

# סטטיסטיקה

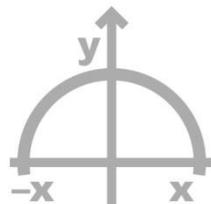


$$\begin{matrix} 1 & \sqrt{2} \\ \diagdown & \diagup \\ 1 & 1 \end{matrix}$$



$$\begin{matrix} + & - & 0 \\ \diagup & \diagdown & \diagdown \\ 0 & 0 & 0 \end{matrix}$$

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



## תוכן העניינים

1.	סטטיסטיקה תיאורית- סיווג משתנים וסולמות מדידה	1
6.	סטטיסטיקה תיאורית- הצגה של נתונים	
14.	סטטיסטיקה תיאורית- סכימה	
(לא ספר)	סטטיסטיקה תיאורית - ממדדי מרכז	
(לא ספר)	סטטיסטיקה תיאורית- ממדדי פיזור	
18.	סטטיסטיקה תיאורית - ממדדי פיזור - טווח בין רבוני	
22.	סטטיסטיקה תיאורית- ממדדי מיקום יהשי-ציוון תקו	
24.	סטטיסטיקה תיאורית-אחוזונים בטבלה בדידה	
26.	סטטיסטיקה תיאורית - טרנספורמציה לינארית	
29.	סטטיסטיקה תיאורית- תרשימים קופסה	
31.	סטטיסטיקה תיאורית- ניתוח פלטימ	
33.	סטטיסטיקה תיאורית שאלות אמריקאיות	
39.	יסודות ההסתברות	
43.	התפלגיות רציפות מיוחדות - התפלגות נורמלית	
50.	התפלגות הדגימה ומשפט הגבול המרכזי	
56.	הסקה סטטיסטית - הקדמה	
59.	מושגי יסוד באמידה	
64.	روح סמך לתוחלת (ממוצע)	
74.	روح סמך להפרש תוחלות (ממוצעים) במדגים בלתי תלויים	
76.	روح סמך לתוחלת (ממוצע) ההפרשים במדגים מזוגים	
78.	מבוא לבדיקה השערות על פרמטרים	
84.	בדיקה השערות על תוחלת (ממוצע)	
113.	בדיקה השערות על הפרש תוחלות במדגים בלתי תלויים	

## תוכן העניינים

24. בדיקת השערות לתוחלת ההפרש במדגמים מזוגיים .....	124
25. הקשר בין רוח סמק לבדיקת השערות להפרש תוחלות .....	134
26. שאלות מסכימות בבדיקה השערות .....	137
27. מבחני חי ברייבור .....	150
28. מודד הקשר ספירמן .....	167
29. מקדם המתאים (מודד קשר) הילינארי ומובהקותו .....	170
30. מודדי קשר - השפעת טרנספורמציה לינארית על פירסון .....	197
31. רגרסיה .....	200
32. מודדי קשר-רגרסיה - שונות מוסברת ושונות לא מוסברת .....	203
33. מודדי קשר - בחירת מודד מתאים .....	206

# סטטיסטיקה

פרק 1 - סטטיסטיקה תיאורית- סיווג משתנים וסולמות מדידה

תוכן העניינים

1.....  
1. סולמות מדידה

## סטטיסטיקה תיאורית – סיווג משתנים וסולמות מדידה:

### רקע:

סטטיסטיקה תיאורית הוא ענף בו לומדים כיצד לאסוף נתונים, להציג אותם ולנתן אותם.

בסטטיסטיקה תיאורית אנו פונים לקבוצה מסוימת. באותה קבוצה אנו אוספים נתונים על הישיותה באותה קבוצה.

משתנה – תכונה שיכולה לקבל מספר ערכים: דעה פוליטית, מקום מגורים, גובה של אדם וכדומה.

כל ישות בקבוצה שאנו צופים בה ואוספים לגבייה נתונים נקראת תצפית. הנתונים שנאספים בדרך כלל מרוכזים בסיס נתונים. בסיס הנתונים כל שורה היא תצפית וכל עמודה מייצגת משתנה.

### דוגמה (פתרון בהקלטה):

למחלקת טראומה הגיעו 5 פצועים מתאונה שקרתה בכיביש החוף. אספו נתונים לגבי אותם פצועים, הנתונים מרוכזים בטבלה הבאה :

מין	גיל	מצב הפצוע	דופק
גבר	26.6	אנוש	40
גבר	24.5	קשה	38
אישה	32.1	קשה	50
גבר	34.9	בינוני	65
אישה	23.1	קל	89

ענו על השאלות הבאות :  
 הגדרו את הקבוצה שבדוגמה.  
 כמה תציפות בקבוצה?  
 כמה משתנים בקבוצה?  
 כמה ערכים יש למשנה "מין"?

את המשתנים במחקר אנו מסוגים ל- "סולמות מדידה" הדבר חשוב בכך שבה ננתה את הנתונים בהמשך.

#### **מיון משתנים לפי סולמות המדידה:**

1. **סולם שמי (nominal)** – משתנה שלعرчиו יש משמעות רק מבחינת הזהות ואין עניין של יותר או פחות לערכיהם שלהם לדוגמה: צבע מועדף.
2. **סולם סדר (ordinal)** – כאשר לערכים של המשתנה בנוסף לשם ישנה גם משמעות לסדר מי יותר או מי פחות אבל אין משמעות לגודל. המשתנה מסוים סדר יכול לקבל ערכים מילוליים או מספריים. למשל: אזוריות ישראלית: ייש או אין.
3. **סולם כמותי (scale)** – משתנה שהייב להיות מספרי, לערכים שלו בנוסף לשם ולסדר בהםם יש משמעות לערך המספרי. המשתנה כמותי הוא משתנה שניינט בדרך כלל למספר או למדוד על ידי מכשיר מדידה. למשל, מספר המחשבים בדירה, שטח הדירה במ"ר.

את המשתנה הכמותי אנו מסוגים לשני סוגים :

**משתנה בדיד :**  
משתנה שערכיו מתקבלים מתוך סידרה של ערכים אפשריים כמו : מספר המחשבים בדירה.

**משתנה רציף :**  
משתנה שערכיו מתקבלים מתוך אינסוףערכים בתחום מסוים, הערכים מתקבלים ברצף ולא קפיצות של ערכים. למשל, שטח הדירה במ"ר.

#### **דוגמה (פתרון בהקלטה):**

למחלקת טראומה הגיעו 5 פצועים מתאונה שקרויה בכביש החוף. אספו נתונים לגבי אותם פצועים, הנתונים מ羅וצים בטבלה הבאה :

מין	גיל	מצב הפסיכואנומטי	דופק
גבר	26.6	אנוש	40
גבר	24.5	קשה	38
אישה	32.1	קשה	50
גבר	34.9	בינוני	65
אישה	23.1	קל	89

סובגו כל משתנה בסיס הנתונים : שמי, סדר, כמותי רציף, כמותי בדיד.

**שאלות:**

**1)** לפניכם טבלה המסכםת נתונים לגבי סקר שנעשה היום :

שם משפחה	מצב משפחתי	מידת דתיות	מספר RCCBITS
כהן	רווק	חילוני	0
חדר	נשיו	חילוני	1
לביא	גרוש	מוסתרני	1
פיינגולד	אלמן	חילוני	2
אבו שוקרא	דתי	נשיו	1
בן חיים	מוסתרני	נשיו	0
רוטשילד	רווק	חילוני	0

א. כמה תצפיות בדוגמה זו?

ב. כמה משתנים בדוגמה זו?

ג. כמה ערכיים ישם ל- "מידת דתיות"?

ד. מהם הערכיים האפשריים למשתנה "מצב משפחתי"?

**2)** סניף מס' 543 של בנק "רואה" בדק ל-80 לקוחות את מספר הפעמים שככל  
לkoń 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 מפעםיים.

א. הגדרו את הקבוצה בדוגמה זו.

ב. כמה תצפיות בדוגמה זו?

ג. הגדרו את המשטנה בדוגמה זו. מהו סולם המדידה שלו?

**3)** במחקר רפואי התעניינו לדעת כיצד מיןון תרופת "קופקס" משפיע על מספר  
 שעות השינה של אדם. במחקר השתתף אדם אחד בשם דני שבכל יום ניתן לו  
 מיןון שונה של התרפיה. הטליה שלහן מתארת בכל יום את מיןון התרפיה  
 במ"ג שקיבל האדם וכמו כן את מספר שעות השינה שלו באותו הלילה :

מספר היום	מיןון התרפיה	מספר שעות שינה
1	12	6
2	14	7
3	16	7.5
4	18	6.5
5	20	8

א. כמה תצפיות נאספו במחקר?

ב. סווגו את סולם המדידה של "מיןון התרפיה" ?

- 4)** לפניכם רשימה של משתנים. ציינו באיזה סולם מדידה מדובר (שמי, סדר, כמותי בדיד, כמותי רציף):
- גובה אדם בס"מ.
  - מספר ילדים למשפחה.
  - מידת חרזה לפני מבחן.
  - шибועות רצון משירותים לקוחות בסקלה מ-1 עד 7 (1 כלל לא מרוצה עד 7 מרוצה מאד).
  - השכלה.
  - מספר אוטובוס.
  - מקום מגוריים.
  - מין (1=גבר ו-2=אישה).
  - מידת נעליהם.
- 5)** לפניכם רשימה של משתנים כמותיים. ציין ליד כל משתנה אם הוא רציף או בדיד :
- שכר עובד ב-₪.
  - ציוון בחינת בגרות.
  - תוצאה בהטלת קובייה.
  - מהירות ריצה במטר לשנייה בתחרות 100 מטר.
  - שיעור התמיכה בממשלה בעיר.
- 6)** גברת לוי החליטה לדגום 25 ימים של נסיעה לעבודה, כאשר בדרך לעבודתה יש 3 צמתים מרומזרים. ב-9 ימים הגיע גברת לוי לעבודה מבלי לעצור באף צומת. ב-9 ימים נוספים היא הצליחה לעبور בשני רמזוריים ירוקים. ב-5 ימים נוספים היא הצליחה לעبور רק בירוק אחד. בשאר הימים, היא לא עברה באף רמזור ירוק. מעוניינים לחקור את מספר הרמזוריים האדומים בהם עצרה גברת לוי.
- מהו המשתנה הנחקר בדוגמה זו?
  - מהם הערכים האפשריים של משתנה זה?
  - כמה ערכים אפשריים יש למשתנה?
  - מהו סולם המדידה של המשתנה?

**תשובות סופיות:**

- (1) א.  $n = 7$ .  
ב. 4.  
ג. 3.  
ד. רווק, נשוי, גרווש, אלמן.
- (2) א. לקוחות סניף 543 של בנק "רואה".  
ב.  $80 = n$ .  
ג.  $X$  = מספר הפעמים בשבוע שלקוח נכנס לסניף. כמותי בדיד.  
ה. כמותי רציף.  
ב. כמותי בדיד.
- (3) א.  $5 = n$ .  
ב. כמותי רציף.  
ג. סדר.  
ה. שמי.  
ז. שמי.  
ט. סדר.
- (4) א. כמותי רציף.  
ג. אין מספיק נתונים.  
ה. אין מספיק נתונים.  
ז. שמי.  
ט. סדר.
- (5) א. רציף.  
ג. בדיד.  
ה. רציף.
- (6) א. מספר הרמזורים בהם עוצרת גברת לוי ביום בדרכן לעבודה.  
ב. 0, 1, 2, 3.  
ג. 4.  
ד. כמותי בדיד.

## סטטיסטיקה

### פרק 2 - סטטיסטיקה תיאורית- הצגה של נתונים

תוכן העניינים

1. כללי .....  
6 .....

## סטטיסטיקה תיאורית – הצגה של נתונים:

---

**רקע:**

דרכים להציג נתונים שנאספו :

**רישימה של תצפיות:**

התצפית היא ערך שנצפה עבור ישות מסוימת בקבוצה. רושמים את התצפיות שהתקבלו כרשומה,יעיל שיש מספר מועט של תצפיות. ההציג הזו רלבנטית לכל סוגים המשתנים. למשל, להלן מספר החדרים בבניין בן 5 דירות : 3,4,3,5,4.

**טבלת שכיחיות בדידה:**

שכיחותיחסית ב אחוזים	שכיחות – $f(x)$	שם המשתנה – $X$
$\frac{f_1}{N} \cdot 100$	$f_1$	$X_1$
$\frac{f_2}{N} \cdot 100$	$f_2$	$X_2$
$\frac{f_3}{N} \cdot 100$	$f_3$	$X_3$
⋮	⋮	⋮
$\frac{f_x}{N} \cdot 100$	$f_k$	$X_k$
100%	$N = \sum_{i=1}^k f_i$	<b>סה"כ</b>

רושמים את התצפיות בטבלה שבה עמודה אחת מבטא את ערכי המשתנה והשנייה את השכיחות. יעיל עבור משתנה איקומי וכמותי בדיד וכ民事 מסטר רב של תצפיות. לא יעיל למשתנה כמותי רציף.

**דוגמה:**

להלן התפלגות הציונים בכיתה מסוימת :

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

שכיחות מצטברת – צבירה של השכיחויות.

השכיחויות  $F_i$  – השכיחות המצטברת נותנת כמה תצפויות קטנות או שותת לערך.

שכיחות יחסית (פרופורציה) – השכיחות מחולקת לכמויות התצפויות הכללי:

$$\frac{f_i}{n} \text{ -- איזה חלק מהתצפויות בקבוצה שותת לערך.}$$
**טבלת שכיחיות בחלוקת:**

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

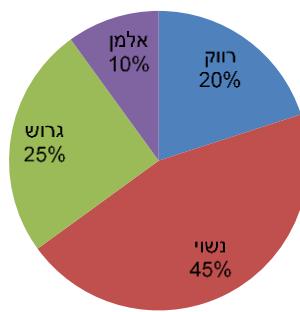
**דוגמה:**

נתנו לקבוצת ילדים לבצע משימה, בדקו את התפלגות זמן הביצוע, בדיקות.  
להלן החתפלגות שהתקבלה :

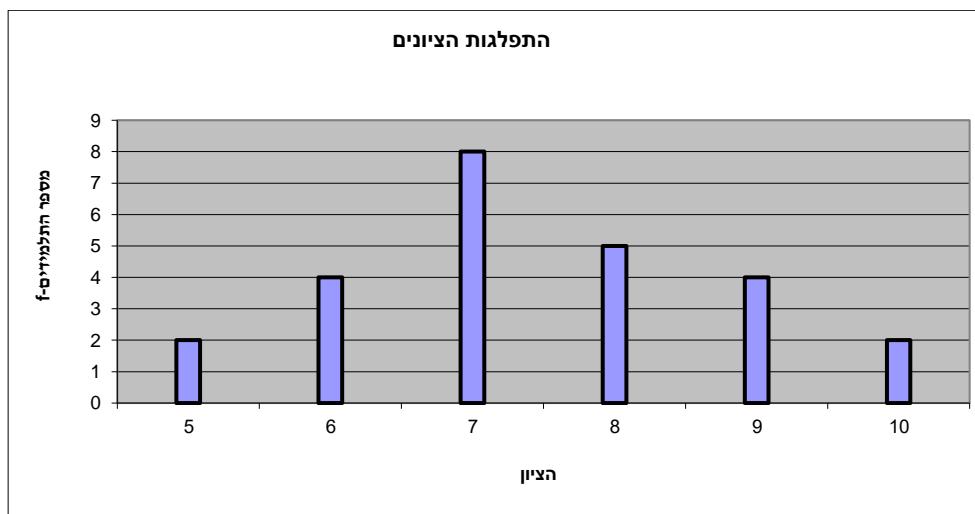
זמן בדיקות	מספר הילדים
20	0.5-3.5
18	3.5-9.5
14	9.5-19.5
8	19.5-29.5

**דיאגרמת עוגה:**

זהו התיאור הגרפי של משתנה איקומי. בדיאגרמת עוגה כל ערך במשתנה מקבל "נתח", שהוא פרופורציוני לשכיחות היחסית של ערך המשתנה בתנאים.

**התפלגות המצב המשפחתי****דיאגרמת מקלות:**

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

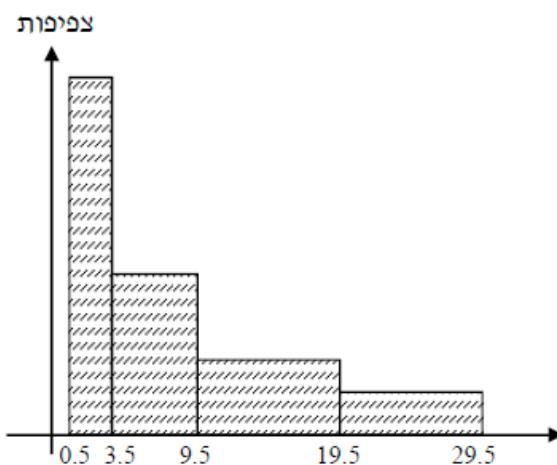


**ההיסטוגרמה:**

ההיסטוגרמה היא הדרך הגרפי כדי לתאר טבלת שכיחיות בחלוקת, והיא רלוונטי למשתנה כמותי רציף.

בහistogramma הציר האופקי הוא הציר של המשטנה והציר האנכי הוא הציר של הצפיפות. הצפיפות מחושבת בכל מחלוקת על ידי חלוקת השכיחות ברוחב של כל המחלוקת, והוא נותנת את מספר התוצאות הממוצעת בכל מחלוקת יחידה. אם המחלוקות הן שוות ברוחב, ניתן לשרטט את הההיסטוגרמה לפי השכיחות ואין צורך בцеיפות.

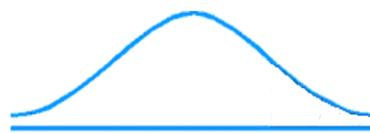
cefipot	cefipot	מצטברת	שכיחות	ממוצע	רוחב	X
6.6667	20	20	2	3	0.5 - 3.5	
3	38	18	6.5	6	3.5 - 9.5	
1.4	52	14	14.5	10	9.5 - 19.5	
0.8	60	8	24.5	10	19.5 - 29.5	

**פוליגון – מצולען:**

אם נחבר את אמצע קצה כל מלבן בקווים ישרים. ניתן לראות חזותי לצורה של התפלגות המשטנה.

**צורות התפלגות נפוצות:****התפלגות סימטרית פעmonoית**

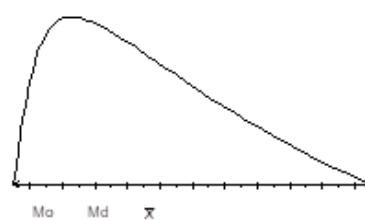
רוב התצפויות במרכז, וככל שנתרחק מהמרכז יהיה פחות תצפויות באופן סימטרי. לדוגמה, ציוני IQ.



ישנן התפלגויות סימטריות שאין פעmonoיות, כגון :

**התפלגות אסימטרית ימנית (חיובית)**

רוב התצפויות מתקבלות ערכים נמוכים ויש מיעוט הולך וקטן של תצפויות שמקבלות ערכים גבוהים קיצוניים. לדוגמה, שכר במשק.

**התפלגות א-סימטרית  
ימנית או חיובית**
**התפלגות אסימטרית שמאלית (שלילית)**

רוב התצפויות מתקבלות ערכים גבוהים ויש מיעוט הולך וקטן של תצפויות שמקבלות ערכים נמוכים קיצוניים. לדוגמה, אורך חיים.



**שאלות:**

- 1)** בסקר צפיה בטלויזיה התקבלו התוצאות הבאות : 25 צפו בערוץ הראשון, 25 צפו בערוץ 10, 75 צפו בערוץ השני, 50 צפו באחד מערוצי הcabלים ו-25 לא צפו בטלויזיה בזמן הסקר.
- רשמו את טבלת השכיחות ואת השכיחות היחסית.
  - תארו את הנתונים באופן גרפי.

- 2)** להלן נתונים על התפלגות המקצוע המועדף של תלמידי שכבה ו' בבית הספר "מעוף":

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

- מהו המשתנה הנחקר?
- מהי פרופורצית התלמידים שمعدיפים תנ"ך?

- 3)** להלן התפלגות ההשכלה במקומות העבודה מסוימים :

השכלה	מספר העובדים
ນמוכה	60
תיכונית	120
אקדמאית	20

- מהו המשתנה הנחקר?  
מאי זה סולם הוא?
- תארו את הנתונים באופן גרפי.

- 4)** להלן רשימת הציונים של 20 תלמידים שנבחנו ב מבחון הבנת הנקרא :
- .7 ,6 ,8 ,5 ,6 ,7 ,8 ,9 ,6 ,7 ,6 ,8 ,7 ,6 ,7 ,6 ,8 ,9 ,6 ,7 ,6 ,8 ,5 ,6

- מהו המשתנה? האם הוא בדיד או רציף?
- תארו את הרשימה בטבלת שכיחויות.
- הוסיפו שכיחיות יחסית לטבלה.
- תארו את הנתונים באופן גרפי.

**תשובות סופיות:**

ב. עיין גרף מלא בסרטון הוידאו.

(1) א. להלן טבלה :

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

ב. 19.6%.

(2) א. מקצוע מועדף.

ב. עיין גרף מלא בסרטון הוידאו.

(3) א. משתנה נחקר : השכלה, סוג : סדר.

ב+ג. להלן טבלה:

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

- (4) א. המשתנה: ציון, משתנה בדיד.  
 ד. עיין גרף מלא בסרטון הויידאו.

## סטטיסטיקה

### פרק 3 - סטטיסטיקה תיאורית- סכימה

תוכן העניינים

1. כללי .....

14 .....

## סטטיסטיקה תיאורית – סכימה:

**רקע:**

בסטטיסטיקה ישנה צורת רישום מקובלת לסקום 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

(הסביר מלא מופיע בסרטוניים באתר).

**שאלות:**

- 1) במבנה 5 דירות. לכל דירה רשמו את מספר החדרים שיש בדירה ( $X$ ), ומספר הנפשות החיים בדירה ( $Y$ )如下：

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

.  $\sum_{i=1}^3 X_i$  . א.

.  $\sum_{i=1}^5 Y_i$  . ב.

.  $\sum_{i=1}^4 X_i$  . ג.

.  $\left( \sum_{i=1}^4 X_i \right)^2$  . ד.

.  $\sum X_i$  . ה.

.  $\sum X_i Y_i$  . ו.

.  $\sum (X_i) \sum (Y_i)$  . ז.

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

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

$$\cdot \sum_{i=1}^4 y_i . \text{א}$$

$$\cdot \sum_{i=1}^6 a . \text{ב}$$

$$\cdot \sum_{i=1}^6 x_i y_i . \text{ג}$$

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

$$\cdot \sum_{i=1}^6 x_i + a . \text{ה}$$

**3)** קבעו לכל זהות האם היא נכונה:

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

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

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

**תשובות סופיות:**

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

## סטטיסטיקה

### פרק 4 - סטטיסטיקה תיאורית - מדי מרכז

תוכן העניינים

1. כללי .....

(ללא ספר) .....

## סטטיסטיקה

### פרק 5 - סטטיסטיקה תיאורית- מדדי פיזור

תוכן העניינים

1. כללי .....  
(ללא ספר) .....

## סטטיסטיקה

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

תוכן העניינים

1. טווח ביןרבוני.....  
18 .....

## סטטיסטיקה תיאורית – מדדי פיזור – טווח בין רביעוני:

**רקע:**

הטווח הבינו-רביעוני (יש הקוראים לו התחום הבינו-רביעוני) נותן את הטווח בין הרבעונים בו נמצאים 50% מהתציפות המרכזית. הרעיון ליצור מדד פיזורי שלא ניתן לתוצאות חריגות ביותר. כדי לחשב את הטווח הבינו-רביעוני יש למצוא את הרבעון התיכון והעליו של התפלגות התציפות.

רביעון תיכון – ערך שמחולק את ההתפלגות לשניים.  
25% מהמקרים נמוכים ממנו או שווים לו ו-75% מהמקרים גבוהים או שווים לו.  
סימון:  $Q_2$ .

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

הטווח הבינו-רביעוני הוא הפער בין שני הרבעונים:  $IQR = Q_3 - Q_1$ .

**שלבים במציאת טווח בין-רביעוני בטבלת שכיחיות:**

שלב א: נמצא את הרבעון תיכון: הוא הערך שהScanner היחסית המctrברת באחוזים עברה לראשונה את 25%.

שלב ב: נמצא את הרבעון עליון: הוא הערך שהScanner היחסית המctrברת באחוזים עברה לראשונה את 75%.

שלב ג: נמצא את הטווח הבינו-רביעוני: 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 הבינו-רביעוני של מספר תוכניות החיסכון בסניף?

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

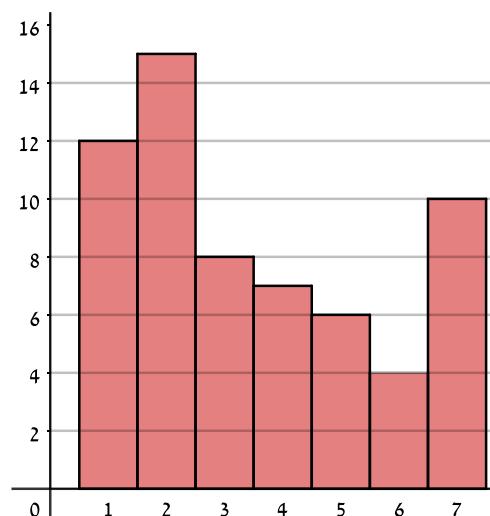
**שאלות:**

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

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

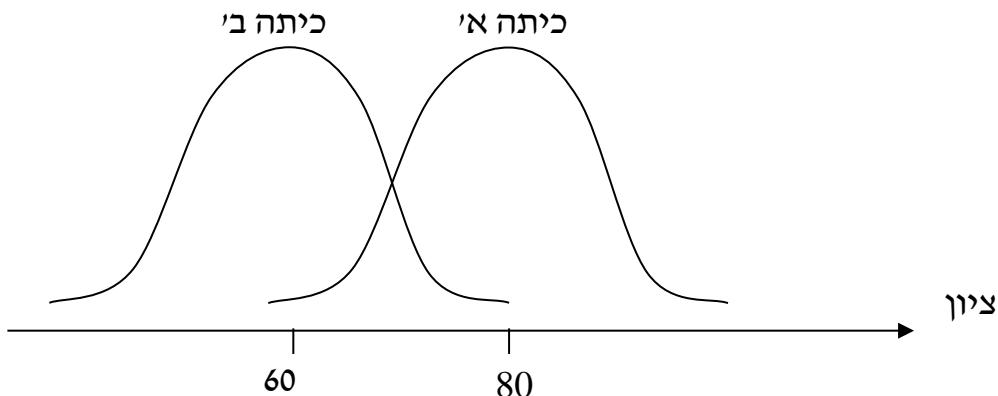
מהו הטווח הבין-רביעוני של מספר המכוניות למשפחה בישוב "הגורן"?

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



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

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



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

4) הוספה גודל קבוע לכל תצפיות סדרת נתונים :

- א. תגדיל את הטוחה הבין-רביעוני.
- ב. תקטין את הטוחה הבין-רביעוני.
- ג. לא תנסה הטוחה הבין-רביעוני.
- ד. לא ניתן לדעת מה יקרה לטוחה הבין-רביעוני.

5) חושב הטוחה הבין-רביעוני עבור התפלגות מסוימת והתקבל התוצאה אפס. לכן :

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

- 6) סניף מס' 543 של בנק "רואה" בדק ל-80 לקוחות את מספר הפעמים שככל  
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 מפעםיים.  
 מהו הטווח הבין-רבוני?  
 א. 60.  
 ב. 2.  
 ג. 50.  
 ד. 1.
- 7) התפלגות הציונים ב מבחון ווקסלר היא סימטרית בכך:  
 א. טווח הציונים הוא אפס.  
 ב. הטווח הבין-רבוני 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 ישנה.
- (3) ג'.
- (4) ג'.
- (5) אי'.
- (6) ד'.
- (7) ד'.

## סטטיסטיקה

### פרק 7 - סטטיסטיקה תיאורית- מדדי מיקום יחס-ציוון תקן

תוכן העניינים

1. כללי .....

- 22 .....

## סטטיסטיקה תיאורית – מדדי מיקום יחסי – ציון תקן:

**רקע:**

המטרה למדוד איך תצפית ממוקמת ביחס לשאר התצפיות בהתפלגות.

**ציון תקן:**

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

ציון התקן נותן כמה סטיות התקן סוטה התצפית מהממוצע. כלומר, ציון התקן מעיד על כמה סטיות התקן התצפית מעל או מתחת לממוצע:

- ציון תקן חיובי אומר שההתצפית מעל הממוצע.
- ציון תקן שלילי אומר שההתצפית מתחת לממוצע.
- ציון תקן אפס אומר שההתצפית בדיק בmäßig.

**דוגמה (פתרון בהקלטה):**

במקומות העבודה מסוימים, ממוצע המשכורות הוא 8 אלף ₪, עם סטיית התקן של אלפיים ₪. באותו מקום העבודה ההשכלה הממוצעת של העובדים הנה 14 שנים, עם סטיית התקן של 1.5 שנים. עורך מרוויח במקום העבודה זה 11 אלף ₪ והשכלהו 16 שנים.  
מה ערך יותר, באופן יחסי, משכיל או משתכר?

**שאלות:**

**1)** תלמידי כיתה ח' ניגשו לבחן בלשון ולבוחן במתמטיקה.  
להלן התוצאות שהתקבלו :

המבחן	סטטיסט Takon	ממוצע
לשון	74	12
מתמטיקה	80	16

עודד קיבל : 68 בלשון ו-70 במתמטיקה.

