' נפשות הגרות בדירה.

להלן התוצאות שהתקבלו:

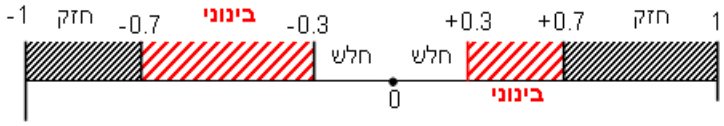
מס' דירה	X	Y
1	3	2
2	2	2
3	4	3
4	3	3
5	5	4

נשרטט מנתונים הללו דיאגרמת פיזור:



© כל הזכויות שמורות לברק קנדל

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 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) אומר שקיים קשר לנארי מלא בו השיפוע b יהיה חיובי ואילו מתאם שלילי מלא אומר שקיים קשר לנארי מלא בו השיפוע b שלילי (מקדם מתאם 1-).
מתאם חיובי חלקי אומר שככל שמשתנה אחד עולה לשני יש נטייה לעלות בערכו אבל לא קיימת נוסחה לינארית שמקשרת את X ל- Y באופן מוחלט.
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 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	ציון
2	80
1	90
0	90
2	70
3	70
4	50

א. שרטט דיאגראמת 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:

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

א. מה הממוצע של כל רמת הורמון?

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

© כל הזכויות שמורות לברק קנדל

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

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

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

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:

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

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.

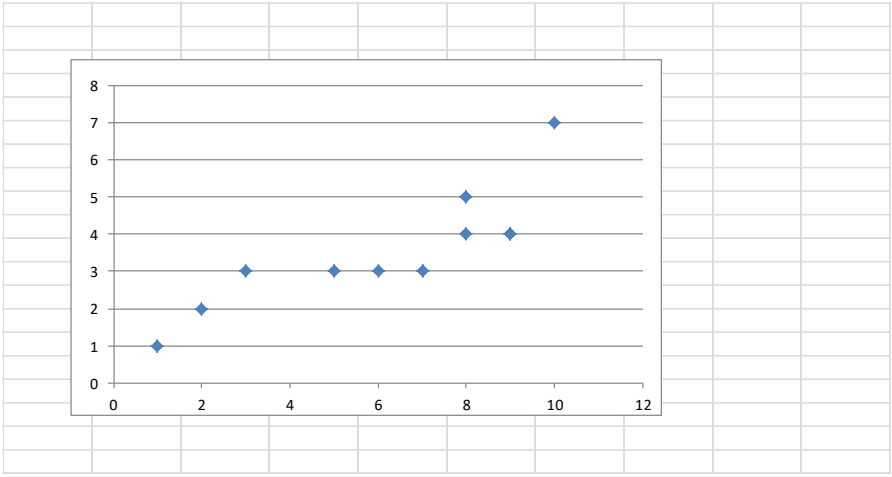
ד. אף אחת מהתשובות לא נכונה.

© כל הזכויות שמורות לברק קנדל

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

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.

© כל הזכויות שמורות לברק קנדל

תשובות סופיות:

1) ב. -0.9325.
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

מבוא לסטטיסטיקה א

פרק 17 - רגרסיה

תוכן העניינים

1. כללי 70

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: $b = r\frac{S_r}{S_x}$, $Y = bX + a$.

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

© כל הזכויות שמורות לברק קנדל

שאלות:

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

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

$$\sum_{i=1}^{20} \left(X_i - \overline{X} \right)^2 = 76 \ , \ \sum_{i=1}^{20} \left(Y_i - \overline{Y} \right) = 76 \ .$$

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

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:

$$\mathrm{COV}\left(X,Y\right) = 7.5 \ , \ \overline{Y} = 8 \ , \ \overline{X} = 14 \ , \ S_Y = 5 \ , \ S_X = 2 \ .$$

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: $\overline{X} = 1.5$, $S_x = S_y = 4$,
וכן שקו הרגרסיה של Y על בסיס X הינו: $Y = -0.2X + 0.5$.
חשבו מהו מקדם המתאם בין X ל- Y.

© כל הזכויות שמורות לברק קנדל

תשובות סופיות:

1) א. 0.8. ב. $Y=0.8X+0.4$. ג. 12.4 אלפי ₪.

2) א. 0.75. ב. 4.25 אלפי ₪. ג. 14.6 שנים.

3) א. 1.2. ב. 29. ג. $Y=1.2X+29$.

4) -0.2.