- א. באיזה מקצוע עודד טוב יותר באופן יחסית לשכבה שלו?  
ב. איזה ציון עודד צריך לקבל במתמטיקה כדי שייהה שקול לציונו בלשון?

**2)** במבצע ליצור מצלבים לרכב בדקו במשך 40 ימים את התפוקה היומית (מספר מצלבים במאוט) ואת מספר הפעלים שעבדו באותו היום.  
להלן טבלה המסכםת את המידע שנאסף על שני המשתנים :

סטטיסט Takon	ממוצע	תפוקה	מספר פעילים
10	48	15	15
Shemittah	10	2	2

באחד הימים מתוך כלל הימים שנבדקו התפוקה הייתה 50 מאות מצלבים ובאותו היום עבדו 13 פעילים.  
מה יותר חריג באותו היום, ייחסית לשאר הימים שנבדקו : נתוני התפוקה או  
כמות הפעלים?  
א. התפוקה.  
ב. כמות הפעלים.  
ג. חריגים באותה מידה.  
ד. חסרים נתונים כדי לדעת זאת.

**3)** הגובה הממוצע של המתגייסים לצבאות הוא 175 סנטימטר עם סטטיסט Takon של 10 סנטימטר. המשקל הממוצע הוא 66 ק"ג עם סטטיסט Takon של 8 ק"ג.  
ערן המתגייס כshaw 180 ס"מ ומשקלנו 59 ק"ג.  
א. כמה ערן חריג יותר ביחס לשאר המתגייסים, גובהו או משקלו?  
ב. כמה ערן אמר לשקלן כדי שמשקלו יהיה שקול לגובהו?

**תשובות סופיות:**

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

## סטטיסטיקה

### פרק 8 - סטטיסטיקה תיאורית-אחסונים בטבלה בדידה

תוכן העניינים

- 24 ..... 1. כללי .....

## **סטטיסטיקה תיאורית – מדדי מיקום יחסי – אחווזוניים בטבלה בדידה:**

**רקע:**

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

**чисוב האחווזון מתוך נתוניים בטבלה שכיחיות בדידה:**

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

**דוגמה (פתרון בהקלטה):**

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

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

א. מצאו את האחווזון ה-25.

ב. מצאו את הערך ש-20% מהמקרים מעליו.

**שאלות:**

**1)** להלן התפלגות של משתנה קלשחו:

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

מצאו להתפלגות את :

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

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

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

חשבו את :

- א. העשרון התחתון.
- ב. האחוזון ה-30.
- ג. הערך ש-20% מהתצפית גודלות ממנו.
- ד. רביעון עליון.

**תשובות סופיות:**

- |       |       |       |       |
|-------|-------|-------|-------|
| .1 .ד | .3 .ג | .1 .ב | .2 .א |
| .4 .ד | .4 .ג | .2 .ב | .1 .א |

## סטטיסטיקה

פרק 9 - סטטיסטיקה תיאורית - טרנספורמציה לינארית

תוכן העניינים

26 ..... 1. כללי .....

## סטטיסטיקה תיאורית – טרנספורמציה לינארית:

**רקע:**

מצב שבו מבצעים שינוי מסווג הוסף (או החסלה) של קבוע, והכפלת (או חילוק) של קבוע, לכל הtcpיות:  $y = a \cdot x + b$ . כך יושפעו המדדים השונים:

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

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

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

**מדדי המרכז:**

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

$$R_y = |a| R_x$$

$$S_y = |a| S_x$$

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

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

$$Z_Y = \frac{a}{|a|} Z_X$$

**שלבי העבודה:**

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

**דוגמה (פתרון בהקלטה):**

השכר הממוצע של עובדים הינו 9000 ש"ח וטוחה 6000 ש"ח. חשבו את המדדים הללו לאחר שהעלו את כל המשכורות ב-10% ולאחר כך קנסו אותן ב-100 ש"ח.

**שאלות:**

- 1) עברו סדרת נתוניים התקבל:  $70, S = 15, MO = 70, \bar{x} = 80$ . הוחלט להכפיל את כל התצפויות ב-4 ולהחסיר מההתוצאה 5. חשבו את המדדים הללו לאחר השינוי.
- 2) בחברה מסוימת השכר הממוצע הוא 40 ש' לשעה עם סטיית תקן של 5 ש' לשעה. הוחלט להעלות את כל המשכורות ב-10%, אך זה לא סיפק את העובדים ולכן הם קיבלו לאחר מכן תוספת של 2 ש' לשעה. מה הממוצע ומהי השונות של השכר לשעה לאחר כל השינויים.
- 3) במחקר מסויים הציון החיצוני היה 73, טווח הציונים היה 40 נקודות והעשורן העליון היה הציון 87. כיוון שהציונים בבחינה היו נמוכים, המורה החליט לתת פקטורי של 4 נקי לכל התלמידים. חשבו את המדדים לאחר הפקטור.
- 4) דגמו מקו ייוצר 50 קופסאות של גפרורים. בדקו בכל קופסה בה יש 40 גפרורים את כמות הגפרורים הפגומים. התקבל שבממוצע יש 3 גפרורים פגומים בקופסה, עם סטיית תקן של 1.5 גפרורים. מה יהיה הממוצע ומה תהיה סטיית התקן של מספר התקינים בקופסה?
- 5) חברת בזק הציעה את ההצעה הבאה: שלושים שקלים דמי מנוי חודשיים קבועים וכן 10 אגרות לכל דקה של שיחת יוצאת. אדם בדק במשך שנה את דקוט השיחות היוצאות שלו, וקיים שבממוצע חודשי יש לו 600 דקוט שיחות יוצאות עם שנות של 2500 דקוט רבעות, כמו כן בחודש ינואר ציון התקן היה 2. חשבו את המדדים הללו עבור חשבון הטלפון החודשי של אותו אדם בשקלים אם היה משתמש בחבילה המוצעת לו על ידי בזק.
- 6) הוכחו שאם כל התצפויות בהתפלגות עברו טרנספורמציה לינארית:  $y_i = a \cdot x_i + b$  אז הממוצע והשונות של כל התצפויות לאחר הטרנספורמציה יהיו בהתאם:  $\bar{y} = a \cdot \bar{x} + b, S_y^2 = a^2 S_x^2$ .

**תשובות סופיות:**

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

## סטטיסטיקה

### פרק 10 - סטטיסטיקה תיאורית- תרשימים קופסא

תוכן העניינים

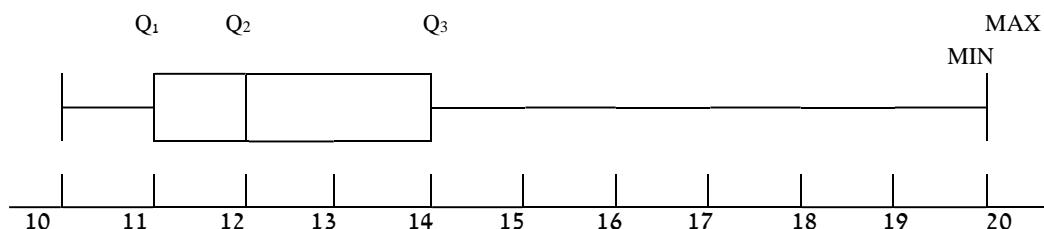
- 29 ..... 1. כללי .....

## סטטיסטיקה תיאורית – תרשימים קופסא (Boxplot):

**רקע:**

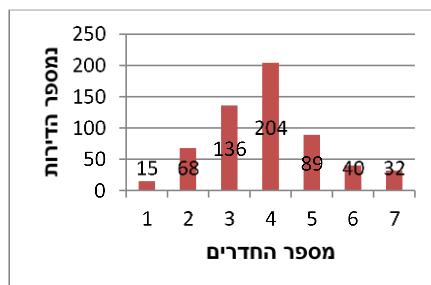
תרשימים קופסא הינו תרשימים שבוצרתו ניתן לבחון :

- 1) את המרכז של ההתפלגות על ידי החציון ( $Q_2$ ).
- 2) את הפיזור של הנתונים (הטוחה והטוחה הבין רבועני).
- 3) את צורת ההתפלגות (סימטריה או אסימטריה שמאלית).



**שאלות:**

**1)** להלן התפלגות מספר החדרים לדירות שנבנו בשנת 2009 בעיר אשדוד :

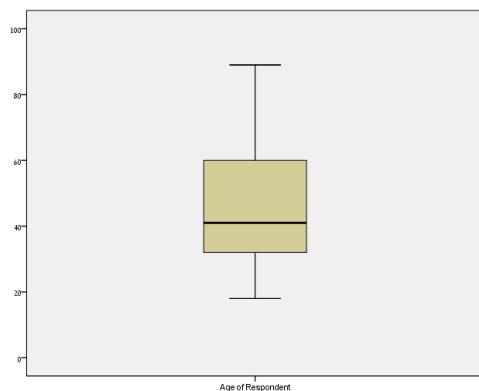


א. מצאו את החציון, הרבעון תחתון והרביעון עליון של ההתפלגות.

ב. שרטטו דיאגרמת קופסה להתפלגות.

ג. מה ניתן לומר על צורת ההתפלגות?

**2)** להלן דיאגרמת קופסה המתארת את התפלגות הגיל (בשנים) באוכלוסייה מסוימת :



א. מה גיל החציון?

ב. מה בערך טווח הגילאים?

ג. מה ניתן להגיד על צורת ההתפלגות?

**תשובות סופיות:**

**1)** א. חציון : 4 , רביעון תחתון : 3 , רביעון עליון : 5 .

ב. ראה גרף מלא בסרטון וידאו. ג. כמעט סימטרית.

**2)** א. חציון : 40 . ב. טווח : 70 . ג. התפלגות אסימטרית ימנית.

## סטטיסטיקה

### פרק 11 - סטטיסטיקה תיאורית- ניתוח פלטים

תוכן העניינים

1. כללי .....

31 .....

## סטטיסטיקה תיאורית – ניתוח פלטימ:

### שאלות:

**1)** להלן פلت על התפלגות הגילאים אוכלוסייה מסוימת :

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

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

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

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

מלאו את הערכים המסומנים בסימני שאלה.

**תשובות סופיות:**

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

## סטטיסטיקה

פרק 12 - סטטיסטיקה תיאורית שאלות אמריקאיות

תוכן העניינים

1. כללי .....

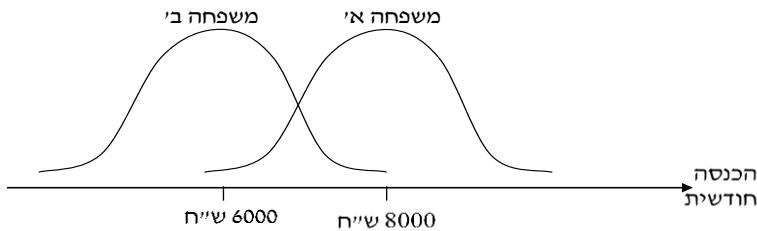
- 33 .....

## סטטיסטיקה תיאורית – שאלות אמריקאיות:

**שאלות:**

**שאלות 3-1 מתייחסות לקטע הבא:**

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



**1)** לאיזו משפחה הכנסה שכיחה גבולה יותר?

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

**2)** באיזו משפחה הכנסה החזיונית שווה להכנסה הממוצעת?

- א. משפחה א'.
- ב. משפחה ב'.
- ג. לשתיهن הכנסה החזיונית שווה להכנסה הממוצעת.
- ד. לא ניתן לדעת – אין מספיק נתונים.

**3)** באיזו משפחה סטיית התקן של הכנסה החודשית גבולה יותר?

- א. משפחה א'.
- ב. משפחה ב'.
- ג. לשתיهن אותה סטיית התקן.
- ד. לא ניתן לדעת – אין מספיק נתונים.

**הנתונים הבאים מתייחסים לשאלות 6-4:**

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

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

**4) השכיח של הנתונים הוא:**

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

**5) חציון הנתונים هو:**

- 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 כמוותי רציף, הערך המתאים למאון ה-30, ציון התקן שלו הוא בהכרח:**

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

8) סדרת נתונים סטטיסטיים מונה 10 תצפויות. נתון כי סדרת הנתונים סימטרית סביבה הממוצע. ממוצע הסדרה-40 ושונות הסדרה-100. בשלב מאוחר יותר נווסף שתי תצפויות נוספות לסדרה: 50 ו-30. השונות של 12 התצפויות:

- א. קטנה.
- ב. גדולה.
- ג. לא משתנה.
- ד. לא ניתן לחשב את השונות ללא ידיעת התצפויות.

9) הוספה גודל קבוע לכל תצפויות סדרת נתונים:

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

#### **הנתונים הבאים מתייחסים לשאלות 11-10:**

להלן נתונים על ציוני תלמידים שנבחנו במועדים שונים בסטטיסטיקה:

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

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

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

11) פנינה נבחנה עם סטף וציוון התקן שלא שווה לציוון התקן של שרית לכן ציונה

הוא:

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

**הנתונים הבאים מתייחסים לשאלות 15-12:**  
**בבדיקה פתע של משרד הבריאות במפעל שוקולד, נמצא ש :**

שוקולד פגום	מס' קופסאות
7 8	6 10 11 13 12 48 63 35

**12)** מהו החציון של מספר הפוגומים בקופסה :

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

**13)** מהו הרביעון התיכון של מספר הפוגומים בקופסה?

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

**14)** מספר הפוגומים בקופסה הוא משתנה :

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

**15)** השכיח של מספר הפוגומים בקופסה :

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

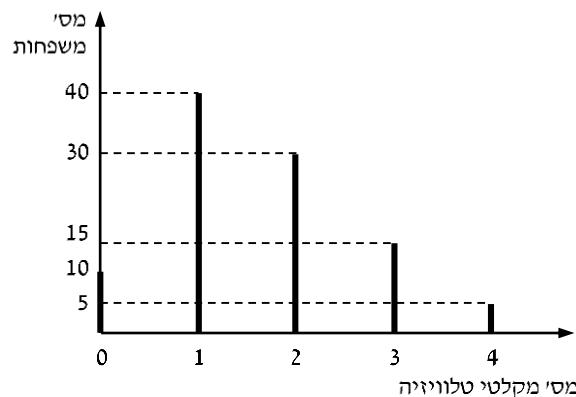
**16)** ביחס לציר המספרים, רוב הערכים בהתקלגות א-סימטרית ימנית נמצאים :

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

- 17) בוצע מחקר על מספר העובדים בחברות מזון לעומת חברות תקשורת. החציון וה ממוצע בשתייהן שווה 8. איזה מהטענות הבאות היא הנכונה והמלואה ביותר :
- השכיחות ב-2 חברות זהה אך שונה מ-8.
  - השכיח ב-2 חברות זהה אך ניתן לדעת מהו.
  - השכיח בשתי חברות הינו בהכרח 8.
  - שכיח בחברה אחת שונה מ-8 ובשנייה הוא 8.
  - אף תשובה אינה נכונה.

**הנתונים הבאים מתיחסים לשאלות 18 עד 22:**

נערך סקר על מספר מקלט טלוויזיה הנמצאים בבית. תוצאות הסקר נתונות בדיאגרמת מקלות הבאה:



18) המשתנה הנחקר כאן הוא :

- משתנהשמי.
- משתנה מסולס סדר.
- משתנה כמותי בדיד.
- משתנה כמותי רציף.

19) הטווח של ההתפלגות הוא :

- .35
- .4
- .3
- .2

**(20)** ממוצע מספר מקלט טלוויזיה למשפחה הוא :

- .1.65
- .1.5
- .1
- .2

**(21)** השכיח של התפלגות זו היא :

- .40
- .1.5
- .1
- .2

**(22)** מסתבר שיש בין 2 ל-5 משפחות נוספות שאין להם מקלט טלוויזיה ויש לצרף את המשפחות הללו להתפלגות. כיצד הנתון זה ישפייע 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'

## סטטיסטיקה

### פרק 13 - יסודות ההסתברות

תוכן העניינים

1. כללי .....

- 39 .....

## הגדירות יסודיות:

**רקע:**

**ניסוי מקרי:** תהליך לו כמה תוצאות אפשריות. התוצאה המתקבלת נודעת רק לאחר ביצוע התהליך. למשל: תוצאה בהטלה קובייה, מזג האויר בעוד שבועיים.

**מרחב מדגם:** כלל התוצאות האפשרות בניסוי המקרי. לדוגמה, בהטלה קובייה:  $\{1, 2, 3, 4, 5, 6\}$ , או: מזג האויר בעוד שבועיים: {נאה, שרבי, מושלג, גשם, מעונן, אקלקית, אביך}.

**מאורע:** תת קבוצה מתוק מרחב המדגם. מסומן באותיות: A, B, C. בהטלה קובייה למשל, המאורע 'לקבל לפחות 5' יסומן:  $A = \{5, 6\}$ . המאורע 'לקבל תוצאה זוגית' יסומן:  $B = \{2, 4, 6\}$ .

**גודל מרחב המדגם:** מספר התוצאות האפשרות למרחב המדגם. בהטלה קובייה למשל נקבע:  $|\Omega| = 6$ .

**גודל המאורע:** מספר התוצאות האפשרות במאורע עצמו. למשל, בהטלה הקובייה האירועים הקודמים יסומנו:  $|A| = 2$ ,  $|B| = 3$ .

**מאורע משלים:** מאורע המכיל את כל התוצאות האפשרות למרחב המדגם פרט לתוצאות במאורע אותו הוא משלים. למשל, בהטלה הקובייה:  $\bar{A} = \{1, 2, 3, 4\}$ ,  $\bar{B} = \{1, 3, 5\}$ .

**מרחב מדגם אחד (סימטרי):** מרחב מדגם בו לכל התוצאות למרחב המדגם יש את אותה עדיפות, אותה סבירות למשל, קובייה הוגנת, אך לא כמו מזג האויר בשבוע הבא.

**הסתברות במרחב מודגם אחיד:** במרחב מודגם אחיד הסיכוי למאורע יהיה :

$$P(A) = \frac{|A|}{|\Omega|}$$

דוגמה : מה הסיכוי בהטלה קובייה לקבל לפחות 5 ?

$$P(A) = \frac{|A|}{|\Omega|} = \frac{2}{6}$$

דוגמה : מה הסיכוי בהטלה קובייה לקבל תוצאה זוגית ?

$$P(B) = \frac{|B|}{|\Omega|} = \frac{3}{6}$$

**הסתברות במרחב לא אחיד:** תחושב לפי השכיחות היחסית :

$$\frac{f}{n}$$

דוגמה :

להלן התפלגות הציונים בכיתה מסוימת :

הציון - x	מספר התלמידים – השכיחות – f
5	2
6	4
7	8
8	5
9	4
10	2

מה ההסתברות שתלמיד אקרי שנבחר בכיתה קיבל את הציון 8 ?

$$\frac{f}{n} = \frac{5}{25} = 0.2$$

מה ההסתברות שתלמיד אקרי שנבחר בכיתה יכשל ?

$$\frac{f}{n} = \frac{2}{25} = 0.08$$

**הסתברות למאורע משלים :** הסתברות לקבלת המשלים של המאורע ביחס למרחב המודגם :

$$P(\bar{A}) = 1 - P(A)$$

להיות מחושב לפי הסיכוי להכשל :

$$P(A) = 1 - \frac{2}{25} = \frac{23}{25}$$

**שאלות:**

- 1)** מהאותיות E, F ו-G יש ליצור מילה בת 2 אותיות, לא בהכרח בת משמעות.  
 א. הרכיבו את כל המילים האפשריות.  
 ב. רשמו את המקרים למאורע:  
 .i. במילה נמצאת האות E.  
 .ii. במילה האותיות שונות.  
 ג. רשמו את המקרים למאורע  $\bar{A}$ .
- 2)** מטילים זוג קוביות.  
 א. רשמו את מרחב המדגם של הניסוי. האם מרחב המדגם אחיד?  
 ב. רשמו את כל האפשרויות לאיורים הבאים:  
 .i. סכום התוצאות 7.  
 .ii. מכפלת התוצאות 12.  
 ג. חשבו את הסיכויים לאיורים שהוגדרו בסעיף ב'.
- 3)** נבחר באקראי ספרה מבין הספרות 0-9.  
 א. מה ההסתברות שהספרה שנבחרה גדולה מ-5?  
 ב. מה ההסתברות שהספרה שנבחרה היא לכל היותר 3?  
 ג. מה ההסתברות שהספרה שנבחרה היא אי זוגית?
- 4)** להלן התפלגות מספר מקלט טלוויזיה עבור כל משפחה ביישוב מסוים:

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

- נבחרה משפחה באקראי מהיישוב.  
 א. מה ההסתברות שאין מקלטים למשפחה?  
 ב. מה ההסתברות שיש מקלטים למשפחה?  
 ג. מה ההסתברות שיש לפחות 3 מקלטים למשפחה?

- 5)** להלן התפלגות מספר המכוניות למשפחה ביישוב "עדן":

מספר משפחות	מספר מכוניות
10	4
30	3
100	2
40	1
20	0

- נבחרה משפחה אקראיית מן היישוב.  
 א. מה ההסתברות שאין לה מכוניות?  
 ב. מה ההסתברות שבבעלות המשפחה לפחות 3 מכוניות?  
 ג. מה הסיכוי שבבעלותה פחות מ-3 מכוניות?

- 6) נתיל מטבע רגיל 3 פעמים. בצד אחד של המטבע מוטבע עץ ובצד השני פלי.  
 א. רשמו את מרחב המדגמים של הניסוי. האם מרחב המדגם הוא אחיד?  
 ב. רשמו את כל האפשרויות לאיורים הבאים:  
 .i. התקבל פעם אחת עץ.  
 .ii. התקבל לפחות פלי אחד.  
 ג. מהו המאורע המשלימים ל-D?  
 ד. חשבו את הסיכויים לאיורים שהוגדרו בסעיפים ב-ג.

### תשובות סופיות:

$$\text{.} \Omega = \{EE, EF, EG, FE, FF, FG, GE, GF, GG\} \quad (1)$$

$$\text{.} A = \{EE, EF, EG, FE, GE\}, B \{EF, EG, FE, FG, GE, GF\}$$

$$\text{.} \bar{A} = \{FF, FG, GF, GG\}$$

$$\text{.} \Omega = \begin{Bmatrix} (1,1) & (2,1) & (3,1) & (5,1) & (4,1) & (6,1) \\ (1,2) & (2,2) & (3,2) & (4,2) & (5,2) & (6,2) \\ (1,3) & (2,3) & (3,3) & (4,3) & (5,3) & (6,3) \\ (1,4) & (2,4) & (3,4) & (4,4) & (5,4) & (6,4) \\ (1,5) & (2,5) & (3,5) & (4,5) & (5,5) & (6,5) \\ (1,6) & (2,6) & (3,6) & (4,6) & (5,6) & (6,6) \end{Bmatrix} \quad (2)$$

$$\text{.} A = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}, C = \{(2,6), (3,4), (4,3), (6,2)\}$$

$$\text{.} \frac{1}{9} \text{ הסיכוי ל-B-A :} \quad \frac{1}{6} : \text{A}$$

$$\text{.} 0.5 \quad \text{.} 0.4 \quad \text{.} 0.4 \quad (3)$$

$$\text{.} 0.32 \quad \text{.} 0.78 \quad \text{.} 0.22 \quad (4)$$

$$\text{.} 0.8 \quad \text{.} 0.2 \quad \text{.} 0.1 \quad (5)$$

$$\text{.} \Omega = \{PPP, PPE, PEP, EPP, PEE, EPE, EEP, EEE\} \quad (6)$$

$$\text{.} A = \{PPE, PEP, EPP\}, D = \{PPP, PPE, PEP, EPP, PEE, EPE, EEP\}$$

$$\text{.} \bar{D} = \{EEE\}$$

$$\text{.} \frac{1}{8} \quad (7)$$

## סטטיסטיקה

### פרק 14 - התפלגיות רציפות מיוחדות - התפלגות נורמלית

תוכן העניינים

1. כללי .....

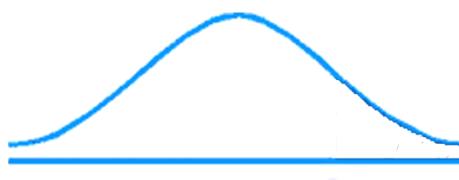
43 .....

## התפלגיות רציפות מיוחדות – התפלגות נורמלית:

---

**רקע:**

התפלגיות נורמלית הינה התפלגות של משתנה רציף. ישנו משתנים רציפים מסוימים שנחוג להתייחס אליהם כנורמליים כדוגמת זמן ייצור, משקל תינוק ביום היולדו ועוד. פונקציית הצפיפות שלהתפלגות הנורמלית נראה כmo פעמון:



לעוקמה זו קוראים גם עוקמת גאוס ועוקמה אחת נבדלת מהשנייה באמצעות המוצע וסטיית התקן שלה.

אליה הם הפרמטרים שמאפיינים אתהתפלגות:  $N(\mu, \sigma^2)$ .

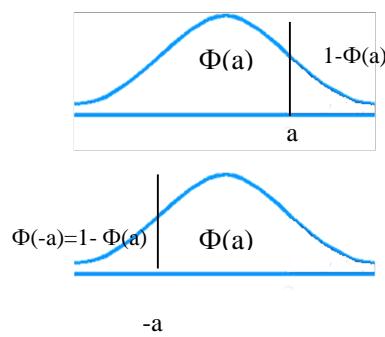
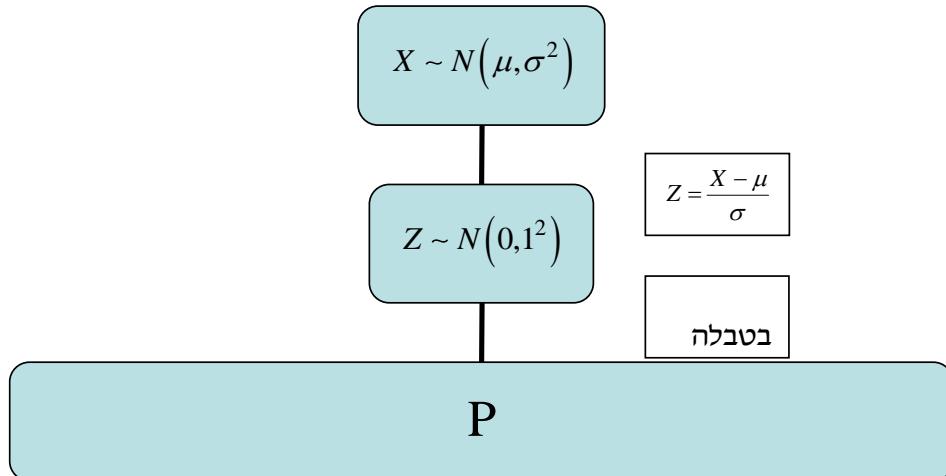
$$\cdot f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} : \text{נוסחת פונקציית הצפיפות}$$

כדי לחשב הסתברויות בההתפלגות נורמלית יש לחשב את השטחים הרלוונטיים שמתוחת לעוקמה. כדי לחשב שטחים אלה נמיר כלהתפלגות נורמלית להתפלגות נורמלית סטנדרטית על ידי תהליך הנקרא תקנון. התפלגות נורמלית סטנדרטית היא התפלגות נורמלית שהממוצע שלה הוא אפס וסטיית התקן היא אחת, והיא מסומן באות  $Z$ :  $N(0, 1^2)$ .

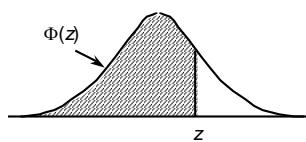
$$\cdot Z = \frac{X - \mu}{\sigma} : \text{תהליך התקנון מבוצע על ידי הנוסחה הבאה}$$

אחרי תקנון מקבלים ערך הנקרא ציון תקן. ציון התקן משמשו בכמה סיטuatיות תקן הערך סוטה מהממוצע.

לאחר חישוב ציון התקן של ערך מסוים נעזרים בטבלה שלהתפלגות הנורמלית הסטנדרטית לחישוב השטח הרצוי, ובאופן כללי בהתאם להסכמה הבאה:



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



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

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

דוגמה (הਪתרון בהקלטה) :

משקל חפיסות שוקולד המיוצרות בחברה מתפלג נורמלית עם ממוצע 100 גרם  
בסטטיסטית 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 בכו הייצור שוקלים פחות מהם?

**שאלות:**

- 1)** הגובה של אנשים באוכלוסייה מסוימת מתפלג נורמלית עם ממוצע של 170 ס"מ וסטיית תקן של 10 ס"מ.
- מה אחוז האנשים שגובהם מתחת ל-182.4 ס"מ?
  - מה אחוז האנשים שגובהם מעל 190 ס"מ?
  - מה אחוז האנשים שגובהם בדיקן 173.6 ס"מ?
  - מה אחוז האנשים שגובהם מתחת ל-170 ס"מ?
  - מה אחוז האנשים שגובהם לכל היותר 170 ס"מ?
- 2)** נתון שהזמן שלוקח לטרופה מסוימת להשפיע מתפלג נורמלית עם ממוצע של 30 דקות ושונות של 9 דקות רביעות.
- מהי פרופורציה המקרים בהן הטרופה תעוזר אחרי יותר משעה?
  - מה אחוז מהמקרים שהבחן הטרופה תעוזר בין 35 ל-37 דקות?
  - מה הסיכוי שהטרופה תעוזר בדיקן תוך 36 דקות?
  - מה שיעור המקרים שהבחן ההשפעה של הטרופה 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.
- מה אחוז האנשים שמשקלם נמוך מ-55 ק"ג?
  - מהי פרופורציה האנשים באוכלוסייה שמשקלם לפחות 50 ק"ג?
  - מהי השכיחות היחסית של האנשים באוכלוסייה שמשקלם בין 60 ל-70 ק"ג?
  - לאיזה חלק מהאוכלוסייה משקל הסוטה מהמשקל הממוצע بلا יותר מ-4 ק"ג?
  - מה הסיכוי שאדם אكري ישקל מתחת ל-140 ק"ג?
- 4)** משקל תינוקות ביום היולדם מתפלג נורמלית עם ממוצע של 3300 גרם וסטיית תקן 400 גרם.
- מצאו את העשרון העליון.
  - מצאו את האחוזון ה-95.
  - מצאו את העשרון התחתון.

- 5) ציוני מבחן אינטלקטואלי מתפלגים נורמליות עם ממוצע 100 ושונות 225.
- מה העשירון העליון של הציונים בבחן האינטלקטואלי?
  - מה העשירון התחתון של ההתפלגות?
  - מהו הציון ש-20% מהנבחנים מקבלים מעליו?
  - מהו האחוזון ה-20?
  - מהו הציון ש-5% מהנבחנים מקבלים מתחתיו?
- 6) נפח משקה בבקבוק מתפלג נורמליות עם סטיית תקן של 20 מ"ל, וננתן ש-33% מהבקבוקים בעלי נפח שעולה על 508.8 מ"ל.
- מה ממוצע נפח משקה בבקבוק?
  - 5% מהבקבוקים המזוהרים עם הנפח הגבוה ביותר נשלחים לבדיקה, החל מאייה נפח שלוחים בקבוק לבדיקה?
  - 1% מהבקבוקים עם הנפח הקטן ביותר נתרמים לצדקה, מהו הנפח המקסימלי לצדקה?
- 7) אורך חיים של מכשיר מתפלג נורמליות. ידוע שמחצית מהמכשירים חיים פחות מ-500 שעות, כמו כן ידוע ש-67% מהמכשירים חיים פחות מ-544 שעות.
- מהו ממוצע אורך חי מכשיר?
  - מהי סטיית התקן של אורך חי מכשיר?
  - מה הסיכוי שמכשיר אקראי יהיה פחות מ-460 שעות?
  - מהו המאיון העליון של אורח חיים מכשיר?
  - 1% מהמכשירים בעלי אורך החיים קצר ביותר נשלחים לבדיקה מעמיקה. מהו אורך החיים המקסימלי לשילוח מכשיר למעבדה?
- 8) להלן שלוש ההתפלגיות נורמליות של שלוש קבוצות שונות ששורתטו באותה מערכת צירים. ההתפלגיות מוספרו כדי להבדיל ביניהן.
- לאיזו ההתפלגות הממוצע הגבוה ביותר?
  - במה מבין המינים הבאים ההתפלגות 1 ו-2 זהות?
    - בעשירון העליון.
    - בממוצע.
    - בשונות.
  - לאיזו ההתפלגות סטיית התקן הקטנה ביותר?
    - .1
    - .2 .ii
    - .3 .iii
    - .iv אין לדעת.
- 

- 9)** הזמן שלוקח לאדם להגיע לעבודתו מתפלג נורמלי עם ממוצע של 40 דקות וסטיית תקן של 5 דקות.
- א. מה ההסתברות שמשך הנסיעה של האדם לעבודתו יהיה לפחות שלושת רביעי השעה?
- ב. אדם יצא לעבודתו בשעה 10:08 מביתו. הוא צריך להגיע לעבודתו בשעה 09:00. מה הסיכוי שיeahר לעבודתו?
- ג. אם ידוע שזמן נסיעתו לעבודה היה יותר משש שעות רביעי השעה. מה ההסתברות שזמן הנסיעה הכלול יהיה פחות מ-50 דקות?
- ד. מה הסיכוי ששבוע (חמשה ימי עבודה) בדיקק פעם אחת יהיה זמן הנסעה לפחות שלושת רביעי השעה?

### תשובות סופיות:

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

## סטטיסטיקה

### פרק 15 - התפלגות הדגימה ומשפט הגבול המרכזי

#### תוכן העניינים

1. התפלגות ממוצע המדגם ומשפט הגבול המרכזי ..... 50

## התפלגות ממוצע המדגם ומשפט הגבול המרכזי:

**רקע:**

$$\bar{x} = \frac{\sum x_i}{n}$$

בפרק זה נדון בהתפלגות של ממוצע המדגם :

מכיוון שמדובר למדגם אנו יכולים לקבל ממוצע מדגם שונה, אזי ממוצע המדגם הוא משתנה מקרי ויש לו ההתפלגות.

גדלים המתארים ההתפלגות כלשוי או אוכלוסייה כלשוי נקרים פרמטרים.  
להלן רישימה של פרמטרים החשובים לפרק זה:  
ממוצע האוכלוסייה נסמן ב-  $\mu$  (נקרא גם תוחלת).

שונות אוכלוסייה נסמן ב-  $\sigma^2$ .  
סטיית תקן של אוכלוסייה:  $\sigma$ .

### תכונות ההתפלגות:

ממוצע כל ממוצעי המדגם האפשריים שווה לממוצע האוכלוסייה:  $E(\bar{x}) = \mu_{\bar{x}} = \mu$   
שונות כל ממוצעי המדגם האפשריים שווה לשונות האוכלוסייה מחולק ב-  $n$ .

$$V(\bar{x}) = \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$$

תמונה זו נconaה רק במדד מקרי:

יש יחס הפוך בין גודל המדגם לבין שונות ממוצעי המדגם.  
אם נוציא שורש לשונות נקבל סטיית תקן של ממוצע המדגם שנקרה גם

$$\sigma(\bar{x}) = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

טעות תקן :

### דוגמה (פתרון בהקלטה) :

השכר הממוצע במשק הינו 9000 לפ עם סטיית תקן של 4000. דגמו באקראי 25 עובדים.

א. מייהי אוכלוסיית המחקר? מהו המשתנה הנחקר?

ב. מהם הפרמטרים של האוכלוסייה?

ג. מה התוחלת ומהי סטיית התקן של ממוצע המדגם?

**דוגמה מההתפלגות נורמללית:**

אם נדגם מתוך אוכלוסייה שהמשתנה בה מתפלג נורמלית עם ממוצע  $\mu$  ושונות  $\sigma^2$ .

$$\text{ממוצע המדגם גם יתפלג נורמלית: } \bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right), Z_{\bar{x}} = \frac{\bar{x} - \mu}{\sigma}$$

**דוגמה (פתרון בהקלטה):**

משקל תינוק ביום הiolדו מתפלג נורמלית עם ממוצע 3400 גרם וסטיית תקן של 400 גרם.

מה ההסתברות שבמדגם של 4 תינוקות אקראיים בעת הולדתם המשקל הממוצע של התינוקות יהיה מתחת ל-3.5 ק"ג?

**משפט הגבול המרכזי:**

אם אוכלוסייה מתפלגת כלשהו עם ממוצע  $\mu$  ושונות  $\sigma^2$  אז עבור מדגם מספיק

$$\text{גדול } (n \geq 30) \text{ ממוצע המדגם מתפלג בקרוב לנורמל}: \bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

**דוגמה (פתרון בהקלטה):**

משקל חפיסת שוקולד בכו ייצור מתפלג עם ממוצע 100 גרם וסטיית תקן של 4 גרם.

דגמו מכו הייצור 36 חפיסות שוקולד אקראיות.

מה ההסתברות שהמשקל הממוצע של חפיסות השוקולד שנדגוño יהיה מתחת ל-102 גרם?

**שאלות:**

- 1)** מתוך כלל הסטודנטים במכלה שסימנו סטטיסטיקה א נדגו שני סטודנטים. נתון שסכום הציונים של כלל הסטודנטים היה 78 עם סטיית תקן של 15.
- מיהי האוכלוסייה?
  - מה המשנה?
  - מהם הפרמטרים?
  - מהו גודל המדגם?
  - מהו תוחלת ממוצע המדגים?
  - מהי טעות התקן?
- 2)** משקל תינוק ביום היולדו מתפלג נורמלית עם ממוצע 3400 גרם וסטיית תקן של 400 גרם.
- מה ההסתברות שתינוק אكري בעת הלידה ישקל פחות מ-3800 גרם?
- נתון כי ביום מסוים נולדו 4 תינוקות.
- מה ההסתברות שהמשקל הממוצע שלהם עלה על 4 ק"ג?
  - מה ההסתברות שהמשקל הממוצע של התינוקות יהיה מתחת ל-2.5 ק"ג?
  - מה ההסתברות שהמשקל הממוצע של התינוקות יהיה רחוק מהתוחלת ללא יותר מ-50 גרם?
  - הסבירו לא חישוב כיצד התשובה לסעיף הקודם הייתה משתנה אם היה מדובר על יותר מ-4 תינוקות?
- 3)** הגובה של המתגיסים לצה"ל מתפלג נורמלית עם תוחלת של 175 ס"מ וסטיית תקן של 10 ס"מ. ביום מסוים התגיסו 16 חילילים.
- מה ההסתברות שהגובה הממוצע שלהם יהיה לפחות 190 ס"מ?
  - מה ההסתברות שהגובה הממוצע שלהם יהיה בדיק 180 ס"מ?
  - מה ההסתברות שהגובה הממוצע שלהם יסטה מהתוחלת הגבוהים בפחות מ-5 ס"מ?
  - מהו הגובה שבהסתברות של 90% הגובה הממוצע של המדגם יהיה נמוך ממנו?

- (4) הזמן הממוצע שלוקח לאדם להגיע לעבודתו 30 דקות עם שונות של 16 דקות רבעות. האדם נוסע לעבודה במשך שבוע 5 פעמיים. לצורך הפתרון הניחו שזמן הנסעה לעבודה מתפלג נורמליות.
- מה ההסתברות שבמשך שבוע משך הנסעה הממוצע יהיה מעל 33 דקות?
  - מהו הזמן שבהסתברות של 90% ממוצע משך הנסעה השבועי יהיה גבוה ממוני?
  - מה ההסתברות שמשך הנסעה השבועי יהיה מרוחק מ-30 דקות לפחות 2 דקות?
  - כיצד התשובה לسؤال הקודם הייתה משתנה אם האדם היה נוסע לעבודה 6 פעמים בשבוע?
- (5) נפח היין בבקבוק מתפלג נורמליות עם תוחלת של 750 סמ"ק וסטיית תקן של 10 סמ"ק.
- בארכוז 4 בקבוקי היין. מה ההסתברות שהנפח הממוצע של הבקבוקים בארכוז יהיה בדיק 755 סמ"ק?
  - בארכוז 4 בקבוקי היין. מה ההסתברות שהנפח הממוצע של הבקבוקים בארכוז יהיה יותר מ-755 סמ"ק?
  - בארכוז 4 בקבוקי היין. מה ההסתברות שהנפח הממוצע של הבקבוקים בארכוז יהיה לפחות 755 סמ"ק?
  - בקבוקי היין שבארכוז נמצאים לקערה עם קיבולת של שלושה ליטר. מה ההסתברות שהיין יגלוש מהקערה?
- (6) משתנה מתפלג נורמליות עם תוחלת 80 וסטיית תקן 4.
- מה ההסתברות שממוצע המדגם יסטה מהתוחלתו ללא יותר מichiיה כאשר גודל המדגם הוא 9?
  - מה ההסתברות שממוצע המדגם יסטה מהתוחלתו ללא יותר מichiיה כאשר גודל המדגם הוא 16?
- הסביר את ההבדל בתשובות של שני הטעיפים.
- (7) לפי הערכות הלשכה המרכזית לסטטיסטיקה השכר הממוצע במשק הוא 8000 ₪ עם סטיית תקן של 3000 ₪. מה ההסתברות שבמדגם מקרי של 100 עובדים השכר הממוצע יהיה יותר מ-8500 ₪?

8) אורך צינור שמבצע מייצר הינו עם ממוצע של 70 ס"מ וסטיית תקן של 10 ס"מ.

- א. נלקחו באקריאי 100 מוטות, מה ההסתברות שסכום אורך המוטות יהיה בין 68 ל 78 ס"מ?

ב. יש לחבר 2 בניינים באמצעות מוטות. המרחק בין שני הבניינים הינו 7200 ס"מ. מה ההסתברות ש 100 המוטות יספיקו למלאה?

- ג. מה צריך להיות גודל המדגם המינימאלי, כדי שהסתברות של 5% ממוצע המדגם יהיה קטן מ-69 ס"מ. הייערו במשפט הגבול המרכזי.

9) נתון  $\bar{X} \sim N(\mu, \sigma^2)$ . דגמו 5 תצפיות מאותה ההתפלגות והתבוננו בממוצע המדגם  $\bar{X}$ . לכן:  $P(\bar{X} > \mu)$  יהיה (בחרו בתשובה הנכונה):

- א. 0.
- ב. 0.5.
- ג. 1.
- ד. לא ניתן לדעת.

10) נתון  $\bar{X}$  מתפלג כלשהו עם תוחלת  $\mu$  ושונות  $\sigma^2$ .

החליטו לבצע מדגם בגודל 200 מתוך ההתפלגות הנתונה לפי משפט הגבול המרכזי מתקיים (בחרו בתשובה הנכונה):

$$\text{א. } X \sim N\left(\mu, \frac{\sigma^2}{200}\right)$$

$$\text{ב. } \mu \sim N\left(\mu, \frac{\sigma^2}{200}\right)$$

$$\text{ג. } \bar{X} \sim N(\mu, \sigma^2)$$

$$\text{ד. } \bar{X} \sim N\left(\mu, \frac{\sigma^2}{200}\right)$$

11) נתון  $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$ . אם נדgos  $n$  תצפיות מתוך ההתפלגות ונגידיר:

אזי (בחרו בתשובה הנכונה):

- א.  $\mu$  ו-  $\bar{X}$  יהיו משתנים מקרים.
- ב.  $\mu$  יהיה משתנה מקרי ו-  $\bar{X}$  קבוע.
- ג.  $\bar{X}$  יהיה משתנה מקרי ו-  $\mu$  קבוע.
- ד.  $\mu$  ו-  $\bar{X}$  יהיו קבועים.

### תשובות סופיות:

- 1) א. כלל הסטודנטים במללה שסימנו סטטיסטיKA א. ב. ציון. ד. 2. ג. ממוצע : 78, סטיתת תקן : 15.
- 2) א. 0.8413 ב. 0.0013 ג. 0.2 .78 ה. 10.6 .1
- 3) א. 0.0 .0 ב. 0.27.71 ג. 0.9544 ד. 178.205 .2
- 4) א. 0.0465 ב. 0.27.71 ג. 0.2628 ד. התשובה הייתה קטנה.
- 5) א. 0.0 ב. 0.1587 ג. 0.5 ד. 0.1587 .5
- 6) א. 0.5468 ב. 0.6826 ג. 0.271 ד. 0.271 .2
- 7) א. 0.0475 ב. 0.9772 ג. 0.0228 ד. 0.0228 .2
- 8) א. 0.9772 ב. 0.0475 ג. 0.5468 ד. 0.0465 .5
- 9) ב. א. 0.0465 ב. 0.0475 ג. 0.5468 ד. 0.0465 .5
- 10) ד. א. 0.0465 ב. 0.0475 ג. 0.5468 ד. 0.0465 .5
- 11) ג. א. 0.0465 ב. 0.0475 ג. 0.5468 ד. 0.0465 .5

## سطטיסטיקה

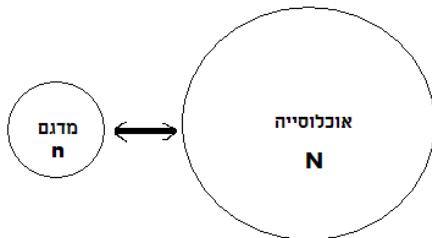
### פרק 16 - השקעה סטטיסטית - הקדמה

תוכן העניינים

- |          |               |
|----------|---------------|
| 56 ..... | 1. כללי ..... |
|----------|---------------|

## הסקה סטטיסטית – הקדמה:

**רקע:**



**אוכלוסייה:**  
קבוצה שאליה מפנים שאלת מחקרית.  
למשל, חברת תרופות שמעוניינת לפתח תרופה  
למחלות הסוכרת מתעניינת באוכלוסיות חוליות  
הסוכרת בעולם.

**مثال:**

חלק מתוך האוכלוסייה.  
למשל, אם נדגים באקראי 10 אנשים מתוך חוליות הסוכרת אז זהו מثال מתוך  
אוכלוסיות חוליות הסוכרת.

במקרים רבים אין אפשרות לחקור את כל האוכלוסייה כיון שאין גישה לכולה,  
היא גדולה מדי, אנו מוגבלים בזמן ובאמצעים טכניים ולכן מבצעים מוגן במטרה  
לבצע הסקה סטטיסטית מהמוגן לאוכלוסייה.  
הדגימה בקורס תהיה דגימה מקראית - הכוונה לדוגמה שבה לכל תצפית באוכלוסייה  
יש את אותו סיכוי להיכל במדגם.

**סטטיטיסטי:**

מודל המוחש בעל המוגן.

**פרמטר:**

מודל המתאר את האוכלוסייה.

### הסימונים לפרמטר וסטטיסטי הם שונים:

סטטיסטי (מדגם)	פרמטר (אוכלוסייה)	
$\mu$	$\bar{X}$	משמעות
$P$	$\hat{p}$	פרופורציה (שכיחות יחסית)

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

### דוגמה (פתרון בהקלטה):

6% מאזרחי המדינה תומכים בהצעת החוק של חבר הכנסת 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 שמתכוונים להוציא ממדגם?
- ו. האם הפרמטר או הסטטיסטי הוא משתנה מקרי?

**שאלות:**

- 1)** מתווך כלל הסטודנטים במכללה שסיוומו סטטיסטיקה א' נדגמו שני סטודנטים. נתון שסכום הציונים של כלל הסטודנטים היה 78 עם סטיית תקן של 15.
- מי האוכלוסייה?
  - מה המשטנה?
  - מהם הפרמטרים?
  - מהו גודל המדגם?
- 2)** להלן התפלגות מספר מקלט הטלוויזיה הטלוויזיה למשפחה בישוב "העוגן".  
נגידר את X להיות מספר המקלטים של משפחה אקראית. מתכנים לדגום מאוכלוסייה זו 4 משפחות ולהתבונן במספר מקלט הטלוויזיה במדגם.
- מיי האוכלוסייה ומהו המשטנה הנחקר?
  - מהו הסטטיסטי שיילקח מהדגם ומה סימונו?

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

- 3)** נתון כי 20% מהשכירים במדינה הם אקדמיים. נבחרו באקראי 10 שכירים באותו אוכלוסייה ומתכנים לפרסם את מספר האקדמאים שנדגמו.
- מיי האוכלוסייה?
  - מה המשטנה באוכלוסייה?
  - מהם הפרמטרים?
  - 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, המשטנה הנחקר : מס' מקלטים.  
ב.  $\bar{X} = \text{ממוצע מדגם.}$
- 3)** א. השכירים במדינה.  
ב. השכלה : אקדמי, לא אקדמי.  
ג. מס' האקדמאים במדגם.      ג. שיעור ההצלחות באוכלוסייה : 0.2.

## סטטיסטיקה

פרק 17 - מושגי יסוד באמידה

תוכן העניינים

1. כללי .....

59 .....

## מושגי יסוד באמידה:

### רקע:

כזכור מהפגש הקודם, פרמטר הוא גודל המתאר את האוכלוסייה או התפלגות מסויימת. כמו ממוצע הגבאים בקרוב מתגisiים לצה"ל -  $\mu$ .

כמו פרופורצית התומכים במשלה בקרוב אזרחי המדינה -  $p$ .

בדרכ כל הפרמטרים הם גדלים שאינם ידועים באמת, ולכן מוצאים מוגדים במטרה לאמוד אותם. אין אפשרות לחשב אותם הניסיון הוא בהערכתו כמה הם שווים ככל שניתן.

- נסמן באופן כללי פרמטר באות  $\theta$  ואומד ב- $\hat{\theta}$ .  $\hat{\theta}$  הוא סטטיסטי המוחשב על המוגדים ובאמצעותו נאמוד את  $\theta$ .
- שגיאת אמידה:  $|\hat{\theta} - \theta|$  - ההפרש בין האומד לאמת (הפרמטר).

### דוגמה (פתרון בהקלטה):

בכנסת ה-19 קיבלת מפלגת העבודה 15 מנדטים. בערוץ 10 ברגע סגירת הקלפיות הערכו את מספר המנדטים של המפלגה להיות 17 מנדטים וזאת על סמך תוצאות מוגדים של העורץ.

- מה הפרמטר בדוגמה זו?
  - מהי טעות האמידה של ערוץ 10?
- $\hat{\theta}$  יהיה אומד חסר הטיה ל- $\theta$  אם התוחלת של  $\hat{\theta}$  תהיה שווה ל- $\theta$ :  $E(\hat{\theta}) = \theta$ .
  - טעות התקן של אומד היא סטיית התקן שלו, כלומר:  $\sigma(\hat{\theta}) = S.E$ .

**פרמטרים מרכזיים ואומדיים שלחה:****ממוצע האוכלוסייה  $\mu$ :**

$$\text{האומד הנקודתי שלו יהיה: ממוצע המדגמים: } \bar{x} = \frac{\sum x}{n}$$

$$\text{. } \sigma(\bar{x}) = \frac{\sigma}{\sqrt{n}} = SE \text{ . } E(\bar{x}) = \mu \text{ . } \text{לכן } \bar{x} \text{ הינו אומר חסר הטיה ל- } \mu \text{ . כמו כן, טעות התקן: } \mu$$

**פרופורציה באוכלוסייה  $p$ :**

$$\text{האומד הנקודתי שלו יהיה: פרופורציה במדגם: } \hat{p} = \frac{y}{n}$$

$$\text{. } \sigma(\hat{p}) = \sqrt{\frac{p \cdot (1-p)}{n}} \text{ . } E(\hat{p}) = p \text{ . } \text{לכן } \hat{p} \text{ הינו אומר חסר הטיה ל- } p \text{ . כמו כן טעות התקן: } E(\hat{p}) = p$$

**שונות האוכלוסייה  $\sigma^2$ :**

$$\text{האומד הנקודתי שלו יהיה: } S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

$$\text{. } S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{\sum x_i^2 - n\bar{x}^2}{n-1} \text{ . } \sigma^2 \text{ . } \text{ולכן } S^2 \text{ הינו אומד חסר הטיה ל- } \sigma^2 \text{ . } E(S^2) = \sigma^2$$

**הערה:** אומד הוא הנוסחה הכללית לאמידת הפרמטר ואומדן הוא הערך הספציפי שהתקבל במדגם מסוים.

**דוגמה (פתרון בהקלטה):**

נדגמו 10 משפחות בתל אביב ונבדק עבור כל משפחה מספר הילדים שלה.

להלן התוצאות שהתקבלו: 1, 3, 2, 1, 4, 5, 2, 1, 3, 2, 1.

אמדו באמצעות אומדיים חסרי הטיה את הפרמטרים הבאים:

1. ממוצע מספר הילדים למשפחה בתל אביב.
2. שונות מספר הילדים למשפחה בתל אביב.
3. פרופורציית המשפחות בנות שני ילדים.

**שאלות:**

- 1)** מתוך 500 טירונים, נמצאו 120 בעלי שברי הליכה. נתנו שהטיסוי שטירון יהיה עם שבר הליכה הוא 0.25.
- מהי האוכלוסייה המוצגת בשאלת? מהם הפרמטרים שלה?
  - מהי טעות התקן של האומד כשהמדגם בגודל 500?
  - מהו האומדן לפרמטר?
  - מהי טעות האמידה?
- 2)** לפי נתונים היכרנו, מקרר צורך ממוצע 2400 וואט לשעה עם סטיית התקן של 500 וואט לשעה.
- במדגם של 25 מקרים של היכרן התקבל ממוצע של 2342 וואט לשעה.
- מהי האוכלוסייה המוצגת בשאלת? מהם הפרמטרים שלה?
  - מהי טעות התקן של האומד?
  - מהו האומדן לפרמטר?
  - מהי טעות האמידה?
- 3)** נדגו עשרה מתגיים לzech"l. גובהם נמדד בס"מ. להלן התוצאות שהתקבלו: 168, 184, 192, 180, 171, 177, 187, 168, 177 ו-175.
- מצאו אומדן חסר הטיה לגובה הממוצע של מתגיisi זה"ל.
  - מצאו אומדן חסר הטיה לשונות הגבהים של מתגיisi זה"ל.
  - מצאו אומדן חסר הטיה לפ羅פורציות המתגיים בגובה של לפחות 180 ס"מ.
- 4)** נדגו 20 שכירים באקראי. עברו כל שכיר נמדד השכਰ באלפי שקלים.
- להלן התוצאות שהתקבלו:  $\sum_{i=1}^{20} X_i^2 = 1502.2$ ,  $\sum_{i=1}^{20} X_i = 162$
- AMDו את השכר הממוצע של השכירים במשק.
  - AMDו את סטיית התקן של שכר השכירים במשק.
- 5)** במטרה לאמוד את ממוצע האוכלוסייה, נדגו תציפות בלתי תלויות מהאוכלוסייה וחישבו את הממוצע שלהם. מהי טעות התקן?
- סטיית התקן של האוכלוסייה.
  - סטיית התקן של ממוצע האוכלוסייה.
  - סטיית התקן של המדגם.
  - סטיית התקן של ממוצע המדגם.

6) משקל הממוצע של אוכלוסייה מסוימת הוא 75 ק"ג עם שונות של 25 . אם יבחרו כל המדגמים האפשריים בגודל 10 מאוכלוסייה זו סטיית התקן של ממוצעי המדגמים תהיה :

- .א. 3.
- .ב. 2.5
- .ג. 1.581
- .ד. אין מספיק נתונים לדעת.

7) במדגם מקרי, متى סכום ריבועי הסטיות מהממוצע,  $\sum_{i=1}^n (x_i - \bar{x})^2$ , מחולק ב- $n-1$ ?  
 א. כאשר  $n$  קטן.  
 ב. כאשר תצפיות המדגם אינן בלתי תלויות.  
 ג. כאשר האוכלוסייה אינה מתפלגת נורמללית.  
 ד. כאשר מעוניינים באומד חסר הטיה לשונות האוכלוסייה ממנה הוצאה המדגם.  
 ה. כאשר מעוניינים לחשב את שונות התפלגות הדגימה של ממוצע המדגם.

8) מדגם מקרי מתוך אוכלוסייה בעלת ממוצע  $\mu$  לא ידוע  $X_1, X_2, \dots, X_{16}$

ושונות :  $s^2 = 64$ . טעות התקן של האומד ל-  $\mu$  היא :

- .א. 16.
- .ב. 8.
- .ג. 4.
- .ד. 2.

9) מהו אומד חסר הטיה?

- א. אומד שערכו שווה לממוצע התפלגות הדגימה שלו.
- ב. אומד שערכו שווה לערך הפרטר באוכלוסייה.
- ג. אומד שממוצע התפלגות הדגימה שלו שווה לערך הפרטר באוכלוסייה.
- ד. אומד שהסיכוי שערכו יהיה גבוה מערך הפרטר באוכלוסייה שווה לשיכוי שהוא נזוק ממנו.

**תשובות סופיות:**

- (1) א. 0.25      ב. 0.019      ג. 0.24      ד. 0.01
- (2) א. אוכלוסייה: מקרים של יצרן, תוחלת: 2400, סטיית תקן: 500.  
                 .58      .2342      ג.      ב. 0.100
- (3) א. 0.4.ג      ב. 64.1      ג. 177.9  
                 .3.16.ב      ב. 8.1
- (4) א.      ב.      ג. 0.1  
                 ד.      ב. 0.1  
                 ג. 0.1
- (5) א. 0.1  
                 ב. 0.1  
                 ג. 0.1
- (6) א. 0.1  
                 ב. 0.1  
                 ג. 0.1
- (7) א. 0.1  
                 ב. 0.1  
                 ג. 0.1
- (8) א. 0.1  
                 ב. 0.1  
                 ג. 0.1
- (9) א. 0.1  
                 ב. 0.1  
                 ג. 0.1

## סטטיסטיקה

### פרק 18 - רוח סמרק לתוחלת (מומוצע)

#### תוכן העניינים

64 .....	1. רוח סמרק כשינוי האוכלוסייה ידועה .....
69 .....	2. קביעת גודל מוגם .....
71 .....	3. רוח סמרק כשינוי האוכלוסייה לא ידועה .....

## רוח סמך כשינויות האוכלוסייה ידועה:

**רקע:**

ממוצע המדגם הוא אומד לממוצע האוכלוסייה, אך לא באמת ניתן להבין ממנו על גודלו של ממוצע האוכלוסייה. ההסתברות שממוצע המדגם יהיה בדיקות כמו הממוצע האמתי הוא אפסי.

מה שנחוג לעשות כדי לאמוד את ממוצע האוכלוסייה, זה לבנות רוח סמך.

בנייה מרוחה בטחון שהסיכוי שהפרט  $\mu$  ייכל בתוכו הוא:  $1 - \alpha$ .

$\alpha - 1$  : נקרא רמת בטחון או רמת סמך. כך ש:  $\alpha - 1 = P(A \leq \mu \leq B)$ .

A - גבול תחתון של רוח הסמך.

B - הגבול העליון של רוח הסמך.

$L = B - A$  - אורך רוח הסמך.