© כל הזכויות שמורות לברק קנדל

מבוא לסטטיסטיקה א

פרק 18 - מדדי קשר-רגרסיה -שונות מוסברת ושונות לא מוסברת

תוכן העניינים

1. כללי 73

מדדי קשר – רגרסיה – שונות מוסברת ושונות לא מוסברת:

רקע:

המטרה ברגרסיה היא להסביר את השונות של המשתנה התלוי.
למשל, להסביר את השונות של המשכורת באמצעות הוותק או להסביר את השוני בציונים באמצעות כמות החיסורים.

r^2 - החלק מהשונות של המשתנה התלוי מוסבר. השונות המוסברת נקראת גם שונות נ'יבויים. השונות הלא מוסברת נקראת גם שונות טעויות.

© כל הזכויות שמורות לברק קנדל

שאלות:

1) נמצא קשר חיובי בעוצמה של 0.7 בין שטח דירה למחירה. כמו כן, נתון שסטיית התקן של מחירי הדירות הינה 200.

א. איזה אחוז מהשונות של מחירי הדירות מוסבר על ידי שטח הדירה?

ב. איזה אחוז מהשונות של מחירי הדירות לא מוסבר על ידי שטח הדירה?

ג. מהי השונות המוסברות ומהי השונות הלא מוסברת של מחירי הדירות?

2) להלן רשימת טענות, לגבי כל טענה קבעו נכון/לא נכון ונמקו!

א. אם שונות הטעויות שווה ל-0 (השונות הלא מוסברת) אז מקדם המתאם של פירסון יהיה 1.

ב. אם מקדם המתאם של פירסון בין שני משתנים הוא 1 אזי שונות הטעויות (השונות הלא מוסברת) תהיה 0.

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

שאלות רב-ברירה:

3) בקשר בין שני משתנים התקבל: $r^2 = 0.64$, לכן:

א. ללא יוצא מן הכלל ככל שערכי משתנה אחד עולה השני יעלה.

ב. 64% מהשונות של משתנה אחד מוסבר על ידי המשתנה השני.

ג. הקשר בין שני המשתנים הוא בעוצמה של 0.64.

ד. כל התשובות נכונות.

4) אם מגדילים את r^2, ניתן לומר כי:

א. אחוז השונות המוסברת יקטן.

ב. אחוז השונות המוסברת יגדל.

ג. אחוז השונות המוסברת יישאר ללא שינוי.

ד. סטיית התקן משתנה.

ה. לא ניתן לדעת.

© כל הזכויות שמורות לברק קנדל

5) בקורס מבוא לכלכלה ניתנו במשך השנה שני מבחנים: מבחן בסוף סמסטר א' X ומבחן בסוף סמסטר ב' Y. כאשר בנו את קו הרגרסיה של הציון במבחן סוף סמסטר ב' לפי הציון במבחן סוף סמסטר א התקבלה שונות טעויות של 80, ושונות ניבויים של 20.
לפי נתונים אלו, מקדם המתאם בין הציון במבחן סוף סמסטר א' לבין הציון במבחן סוף סמסטר ב' הוא:

א. 0.44.

ב. 0.44 -.

ג. עוצמת ההקשר הלינארי היא 0.44, אך אין אפשרות לדעת את סימנה.

ד. אין אפשרות לחשב את מקדם המתאם.

ה. 0.35.

תשובות סופיות:

1) א. 49%. ב. 51%.
ג. שונות מוסברת: 19,600, שונות לא מוסברת: 20,400.

2) א. לא נכון. ב. נכון. ג. נכון.

3) ב'.

4) ב'.

5) ג'.

© כל הזכויות שמורות לברק קנדל

מבוא לסטטיסטיקה א

פרק 19 - מדדי קשר - בחירת מדד מתאים

תוכן העניינים

1. בחירת מדד מתאים 76

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

© כל הזכויות שמורות לברק קנדל

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.

© כל הזכויות שמורות לברק קנדל

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.

© כל הזכויות שמורות לברק קנדל

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-₪ 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-₪ 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.

© כל הזכויות שמורות לברק קנדל

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.

© כל הזכויות שמורות לברק קנדל

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

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

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 ראליטי 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:

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

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:

שכר	12	18	20	15	30
מוטיבציה	4	4	5	3	5

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

א. פירסון.

ב. ספירמן.

ג. קרמר.

ד. אין מספיק נתונים כדי לדעת.

© כל הזכויות שמורות לברק קנדל

14) להלן טבלה על נתונים שנאספו על מספר תצפיות:

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

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'	

© כל הזכויות שמורות לברק קנדל

מבוא לסטטיסטיקה א

פרק 20 - שאלות אמריקאיות מסכמות על כל חומר הלימוד

תוכן העניינים

1. כללי ..... (ללא ספר)