**דוגמה (פתרון בהקלטה):**

חווק דגם 25 חיילים שנבחנו ב מבחון הפסיכומטרי. הוא בנה רוח סמך לממוצע הציונים ב מבחון הפסיכומטרי ב קרב אוכלוסיית החיילים ו קיבל בין 510 ל-590. רוח הסמך בונה ברמת סמך של 95%.

1. מהי אוכלוסיית המחקר?
2. מה המשתנה באוכלוסייה?
3. מה הפרט שהחווק רצה לאמוד?
4. מהו רוח הסמך?
5. מה אורך רוח הסמך?
6. מהי רמת הביטחון של רוח הסמך?

בפרק זה נרצה לבנות רוח סמך לתוחלת ( $\mu$ ) במקהה ש- $\sigma^2$  (שונות האוכלוסייה) ידועה. פרמטר אותו נרצה לאמוד:  $\mu$ .

אומד נקודתי:  $\bar{x}$ .  
תנאים לבניית רוח הסמך:  $N \sim X$  או  $n \geq 30$ .

$\sigma^2$  (שונות האוכלוסייה) ידועה.

$$\text{נוסחה לרוח הסמך: } \bar{x} \pm Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

#### דוגמה (פתרון בהקלטה):

על פי נתונים היצרנו אורך חיי סוללה מתפלג נורמלית עם סטיית תקן של 1 שעה. מעוניינים לאמוד את תוחלת חיי סוללה. נציגו באקראי 4 סוללות, אורך החיים הממוצע שהתקבל הוא 13.5 שעות. בנו רוח סמך ברמת סמך של 95% לתוחלת אורך חיי סוללה.

$$\text{שגיאת האמידה המקסימלית: } \varepsilon = Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

ע - נותן את שגיאת האמידה המקסימלית, דבר שנקרה גם טעות סטטיסטית, טעות דגימה.

#### דוגמה (פתרון בהקלטה):

בשימוש לשאלת עם הסוללות. מה ניתן להגיד בביטחון של 95% על שגיאת האמידה?

קשרים מתמטיים ברוח הסמך:

- אורך רוח הסמך הוא פערם שגיאת האמידה המקסימלית:  $L = 2\varepsilon$ .
- ממוצע המדגמים נופל תמיד באמצע רוח הסמך:  $\bar{X} = \frac{A+B}{2}$ .
- ככל שמספר התצפיות ( $n$ ) גבוהה יותר, כך יש יותר אינפורמציה ולכן האומד יותר מדויק, ולכן מקבל רוח סמך יותר קצר.
- ככל שרמת הביטחון ( $\alpha-1$ ) גבוהה יותר, כך:  $\varepsilon = z_{1-\frac{\alpha}{2}}$  גבוהה יותר, ורוח הסמך יותר ארוך.

**שאלות:**

- 1)** חוקר התענין למד את השכר המומוצע במשק. על סמך מוגם הוא קבע שבביטחון של 95% כי השכר המומוצע במשק נع בין 9200 ל-9800 ₪.
- מי האוכלוסייה במחקר?
  - מה המשנה הנחקר?
  - מה הפרמטר שאותו רוצים למד?
  - מה רוח הסמך לפרמטר?
  - מה רמת הסמך לפרמטר?
  - מה אורך רוח הסמך?
  - מה הסיכוי שטעות הדגימה תעלה על 300 ₪?
- 2)** מעוניינים למד את התפוקה היומית המומוצעת של מפעל מסוים ברמת סמך של 95%. בדוגמאות אקראי של 100 ימים התקבלה תפוקה ממוצעת 4950 מוצרים ביום. לצורך פתרון הנח שטית התקן האמצעית ידועה ושויה 150 מוצרים ביום. בנו את רוח הסמך.
- 3)** מעוניינים למד את ממוצע אורך החיים של מכשיר. מנתוני היצרך ידוע שאורך החיים מתפלג נורמלית עם סטיית תקן של 20 שעות. נגמו 25 מכשירים ונמצא כי ממוצע אורך החיים שלהם היה 230 שעות.
- בנו רוח סמך ברמת סמך של 90% לאורך החיים המומוצע של מכשיר.
  - בנו רוח סמך ברמת סמך של 95% לאורך החיים המומוצע של מכשיר.
  - הסבירו כיצד ומדוע השתנה רוח הסמך.
- 4)** נגמו 200 עובדים מהמשק הישראלי. השכר המומוצע שלהם היה 9700 ₪. נניח שטית התקן של השכר במשק היא 3000 ₪.
- בנו רוח סמך ברמת סמך של 95% לתוחלת השכר במשק.
  - מה ניתן לומר בביטחון של 95% על הסטייה המרבית בין ממוצע המוגם לתוחלת השכר?
  - מה היה צריך להיות גודל המוגם אם היו רוצחים להקטין את רוח הסמך ב-50%?
  - אם היינו מגדילים את גודל המוגם ובונים רוח סמך באותה רמת סמך האם היה ניתן לטעון בביטחון רב יותר שרוח הסמך מכיל את הפרמטר?

- 5) בנו רוח סמך לממוצע הציוניים של מבחן אינטלייגנציה. ידוע שסטיטית התקן היא 15 והמדד מtabסס על 100 תוצאות. רוח הסמך שהתקבל הוא (105,99). שזרו את :
- ממוצע המדגמים.
  - שגיאת האמידה המקסימאלית.
  - רמת הסמן.
- 6) זמן החלמה מאנגינה מתפלג עם סטיטית התקן של יומיים. חברת תרופות מעוניינת לחקור אנטיביוטיקה חדשה שהיא פיתחה. במחקר השתתפו 60 אנשים שחלו באנגינה וקיבלו את האנטיביוטיקה החדשה. בממוצע הם החלימו לאחר 4 ימים.
- בנו רוח סמך לתוחלת זמן ההחלמה תחת האנטיביוטיקה החדשה ברמת סמך של 90%.
  - מה הייתה קורה לאורך רוח הסמן אם תקציב להגדלת גודל המדגמים פי 4? הסבירו.
  - מה הייתה קורה לאורך רוח הסמן אם היינו בונים את רוח הסמן ברמת סמך גדולה יותר? הסבירו.
- 7) חוקר בנה רוח סמך לממוצע וקיבל את רוח הסמן הבא :  $\mu = 82$ . נתון שסטיטית התקן בהתפלגות שווה ל-10 ושהמדד מtabסס על 16 תוצאות. התפלגות המשתנה היא נורמללית.
- מהו ממוצע המדגמים?
  - מהי רמת הסמך של רוח הסמן שנבנה?
  - מה הסיכוי ששגיאת האמידה באמידת ממוצע האוכלוסייה תעלה על 5?
- 8) חוקר בנה רוח סמך לתוחלת כאשר השונות בהתפלגות ידועה ברמת סמך של 95%. אם החוקר כעת יבנה על סמך אותו נתונים רוח סמך ברמת סמך קטנה מ-95%, איזה מהמשפטים הבאים לא יהיה נכון.
- אורך רוח הסמן החדש יהיה קטן יותר.
  - גודל המדגמים יהיה כעת קטן יותר.
  - המරחק בין ממוצע המדגם לקצota רוח הסמן יהיה קטנים יותר ברוח הסמן החדש.
  - רמת הביטחון לבנות רוח הסמן החדש תהיה קטנה יותר.

9) חוקר בנה רוח סמך ל-  $\mu$  וקיבל:  $48 < \mu < 54$ . מה נכון בהכרח:

- א.  $\mu = 51$ .
- ב.  $\bar{X} = 6$ .
- ג.  $\bar{X} = 51$ .
- ד. אורך רוח הסמך הינו 3.

10) אייזה מהגורםים הבאים אינו משפיע על גודלו של רוח בר סמך, כאשר שונות האוכולוסייה ידועה (בחרו בתשובה הנכונה):

- א. רמת הביטחון.
- ב. סטיית התקן באוכולוסייה.
- ג. מספר המשתתפים.
- ד. סטיית התקן במדגם.

### תשובות סופיות:

- |                               |  |
|-------------------------------|--|
| 1) א. העובדים במשק.           | ב. שכר ב-₪.  |
| ה. $\mu < 9800$ .             | ג. $\mu > 9200$ .  |
| ד. $\mu > 9200$ .             | צ. 0.05.   |
| 2) א. $4979.4 < \mu < 4920.6$ | ב. $0.95 < \mu < 0.600$                                      |
| ב. $237.84 < \mu < 222.16$    | ג. $236.58 < \mu < 223.42$                                   |
| 3) א. ראה סרטון.              |  |
| ב. $10,116 < \mu < 9284$ .    | ג. $416 < \mu < 95%$ .                                       |
| 4) א. לא.                     | ב. המירבית בין $\bar{x}$ ל- $\mu$ היא 416 נט בביטחון של 95%. |
| ב. 0.800.                     | ג. 0.9544.   |
| ג. 0.102.                     | ד. 0.3.  |
| ד. 0.102.                     | א. 0.9544.   |
| 5) א. 0.9544.                 | ב. פיקטן פיאר.   |
| ב. גדל.                       | ג. קטן.  |
| 6) א. 4.42 < $\mu < 4.83$ .   | ב. 0.5.  |
| ב. 0.87.                      | ג. 0.87.   |
| 7) א. 0.87.                   | ב. 0.87.   |
| ב. 0.87.                      | ג. 0.87.   |
| 8) א. 0.87.                   |  |
| ב. 0.87.                      |  |
| 9) א. 0.87.                   |  |
| ד. 0.87.                      |  |
| 10) א. 0.87.                  |  |

## קביעת גודל מוגן:

**רקע:**

אם מעוניינים לאמוד את ממוצע האוכלוסייה כאשר סטיטית התקן של האוכלוסייה ידועה:  $\sigma$  ברמת סמך של  $\alpha=1$  ושיגיאת אמידה שלא עלתה על  $\epsilon$  מסויים, נציב

$$n \geq \left( \frac{z_{1-\frac{\alpha}{2}} \cdot \sigma}{\epsilon} \right)^2$$

בנוסחה הבאה:

כדי להציג בנוסחה צרייך שהמשתנה הנחקר יתפלג נורמלית או שהמוגן ייצא בגודל של לפחות 30 תצפיות.

**דוגמה (פתרון בהקלטה):**

חברת תעופה מעוניינת לאמוד את תוחלת משקל המטען של נוסע. נניח שמשקל מטען של נוסע מתפלג נורמללית עם סטיטית התקן של 2 ק"ג. כמה נוסעים יש לדוגם אם מעוניינים שבבביחוון של 98% הסטייה המרבית בין ממוצע המוגן לממוצע האמתי לא עליה על 0.5 ק"ג? (תשובה: 87).

**שאלות:**

- (1)** משתנה מקרי מתפלג נורמללית עם סטיטית תקן ידועה 12. מה צריך להיות גודל המדגם כדי לבנות רוח סמך ברמת סמך של 98% שאורכו לא עולה על 2?
- (2)** מעוניינים לאמוד את הדופק הממוצע של מתגייסים לצבאי. מעוניינים שבביטחון של 95% שגיאת האמידה המרבית תהיה 0.5. נניח שהדופק מתפלג נורמלית על סטיטית תקן של 3 פעימות לדקה.
- כמה מתגייסים יש לדוגום?
  - אם ניקח מדגם הגדל פי 4 מהמדד של סעיף א' ונאמוד את הממוצע באותה רמת סמך כיצד הדבר ישפייע על שגיאת האמידה?
- (3)** יהיו  $X$  משתנה מקרי עם ממוצע  $\mu$  וסטיטית תקן  $\sigma$ . חוקר רוצה לבנות רוח בר סמך ל- $\mu$  ברמת ביטחון של 0.95, כך שהאורך של הרוח יהיה  $\sigma = 0.5$ . מהו גודל המדגם הנדרש?

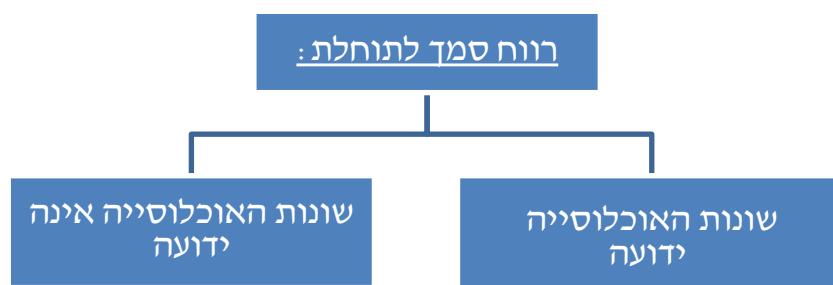
**תשובות סופיות:**

- (1) .780  
 (2) א. 139  
 (3) .  $n = 62$
- ב. הדבר יקטין את  $\sigma$  פי 2.

## רוח סמך כשיעוריות האוכלוסייה לא ידועה:

**רקע:**

בבואהנו לבנות רוח סמך לתוחלת אנו צריכים להתמקד בשני המצביעים הבאים:



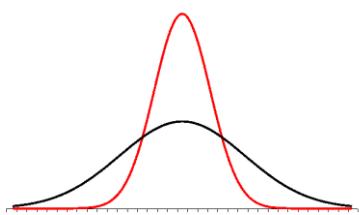
בפרק זה עוסק במקרה ש**שונות האוכלוסייה** אינה ידועה לנו.

מקרה יותר פרקטיבי.

**התנאי:**  $N \sim X$  או שהמדגם גדול.

**רוח סמך:**  $\bar{X} \pm t_{\frac{n-\alpha}{2}} \cdot \frac{S}{\sqrt{n}}$ .

$$\text{האומד לשונות: } S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} = \frac{\sum_{i=1}^n X_i^2 - n\bar{X}^2}{n-1}$$



**התפלגות T:**

הינה התפלגות סימטרית בעומניות שהתוחלת שלה היא 0. ההתפלגות דומה

להתפלגות Z רק שהיא רחבה ולכון הערכים שלה יהיו יותר גבוהים.

התפלגות T תלויות במושג שנקרוא דרגות החופש. דרגות החופש הן:  $df = n - 1$ .

כל שדרגות החופש עלות התפלגות הופכת להיות יותר גבוהה וצרה.

שדרגות החופש שואפות לאינסוף התפלגות T שואפת להיות כמו התפלגות Z.

**דוגמה (פתרון בהקלטה):**

הזמן שלוקח לפתרון שאלה מסוימת ביחסו מתפלג אצל תלמידי כיתה ח' נורמלית.

במטרה לאמוד את תוחלת זמן הפתרון נדגומו 4 תלמידים בכיתה ח'. להלן התוצאות

שהתקבלו בבדיקות: 4.7, 5.2, 4.6, 5.3.

בנו רוח סמך ברמת סמך של 95% למספר זמן הפתרון לשאלת קרב תלמידי כיתה ח'.

**שאלות:**

- 1)** מחקר מעוניין לדעת כיצד תרופה מסוימת משפיעה על קצב פעימות הלב.  
ל-5 אנשים שנטלו את התרופה מדדו את הדופק והתקבל מספר פעימות לדקה : 84, 84, 88, 79, 89.  
הערה : לצורך פתרון הנח שקצב פעימות הלב מתפלג נורמלית בקירוב.  
א. בנו רוח סמך ברמת סמך של 95% לתוחלת הדופק של נוטלי התרופה הניל.  
ב. נתון שהדופק הממוצע ללא לקיחת התרופה הינו 70. לאור זאת, האם בביטחון של 95% התרופה משפיעה על הדופק?  
ג. בהמשך לסעיף א', אם היינו בונים את רוח הסמך ברמת ביטחון של 99%, כיצד הדבר היה משפיע על רוח הסמך?
- 2)** במדגם שנעשה על 25 מתגייסים לצבע האמריקאי התקבל כי גובה ממוצע של חיל הינו 178 ס"מ עם סטיית תקן :  $S = 13$  ס"מ.  
בנו רוח סמך ברמת סמך של 90% לתוחלת גובה המתגייסים לצבע האמריקאי. מה יש להניח לצורך פתרון?
- 3)** אדם מעוניין לאמוד את זמן הנסיעה הממוצע שלו לעבודה. לצורך כך הוא דוגם 5 ימים שזמן הנסעה בהם בבדיקות הוא : 27, 34, 32, 40, 30.  
א. ברמת ביטחון של 95% אמוד את זמן הנסעה הממוצע. מהי ההנחה הדורשahn לorzuch פתרון?  
ב. איך גודל רוח הסמך היה משתנה אם היו דוגמים עוד ימים?
- 4)** ציוני מבחן אינטיליגנציה מתפלגים נורמלית. נדגו 25 מבחנים והתקבל ממוצע ציונים 102 וסטיית תקן מדגמית 13.  
א. בנו רוח סמך לממוצע הציונים באוכלוסייה ברמת ביטחון של 95%.  
ב. חזרו על סעיף א' אם סטיית התקן הינה סטיית התקן האמתית של כלל הנבחנים.  
ג. הסבירו את ההבדלים בין שני השעיפים הניל.
- 5)** נשקלו 60 תינוקות אשר נולדו בשבוע ה-40 של ההריון. המשקל נמדד בKİLOGRAMIM. להלן התוצאות שהתקבלו :  $\sum_{i=1}^{60} X_i^2 = 643.19$  ,  $\sum_{i=1}^{60} X_i = 195$ .  
בנו רוח סמך ברמת סמך של 95% לתוחלת משקל תינוק ביום היולדו.

- 6) נדגו 120 אנשים אקראים מעל גיל 50. עבור כל אדם נבדק מספר שנות השכלתו. להלן התוצאות שהתקבלו :  $S = 2$  ,  $\bar{x} = 13.8$  .  
בנו רוח סמך ברמת סמך של 96% למומוצע ההשכלה של אזרחים מעל גיל 50.
- 7) שני סטטיסטיקים בנו רוח בר-סמך לאותו פרמטר  $\mu$  .  
 לכל אחד מהסטטיסטיקים מדגם אחר, אך באותו גודל 10.  
 שניהם קבעו אותה רמת סמך.  
 סטטיסטיκי א : הניח  $20 = \sigma$  .  
 סטטיסטיκי ב : חישב לפיה המדגם וקיבל  $20 = S$  .  
 למי משני הסטטיסטיקים יהיה רוח סמך ארוך יותר?  
 א. סטטיסטיκי א.  
 ב. סטטיסטיκי ב.  
 ג. אותו ארוך רוח סמך לשני הסטטיסטיקים.  
 ד. תלוי בתוצאות המדגם של כל סטטיסטיκי.
- 8) נתון ש :  $N(\mu, \sigma^2) \sim X$  ביצעו מדגם בגודל 16 וקיבלו סטיית תקן מדגימות 10.  
 אורך רוח הסמך שהתקבל הוא : 8.765. מהי רמת הביטחון של רוח הסמך?

### תשובות סופיות:

- (1) א.  $\mu < 89.72$       ב.  $\mu > 79.88$   
 ג. הוא היה גדול.      ד. ראה בסרטון.
- (2) א.  $\mu < 107.37$       ב.  $\mu > 96.63$   
 ג. לא ניתן לדעת.      ד. צרייך להניח שהמשתנה מתפלג נורמלית.
- (3) א.  $\mu < 107.37$       ב.  $\mu > 96.63$   
 ג. ראה בסרטון.      ד.  $3.149 < \mu < 3.351$
- (4) א.  $\mu > 14.18$       ב.  $\mu < 13.42$   
 ג.  $\mu > 90\%$       ד.  $\mu < 14.18$
- (5) א.  $\mu > 13.42$       ב.  $\mu < 14.18$   
 ג.  $\mu > 90\%$       ד.  $\mu < 14.18$
- (6) א.  $\mu > 13.42$       ב.  $\mu < 14.18$   
 ג.  $\mu > 90\%$       ד.  $\mu < 14.18$
- (7) א.  $\mu > 13.42$       ב.  $\mu < 14.18$   
 ג.  $\mu > 90\%$       ד.  $\mu < 14.18$
- (8) א.  $\mu > 13.42$       ב.  $\mu < 14.18$   
 ג.  $\mu > 90\%$       ד.  $\mu < 14.18$

## סטטיסטיקה

פרק 19 - רוח סמר להפרש תוחלות (ממוצעים) במדגים בלתי תלויים

תוכן העניינים

1. כישרונות האוכלוסייה לא ידועות ובחינת שווין שונות.....  
74 .....

## כשונות האוכלוסייה לא ידועות ובהנחה שוויון שונויות:

**רקע:**

המטרה היא לאמוד את פער התוחולות:  $\mu_2 - \mu_1$ , כולם ההבדלים של הממוצעים בין שתי האוכלוסיות.

**האומד נקודתי:**  $\bar{x}_1 - \bar{x}_2$ .

**התנאים לבניית רוח סמך:**

$$\cdot \sigma^2_1 = \sigma^2_2 \cdot 1$$

$$\cdot X_1, X_2 \sim N \cdot 2$$

3. מוגדים בلتוי תלויים.

**השונות המשוקלلات:** כיון שאנו מניחים שבין שתי האוכלוסיות השונות שווות אנו אומדים את השונות הזו על ידי שקולן שתי השונות של שני המוגדים על ידי

$$\text{הנוסחה הבאה: } S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

**דרגות החופש:**  $d.f = n_1 + n_2 - 2$

$$\text{רוח סמך: } (\bar{x}_1 - \bar{x}_2) \pm t_{1-\frac{\alpha}{2}}^{n_1+n_2-2} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$$

אם הערך אפס נופל בגבולות רוח סמך נגד שבביטחון של  $\alpha - 1$ , לא קיים הבדל בין התוחולות.

**דוגמה (פתרון בהקלטה):**

מחקר מעוניין לבדוק האם קיים הבדל בין תל אביב לברק שבע מבחן הכנסה הממוצעת של אקדמיים. להלן תוצאות המוגדים שנעשה:

טל אביב	ברק שבע	מספר האקדמאים
20	10	
11,000	9500	ממוצע הכנסות של אקדמאים
200	250	סטיית התקן של הכנסות אקדמיים

בנו רוח סמך ברמת ביטחון של 90% להפרש תוחולות הכנסה בשני האזוריים. הניחו שהשכר מתפלג נורמלית עם אותה שוננות בכל אחד מהאזורים.

**שאלות:**

- 1) נדגמו 15 ישראלים ו-15 אמריקאים. כל הנדגמים נגשו ל מבחן IQ.  
להלן תוצאות המדגמים:

ארה"ב	ישראל	המדינה
15	15	גודל המדגם
1470	1560	סכום הציונים
147,560	165,390	סכום ריבועי הציונים

מצאו רוח סמך ברמת סמך של 95% לסתיטה בין ממוצע הציונים בישראל לממוצע הציונים בארה"ב. רשמו את כל הנקודות הדרושות לצורך פתרון התרגילים.

- 2) להלן 4 תוצאות על משתנה  $X$  שמתפלג:  $N(\mu_x, \sigma^2)$ , ומשתנה  $Y$  שמתפלג:  $N(\mu_y, \sigma^2)$ .

X	22	20	21	25
Y	18	25	17	12

חשבו רוח סמך ל-  $\mu_x - \mu_y$  ברמת הסמך 90%, בהנחה שני המדגמים בלתי תלויים.

**תשובות סופיות:**

- 1) הנקודות:  
 1. השונות שווה.  
 2. שהציונים מתפלגים נורמלית.  
 3. המדגמים אינם תלויים זה בזה.  
 $-5.52 < \mu_1 - \mu_2 < 17.52$   
 $-9.6 < \mu_y - \mu_x < 1.6$  (2)

## סטטיסטיקה

פרק 20 - רוח סマー לתוכלת (מומוצע) ההפרשים במדגים מזוגים

תוכן העניינים

1. רוח סマー לתוכלת (מומוצע) ההפרשים במדגים מזוגים ..... 76

## רוח סמך לתוחלת (ממוצע) ההפרשים במדגים مزוגים:

---

**רקע:**

**מבחן מזוג:** מבחן אחד שבו יש  $n$  צמדים. כל תצפית במדגם תנפק זוג ערכים:  $X$  ו- $Y$ . ניצור משתנה חדש:  $D = y - x$ . הפרמטר שנרצה לאמוד:  $\mu_D$ . התנאים לבניית רוח הסמך:  $x, y \sim N(1)$ .

2. המבחן מזוג.

נוסחת רוח הסמך:  $\bar{D} \pm t_{\frac{\alpha}{2}}^{n-1} \frac{S_D}{\sqrt{n}}$ . כאשר דרגות החופש:  $df = n - 1$ .

**דוגמה (פתרון בהקלטה):**

מעוניינים לבדוק האם יש הבדל בין מהירות הריצות של שתי תוכנות מחשב. נלקחו 5 קבצים אקראיים והריצו אותם בשתי התוכנות:

5	4	3	2	1	הקובץ
38	46	49	48	25	זמן בתוכנה הראשונה
48	40	42	46	27	זמן בתוכנה השנייה

הניחו כי זמי הריצות מתפלגים נורמלית. נמצא רוח סמך של 95% להפרש תוחלת הזמן בין שתי התוכנות.

**שאלות:**

- 1)** נדגמו 5 סטודנטים ששסיימו את הקורס סטטיסטיקה ב'. להלן הציונים בסמסטר א' ו-ב':

82	75	90	68	74	סמסטר א'
100	76	87	84	80	סמסטר ב'

- נניח שהציונים מתפלגים נורמללית.
- א. בנו רוח סמך ברמת סמך של 95% לתוחלת פער הציונים בין סמסטר א' לבין סמסטר ב'.
- ב. האם על סמך רוח הסמך קיים הבדל בין הסמסטרים מבחינת תוחלת הציונים?
- ג. מה צריך לשנות בתנאים כדי שהمدגמים יהיו בלתי תלויים?
- 2)** במטרה לבדוק האם קיים הבדל בין קווי זהב לבזק מבחינת ממוצע המחרירים לשיחות בין"יל. נדגמו באקראי 7 מדינות ועבור כל מדינה נבדקה עלות דקט שיחת. להלן התוצאות:

חברה/ מדינה	ארה"ב	קווי זהב - Y	בזק - X	יפן	סין	מצרים	פולין	הולנד	קנדה
4.2	3.2	3.5	3	2.2	2.1	1.5			
4.2	3.2	3.3	3.1	1.9	2	1.4			

בහנה והחרירים מתפלגים נורמלית עבור כל חברה, בנו רוח סמך ברמת סמך של 90% לתוחלת הפרש המחרירים של שתי החברות.

**תשובות סופיות:**

- 1)** א.  $\mu_0 < 38$ .      ב. בביטחון של 95% לא קיים הבדל.  
ג. ראה הסבר בסרטון.
- 2)**  $-0.013 < \mu < 0.185$ .

## סטטיסטיקה

### פרק 21 - מבוא לבדיקה השערות על פרמטרים

#### תוכן העניינים

78 .....	1. הקדמה .....
82 .....	2. סוגים טעויות .....

## הקדמה:

**רקע:**

תהליך של בדיקת השערות הוא תהליך מאד נפוץ בעולם הסטטיסטי. בבדיקה השערות על פרמטרים עוסcid לפיה שלבים הבאים:

**שלב א:** נזהה את הפרמטר הנחקר.

**שלב ב:** נרשום את השערות המחקר. השערת האפס המסומנת ב- $H_0$ .

בדרך כלל השערת האפס מסמלת את אשר היה מקובל עד עכשו, את השגרה הנורמה.

השערה אלטרנטיבית (השערת המחקר) המסומנת ב- $H_1$ .

ההשערה האלטרנטיבית מסמלת את החדשנות בעצם ההשערה האלטרנטיבית בדברת על הסיבה שהמחקר נעשה היא שאלת המחקר.

**שלב ג:** נבדוק האם התנאים לביצוע התהליך מתקיימים ונניח הנחות במידת הצורך.

**שלב ד:** נרשום את כל ההכרעה. בתהליך של בדיקת השערות יוצרים כלל שנראה כלל הכרעה. הכלל יוצר אзорים שנראים:

1. **אזור דחיה:**

דחיה של השערת האפס כלומר קבלה של האלטרנטיבית.

2. **אזור קבלה:**

קיבלה של השערת האפס ודחיה של האלטרנטיבית. כלל ההכרעה מתבסס על איזשהו סטטיסטי. אזור הדחיה מוכתב על ידי סיכון שלוקח החוקר מראש שנראה רמת מובהקות ומסומן ב- $\alpha$ .

**שלב ה:** בתהליך יש ללקת לתוצאות המדגם וליחס את הסטטיסטי המתאים ולבדוק האם התוצאות נופלות באזור הדחיה או הקבלה.

**שלב ו:** להסיק מסקנה בהתאם לתוצאות המדגם.

**דוגמה (פתרון בהקלטה):**

משרד הבריאות פרסם משקל ממוצע של תינוקות ביום לידתם בישראל 3300 גרם. משרד הבריאות רוצה לחזור את הטענה שנשים מעשנות בזמן ההריון יולדות תינוקות במשקל נמוך מהממוצע. במחקר השתתפו 20 נשים מעשנות בהריון. להלן תוצאות המדגם שבדק את המשקל של התינוקות בעת הלידה:

$$\bar{X} = 3120, S = 280, n = 20.$$

- א. מהי אוכלוסיות המחקר?
- ב. מה המשתנה הנחקר?
- ג. מה הפרמטר הנחקר?
- ד. מהן השערות המחקר?

**שאלות:****בשאלות הבאות, ענו על הטעיפים הבאים:**

- א. מהי אוכלוסיית המחקר?
- ב. מה המשנה הנחקר?
- ג. מה הפרמטר הנחקר?
- ד. מהן השערות המחקר?

- (1) ממוצע הציונים בבחינת הבגרות באנגלית הנו 72 עם סטיטית תקן 15 נקודות. מורה טוען שפיתחה שיטה לימוד חדשה שתעלה את ממוצע הציונים. משרד החינוך החליט לתת למורה 36 תלמידים אקראים. ממוצע הציונים של אותם תלמידים לאחר לימודו בשיטתו היה 75.5.
- (2) לפי הצהרת היিירן של חברת משקאות מסויימת נפח הנוזל בבקבוק מתפלג נורמלית עם תוחלת 500 סמ"ק וסטיטית תקן 20 סמ"ק. אגודה הרכנים מתלוננת על הפחתת נפח המשקה בבקבוק מהכמות המוצחרת. במדוג שעשתה אגודה הרכנים התקבל נפח ממוצע של 492 סמ"ק במדוג בגודל 25.
- (3) במשך שנים אחדו המועמדים שהתקבלו לפיקולטה למשפטים היה 25%. השנה מתוך מדגם של 120 מועמדים התקבלו 22. מחקר מעוניין לבדוק האם השנה מקשים על הקבלה לפיקולטה למשפטים.
- (4) בחודש ינואר השנה פורסם שאחדו האבטלה במשק הוא 8% במדוג עכשווי התקבל שמתוך 200 אנשים 6.5% מובטלים. רוצחים לבדוק ברמת מובהקות של 5% האם אחדו האבטלה הוא כמו בתחילת השנה.

**תשובות סופיות:**

- ב. ציון.
- 1) א. נבחנים בברירות באנגלית.  
 $H_0: \mu = 72$       ג. ממוצע הציונים בשיטת לימוד חדשה.  
 $H_1: \mu > 72$
- ב. נפח משקה בבקבוק של חברת מסויימת.
- 2) א. משקאות בבקבוק של חברת מסויימת.  
 $H_0: \mu = 500$       ג. ממוצע נפח המשקה בבקבוק.  
 $H_1: \mu < 500$
- ב. משתנה דיכוטומי (התקבל, לא התקבל).
- 3) א. מועמדים לפיקולטה למשפטים.  
 $H_0: p = 0.25$       ג. אחוז הקבלה.  
 $H_1: p < 0.25$
- ב. משתנה דיכוטומי (מובטל, עובד).
- 4) א. אזרחים בוגרים במשק.  
 $H_0: p = 0.08$       ג. אחוז האבטלה ביום.  
 $H_1: p \neq 0.08$

## סוגי טעויות:

### רַקֵּעַ:

בתחילת בדיקת השערות יוצרים כלל שנקרא כלל הכרעה. הכלל יוצר אזורים שנקראים:

1. אזור דחיה – דחיה של השערת האפס כלומר קבלה של האלטרנטיבה.
2. אזור קבלה – קבלה של השערת האפס ודחיה של האלטרנטיבה.

כל הכרעה מתבסס על איזשהו סטטיסטי.

בתחילת יש ל选取 תוצאות המדגם ולבזוק האם התוצאות נופלות באזור הדחיה או הקבלה וכן להגעה למסקנה – המסקנה היא עירובן מוגבל כיוון שהיא תלולה בכל הכרעה ובתוצאות המדגם. אם נשנה את כלל הכרעה אז אנחנו יכולים לקבל מסקנה אחרת. אם נבצע מדגם חדש אז אנחנו עלולים לקבל תוצאה אחרת. לכן יתכונו טעויות במסקנות שלנו:

		הכרעה	
מציאות		$H_0$	$H_1$
	$H_0$	אין טעות 1	טעות מסוג 1
	$H_1$	טעות מסוג 2	אין טעות

### הגדרת הטעויות:

טעות מסוג ראשון: להכריע לדוחות את  $H_0$  למראות שבמציאות  $H_0$  נכונה.

טעות מסוג שני: להכריע לקבל את  $H_0$  למראות שבמציאות  $H_1$  נכונה.

### דוגמה (פתרון בהקלטה):

אדם חשוד בביוץ עבירה ונتابע בבית המשפט.

אילו סוגי טעויות אפשריות בהכרעת הדין?

**שאלות:**

- 1)** לפי הצהרת היכרן של חברת משקאות מסוימת נפח הנוזל בבקבוק מתפלג נורמלית עם תוחלת 500 סמ"ק וסטיית תקן 20 סמ"ק. אגודת הcrcנים מתלוננת על הפחתת נפח המשקה בבקבוק מהכמות המוצחרת. במדוג שעשתה אגודת הcrcנים התקבל נפח ממוצע של 492 סמ"ק במדוג בגודל 25. בסופו של דבר הוחלט להזכיר לטובת חברת המשקאות.
- רשמו את השערות המחקר.
  - מה מסקנת המחקר?
  - אייזו סוג טעות יתכן וביצעו במחקר?
- 2)** במחקר על פרמטר מסוים הוחלט בסופו של דבר לדוחות את השערת האפס.
- אם ניתן לדעת אם בוצע טעות במחקר?
  - מה סוג הטעות האפשרית?
- 3)** לפי נתוני משרד הפנים בשנת 1980 למשפחה ממוצעת היה 2.3 ילדים למשפחה עם סטיית תקן 0.4. ישנה טענה שכיוום ממוצע מספר הילדים במשפחה קטן יותר. לצורך כך הוחלט לדוגם 121 משפחות. במדוג התקבל ממוצע 2.17 ילדים למשפחה. על סמך תוצאות המדוג נקבע שלא ניתן לקבוע שבאופנו מובהק תוחלת מספר הילדים למשפחה קטנה כיום.
- מהי אוכלוסיות המחקר?
  - מה המשנה הנחקרה?
  - מה הפרמטר הנחקר?
  - מה השערות המחקר?
  - מה מסקנת המחקר?
  - מי סוג הטעות האפשרית במחקר?

**תשובות סופיות:**

- 1)** א.  $\mu = 500$ .  
ב.  $\mu < 500$ .
- 2)** א. לא ניתן לדעת.  
ב. טעות מסווג ראשון.
- 3)** א. משפחות כיום.  
ב. מס' הילדים.
- ג. תוחלת מספר הילדים למשפחה כיום.  
ה. לא לדוחות את  $H_0$ . ו. טעות מסווג שני.
- $H_0 : \mu = 2.3$   
 $H_1 : \mu < 2.3$

## סטטיסטיקה

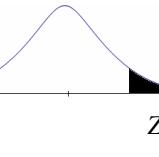
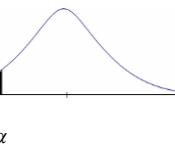
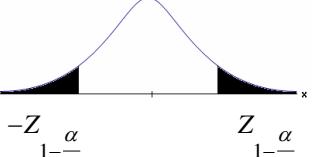
### פרק 22 - בדיקת השערות על תוחלת ( ממוצע )

#### תוכן העניינים

1. בדיקת השערות על תוחלת ( ממוצע ) כשינויים האוכלוסייה ידועה.	84
2. סיכון לטעויות ועוצמה (שינויים האוכלוסייה ידועה)	88
3. מובהקותות תוצאה - אלף מינימלית (שינויים האוכלוסייה ידועה)	94
4. בדיקת השערות על תוחלת ( ממוצע ) כשינויים האוכלוסייה לא ידועה.	99
5. מובהקותות תוצאה - אלף מינימלית (שינויים האוכלוסייה לא ידועה).	103
6. הקשר בין רוח סמך לבדיקה השערות על תוחלת ( ממוצע )	106
7. ניתוח פלטים.	108

## בדיקות השערות על תוחלת (ממוצע) כשבונות האוכלוסייה ידועה:

**רקע:**

$H_0 : \mu \leq \mu_0$	$H_0 : \mu \geq \mu_0$	$H_0 : \mu = \mu_0$	השערת האפס: השערת אלטרנטיבית:
$H_1 : \mu > \mu_0$	$H_1 : \mu < \mu_0$	1. $\sigma$ ידועה או מוגן מספיק גדול $X \sim N$ .2	
$Z_{\bar{x}} > Z_{1-\alpha}$ 	$Z_{\bar{x}} < -Z_{1-\alpha}$ 	$Z_{\bar{x}} < -Z_{\frac{1-\alpha}{2}}$ או $Z_{\bar{x}} > Z_{\frac{1-\alpha}{2}}$ 	<b>כלל</b> ההכרעה: אזור הדחיה של $H_0$
$H_0$ -דוחים את 	$H_0$ -דוחים את 	$H_0$ -דוחים את 	

**סטטיסטי המבחן:**  $Z_{\bar{x}} = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$

**חלופה אחרת לכל הכרעה:**

$\bar{X} > \mu_0 + Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$	$\bar{X} < \mu_0 - Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$	$\bar{X} > \mu_0 + Z_{\frac{1-\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$ או $\bar{X} < \mu_0 - Z_{\frac{1-\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$	<b>נתקיימת <math>H_0</math> אם</b>
--	--	--	------------------------------------

**דוגמה:**

יבול העגבנייהות מתפלג נורמלית עם תוחלת של 10 טון לדונם וסטיית תקן של 2.5 טון לדונם בעונה. משערים ששיטת זיוב חדשת تعالה את תוחלת היבול לעונה מבלי לשנות את סטיית התקן. נדגמו 4 חלוקות שזובלו בשיטה החדשת. היבול הממוצע שהתקבל היה 12.5 טון לדונם. בדקו את ההשערה ברמת מובהקות של 1%.

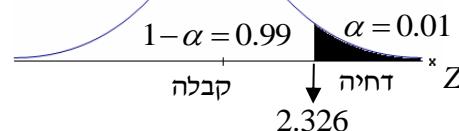
**פתרונות:**אוכלוסייה: עגבנייהות.המשתנה:  $X$  = יבול העגבנייהות בטון לעונה.הפרמטר:  $\mu$  = תוחלת היבול בשיטה החדשת.

$$\begin{aligned} H_0 : \mu &= 10 \\ H_1 : \mu &> 10 \end{aligned}$$

תנאים:

1.  $X \sim N$ .

2.  $\sigma = 2.5$ .

כל הכלעה:נדחה את  $H_0$  אם  $Z_{\bar{x}} > 2.326$ תוצאות:  $n = 4$ ,  $\bar{x} = 12.5$ 

$$\text{סטטיסטי המבחן} : Z_{\bar{x}} = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

$$\text{נzieb} : Z_{\bar{x}} = \frac{12.5 - 10}{\frac{2.5}{\sqrt{4}}} = 2 < 2.326$$

מסקנה:לא נדחה  $H_0$  (נקבל  $H_0$ ).

ברמת מובהקות של 1% לא נוכל לקבל את הטענה ששיטה החדשת היבול מעלה את תוחלת היבול של העגבנייהות.

**שאלות:**

- 1)** מモוצע הציונים בבחינות הבגרות באנגלית הנו 72 עם סטיית תקן 15 נקודות. מורה טוען שפיתח שיטת לימוד חדשה שתעלה את ממוצע הציונים. משרד החינוך החליט לתת למורה 36 תלמידים אקראים. מモוצע הציונים של אותם תלמידים לאחר שלמדו בשיטתו היה 75.5. בהנחה שגם בשיטתו סטיית התקן תהיה 15 מה מסקנתכם ברמת מובהקות של 5%?
- 2)** לפי הצהרת היצרן של חברת משקאות מסוימת נפח הנוזל בבקבוק מתפלג נורמלית עם תוחלת 500 סמ"ק וסטיית תקן 20 סמ"ק. אגודות היצרנים מתלוננת על הפחתת נפח המשקה בבקבוק מהכמות המומוצרת. במדוגם שעשתה אגודות היצרנים התקבל נפח ממוצע של 492 סמ"ק בוגודל 25.
- מה מסקנתכם ברמת מובהקות של 2.5%?
  - האם ניתן לדעת מה תהיה המסקנה עבור רמת מובהקות גבוהה מ-5%?
- 3)** מהנדס האיכות מעוניין לבדוק אם מכונה מכילה (מאופסת). המכונה כוננה לחתווך מוטות באורך 50 ס"מ. לפי נתוני היצרן סטיית התקן בחיתוך המוטות היא 0.5 ס"מ. במדוגם של 50 מוטות התקבל ממוצע אורך המוט 50.93 ס"מ. מה מסקנתכם ברמת מובהקות של 5%?
- 4)** המשקל המומוצע של הספורטאים בתחום ספורט מסוים הוא 90 ק"ג, עם סטיית תקן 8 ק"ג. לפי דעת מומחים בתחום יש צורך בהורדת המשקל ובשימוש בדיאטה מסוימת לצריכה להביא להורדת המשקל. לשם בדיקתיעילות הדיאטה נלקח מדגם מקורי של 50 ספורטאים ובתום שנה של שימוש בדיאטה התברר שהמשקל המומוצע במדוגם זה היה 84 ק"ג. יש לבדוק בר"מ של 10%, האם הדיאטה גורמת להורדת המשקל.
- 5)** לפי מפרט נתון, על עובי בורג להיות 4 מ"מ עם סטיית תקן של 0.2 מ"מ. במדוגם של 25 ברגים העובי המומוצע היה 4.07 מ"מ. קבעו ברמת מובהקות 0.05, האם עובי הברגים מתאים למפרט. הניחו כי עובי של בורג מתפלג נורמלית וסטיית התקן של עובי בורג היא אכן 0.2 מ"מ.
- 6)** במחקר נמצא שתוצאה היא מובהקת ברמת מובהקות של 5% מה תמיד נכון? בחרו בתשובה הנכונה.
- הגדלת רמת המובהקות לא תנסה את מסקנת המחקר.
  - הגדלת רמת המובהקות תנסה את מסקנת המחקר.
  - הקטנת רמת המובהקות לא תנסה את מסקנת המחקר.
  - הקטנת רמת המובהקות תנסה את מסקנת המחקר.

7) חוקר ערך מבחן דו צדי ברמת מובהקות של  $\alpha$  והחליט לדחות את השערת האפס.

אם החוקר היה עורך מבחן דו צדי ברמת מובהקות של  $\frac{\alpha}{2}$  אז בהכרח:

- א. השערת האפס הייתה נדחתה.
- ב. השערת האפס הייתה לא נדחתה.
- ג. לא ניתן לדעת מה תהיה מסקנתו במקרה זה.

8) שני סטטיסטיקים בדקו השערות:  $H_1: \mu > \mu_0$ ,  $H_0: \mu = \mu_0$  נגדן עברו שנות ידועה ובאותה רמת מובהקות. שני החוקרים קיבלו אותו ממוצע במדגם אך לחוקר א' היה מדגם בגודל 100 ולחוקר ב' מדגם בגודל 200.

- א. אם חוקר א' החליט לדחות את  $H_0$ , מה יהיה חוקר ב'? נמקו.
- ב. אם חוקר א' יחליט לא לדחות את  $H_0$ , מה יהיה חוקר ב'? נמקו.

### תשובות סופיות:

- 1) קיבל  $H_0$ , בר"מ של 5% לא קיבל את הטענה של המורה ששיטת הלימוד שלו מעלה את ממוצע הציונים.
- 2) א. נדחה  $H_0$ , בר"מ של 2.5% קיבל את תלונת אגודות הרכנים בדבר הפחיתה נפח המשקה בבקבוק.  
ב. הגדלנו את רמת המובהקות לכן אנחנו נשארים בדוחיה של  $H_0$  והמסקנה לא משתנה.
- 3) נדחה  $H_0$ , בר"מ של 5% נקבע שהמכונה לא מאופסת.
- 4) נדחה  $H_0$ , בר"מ של 0.1 קיבל את הטענה שהדיאטה עיליה ומפחיתה את המשקל הממוצע.
- 5) קיבל  $H_0$ , בר"מ של 0.05 נזכיר שתוחלת עובי הבורג מתיים למפרט.
- 6) א'.
- 7) ג'.
- 8) א. לדחות.  
ב. לא ניתן לדעת.

## סיכום לטעויות ועוצמה (שינוי האוכלוסייה ידועה):

רקע:

		הכרעה	
		$H_0$	$H_1$
מציאות	$H_0$	אין טעות 1	טעות מסוג 1
	$H_1$	טעות מסוג 2	אין טעות

הגדרת הסתברויות:

הסיכוי לבצע טעות מסוג 1 (רמת מובהקות) :  
 $(\text{לדוחות } H_0 = P_{H_0} (H_0 \text{ נכונה}) | \text{ לדוחות את } H_0)$

הסיכוי לבצע טעות מסוג 2 :  
 $(\text{לקבל } H_0 = P_{H_1} (H_1 \text{ נכונה}) | \text{ לקבל את } H_1)$

רמת בתרון :  
 $(\text{לקבל } H_0 = P_{H_0} (H_0 \text{ נכונה}) | \text{ לקבל את } H_0)$

עוצמה :  
 $(\text{לדוחות את } H_1 = P_{H_1} (H_1 \text{ נכונה}) | \text{ לדוחות את } H_0)$

**התהlixir לחישוב סיכוי לטעות מסוג שני:**

$H_0 : \mu = \mu_0$	$H_0 : \mu = \mu_0$	$H_0 : \mu = \mu_0$	<b>השערת האפס: השערת אלטרנטיבתית:</b>
$H_1 : \mu > \mu_0$	$H_1 : \mu < \mu_0$	$H_1 : \mu \neq \mu_0$ <b>תנאים:</b> 1. $\sigma$ ידועה 2. או מדגם מספיק גדול $X \sim N$ .	
$\bar{X} > \mu_0 + Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$	$\bar{X} < \mu_0 - Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$	$\bar{X} > \mu_0 + Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$ או $\bar{X} < \mu_0 - Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$	<b>כל הכרעה: אזור הדחיה של <math>H_0</math>:</b>
$P_{H_0} \left( \bar{X} < \mu_0 + Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}} \right)$	$P_{H_0} \left( \bar{X} > \mu_0 - Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}} \right)$	$P_{H_1} \left( \mu_0 - Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}} < \bar{X} < \mu_0 + Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}} \right)$	<b>חישוב <math>\beta</math>:</b>

**התפלגות ממוצע המדגמים:**  $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

$$\text{התקנון: } Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

**דוגמה:**

בתחילת השנה חשבו הטלפון הסלולארי הממוצע לאדם היה 200 נק' עם סטיית תקן של 80 נק' לחודש. בעקבות כניסה של חברות טלפון סלולארית חדשות מעונייניות לבדוק האם ביום ממוצע חשבו הטלפון הסלולארי פחות. לצורך בדיקה דגמו באקראי 36 אנשים וחשבו הטלפון הסלולاري שלהם היה 150 נק' בממוצע לחודש.

- רשמו את השערות המחקר ובנו כלל הכרעה במנוחי חישוב ממוצע מדגמי ברמת מובהקות של 5%.
- מה מסקנתכם? איזה סוג טעות אפשרית במסקנה?
- נניח שבמציאות ביום החשבון הממוצע הוא 160 נק'. מה הסיכוי לבצע טעות מסוג שני?
- אם נקבע את רמת המובהקות מסעיף א', כיצד הדבר ישפיע על התשובה מסעיף ג'?

**פתרונות:**א. אוכלוסייה: משלמי חשבון טלפון סלולאר Cioms.המשתנה :  $X = \text{חשבון הטלפון החדש שקלים}$ .הפרמטר :  $\mu$ .

$$\begin{array}{l} H_0: \mu = 200 \\ H_1: \mu < 200 \end{array} \quad \text{השערות:}$$

תנאים :

$$\cdot \mu = 200 \cdot 1$$

$$\cdot n = 36 \cdot 2$$

$$\cdot \bar{X} < \mu_0 - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \quad K = \mu_0 - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

$$\alpha = 0.05$$

$$Z_{1-\alpha} = Z_{0.95} = 1.645$$

$$\cdot K = 200 - 1.645 \cdot \frac{80}{\sqrt{36}} = 178.07$$

ככל הבדיקה: דחה את  $H_0$  אם שקלים  $\bar{X} < 178.07$ 

ב. ברמת מובהקות של 5% נזכיר שאכן ממוצע חשבון הטלפון הסלולרי פחת מתחילת השנה.

$$\begin{array}{l} H_0: \mu_0 = 200 \\ H_1: \mu < 200 \end{array} \quad \text{ג. השערות:}$$

ככל הבדיקה: נדחה את  $H_0$  אם  $\bar{X} < 178.07$ 

$$\cdot H_1: \bar{X} \sim N\left(160, \frac{80^2}{36}\right)$$

$$Z = \frac{178.07 - 160}{\frac{80}{\sqrt{36}}} = 1.36$$

$$\beta = P_{H_1} \left( \bar{X} > 178.07 \mid H_0 \right) = P_{H_1} \left( \bar{X} > 178.07 \right) = 1 - \phi(1.36) = 1 - 0.9131 = 0.0869$$

ד. הקטנת  $\alpha$  מגדילה את  $\beta$ .

**שאלות:**

**1)** נתון ש:  $X \sim N(\mu, \sigma^2 = 1)$ .

להלן השערות של חוקר לגבי הפרמטר  $\mu$ :  $H_0: \mu = 5$ ,  $H_1: \mu = 7$ . מעוניינים ליצור כל הכרעה המתבסס על הסמך תצפית בודדת כך שרמת המובייקות תהיה 5%.

א. עבור אילו ערכים של  $X$  שידגום נדחת השערת  $H_0$ ?

ב. מה הסיכוי לבצע טעות מסוג שני?

ג. אם במדגם התקבל ש-  $X = 6.9$  מה תהיה המסקנה ומה הטעות האפשרית?

**2)** לפי נתוני משרד הפנים בשנת 1980 למשפחה ממוצעת היה 2.3 ילדים למשפחה עם סטטיסטיקת תקן 0.4. מעוניינים לבדוק אם כיוון ממוצע מספר הילדים למשפחה קטן יותר. לצורך כך הוחלט לדגום 121 משפחות. במדגם התקבל ממוצע 2.17 ילדים למשפחה.

א. רשמו כלל הכרעה במונחי ממוצע מדגם קרייטי ברמת מובייקות של 5%.

ב. בהמשך לסעיף א' מה תהיה המסקנה ומהי הטעות האפשרית במסקנה?

ג. אם באמצעות ממוצע מספר הילדים במשפחה פחות לכדי 2.1 מהי העצמה של הכלל מסעיף א'?

**3)**להלן נתונים על תהליכי בדיקת השערות על תוחלת:

$n = 30$ ,  $\sigma = 30$ ,  $H_1: \mu \neq 200$ ,  $H_0: \mu = 225$ .

א. רשמו כלל הכרעה במונחי ממוצע מדגם קרייטי וברמת מובייקות של 10%.

ב. בהמשך לסעיף א', מהי העצמה אם התוחלת שווה ל-195?

ג. הסבירו, ללא חישוב, איך העצמה תשנה אם רמת המובייקות תהיה 5%?

**4)** מפעל לייצור צינורות מייצר צינור שקוטרו מתפלג נורמלית עם תוחלת של 50 מ"מ וסטטיסטיקת תקן של 6 מ"מ. במחלקת ביקורת האיכות דוגמים בכל יום 81 צינורות ומודדים את קוטרם, בצד בדוק, בעזרת מבחן סטטיסטי, האם מכונת הייצור מכוקית כנדרש או שקוטר הצינורות קטן מהדרוש.

א. רשמו את ההשערות ואת כלל ההכרעה ברמת מובייקות של 5%.

ב. אם ביום כלשהו מכונת הייצור התקללה והיא מייצרת את הצינורות שתקלה לא תגלה בבדיקה האיכות? כיצד נקבעת הסתברות זו?

ג. הסבירו ללא חישוב כיצד התשובה לשעיף ב' תשנה אם רמת המובייקות גדל.

ד. הסבירו ללא חישוב כיצד התשובה לשעיף ב' תשנה אם התוחלת האמיתית היא 47 ולא 48 מ"מ.

- 5) להלן השערות של מחקר:  $H_0: \mu = 50$ ,  $H_1: \mu = 58$ .  
 מעוניינים לדגום 100 תכפיות. ידוע שטטיות התקן של ההתפלגות הינה 20.  
 א. בנו כלל הכרעה שהסיכוי לטעות מסוג שני בו הוא 10%.  
 מהי רמת המובהקות?  
 ב. כיצד הייתה משתנה רמת המובהקות אם (כל סעיף בפני עצמו)?  
 i. טטיות התקן הייתה יותר גדולה.  
 ii. הסיכוי לטעות מסוג שני גדול יותר.

**השאלות שלහן הן שאלות רב-ברירה, בחרו בתשובה הנכונה ביותר:**

- 6) אם חוקר החליט להגדיל את רמת המובהקות במחקר שלו אז:  
 א. הסיכוי לטעות מסוג ראשון גדול.  
 ב. העוצמה של המבחן גדולה.  
 ג. הסיכוי לטעות מסוג שני גדול.  
 ד. תשובות א' ו-ב' נכונות.
- 7) חוקר ביצע מחקר ובו עשה טעות מסוג שני בכך:  
 א. השערת האפס נכונה.  
 ב. השערת האפס נדחתה.  
 ג. השערת האפס לא נדחתה.  
 ד. אף אחת מהתשובות לא נכונה בהכרח.

- 8) מה המצב הרצוי לחוקר המבצע בבדיקה השערה:

$\alpha$	$1 - \beta$
א. גדולה	קטנה
ב. גדולה	קטנה
ג. קטנה	גדולה
ד. קטנה	קטנה

- 9) נערך שינוי בכלל ההחלטה של בדיקת השערה מסוימת ובעקבותיו איזור דחיתת  $H_0$  קטן. כל שאר הגורמים נשארו ללא שינוי. כתוצאה לכך:  
 א. הוא  $\alpha$ , והוא  $\beta - 1$ , יקטנו.  
 ב.  $\alpha$  יישאר ללא שינוי ואילו  $\beta - 1$  יגדל.  
 ג.  $\alpha$  יגדל ואילו  $\beta - 1$  יקטן.  
 ד. הוא  $\alpha$  והוא  $\beta - 1$  יגדל.

**10)** ידוע כי לחץ דם תקין באוכלוסייה הוא 120. רופא מניח של לחץ הדם בקרוב עיתונאים גבוה יותר מה ממוצע באוכלוסייה. הואלקח מדגם של 60 עיתונאים וקיים ממוצע 137. על סמך המדגם, הוא בודק טענתו ברמת מובהקות 0.02 ומסיק של לחץ הדם בקרוב העיתונאים אינו גבוה יותר. מה הטעות האפשרית שהרופא עושה?

- א. טעות מסוג ראשון.
- ב. טעות מסוג שני.
- ג. טעות מסוג שלישי.
- ד. אין טעות במסקנותו.

### תשובות סופיות:

- (1) א. מעל 5.646.      ב. 0.3594.      ג. דחינו את  $H_0$ , ת騰ן טעות מסוג ראשון.
- (2) א. נדחה  $H_0$  אם  $\bar{X} < 2.24$ .      ב. 1.  $\bar{X} < 2.24$ .      ג. תקתו.
- (3) א. נדחה  $H_0$  אם  $\bar{X} > 203.29$  או  $\bar{X} < 196.71$ .      ב. 0.8051.      ג. תקתו.
- (4) א. נדחה  $H_0$  אם  $\bar{X} < 48.9$ .      ב. 0.0885.      ג. תקתו.      ד. תקתו.
- (5) א. 0.0033.      ב. נ. רמת המובהקות הייתה קטנה.      ג. נ. רמת המובהקות הייתה גדולה.
- (6) ד. נ.
- (7) ג. נ.
- (8) ג. נ.
- (9) א. נ.
- (10) ב. נ.

## mobekot\_tozacha - alfa minimilit (shonot) haoculosiya idouha):

**רקע:**

דרך נוספת להגעה להכרעות שלא דרך כלל הכרעה, היא דרך חישוב מובהקות התוצאות :

באמצעות תוצאות המדגם מחשבים את מובהקות התוצאה שמסומן ב-  $p_v$ .  
את רמת המובהקות החוקר קובע מראש לעומת זאת, את מובהקות התוצאה החוקר יוכל לחשב רק אחרי שייהיו לו את התוצאות.

המסקנה של המחקר תקבע לפי העיקרונו הבא : אם  $\alpha \leq p_v$ , דוחים את  $H_0$ .  
mobekot\_tozacha זה הסיכוי לקבלת תוצאות המדגם וקיומו מתוצאות אלה בהנחה השערת האפס.

(לקבל את תוצאות המדגם וקיומו)  $\cdot p_v = P_{H_0}$

אם ההשערה היא דו צדדיות :

(לקבל את תוצאות המדגם וקיומו)  $\cdot p_v = 2P_{H_0}$

mobekot\_tozacha היא גם האלפא המינימלית לדחיתת השערת האפס.

$H_0: \mu = \mu_0$	$H_0: \mu = \mu_0$	$H_0: \mu = \mu_0$	$H_1: \mu > \mu_0$	$H_1: \mu < \mu_0$	$H_1: \mu \neq \mu_0$	השערת האפס : השערה אלטרנטיביה :
$H_0: \mu = \mu_0$	$H_0: \mu = \mu_0$	$H_0: \mu = \mu_0$	. $\sigma$ ידועה			
$H_1: \mu > \mu_0$	$H_1: \mu < \mu_0$	$H_1: \mu \neq \mu_0$	או מדגם מספיק גדול $X \sim N$ .2			
$P_{H_0}(\bar{X} \geq \bar{x})$	$P_{H_0}(\bar{X} \leq \bar{x})$	$2 \cdot P_{H_0}(\bar{X} \geq \bar{x}) \iff \bar{x} > \mu_0$ $2 \cdot P_{H_0}(\bar{X} \leq \bar{x}) \iff \bar{x} < \mu_0$	<b>Tנאים :</b>			
			<b>p-value</b>			

כאשר בהנחה השערת האפס :  
 $Z_{\bar{x}} = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} , \bar{X} \sim N\left(\mu_0, \frac{\sigma^2}{n}\right)$

**דוגמה:**

המשקל הממוצע של מתגייסים לצבע לפני 20 שנה היה 65 ק"ג. מחקר מעוניין לבדוק האם כיום המשקל הממוצע של מתגייסים גבוה יותר. נניח שהמשקל המתגייסים מתפלג נורמלית עם סטטיסטיקה של 12 ק"ג. במדגם של 16 מתגייסים התקבל משקל ממוצע של 71 ק"ג.

- מהי מובהקות התוצאה?
- מה המסקנה אם רמת המובהקות היא 5% ואם רמת המובהקות היא ?!

**פתרון:**

a. אוכלוסייה: המתגייסים לצבע ביום.

משתנה:  $X$  = משקל בק"ג.

פרמטר:  $\mu$ .

השערות:  
 $H_0: \mu = 65$   
 $H_1: \mu > 65$

תנאים:

.  $X \sim N$ . 1

.  $\sigma = 12$ . 2

תוצאות מדגם:

$$n = 16$$

$$\bar{X} = 71$$

$$P_V = P_{H_0} \left( \text{لتוצאות המדגמים וקיצוני} \right) = P_{H_0} (\bar{X} \geq 71) = 1 - \phi(2) = 1 - 0.9772 = 0.0228$$

$$Z_{\bar{x}} = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{71 - 65}{12 / \sqrt{16}} = 2$$

$$\alpha_{\min} = 0.0228$$

**שאלות:**

- 1)** להלן השערות של מחקר:  $H_0: \mu = 70$ ,  $H_1: \mu > 70$ .  
 המשתנה הנחקר מתפלג נורמלית עם סטיטית תקן 20.  
 במדגם מאותה אוכלוסייה התקבלו התוצאות הבאות:  $\bar{x} = 74$ ,  $n = 100$ .  
 מהי מובהקות התוצאה?
- 2)** השכר הממוצע במשק בשנת 2012 היה 8800 נס' עם סטיטית תקן 2000. במדגם שנעשה אטמול על 100 עובדים התקבל שכר ממוצע 9500 נס'. מטרת המחקר היא לבדוק האם כיים חלה עלייה בשכר. עבור אילו רמות מובהקות שיבחר החוקר יוחלט שחלла עלייה בשכר הממוצע במשק?
- 3)** אדם חושד שהברת ממתקים לא עומדת בהתחייבותה, ומשקלו של חטייף מסוים אותו הוא קונה מדי בוקר נזוק מ-100 גרם.  
 חברת הממתקים טוענת מצידה שהיא אכן עומדת בהתחייבותה. ידוע כי סטיטית התקן של משקל החטייף היא 12 גרם. האדם מתכוון לשקלול 100 חפיפות חטייפים ולאחר מכן מכון להגיע להחלטה.  
 לאחר הבדיקה הוא קיבל משקל הממוצע של 98.5 גרם.  
 א. רשמו את השערות המחקר.  
 ב. מהי רמת המובהקות המינימלית עבורה דוחים את השערת האפס?  
 ג. מהי רמת המובהקות המקסימלית עבורה קיבל את השערת האפס?  
 ד. מה המסקנה ברמת מובהקות של 5?
- 4)** מכונה לחישוק מוטות בפעול חותכת מוטות באורך שמתפלג נורמלית עם תוחלת אליה כוונה המכונה וסטיטית תקן 2 ס"מ. ביום מסוים כוונה המכונה לחישוק מוטות באורך 80 ס"מ. אחראי האיכות מעוניין לבדוק האם המכונה מכילה. לצורך כך נדרגו מקו הייזור 16 מוטות שנחתכו אורכו הממוצע היה 81.7 ס"מ.  
 א. מהי רמת המובהקות המינימלית עבורה נカリע שהמכונה לא מכילה?  
 ב. אם נסיף עוד ציפוי שערכה יהיה 82 ס"מ, כיצד הדבר ישפיע על התשובה של הסעיף הקודם?  
 ג. הכרע ברמת מובהקות של 5% האם המכונה מכילה.
- 5)** אם מקבלים בחישובים לפחות מינימלית (value P) קטנה מאד, סביר להניח כי החוקר ידחה את השערת האפס בקלות. נכון/לא נכון? נמק.

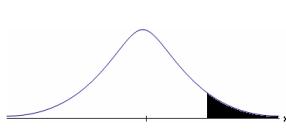
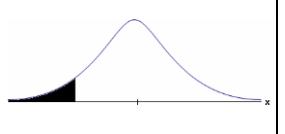
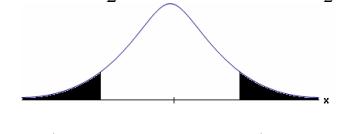
- 6) בבדיקה השערות התקבל שה-  $p-value = 0.02$ . מה תהיה מסקנת חוקר המשמש ברמת מובהקות 1%? בחרו בתשובה הנכונה.
- יקבל את השערת האפס בכל מקרה.
  - ידחה את השערת האפס מקרה.
  - ידחה את השערת האפס רק אם המבחן הנו דו צדדי.
  - לא ניתן לדעת כי אין מספיק נתונים.
- 7) מובהקות התוצאה (PV) היא גם (בחרו בתשובה הנכונה):
- רמת המובהקות המינימאלית לדחינת השערת האפס.
  - רמת המובהקות המקסימאלית לדחינת השערת האפס.
  - רמת המובהקות שנקבעה מראש על ידי החוקר שטרם קיבל את תוצאות המחקר.
  - רמת המובהקות המינימאלית לאי דחינת השערת האפס.
- 8) בבדיקה השערות מסוימת התקבל:  $p value = 0.0254$  לכן (בחרו בתשובה הנכונה):
- ברמת מובהקות של 0.01 אך לא של 0.05 נדחה את  $H_0$ .
  - ברמת מובהקות של 0.01 ושל 0.05 לא נדחה את  $H_0$ .
  - ברמת מובהקות של 0.05 אך לא של 0.01 נדחה את  $H_0$ .
  - ברמת מובהקות של 0.01 ושל 0.05 נדחה את  $H_0$ .

**תשובות סופיות:**

- (1) 0.0228.
- (2) עבר כל רמת מובהקות סבירה.
- (3)  $H_0: \mu = 100$ .  
 $H_1: \mu < 100$ .
- א. 0.1056.      ב. 0.1056.      ג. נכון.
- ד. נכريع שיש עמידה בהתחייבות של החברה.
- (4) א. 0.0006.      ב. יקטן.      ג. נכريع שאין כיול.
- (5) נכון.
- (6) א'.
- (7) א'.
- (8) ג'.

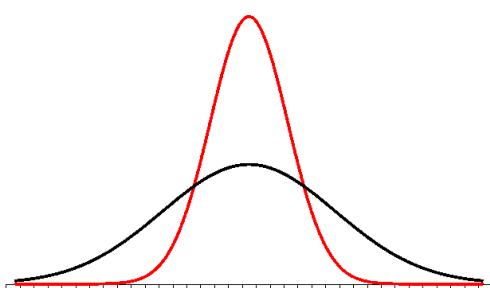
## בדיקות השערות על תוחלת (ממוצע) כשבונות האוכלוסייה לא ידועה:

**רקע:**

$H_0 : \mu = \mu_0$	$H_0 : \mu = \mu_0$	$H_0 : \mu = \mu_0$	<b>השערת האפס: השערה אלטרנטיבית:</b>
$H_1 : \mu > \mu_0$	$H_1 : \mu < \mu_0$	$H_1 : \mu \neq \mu_0$	
.1. $\sigma$ אינה ידועה או מוגן מספיק גדול $X \sim N$ .2			<b>תנאים:</b>
$t_{\bar{x}} > t_{1-\alpha}^{(n-1)}$  $t_{1-\alpha,n-1}$	$t_{\bar{x}} < -t_{1-\alpha}^{(n-1)}$  $-t_{1-\alpha,n-1}$	$t_{\bar{x}} < -t_{\frac{1-\alpha}{2},n-1}^{(n-1)}$ או $t_{\bar{x}} > t_{\frac{1-\alpha}{2},n-1}^{(n-1)}$  $-t_{\frac{1-\alpha}{2},n-1}$ $t_{\frac{1-\alpha}{2},n-1}$	<b>כל הנסיבות: אזור הדחיה של <math>H_0</math>:</b>
$H_0$ - דוחים את ■	$H_0$ - דוחים את ■	$H_0$ - דוחים את ■	
$\bar{X} > \mu_0 + t_{1-\alpha}^{n-1} \cdot \frac{S}{\sqrt{n}}$	$\bar{X} < \mu_0 - t_{1-\alpha}^{n-1} \cdot \frac{S}{\sqrt{n}}$	$\bar{X} > \mu_0 + t_{\frac{1-\alpha}{2}}^{n-1} \cdot \frac{S}{\sqrt{n}}$ או $\bar{X} < \mu_0 - t_{\frac{1-\alpha}{2}}^{n-1} \cdot \frac{S}{\sqrt{n}}$	<b>חלופה לכל הנסיבות: נדחה <math>H_0</math> אם מתקיים:</b>

$$\text{סטטיטיסטי המבחן: } t_{\bar{x}} = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}}$$

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

**התפלגות T:**

הינה התפלגות סימטרית בעומניות שהתוחלת שלה היא 0. ההתפלגות דומה לתפלגות Z רק שהיא יותר רחבה ולכן הערכים שלה יהיו יותר גבוהים. התפלגות T תלויות במושג שנקרא דרגות החופש.

דרגות החופש הן:  $df = n - 1$ .

כל שדרגות החופש עלות התפלגות הופכת להיות יותר גבוהה וצרה. כסדרות החופש שוואפות לאינסוף התפלגות T שואפת להיות כמו התפלגות Z.

**דוגמה (פתרון בהקלטה):**

מפעל קיבל הזמנה לייצור משטחים בעובי של 0.1 ס"מ. כדי לבדוק האם המפעל עומד בדרישה נדגו 10 משטחים ונמצא שהעובי הממוצע הוא 0.104 עם אומדן לסטיתת תקן 0.002 ס"מ.

- מהו השערות המתקי?
- מה ההנחה הדורשא לצורך פתרון?
- בודק ברמת מובהקות של 5%.

**שאלות:**

- 1)** משך זמן ההחלמה בלקיחת אנטיביוטיקה מסויימת הוא 120 שעות בממוצע עם סטיית תקן לא ידועה. מעוניינים לבדוק האם אנטיביוטיקה אחרת מקטינה את משך זמן ההחלמה. במדגם של 5 חולים שלקחו את האנטיביוטיקה האחראית התקבלו זמני ההחלמה הבאים: 125, 100, 95, 90, 80 שעות. מה מסקנתכם ברמת מובהקות של 5% מהי ההנחה הדרושה לצורך הפתרון?
- 2)** משרד הבריאות פרסם שמשקל ממוצע של תינוקות ביום היולדות בישראל 3300 גר'. משרד הבריאות רוצה לחקור את הטענה שנשים מעשנות בזמן ההריון يولדות תינוקות במשקל נמוך מהתמוצע. במחקר השתתפו 20 נשים מעשנות בהריון. להלן תוצאות המדגם שבדק את המשקל של התינוקות בעת הלידה:
- $$n = 20$$
- $$\bar{x} = 3120$$
- $$S = 280$$
- מה מסקנתכם ברמת מובהקות של 5% מה יש להניח לצורך פתרון?
- 3)** ציוני מבחן אינטילגנציה מתפלגים נורמלית. באלה"ב ממוצע הציונים הוא 100. במדגם שנעשה על 23 נבחנים ישראלים, התקבל ממוצע ציונים 104.5 וסטיית התקן המדגמית 16. האם בישראל ממוצע הציונים שונה מאשר באלה"ב? הסיקו ברמת מובהקות של 5%.
- 4)** באוכלוסייה מסוימת נדגמו 10 תכפיות והתקבלו התוצאות הבאות:
- $$\sum_{i=1}^{10} X_i = 750$$
- $$\sum_{i=1}^{10} (X_i - \bar{X})^2 = 900$$
- נתון שההתפלגות היא נורמלית.  
בדוק ברמת מובהקות של 5% האם התוחלת של ההתפלגות שונה מ-80.

- 5) ליאור ורוני העלו את אותן השערות על ממוצע האוכלוסייה. כמו כן הם התבפסו על אותן תוצאות של מדגם. ליאור השתמש בטבלה של התפלגות Z. רוני השתמשה בטבלה של התפלגות t. מה נוכל לומר בנוגע להחלטת המחקר שלהם? בחר בתשובה הנכונה.
- אם ליאור ידחה את השערת האפס אז גם בהכרח רוני.
  - אם רוני תדחה את השערת האפס אז גם בהכרח ליאור.
  - שני החוקרים בהכרח הגיעו לאותה מסקנה.
  - לא ניתן לדעת על היחס בין דמיון השערת האפס של שני החוקרים.

- 6) נתון ש:  $H_0: \mu = \mu_0$  ו-  $H_1: \mu < \mu_0$ .  $X \sim N(\mu, \sigma^2)$  כמו נטוונות ההשעורות הבאות:
- חוקר בדק את ההשעורות הללו על סמך מדגם שככל 10 תצפיות.  $\sigma^2$  לא הייתה ידועה לחוקר. החוקר החליט לדחות את השערת האפס ברמת מובייקות של 5% לאחר מכן כדי לחזק את קביעתו הוא דגם עוד 5 תצפיות וشكلל את תוצאות אלה גם למדגם כך שככל עכשו 15 תצפיות. בחר בתשובה הנכונה:
- cut בברור הוא ידחה את השערת האפס.
  - cut הוא דוקא קיבל את השערת האפס.
  - cut לא ניתן לדעת מה תהיה מסקנתו.

### תשובות סופיות:

- 1) נדחה  $H_0$ .
- 2) נדחה  $H_0$ .
- 3) קיבל  $H_0$ .
- 4) קיבל  $H_0$ .
- 5) ב'.
- 6) ג'.

## mobekot\_tozacha - alfa\_minimalit (shevona) האוכלוסייה לא ידועה):

**רקע:**

נזכיר שהמסקנה של המבחן תיקבע לפי העיקרון הבא: אם  $\alpha \leq p_v$  דוחים את  $H_0$ .  
mobekot\_tozacha היא הסיכוי לקבלת תוצאות המדגם וקיצוני מהתוצאות אלה בהנחה השערת האפס.  
לקבל את תוצאות המדגם וקיצוני) .  
אם ההשערה היא דו צדדית :  
לקבל את תוצאות המדגם וקיצוני) .

mobekot\_tozacha היא גם האלפא המינימלית לדחינת השערת האפס.

$H_0: \mu = \mu_0$	$H_0: \mu = \mu_0$	$H_0: \mu = \mu_0$	$H_0: \mu = \mu_0$	השערת האפס: השערה אלטרנטיבית:	
$H_1: \mu > \mu_0$	$H_1: \mu < \mu_0$	1. $\sigma$ אינה ידועה או 2. מדגם מספיק גדול $X \sim N$			
$P_{H_0}(\bar{X} \geq \bar{x})$	$P_{H_0}(\bar{X} \leq \bar{x})$	$2 \cdot P_{H_0}(\bar{X} \geq \bar{x}) \leftarrow \bar{x} > \mu_0$	$2 \cdot P_{H_0}(\bar{X} \leq \bar{x}) \leftarrow \bar{x} < \mu_0$	p-value	

$$t_{\bar{x}} = \frac{\bar{x} - \mu_0}{\frac{S}{\sqrt{n}}}$$

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

$$d.f = n-1$$

**דוגמה:**

ממוצע זמן הנסיעה של אדם לעובדה הינו 40 דקות. הוא מעוניין לבדוק דרך חלופית שאמורה להיות יותר מהירה. לצורך כך הוא דוגם 5 ימים שבהם הוא נוסע בדרך החלופית. זמני הנסיעה שקיבל בדיקות הם: 34, 40, 30, 32, 27. הנicho שזמן הנסיעה מתפלג נורמלית.

- א. רשמו את השערות המחקר.
- ב. מצאו חסמים לモבಹקות התוצאה.
- ג. מה המסקנה ברמת מובהקות של 5%?

**פתרונות:**

אוכלוסייה: כלל הנסיעות לעובדה בדרך החלופית.

משתנה:  $X =$  זמן נסעה בדיקות.

תנאים:  $X \sim N$ .

פרמטר:  $\mu$ .

א. השערות:  
 $H_0: \mu = 40$   
 $H_1: \mu < 40$

ב. תוצאות המדגם:

$$n = 5, \bar{X} = \frac{\sum X_i}{n} = \frac{34 + 40 + \dots}{5} = 32.6$$

$$S^2 = \frac{\sum X_i^2 - n \cdot \bar{X}^2}{n-1} = \frac{34^2 + 40^2 + \dots - 5 \cdot 32.6^2}{5-1} = 23.4$$

$$S = \sqrt{23.4}$$

$$t_{\bar{X}} = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} = \frac{32.6 - 40}{\frac{4.88}{\sqrt{5}}} = -3.39$$

$$P_V = P_{H_0} = (\bar{X} \leq 32.6) = P(t \leq -3.39)$$

$$d.f = 5 - 1 = 4$$

$$1\% < P_V < 2.5\%$$

$P_V < \alpha = 0.05$ , לכן דוחים את  $H_0$ .

מסקנה: בר"מ של 5% נכרייע שהדרך החלופית מהירה יותר.

**שאלות:**

- 1)** קוו ייצור אריזות סוכר נארזות כך שהמשקל הממוצע של אריזות הסוכר צריך להיות אחד קילוגרם. בכל יום דוגמים מקו הייצור 5 אריזות במטרה לבדוק האם קו הייצור תקין. בבדיקה דגמו 5 אריזות סוכר ולהלן משקלן בגרמים: 1024, 996, 1005, 997, 1008.
- רשמו את השערות המחקר.
  - מהי מובಹקות התוצאות? הצג חסמים.
  - מה המסקנה ברמת מוב hawkות של 5%?
- 2)** חוקר בדק את הטענה כי פועלים העובדים במשמרתليل איטיים יותר מפועלים העובדים ביום. ידוע כי משך הזמן הממוצע הדרוש לייצר מוצר מסוים ביום הוא 6 שעות. בדוגמא מיקרי של 25 פועלים שעבדו במשמרתليل נמצא כי הזמן הממוצע לייצר אותו מוצר הוא 7 שעות עם סטיית תקון של 3 שעות. מהי  $\alpha$ -המינימלית שלפיה ניתן להחליט שacen העובדים במשמרתليل איטיים יותר?
- 3)** הגובה של מתגייםים לצה"ל מתפלג נורמלית. בדוגמא של 25 מתגייםים מדדו את הגבהים שלהם בס"מ והתקבלו התוצאות הבאות:
- $$\sum (x_i - \bar{x})^2 = 2832, \bar{x} = 176.2$$
- מטרת המחקר היא לבדוק האם תוחלת הגבהים של המתגייםים גבוהה מ-174 ס"מ באופן מובהק. מהי בקרוב מוב hawkות התוצאות ועל פייה מה תהיה המסקנה ברמת מוב hawkות של 6%?

**תשובות סופיות:**

- 1)** א.  $H_0: \mu = 1000$       ב.  $20\% \leq P_v \leq 50\%$   
 $H_1: \mu \neq 1000$
- ג. ברמת מוב hawkות של 5% לא נוכל לקבוע שקו הייצור אינו תקין.
- 2)**  $.10\%$
- 3)** נקבל את  $H_0, 1.01$

## הקשר בין רוח סמך לבדיקה השערות על תוחלת (מומוצע):

**רקע:**

ניתן לבצע בדיקת השערות דו צדדיות ברמת מובהקות  $\alpha$  על  $\mu$  :

$$\mu_0 : \mu = \mu_1 , H_0 : \mu \neq \mu_0$$

על ידי בניית רוח סמך ברמת סמך של  $\alpha - 1$  ל-  $\mu$  :

אם  $\mu_0$  נופל ברווח  $\leftarrow$  קיבל את  $H_0$ .

אם  $\mu_0$  לא נופל ברווח  $\leftarrow$  נדחה את  $H_0$ .

**דוגמה:**

חוקר ביצע בדיקת השערות לתוחלת. להלן השערותיו :

$$H_0 : \mu = 80 , H_1 : \mu \neq 80 , \alpha = 5\%$$

החוקר בנה רוח סמך ברמה של 90% וקיבל:  $84 < \mu < 79$ .

האם אפשר לדעת מה מסקנתו, ואם כן מהי?

**פתרון (פתרון מלא בהקלטה):**

רוח הסמך ברמת סמך של 90% מכיל "80".

ברמת סמך של 95% רוח הסמך יגדל וכייל "80".

לכן, ברמת מובהקות של 5% קיבל  $H_0$ .

**שאלות:**

- 1)** חוקר רצה לבדוק את ההשערות הבאות:  $H_0: \mu = 90$ ,  $H_1: \mu \neq 90$ . החוקר בנה רוחח סמך לתוכלת ברמת סמך של 95% וקיבל את רוחח הסמך הבא: (87, 97). אם החוקר מעוניין לבצע בדיקת השערות ברמת מובהקות של 1% האם ניתן להגיע למסקנה ע"י רוחח הסמך? נמקו.
- 2)** חוקר מעוניין לבדוק השפעת דיאטה חדשה על רמת הסוכר בدم. ידוע כי מספר מיליגרים הסוכר בסמ"ק דם הוא משתנה מקרי שמתפלג נורמלית עם סטיית תקן 10.4 מ"ג. נלקח מדגם של 60 נבדקים שניזונו מדיאטה זו. נמצא כי ממוצע מספר המיליגרים סוכר היה 115.5 מ"ג לסמ"ק.
- א. בנה רוחח סמך ברמת סמך 95% לתוכלת רמת הסוכר בדם אצל הניזונים מדיאטה זו.
- ב. ידוע שתוחלת רמת הסוכר בדם באוכלוסייה היא 90 מ"ג לסמ"ק. האם לדעתך ניתן להסיק על סמך תוצאת סעיף א' שהדיאטה משפיעה על רמת הסוכר בדם? הסבירו.
- 3)** יצרן אנטיביוטיקה רושם על גבי התרופות שכמות הפנצליין היא 200 מ"ג لكפסולה. משרד הבריאות ביצע מדגם של 8 קפסולות אקרראיות מקו הייצור ומצא שבממוצע יש 196 מ"ג פנצליין لكפסולה עם סטיית תקן מדגמית של 5 מ"ג. בהנחה וכמות הפנצליין בקפסולה מתפלגת נורמלית.
- א. בנו רוחח סמך ברמת סמך של 95% למומוצע כמות הפנצליין لكפסולה המיוצרת על ידי יצרן האנטיביוטיקה.
- ב. בדקו ברמת מובהקות של 5% האם יש אמת באינפורמציה המופיעukt על ידי הייצן.

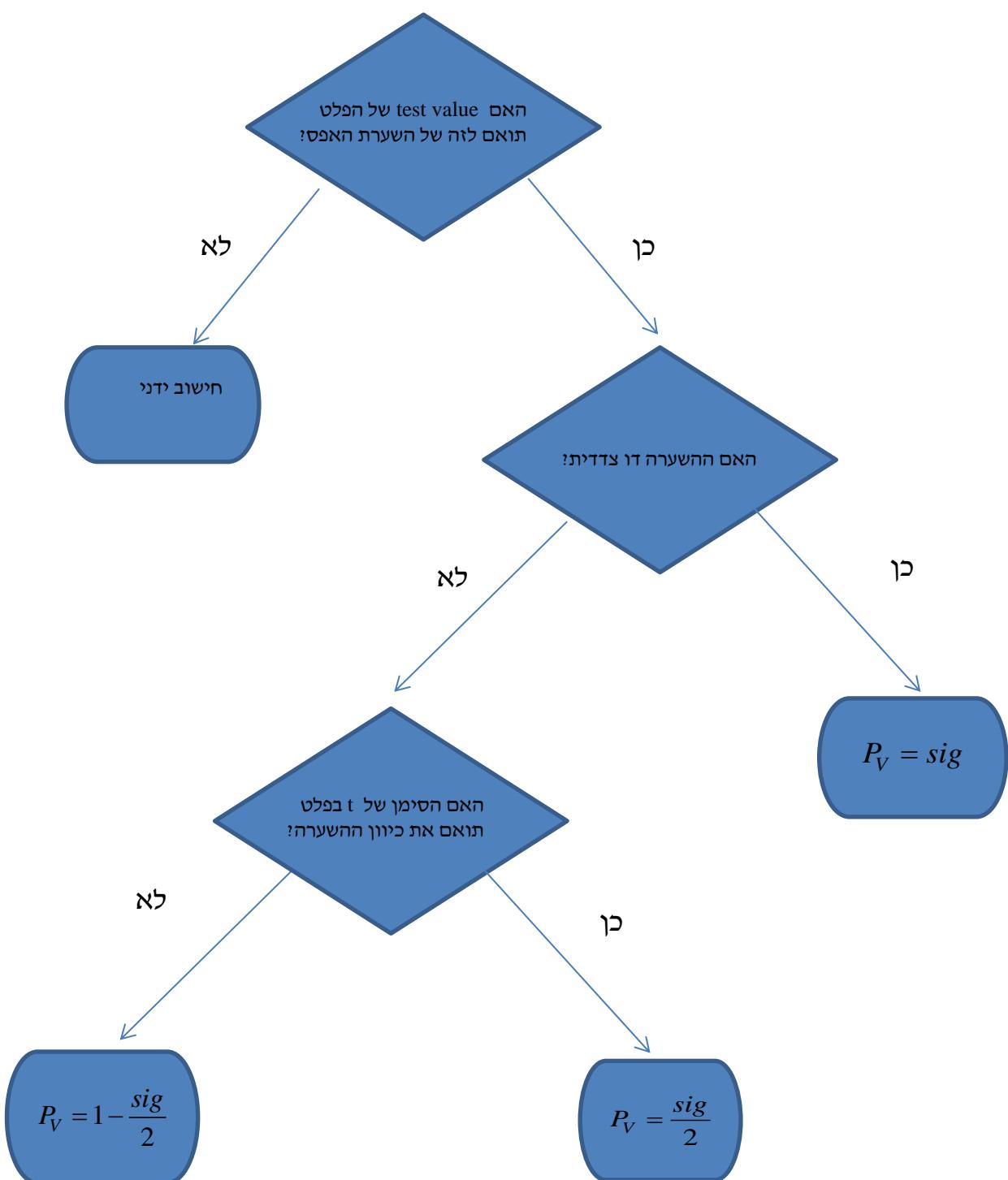
**תשובות סופיות:**

- 1)** קיבל השערת.
- 2)** א.  $112.87 \leq \mu \leq 118.13$   
ב. נזכיר שהדיאטה משפיעה על תוכלת רמת הסוכר בדם.
- 3)** א.  $200.2 \leq \mu \leq 191.8$ . ב. נזכיר שיש אמת בפרסום.

## ניתוח פלטים:

רעיון:

חישוב מובהקות התוצאה באמצעות פלט תוכנת SPSS :



**דוגמה (פתרון בהקלטה):****One-Sample Statistics**

	N	Mean	Std. Deviation	Std. Error Mean
X	25	87.6400	64.90434	12.98087

**One-Sample Test**

X	Test Value = 60					
	t	df	Sig. (2-tailed)	Mean	95% Confidence Interval of the Difference	
				Difference	Lower	Upper
X	2.129	24	.044	27.64000	.8488	54.4312

ממוצע הציונים ב מבחן המיצב בחשבון הוא 60. הוחלט לדגם כייתה אקראית של 25 תלמידים ולמד אותם בשיטת לימוד חדשה.

- א. מהו רוח הסמק לVERAGE הציונים בחשבון אם יוחלט ליישם את שיטת הלימוד החדש?
- ב. מהו  $P_{\text{value}}$  לבדיקהיעילותה של שיטת הלימוד החדש?
- ג. מה יוכרע ברמת מובהקות של 5% לגבי יעילותה של שיטת הלימוד החדש?

**שאלות:**

- 1) באוניברסיטה גודלה גיל הסטודנטים לתואר ראשון מתפלג נורמלית.  
בעבר פורסם שהגיל הממוצע של הסטודנטים הינו 23. להלן פلت תוכנת SPSS על מדגם של 16 סטודנטים אקראים מתואר ראשון:

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
age	16	23.4375	2.50250	.62562

One-Sample Test

	Test Value = 23					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
age	.699	15	.495	.43750	-.8960	1.7710

- א. מהו המבחן הסטטיסטי שנעשה כאן?  
 ב. מה ערכו של הפרמטר לפי השערת האפס?  
 ג. רשום רוח סמך ברמת סמך של 95% לתוחלת גיל הסטודנטים באוניברסיטה לתואר ראשון.  
 ד. בדוק ברמת מובהקות של 5% האם גיל הממוצע ביום שונה מבעבר?
- 2) קבוצת ילדים בגיל 6 קיבלה משימה לביצוע. עברו כל ילד בדקו כמה זמן לוקח לו לסיים את המשימה בבדיקות. להלן תוצאות הניטוח הסטטיסטי:

One-Sample Test

	Test Value = 4.5					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
time	-1.853	24	.076	-.09200	-.1944	.0104

- א. כמה ילדים השתתפו במחקר?  
 ב. מצא רוח סמך ברמת סמך של 95% לתוחלת זמן ביצוע המשימה עברו ילדים בני 6.  
 ג. מה יש להניח כדי שרוח הסמך מסעיף אי יהיה תקף?  
 ד. בדוק ברמת מובהקות של 5% שזמן ביצוע המשימה הממוצע נמור מ-4.5 דקות.

(3) להלן פלט מחשב עבור ניתוח סטטיסטי שנעשה בתוכנת SPSS. הניתוח הוא עבור מבחן אקראי של קבוצת נבחנים בבריות באנגלית.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
grade	???	???	19.62787	2.95901

One-Sample Test

	Test Value = 75					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
grade	???	43	.017	-7.34091	-13.3083	????

- א. השלימו את הגדרים החסרים המסומנים בסמני שאלה בפלט.
- ב. מהי מובאות התוצאות לבדיקת ההשערה שהתוחלת של הציונים שונה מ-75?
- ג. מהי מובאות התוצאות לבדיקת ההשערה שהתוחלת של הציונים קטנה מ-75?
- ד. מהי מובאות התוצאות לבדיקת ההשערה שהתוחלת של הציונים גדולה מ-75?

(4) יצרו סיגריות מפרסם כי תוחלת הניקוטין בסיגריות שהוא מיצר קטנה מ-27 מ"ג. בבדיקה מקרית של 5 סיגריות מתוצרתו נמצאו כמותות הניקוטין הבאות : 21, 20, 24, 22 מ"ג. הניחו כי כמות הניקוטין בסיגריות מפולג נורמלי.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
nicotine	5	21.6000	1.51658	.67823

One-Sample Test

	Test Value = 27					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
nicotine	-7.962	4	.001	-5.40000	-7.2831	-3.5169

- א. האם ברמת מובהקות של 5% ניתן להסיק שיש אמת בפרסום?
- ב. אם היינו מוסיפים עוד תצפית שערך 20. כיצד הדבר היה משנה על ערך Sig ועל המסקנה?
- ג. בדקו האם ניתן להגיד שתוחלת רמת הניקוטין שונה מ-26 ברמת מובהקות של 5%.

**תשובות סופיות:**

- (1) א. הסקה של תוחלת אחת. ב. 23. ד. קיבל  $H_0$ .
- ב. (2). 25. א. המשטנה מתפלג נורמלית.
- . (3) א.  $upper = -1.3735$ ,  $t = -2.48$ ,  $\bar{X} = 67.66$ ,  $n = 44$ . ב. 0.017.
- .0.9915. ד. 0.0085. ג. נדחה  $H_0$ .
- (4) א. נכרייע שיש אמת בפרסום. ב. המסקנה לא תשתנה. ג. נכרייע שהתוחלת שונה מ-26.

## סטטיסטיקה

פרק 23 - בדיקת השערות על הפרש תוחלות במדגים בלתי תלויים

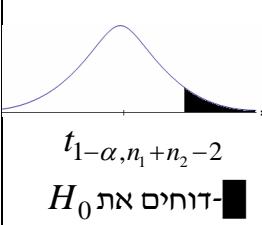
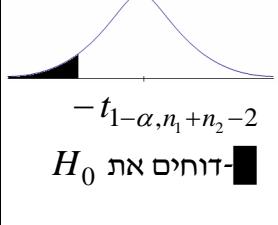
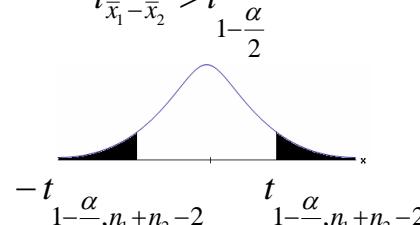
### תוכן העניינים

1.uschoniyot ha'ocelosia la'iduot v'menichim shehn shovot.	113
2. Nithach platisim.	117

## בדיקות השערות על הפרש תוחלות במדגמים בלתי תלויים

---

### כששוניות האוכלוסייה לא ידועות ומינחים שהן שווים – רקע

$H_0 \quad \mu_1 - \mu_2 = c$	$H_0 \quad \mu_1 - \mu_2 = c$	$H_0 \quad \mu_1 - \mu_2 = c$	השערת האפס: השערת אלטרנטיבית:
$H_1 \quad \mu_1 - \mu_2 > c$	$H_1 \quad \mu_1 - \mu_2 < c$	1. מדגמים בלתי תלויים 2. $\sigma_1, \sigma_2$ לא ידועות אך שווות 3. המשתנים בכל אוכלוסייה מתפלגים נורמלית	
$t_{\bar{x}_1 - \bar{x}_2} > t_{1-\alpha}^{(n_1+n_2-2)}$  $t_{1-\alpha, n_1+n_2-2}$ $H_0$ -דוחים את ■	$t_{\bar{x}_1 - \bar{x}_2} < -t_{1-\alpha}^{(n_1+n_2-2)}$  $-t_{1-\alpha, n_1+n_2-2}$ $H_0$ -דוחים את ■	$t_{\bar{x}_1 - \bar{x}_2} < -t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)}$ או $t_{\bar{x}_1 - \bar{x}_2} > t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)}$  $-t_{1-\frac{\alpha}{2}, n_1+n_2-2}$ $t_{1-\frac{\alpha}{2}, n_1+n_2-2}$ $H_0$ -דוחים את ■	אזור הדחיה של $H_0$

**סטטיסטי המבחן:**

$$t_{\bar{x}_1 - \bar{x}_2} = \frac{(\bar{x}_1 - \bar{x}_2) - c}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}}$$

**השונות המשוקלلت:**

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2}$$

**חלופה אחרת לכל הכרעה:**

נדחה $H_0$ אם מתקיים :	
$\bar{x}_1 - \bar{x}_2 < c - t_{1-\alpha}^{(n_1+n_2-2)} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$	$\bar{x}_1 - \bar{x}_2 > c + t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$ או $\bar{x}_1 - \bar{x}_2 < c - t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$
$\bar{x}_1 - \bar{x}_2 > c + t_{1-\alpha}^{(n_1+n_2-2)} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$	

**דוגמה (פתרון בהקלטה) :**

חברה המייצרת מוצרי בנייה טוענת שפיתחה סגסוגת (תערובת מתכות) שטמפרטורת ההתקה שלה גבוהה משמעותית מטמפרטורת ההתקה של הסגסוגת לבנייה שימושים בה כיום לבניית בניינים. לצורך בדיקת טענתה המחקר נדגמו 10 יחידות של מתכוות מהסוג היין ו-12 יחידות של מתכוות מהסוג החדש. להלן תוצאות המדגם:

טמפרטורת ההתקה הממוצעת במתכת היינה 1170 מעלות עם אומד חסר הטיה לשונות  $S^2 = 200$ .

טמפרטורת ההתקה הממוצעת במתכת החדשה 1317 מעלות עם אומד חסר הטיה לשונות  $S^2 = 260$ .  
 נניח לצורך פתרון שטמפרטורת ההתקה מתפלגת נורמללית עם אותה שונות במתכוות השונות. בדקו ברמת מובהקות של 5%.

## שאלות

**1)** להלן נתונים של שטחי דירות מtower דירות שנבנו בשנת 2012 ובשנת 2013 (במ"ר) :

120	94	90	130	95	112	120	<b>2012</b>
69	74	105	91	82	100		<b>2013</b>

בדקו שבסנת 2013 הייתה ירידה משמעותית בשטחי הדירות לעומת שנת 2012  
 עבור רמת מובהקות של 5%.  
 הניתנו שטחי הדירות בכל שנה מתפלגים נורמלית עם אותה שוננות.

**2)** נדגמו 15 ישראלים ו-15 אמריקאים. כל הנדגמים נגשו ל מבחון IQ. להלן תוצאות

הדגם :	המדינה	ישראל	ארה"ב
	גודל המדגם	15	15
	סכום הציונים	1560	1470
	סכום ריבועי הציונים	165,390	147,560

בדקו ברמת מובהקות של 5% האם קיים הבדל של נקודה בין ישראלים  
 לאמריקאים מבחינת ממוצע הציונים ב מבחון-h-IQ לטובת ישראל.  
 רשמו את כל ההנחות הדרושים לצורך פתרון התרגיל.

**3)** להלן תוצאות מבחן אורך חיים של נורות מסוג W60 ומסוג W100.

אורך החיים נמדד בשעות.

הקבוצה	60W	100W
$\bar{x}$	1007	956
S	80	72
n	13	15

- א. בדקו ברמת מובהקות של 5% האם נורות מסוג W60 דולקיות בממוצע יותר מאשר נורות מסוג W100. רשמו את כל ההנחות הדרושים לפתרון.
- ב. עבור איזו רמת מובהקות ניתן לקבוע שנורות מסוג W60 דולקיות בממוצע יותר מאשר נורות מסוג 100?
- ג. בדקו ברמת מובהקות של 5% האם נורות מסוג W60 דולקיות יותר מאשר נורות מסוג 1000 שעות. רשמו את כל ההנחות הדרושים.

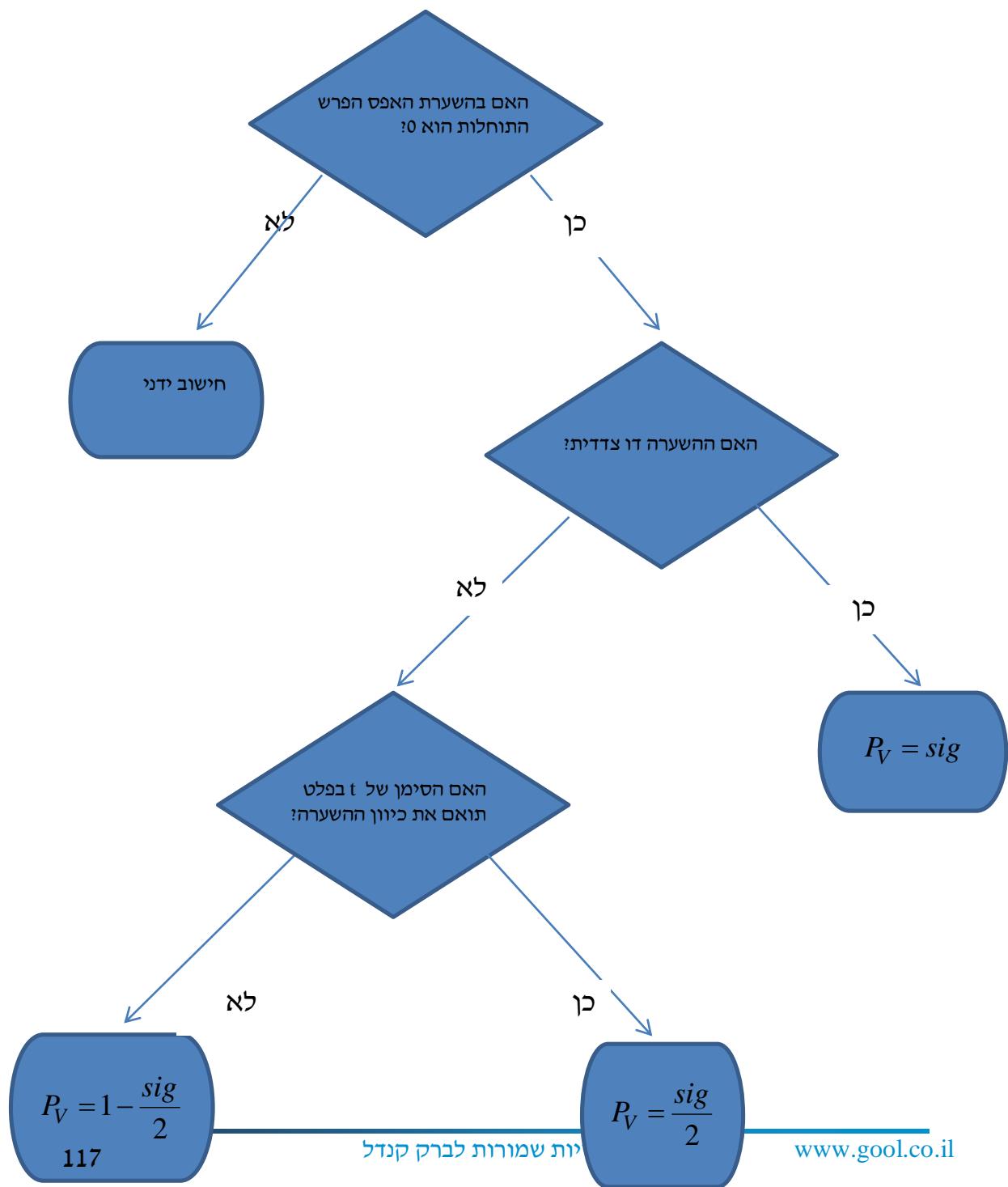
## תשובות סופיות

- 1) נדחה את  $H_0$ .
- 2) הנחות:  
1. סטיות התקן שוות.  
2. המשתנים מתפלגים נורמלית.
- נקבל את  $H_0$ .
- 3) א. נדחה את  $H_0$ .  
ב. רמת מובהקות של לפחות 5%.  
ג. לא נדחה את  $H_0$ .

## בדיקת השערות על הפרש תוחלות במדגים בלתי תלויים

### ניתוח פלטיט – רקע

MOVHECKOTHTOTOCHEA UL SMOK HAPLIT:



**דוגמה (פתרון בהקלטה) :**

בסקר שנערך בארה"ב בשנת 1993 נשאלו נסקרים שניים אזוריים שונים במדינה על מס' האחים והאחיות שלהם. להלן הפלט שהתקבל:

**Group Statistics**

	Region of the United States	N	Mean	Std. Deviation	Std. Error Mean
Number of Brothers and Sisters	North East	676	3.76	2.939	.113
	South East	410	4.05	2.993	.148

**Independent Samples Test**

		Levene's Test for Equality of Variances		t-test for Equality of Means						95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower		
Number of Brothers and Sisters	Equal variances assumed Equal variances not assumed	.173	.677	-1.583 -1.576	1084 850.945	.114 .115	-.293 -.293	.185 .186	-.657 -.658	.070 .072	

- מהו המבחן הסטטיסטי שנעשה כאן?
- בדוק ברמת מובהקות של 5% האם קיים שוויזון שוניות בין שני האזורים?
- בדוק האם קיים הבדל בין-"South East" ל-"North East" ברמת מובהקות של 5% מבחן מס' האחים והאחיות הממוצע.
- מתי מובהקות התוצאה לבדיקת הטענה שההפרש הממוצע בין-"South East" לבין-"North East" חיובי??

**שאלות**

**1)** להלן פلت מתוכנת SPSS מתוך מחקר שבחן את רמת האופטימיות של גברים ונשים. רמת האופטימיות נמדזה בסולם ציונים של 1 עד 5.

**Group Statistics**

		GENDER	N	Mean	Std. Deviation	Std. Error Mean
optimizm	MALE		633	2.6053	.49781	.01979
	FEMALE		568	2.5503	.48483	.02034

**Independent Samples Test**

	Levene's Test for Equality of Variances		t-test for Equality of Means							
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference		
								Lower	Upper	
optimizm	Equal variances assumed	.383	.536	1.935	1199	.053	.05500	.02842	-.00076	?
	Equal variances not assumed			1.938	1190.977	.053	.05500	.02838	-.00068	.11067

- א. האם ניתן להניח ששיעור האופטימיות של נשים וגברים שווה ברמת מובהקות של 5%?
- ב. ברמת מובהקות של 5% האם קיים הבדל בין הנשים לגברים ברמת האופטימיות הממוצעת שלהם?
- ג. מצא את הגבול העליון של רוח הסמק המשומן בסימן שאלה בפלט. דיקק עד 5 ספרות אחרי הקודעה.
- ד. בנה רוח סמק לתוחלת רמת האופטימיות של הגברים ברמת סמק של 95%.

2) פסיכולוגים טוענים שאנשים שניגשים לבחן אינטלקגנציה יותר מפעם אחת נוטים לקבל ציונים גבוהים יותר. להלן הפלט שהתקבל:

### Group Statistics

		N	Mean	Std. Deviation	Std. Error Mean
grade	A	9	96.8889	9.40006	3.13335
	B	11	108.4545	11.46616	3.45718

Independent Samples Test

	Levene's Test for Equality of Variances			t-test for Equality of Means							
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference			
								Lower	Upper		
grade	.206	.656	-2.428	18	.026	-11.56566	4.76333	-21.57304	-1.55828		
			-2.479	17.997	.023	-11.56566	4.66583	-21.36832	-1.76299		

### T-Test

מקרה :

A = נגשו פעם אחת.

B = נגשו יותר מפעם אחת.

א. רשמו את השערות המחקר והסבירו מהו המבחן המתאים כאן.

ב. כיצד הייתה משתנה התשובה לסעיף הקודם אם היה מדובר על אותם אנשים שציוונים נבדק פעמי אחדי המבחן הראשוני שעשו ופעמי אחריה המבחן השני?

ג. האם ניתן לומר כי מידת הפיזור של ציוני האנשים הנבחנים בפעם הראשונה שונה ממידת הפיזור של ציוני האנשים אשר נבחנים בפעם השנייה. בדוק ברמת מובהקות של  $\alpha = 0.05$ .

ד. האם נכונה טענת הפסיכולוגים ברמת מובהקות של  $\alpha = 0.01$ .

(3) חלק מחקר בנושא הנישואין בישראל, אחד החוקרים העלה השערה שיש הבדל בממוצע גיל הנישואין (הראשונים), בין נשים הגרות בעירים מרכזיות לבין נשים הגרות בעירים מרוחקות מהמרכז.  
לשם כך נדגו 50 כלות מכל אחת משתי ערים עיר א'-מרכזית ועיר ב'-מרוחקת ונרשם גילן. תוצאות עיבוד הנתונים מופיעות בטבלאות שללון:

**T-Test****Group Statistics**

מקום המגורים	N	Mean	Std. Deviation	Std. Error Mean
גיל הנישואין	עיר א	24.8072	1.38978	.19654
	עיר ב	23.0131	1.62070	.22920

**Independent Samples Test**

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
גיל הנישואין	Equal variances assumed	.330	.567	5.942	98	.000	1.79415	.30193	1.19497 2.39332
				5.942	95.772	.000	1.79415	.30193	1.19480 2.39350

- א. מהו המבחן הסטטיסטי שנעשה כאן?
- ב. מצא רוח סמך ברמת סמך של 95% להפרש בין עיר א לעיר ב מבחינת גיל הנשים הממוצע בנישואין הראשונים.
- ג. האם ניתן לומר ברמת מובהקות של 1% שנשים בעירים מרכזיות מתחנות בגיל מאוחר יותר מאשר נשים הגרות בעירים מרוחקות?

(4) להלן פלט של תוכנת SPSS :

**T-Test**

	N	Mean	Std. Deviation	Std. Error Mean
X	26	36.3077	13.23259	2.59513
Y	24	46.4583	20.96369	4.27920

**Independent Samples Test**

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variances assumed	4.446	.040	-2.164	???	.044	-10.15064	???	-20.03781	-.26347
Equal variances not assumed			-2.038	38.267	.048	???	5.00462	-20.27964	-.02164

- א. השלימו את סימני השאלה בטבלה.  
 ב. מהי מובוקות התוצאה לבדיקה הטענה שקיים הבדל בין השונות של  $X$  לעומת  $Y$ ?  
 ג. מהי מובוקות התוצאה לבדיקה הטענה שהתוחלת של  $X$  גדולה מהתוחלת של  $Y$ ?  
 ד. מהי מובוקות התוצאה לבדיקה הטענה שהתוחלת של  $X$  קטנה מהתוחלת של  $Y$ ?

**תשובות סופיות**

- 1)** א. קיבל את  $H_0$  ונכarius שיש שוויון שוניות.  
 ב. נקבע שלא קיים הבדל בין נשים לגברים מבחינת האופטימיות המומוצעת.  
 ג.  $2.5665 \leq \mu \leq 2.6441$ .  
 ד.  $0.11076$ .
- 2)** א. מבחן T להפרש ממוצעים במדגים בלתי תלויים.  
 ב. מבחן T למדגים מזוגיים.  
 ג. קיבל את  $H_0$ , נקבע לקיום שוויון שוניות.  
 ד. קיבל את  $H_0$ , לא קיבל את טענת הפסיכולוגים.
- 3)** א. מבחן T להשוואת תוחלת במדגים בלתי תלויים.  
 ב.  $1.19497 \leq \mu_1 - \mu_2 \leq 2.39332$  ג. כן.  
 ד. 0.022.
- 4)** א. 10.15, 4.69,-48  
 ב. 0.04  
 ג. 0.978

## סטטיסטיקה

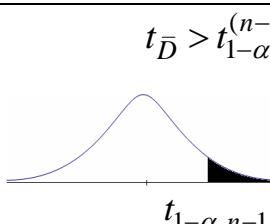
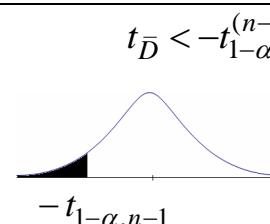
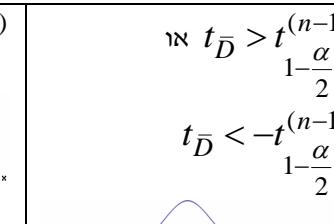
פרק 24 - בדיקת השערות לתוכלת ההפרש במדגמים מזוגים

### תוכן העניינים

124 .....	1. בדיקת השערות למדגמים מזוגים .....
128 .....	2. ניתוח פלטימ.

## בדיקות השערות על תוחלת הרפישים במדגמים مزוגים (תלויים)

### בדיקות השערות למדגמים מזוגים – רקע

$H_0: \mu_D = C$	$H_0: \mu_D = C$	$H_0: \mu_D = C$	השערת האפס: השערת אלטרנטיבית:
$H_1: \mu_D > C$	$H_1: \mu_D < C$	$H_1: \mu_D \neq C$	תנאים:
		1. $\sigma_D$ אינה ידועה 2. או מדגם מספיק גדול $D \sim N$	כלל ההכרעה: ازור הדחיה של $H_0$
$t_{\bar{D}} > t_{1-\alpha}^{(n-1)}$  $t_{1-\alpha, n-1}$	$t_{\bar{D}} < -t_{1-\alpha}^{(n-1)}$  $-t_{1-\alpha, n-1}$	$t_{\bar{D}} > t_{\frac{1-\alpha}{2}, n-1}^{(n-1)}$ או $t_{\bar{D}} < -t_{\frac{1-\alpha}{2}, n-1}^{(n-1)}$  $-t_{\frac{1-\alpha}{2}, n-1} \quad t_{\frac{1-\alpha}{2}, n-1}$	
$\bar{D} > C + t_{1-\alpha}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$	$\bar{D} < C - t_{1-\alpha}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$	$\bar{D} > C + t_{\frac{1-\alpha}{2}}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$ ו $\bar{D} < C - t_{\frac{1-\alpha}{2}}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$	חלופה לכלל הכרעה: נדחה $H_0$ אם מתתקיים:

$$S_D^2 = \frac{\sum_{i=1}^n (D_i - \bar{D})^2}{n-1} = \frac{\sum_{i=1}^n D_i^2 - n\bar{D}^2}{n-1}, \quad t_{\bar{D}} = \frac{\bar{D} - \mu_D}{\frac{S_D}{\sqrt{n}}}$$

**סטטיסטי המבחן:**

דוגמה (פתרו בהקלטה):

חברה שיווקית מעוניינת לבדוק את טענת רשות השיווק "מגה בעיר" הטענה שמחירים נמוכים מהמחירים מרשות השיווק "שופרסל". לצורך הבדיקה נבחרו באקראי 4 מוצרים שונים. המחירים נבדקו בשתי הרשותות. להלן המחירים:

ה מוצר / רשות	מגה בעיר	שופרסל
18	17	שמפו
57	48	gil כביסה
35	35	עוגת גבינה
10	12	לחם
47	49	קפה נמס
142	113	בקבוק יין
26	20	גבינה בולגרית

בහנחה והמחירים מתפלגים נורמלית, בדקנו ברמת מובהקות של 5% את טענת רשות "מגה בעיר".

**שאלות**

- 1)** במטרה לבדוק האם קיימים הבדל בין חברות  $X$  ו- $Y$  מבחינת המחיר לשיחות בין-יל. נדגמו באקראי 7 מדינות ועבור כל מדינה נבדקה עלות דלקת שיחה. להלן התוצאות:

חברה/ מדינה	ארה"ב	קנדה	הולנד	פולין	מצרים	סין	יפן
$X$	1.5	2.1	2.2	3	3.5	3.2	4.2
$Y$	1.4	2	1.9	3.1	3.2	3.2	4.2

בנהנה והמקרים מתפלגים נורמלית בכל חברת, בדקו ברמת מובהקות של 5% האם קיימים הבדל בין החברות מבחינת המחיר במומוץ?

- 2)** מכון המRAIN לפסיכוןטורי טוען שהוא מעלה את ממוצע הציונים ביוטר מ-30 נקודות. 8 נבחנים נבדקו לפני ואחרי שהם למדו במכון. להלן התוצאות שהתקבלו:

אחרי	לפני	590	500	390	670	640	420	470	506
580	520	510	680	610	430	540	570	570	506

מה מסקנתכם ברמת מובהקות 5%? הניחו שציוני פסיכוןטורי מתפלגים נורמלית.

- 3)** נדגמו 5 סטודנטים שישימנו את הקורס סטטיסטיקה ב'. להלן הציונים שלהם בסMASTER א' ו- ב':

סטטיסטיקה ב'	סטטיסטיקה א'	82	75	90	68	74
סטטיסטיקה ב'	סטטיסטיקה א'	100	76	87	84	80

פורסם שתלמידים שמשיכים את סטטיסטיקה ב' מושפרים ממוצע את הציונים ב-5 נקודות לעומת סטטיסטיקה א'. הניחו שהציונים מתפלגים נורמלית.

- א. מהי מובהקות התוצאה לבדיקת הטענה שהשיפור הוא יותר מ 5 נקודות?  
 ב. על סמך הסעיף הקודם, מהי רמת המובהקות המינימלית להכרעה שהשיפור הוא יותר מ- 5 נקודות?  
 ג. לאור זאת, מה המסקנה ברמת מובהקות של 10%?

- 4)** לצורך בדיקת השפעת היפנוזה על לימוד אנגלית, נבחרו 10 זוגות תאומים זהים. אחד התאומים למד אנגלית בהשפעת היפנוזה, והשני לא היפנוזה. לאחר מכן נערך לשניהם מבחון באנגלית. נניח שציוני המבחן מתפלגים נורמלית ללא ידיעת השונות האמתית. המבחן שיש לבצע כאן הוא:

- א. מבחן  $Z$  למדגם יחיד.  
 ב. מבחן  $T$  למדגם יחיד.  
 ג. מבחן  $T$  למדגמים בלתי תלויים.  
 ד. מבחן  $T$  למדגמים מזוגיים.

5) בتحقנת טיפת חלב מסויימת יש שני מכשירי שקילה. על מנת להשוות בין שני המשקלים נדגמו 4 תינוקות. כל תינוק בן חודשיים נשקל בכל אחד מהמשקלים.

להלן תוצאות השקליה (בק"ג) :

	משקל במכשיר 1	4.5	9.6	0.7	2.5
	משקל במכשיר 2	3.5	6.9	1.7	0.5

נניח שהמשקלים מתפלגים נורמלית, המבחן שיש לבצע כאן הוא :

- א. מבחן Z למדגם יחיד.
- ב. מבחן T למדגם יחיד.
- ג. מבחן T למדגים בלתי תלויים.
- ד. מבחן T למדגים מזוגים.

6) כדי להשוות בין שני אצנים נדגמו 5 תוצאות מריצת 100 מטר של כל אצן. זמני הריצה נרשמו ויש להניח שמתפלגים נורמלית. המטרה להשוות בין האצנים.

המבחן שיש לבצע כאן הוא :

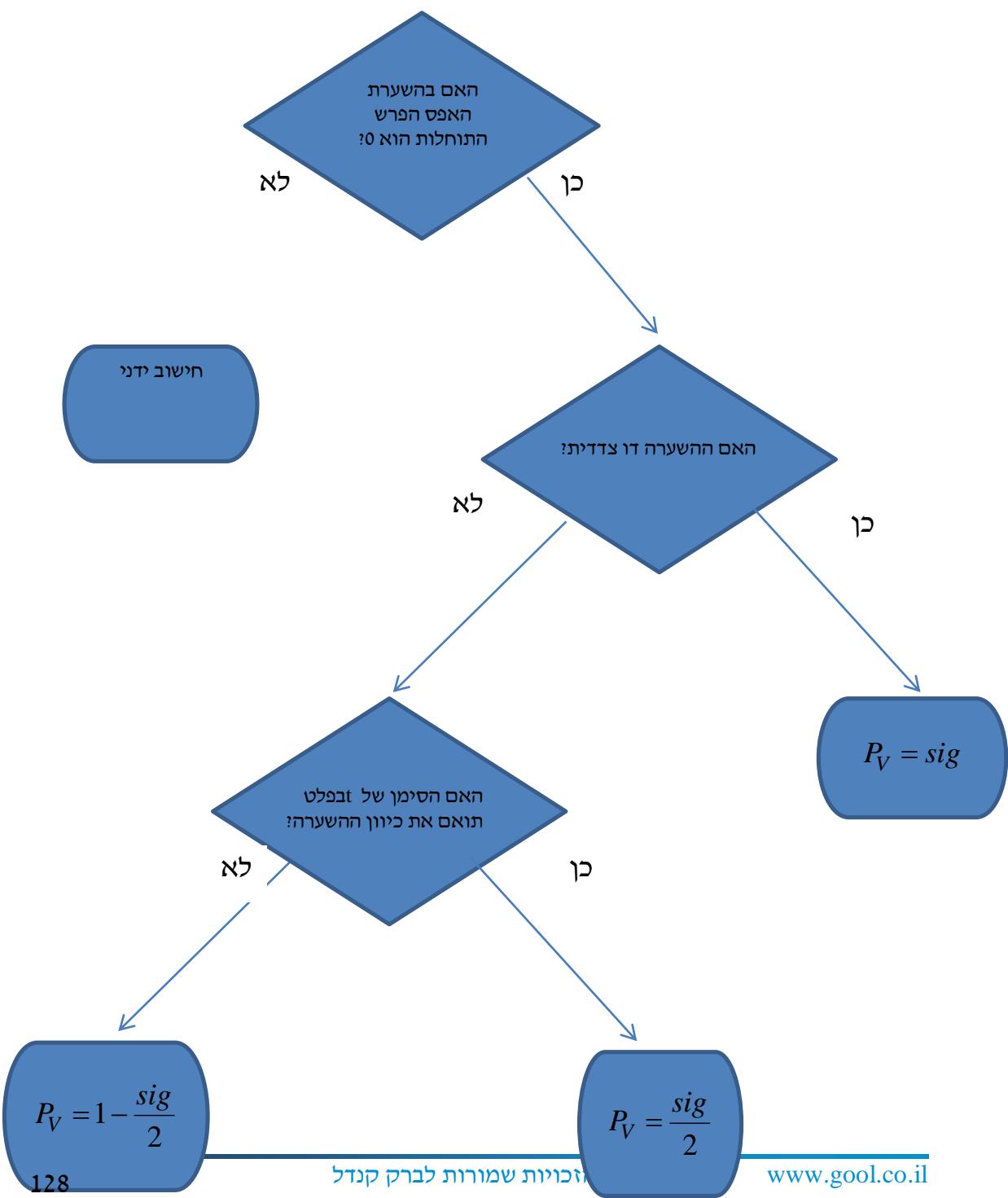
- א. מבחן Z למדגם יחיד.
- ב. מבחן T למדגם יחיד.
- ג. מבחן T למדגים בלתי תלויים.
- ד. מבחן T למדגים מזוגים.

## תשובות סופיות

- 1) לא נדחה  $H_0$ .
- 2) לא נדחה  $H_0$ .
- 3) א. לא נדחה  $H_0$ .      ב.  $0.5 \leq pv \leq 0.5$       ג. לא נדחה  $H_0$ .
- 4) ד'.
- 5) ד'.
- 6) ג'.

## בדיקות השערות על תוחלת הרפישים במדגמים مزוגים (תלוים)

### מדגמים מזוגים – ניתוח פלטיט – רקע



**דוגמה (פתרון בהקלטה) :**

כדי לבדוק את ההשפעה של קורס לגמילה מעישון נלקח מדגם מקרי של 5 נבדקים. עבור כל אחד מהם נמדדה צריכת הסיגריות היומיית לפני הקורס וחודשיים אחריו. הניחו שצריכת הסיגריות מתפלגת נורמלית. להלן התוצאות:

5	4	3	2	1	נבדק
30	28	25	22	40	לפני
12	10	13	24	30	אחרי

**Paired Samples Statistics**

	Mean	N	Std. Deviation	Std. Error Mean
Pair 1 BEFORE	29.0000	5	6.85565	3.06594
AFTER	17.8000	5	8.72926	3.90384

**Paired Samples Test**

	Paired Differences					t	df	Sig. (2-tailed)			
	Mean	Std. Deviation	Std. Error Mean	90% Confidence Interval of the Difference							
				Lower	Upper						
Pair 1 BEFORE - AFTER	11.20000	8.19756	3.66606	3.38452	19.01548	3.055	4	.038			

בדקו ברמת מובהקות של 5% האם הקורס יעיל.

## שאלות

**1)** בסקר שנערך באראה"ב בשנת 1993 נשאלו נסקרים על השכלה הוריהם, להלן הפלט שהתקבל:

Paired Samples Test											
	Paired Differences					t	df	Sig. (2-tailed)			
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference							
				Lower	Upper						
Pair 1	Highest Year School Completed, Father - Highest Year School Completed, Mother	-.007	3.115	.100	-.203	.189	-.072	.973	.943		

- א. תננו אומדן להפרש הממוצעים.
- ב. תננו אומדן לטיעות התקן של הפרש הממוצעים.
- ג. האם קיים הבדל מובהק בין השכלה האבות להשכלה האימהות ברמת מובהקות של 5%?

**2)** בתחרות קפיצה למים שופטים באופן קבוע שופט איטלקי ושופט דרום קורייני. להלן פלט המנתה את הציונים לשופטים אלה נתנו בתחרויות השונות:

Paired Samples Statistics					
		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Italy South Korea	???	300	.86742	.05008
		8.9183	???	.81992	.04734

Paired Samples Test											
	Paired Differences					t	df	Sig. (2-tailed)			
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference							
				Lower	Upper						
Pair 1	Italy - South Korea	-.42233	.36153	.02087	-.46341	-.38126	-20.234	???	???		

- א. השלימו את החלקים החסרים בפלט (מסומנים בסימני שאלה).
- ב. בדקו את הטענה שהשופט הדרום קורייני נותן בממוצע 0.2 נקודות יותר מאשר השופט האיטלקי ברמת מובהקות של 5%.
- ג. מהו רוחח הסמך ברמת סמך של 95% ל佗חת פער הציונים בין השופטים?
- ד. בנו את הרוחח כת ערך ברמת סמך של 98% ל佗חת פער הציונים בין השופטים.

**3)** בדקו את ציוניהם של 44 נבדקים אקראים ב מבחן הפסיכומטרי. פעם אחת לפני הכנה (Before) ופעם אחת אחרי הכנה (After).

Paired Samples Test											
	Paired Differences					t	df	Sig. (2-tailed)			
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference							
				Lower	Upper						
Pair 1 Before - After	-7.45455	19.28303	2.90703	-13.31712	-1.59197	-2.564	43	.014			

- א. רשמו מהו המבחן הסטטיסטי ונשח את ההשערות אליהם מתייחס הפלט.
- ב. בדקו את ההשערה שממוצע ציונים משתפרים לאחר ההכנה ברמת מובהקות של .5%.
- ג. בדקו את ההשערה שממוצע ציונים משתפרים לאחר ההכנה ביותר מ-5% נקודות ברמת מובהקות של .5%.
- ד. מצאו רוח סמך ל佗ולת שיפור ממוצע הציונים לאחר ההכנה ברמת ביטחון של .95%.

(4) להלן פלט של תכנת SPSS :

**T-Test****Paired Samples Statistics**

	Mean	N	Std. Deviation	Std. Error Mean
Pair 1	54.0000	6	5.86515	2.39444
x y	46.5000	6	10.72847	4.37988

**Paired Samples Test**

	Paired Differences					t	df	Sig. (2-tailed)			
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference							
				Lower	Upper						
Pair 1	7.50000	??	4.72405	-4.64356	19.64356	??	5	.173			
x - y											

- א. מלא את החלקים החסרים בטבלה.
- ב. מהי רמת המובהקות המינימלית לקבלת הטענה שיש הבדל בין  $X$  ל- $Y$  בממוצע?
- ג. האם התשובה לסעיף הקודם הייתה משתנה, ואם כן גדלתה או קטנה, אם היינו מושיעים עוד תצפית שההפרש בין  $X$  ל- $Y$  הוא 0.
- ד. מהי מובהקות התוצאה לבדיקת הטענה ש  $X$  גדול מ-  $Y$  בממוצע?
- ה. מהי מובהקות התוצאה לבדיקת הטענה ש  $X$  קטן מ-  $Y$  בממוצע?
- ו. בנו רוח סמן ל佗ולת של  $X$  ברמת סמן של 90%.

**תשובות סופיות**

- (1) א.  $-0.007$       ב. 0.1      ג. אין הבדל מובהק.
- (2) א.  $d.f = 299$       ב. נדחה את  $H_0$       ג.  $\bar{X} = 8.496$ ,  $n = 300$
- (3) א. ראה וידאו.      ב. נדחה את  $H_0$       ג. לא נדחה את  $H_0$
- (4) א. 11.5715, 1.5876      ב. 0.173      ג. יגדל.  
 ד.  $0.0865 < \mu < 49.18$

## סטטיסטיקה

פרק 25 - הקשר בין רוח סמרק לבדיקת השערות להפרש תוחלות

תוכן העניינים

1. הקשר בין רוח סמרק לבדיקת השערות להפרש תוחלות.....  
134 .....

## הקשר בין רוח סמך לבדיקה השערות על הפרש תוחלות

### רקע

ניתן לבצע בדיקת השערות דו צדדיות ברמת מובהקות  $\alpha$  על  $\mu_1 - \mu_2$ :

$$H_0: \mu_1 - \mu_2 = C, \quad H_1: \mu_1 - \mu_2 \neq C$$

על ידי בניית רוח סמך ברמת סמך של  $\alpha - 1$  על  $\mu_1 - \mu_2$ :

אם  $C$  נופל ברווח  $\leftarrow$  קיבל את  $H_0$ .

אם  $C$  לא נופל ברווח  $\leftarrow$  נדחה את  $H_0$ .

### דוגמה (פתרו החלטה):

חוקר ביצע בדיקת השערות לתחוללת ההפרש במדגים מזוווג.

להלן השערותיו:  $H_0: \mu_D = 80, \quad H_1: \mu_D \neq 80, \quad \alpha = 5\%$ .

החוקר בנה רוח סמך ברמה של  $90\% < \mu_D < 78$ .

האם אפשר לדעת מה מסקנתו, ואם כן מהי?

## שאלות

**1)** נדגמו 5 סטודנטים שסיימו את הקורס סטטיסטיKA ב'. להלן ציוניהם בסMASTER A' ו-B':

סMASTER A	סMASTER B
80	74
84	68
87	90
76	75
100	82

א. בנו רוח סמך ברמת סמך של 95% לתוחלת פער הציונים בין סMASTER A' לבין סMASTER B'.

ב. פורסם שתלמידים ממשיכים את סMASTER B' משפרים ממוצע את הציונים ב-5 נק' לעומת סMASTER A'. האם יש אמת בפרסום?

**2)** הוחלט להשוות הציונים אצל מרצה X ואצל מרצה Y. נבחרו באקראי 6 סטודנטים, 3 סטודנטים של מרצה X ו-3 סטודנטים של מרצה Y, עבורם התקבלו הציונים הבאים:

מרצה X	82	90	68
מרצה Y	68	68	81

א. חשבו רוח סמך ברמת סמך 90% להפרש בין התוחלות של הציונים אצל שני המרצים.

ב. האם ברמת מובהקות של 10% נכיר שיש הבדל בין תוחלות הציונים אצל שני המרצים?

### שאלות רב-ברירה:

**3)** סטטיסטיKA נתקש לאמוד את הפרש הממוצעים של שני טיפולים לפי שני מוגדים מקרים בלתי תלויים.

הוא חישב רוח סמך להפרש ברמת סמך 0.98, וקיים את הרווח  $\mu_2 - \mu_1 < 2$ . אילו יתבקש החוקר לבדוק לפי אותן נתוניות את השערות:

$H_0: \mu_1 - \mu_2 = 0$ ;  $H_1: \mu_1 - \mu_2 > 0$ , מסקנתו תהיה:

א. לדחות את השערת האפס.

ב. לא לדחות את השערת האפס.

ג. שלא ניתן לדעת את המסקנה עבור רמת מובהקות 0.05.

ד. שלא נתנות בשאלת סטיות התקן של האוכלוסיות, ולכן לא ניתן להסיק דבר.

- (4) במטרה לבדוק האם קיימים הבדל בין קווי זהב לבזק מבחינות ממוצע המחרירים לשיחות ביןיל. נגדמו באקראי 7 מדינות ועבור כל מדינה נבדקה עלות דקט שיחה. בהנחה והמחקרים מתפלים נורמלית בנו רוח סמך ממוצע ההפרשנים וקיים :  $\mu_D < 0.0293 - 0.2145$ , רוח הסמך הוא ברמת סמך של 95% .  
לכן מסקנת המחקר היא :
- א. ברמת מובהקות של 5% לא נוכל לקבוע שקיימים הבדל בין החברות.
  - ב. ברמת מובהקות של 5% נקבע שקיימים הבדל מובהק בין החברות.
  - ג. לא ניתן לדעת מה המסקנה ברמת מובהקות של 5% כיון שלא נאמר מה ההגדרה של  $D$  .

### תשובות סופיות

- |                           |  |
|---------------------------|--|
| ב. נכרייע שיש אמת בפרסום. | א. $-3.8 \leq \mu_D \leq 19$           |
| ב. נכרייע שאין הבדל.      | א. $-8.5 \leq \mu_X - \mu_Y \leq 26.5$ |
|                           | ג. ✓                                   |
|                           | אי. ✓                                  |

## סטטיסטיקה

### פרק 26 - שאלות מסכמת בבדיקה השערות

תוכן העניינים

1. שאלות רב ברירה ( אמריקאיות) .....

137 .....

## שאלות סיוכם – שאלות רב ברירה על בדיקת השערות

- 1)** בבדיקה השערה חד-צדדית ימנית ברמת מובהקות  $\alpha = 0.01$ , נדחתה השערת האפס. מה הייתה המסקנה לו נבדקה אותה ההשערה באמצעות נתונים ברמת מובהקות  $\alpha = 0.05$  ?
- השערת האפס הייתה נדחתה.
  - השערת האפס לא הייתה נדחתה.
  - השערה המקרית הייתה נדחתה.
  - בHUDER נתונים נוספים, לא ניתן לדעת.
- 2)** לצורך בדיקת השפעת היפנוזה על לימוד אנגלית, נבחרו 10 זוגות תאומים זהים. אחד התאומים למד אנגלית בהשפעת היפנוזה, והשני לא היפנוזה. לאחר מכן נערך לשניהם מבחן באנגלית. נניח שצינוי המבחן מתפלגים נורמלית ללא ידיעת השונות האמיתית. המבחן שיש לבצע כאן הוא :
- מבחן Z למדגם יחיד.
  - מבחן Z למדגמים רבים.
  - מבחן T למדגמים בלתי תלויים.
  - מבחן T למדגמים מזוגיים.
- 3)** כדי לבדוק את הטענה שגברים רווקים שוקלים פחות מגברים נשואים לעומת חוקר מדגם מקורי של 4 גברים ומדד את משקלם לפני נישואיהם ולאחר נישואיהם. הנה התוצאות :  
 מהן ההשערות הנבדקות? (ההפרש חושב  $Y - X$ )
- |    |    |    |    |   |                 |
|----|----|----|----|---|-----------------|
| 68 | 82 | 93 | 69 | X | לפני הנישואין - |
| 71 | 84 | 88 | 80 | Y | לאחר הנישואין - |
- $H_1: \mu_d < 0, H_0: \mu_d = 0$
  - $H_1: \mu_X - \mu_Y < 0, H_0: \mu_X - \mu_Y = 0$
  - $H_1: \mu_X - \mu_Y < 0, H_0: \mu_X - \mu_Y = 0$
  - $H_1: \mu_d > 0, H_0: \mu_d = 0$
- 4)** חוקר ביצע מחקר ובו עשה טעות מסווג שני לכך :
- השערת האפס נכונה.
  - השערת האפס נדחתה.
  - השערת האפס לא נדחתה.
  - אף אחת מהතושבות לא נכונה בהכרח.

- 5)** ידוע כי ילד בגיל שנתיים ישן ממוצע 9 שעות בלילה. במדגם של 20 תינוקות בני שנתיים המתגוררים בצפון נמצא, כי ממוצע שעות השינה בלילה הינו 10 עם סטיית תקן של 1.1. במדגם של 10 תינוקות בדרום נמצא, כי ממוצע שעות השינה בלילה הינו 7.9 עם סטיית תקן של 1.1. על מנת להשוות בין ממוצע שעות השינה של ילדים לבין המתגוררים בצפון יש לעורך \_\_\_\_\_. ועל מנת להשוות בין ממוצע שעות השינה של ילדים מהדרום לבין של ילדים יש להניח שההנחות הדומות מתקיימות.
- מבחן Z למדגם יחיד ; מבחן T למדגם יחיד.
  - מבחן T למדגם יחיד ; מבחן T למדגים תלויים.
  - מבחן T למדגם יחיד ; מבחן T למדגים בלתי תלויים.
  - מבחן T למדגים בלתי תלויים ; מבחן T ממוצע יחיד.
- 6)** מובקות התוצאה (PV) היא גם :
- רמת המובקות המינימאלית לדוחות השערת האפס.
  - רמת המובקות המקסימאלית לדחית השערת האפס.
  - רמת המובקות שנקבעת מראש על ידי החוקר טרם קיבל את תוצאות המחקר.
  - רמת המובקות המינימאלית לאי דחית השערת האפס.
- 7)** כדי לבדוק את הטענה שגברים רווקים שוקלים פחות מאשר נשים לעומת זאת חוקר מגדם מקרי של 4 גברים ומדד את משקלם לפני נישואיהם ולאחר נישואיהם. הנה התוצאות :
- |    | לפני הנישואין | לאחר הנישואין |
|----|---------------|---------------|
| 68 | 82            | 93            |
| 71 | 84            | 88            |
|    | 69            | 80            |
- באיזה התפלגות משתמשים בבדיקה השערות, ובכמה דרגות חופש :
- התפלגות Z ללא דרגות חופש.
  - התפלגות T ו-3 דרגות חופש.
  - התפלגות T ו-6 דרגות חופש.
  - התפלגות  $\chi^2$  ו-3 דרגות חופש.

- 8) שני סטטיסטיקים בודקים השערות ברמת מובהקות  $\alpha = 0.05$  על סמך אותו מבחן. סטטיסטיקי א' בודק את ההשערה:  $H_0: \mu = 20$  נגד האלטרנטיבית  $H_1: \mu \neq 20$  ומחליט לא לדוחות את השערת האפס.
- סטטיסטייקי ב' בודק את ההשערה  $H_0: \mu \leq 20$  נגד האלטרנטיבית  $H_1: \mu > 20$ : מה יחליט סטטיסטייקי ב'?
- לדוחות את השערת האפס.
  - לא לדוחות את השערת האפס.
  - לא נתוניים נוספים אי אפשר לדעת מה יחליט.
- 9) חוקר בדק השערה מסויימת והחליט לדוחות את השערת האפס ברמת מובהקות 5%. מה נכון לו?
- הוא בודאות ידחה את השערת האפס ברמת מובהקות 9% ואילו ברמת מובהקות 2% יש לבדוק מחדש.
  - הוא בודאות לא ידחה את השערת האפס ברמת מובהקות 9% ואילו ברמת מובהקות 2% יש לבדוק מחדש.
  - הוא בודאות ידחה את השערת האפס ברמת מובהקות 9% וברמת מובהקות 2%.
  - הוא בודאות לא ידחה את השערת האפס ברמת מובהקות 9% ואילו ברמת מובהקות 2% יש לבדוק מחדש.
- 10) רמת הcolesterol בدمם של אנשים מתפלג נורמלית עם תוחלת של 180 מ"ג (ל 100 סמ"ק דם). וטיפות התקן של 10 מ"ג. מעוניינים לבדוק את הטענה שצמחיים הם בעלי רמתコレsterol נמוכה יותר. נניח שטיפות התקן אצל צמחיים זהה לטיפות התקן של כלל האנשים. בדוגמה של 20 צמחיים התקבל ממוצע רמתコレsterol 174.5 מ"ג. אם הוחלט לקבל את הטענה שצמחיים הם בעלי רמתコレsterol נמוכה יותר איזה סוג טעות אפשרית במסקנה?
- טעות מסוג ראשון.
  - טעות מסוג שני.
  - טעות מסוג שלישי.
  - לא ניתן לדעת כיון שאין לנו ידיעות מה התוחלת האמיתית אצל הצמחיים.

- (11)** שני חוקרים העוסקים בתחום מחקרי משותף החליטו להסתמך על נתונים של מדגם שפורסם על ידי הלשכה המרכזית לסטטיסטיקה.  
 חוקר א' ניסח השערה זו צדדית ואילו חוקר ב' ניסח השערה חד צדדית.  
 מסקנתו של איזה מבין המשפטים הבאים הוא הנכון בנוגע למסקנות החוקרים?  
 א. אם חוקר א' ידחה את השערת האפס לא ניתן לדעת מה יחליט חוקר ב'  
 באותה רמת מובהקות.  
 ב. אם חוקר א' יקבל את השערת האפס גם חוקר ב' יקבל את השערת האפס  
 באותה רמת מובהקות.  
 ג. אם חוקר ב' ידחה את השערת האפס גם חוקר א' ידחה את השערת  
 האפס באותה רמת מובהקות.  
 ד. אם חוקר א' ידחה את השערת האפס גם חוקר ב' ידחה את השערת  
 האפס בתנאי שרמת המובהקות כפולה בגודלה.
- (12)** ידוע מנתוני ה עבר כי תוכלת הציוניים בבחינה בפסיכולוגיה היא 79. ה问题是  
 השערה כי תוכלת הציוניים בקרב העולים החדשניים נמוכה יותר. לצורך בדיקת  
 הטענה נלקח מדגם מקרי של 47 סטודנטים עולים ונמצא ממוצע של 75.  
 מה משמעות הפרמטר בניסוח השערות?  
 א. תוכלת ציוני העולים באוכלוסייה.  
 ב. ממוצע ציוני העולים במדגם.  
 ג. תוכלת ציוני האוכלוסייה מנתוני ה עבר.  
 ד. ממוצע ציוני שאר האוכלוסייה במדגם.
- (13)** חוקר ביצע מחקר וידוע כי עשה טעות מסווג 1. מה מהබאים נכון?  
 א. החוקר דחה את השערת  $H_0$  כאשר היא הייתה נכונה.  
 ב. החוקר דחה את השערת  $H_1$  כאשר היא הייתה נכונה.  
 ג. החוקר לא דחה את השערת  $H_0$  כאשר היא הייתה לא נכונה.  
 ד. המדגם של החוקר שיק בפועל להתפלגות הדגימה של  $H_1$ .
- (14)** חוקר ביקש לבדוק האם תאומים זהים אשר הופרדו בילדותם שונים  
 מתאימים זהים אשר גדלו יחדיו מבחן מידת הפער בין התאומים בלחץ  
 הדם. הוא דגם 20 זוגות תאומים מכל אוכלוסייה ומדד את הפרש בין לחץ  
 הדם בכל זוג תאומים. מהו המבחן הסטטיסטי המתאים?  
 א. מבחן T למדגים בלתי תלויים עם 38 דרגות חופש.  
 ב. מבחן T למדגים מזוווגים, עם 39 דרגות חופש.  
 ג. מבחן T למדגים בלתי תלויים עם 39 דרגות חופש.  
 ד. מבחן T למדגים מזוווגים עם 38 דרגות חופש.

- 15)** שלושה חוקרים רצו לבדוק את השפעתו של שידור פרסומות נגד תאונות דרכים על מהירות הנהיגה של נהגים בישראל (השונות של מהירות הנהיגה בישראל אינה ידועה). עידו השווה את מהירות הנהיגה של קבוצת נהגים אחת, חדש לפני שידור הפרסומות וחודש לאחר שידור הפרסומות. רון השווה את מהירות הנהיגה של קבוצת נהגים, שראו את הפרסומות, ל מהירות הנהיגה של קבוצת נהגים, שלא ראו את הפרסומות. יואב השווה את מהירות הנהיגה הממוצעת בישראל על פי נתוני משרד התחבורה. ה מבחנים בהם צריכים החוקרים להשתמש הם :
- שלושתם בבחן T לדוגמים בלתי תלויים.
  - עידו בבחן T לדוגמים מזוגים, רון ויואב בבחן T לדוגמים בלתי תלויים.
  - עידו בבחן T לדוגמים מזוגים, רון ויואב בבחן T לדוגמה יחיד.
  - עידו בבחן T לדוגמים מזוגים, רון ויואב בבחן T לדוגמה יחיד.

- 16)** במחקר נמצא שתוצאה היא מובהקת ברמת מובהקות של 5%. מה תמיד נכון?
- הגדלת רמת המובהקות לא תשתנה את מסקנת המחקר.
  - הגדלת רמת המובהקות תנסה את מסקנת המחקר.
  - הקטנת רמת המובהקות לא תנסה את מסקנת המחקר.
  - הקטנת רמת המובהקות תנסה את מסקנת המחקר.
- 17)** חוקן ערך בבחן דו צדי ברמת מובהקות של  $\alpha$  והחליט לדוחות את השערת האפס. אם החוקן היה עורך בבחן חד צדי ברמת מובהקות של  $\frac{\alpha}{2}$  איזי בהכרח:
- השערת האפס הייתה נדחתה.
  - השערת האפס הייתה לא נדחתה.
  - לא ניתן לדעת מה תהיה מסקנתו במקרה זה.

- 18)** ליואר ורוני העלו את אותן השערות על ממוצע האוכלוסייה. כמו כן הם התבוססו על אותן נתונים של דוגם.
- ליואר השתמש בטבלה של התפלגות Z.
- רוני השתמשה בטבלה של התפלגות T.
- מה יוכל לומר בנוגע להחלטת המחקר שלהם?
- אם ליואר ידחה את השערת האפס אז גם בהכרח רוני.
  - אם רוני ידחה את השערת האפס אז גם בהכרח ליואר.
  - שני החוקרים בהכרח יגיעו לאותה מסקנה.
  - לא ניתן לדעת על היחס בין דחינת השערת האפס של שני החוקרים.

.  $H_0: \mu = \mu_0$ ,  $H_1: \mu < \mu_0$  נטען ש  $(\sigma^2, \mu) \sim N$  כmo כn נתונות ההשערות הבאות:

- חוקר בדק את ההשערות הללו על סמך מדגם שככל 10 תוצאות.  $\sigma^2$  לא הייתה ידועה לחוקר. החוקר החליט לדוחות את השערת האפס ברמת מובהקות של 5% לאחר מכן כדי לוכיח את קביעתו הוא דגם עוד 5 תוצאות וشكל את תוצאות אלה גם למדגם כך שככלו עשו 15 תוצאות.
- cut בברור הוא ידחה את השערת האפס.
  - cut הוא דווקא קיבל את השערת האפס.
  - cut לא ניתן לדעת מה תהיה מסקנתו.

(20) אם חוקר החליט להגדיל את רמת המובהקות במחקר שלו אז:

- הסיכוי לטעות מסווג ראשון גדול.
- העוצמה של המבחן קטנה.
- הסיכוי לטעות מסווג שני גדול.
- תשובות A ו-B נכונות.

(21) חוקר ביצע מחקר ובו עשה טעות מסווג שני בכך:

- השערת האפס נכונה.
- השערת האפס נדחתה.
- השערת האפס לא נדחתה.
- אף אחת מהתשובות לא נכונה בהכרח.

(22) מה המצב הרצוי לחוקר המבצע בבדיקה השערה:

- |             |          |
|-------------|----------|
| $1 - \beta$ | $\alpha$ |
| א. גדולה    | קטנה     |
| ב. גדולה    | גדולה    |
| ג. קטנה     | גדולה    |
| ד. קטנה     | קטנה     |

(23) נערך שינוי בכלל החלטה של בדיקת השערה מסוימת ובקבוקותיו איזור דחיה  $H_0$  קטן. כל שאר הגורמים נשארו ללא שינוי. כתוצאה לכך:

- חן  $\alpha$ , והן  $(\beta - 1)$ , יקטנו.
- $\alpha$  יישאר ללא שינוי ואילו  $(\beta - 1)$  יגדל.
- $\alpha$  יגדל ואילו  $(\beta - 1)$  יקטן.
- חן  $\alpha$  והן  $(\beta - 1)$  יגדלו.

**(24)** ידוע כי לחץ דם תקין באוכלוסייה הוא 120. רופא מניה שלחץ הדם בקרוב עיתונאים גבוה יותר מה ממוצע באוכלוסייה. הוא לκת מדגם של 60 עיתונאים וקיים ממוצע 137. על סמך המדגם, הוא בודק טענתו ברמת מובהקות 0.02 ומסיק שלחץ הדם בקרוב העיתונאים אינו גבוה יותר. מה הטעות האפשרית שהרופא עושה?

- א. טעות מסוג ראשון.
- ב. טעות מסוג שני.
- ג. טעות מסוג שלישי.
- ד. אין טעות בטענתו.

**(25)** בבדיקה השערות התקבל שה-  $p-value = 0.02$ . מה תהיה מסקנת חוקר המשמש ברמת מובהקות 1%? בחר בתשובה הנכונה:

- א. קיבל את השערת האפס בכל מקרה.
- ב. ידחה את השערת האפס מכל מקרה.
- ג. ידחה את השערת האפס רק אם המבחן הנו דו צדי.
- ד. לא ניתן לדעת כי אין מספיק נתונים.

**(26)** מובהקות התוצאה (PV) היא גם :

- א. רמת המובהקות המינימאלית לדוחות השערת האפס.
- ב. רמת המובהקות המקסימאלית לדחית השערת האפס.
- ג. רמת המובהקות שנקבעת מראש על ידי החוקר טרם קיבל את תוצאות המחקר.
- ד. רמת המובהקות המינימאלית לאירוע דחית השערת האפס.

**(27)** בבדיקה השערות מסוימת התקבל שה-  $p-value = 0.0254$ , לכן :

- א. ברמת מובהקות של 0.01 אך לא של 0.05 נדחה את  $H_0$ .
- ב. ברמת מובהקות של 0.01 ושל 0.05 לא נדחה את  $H_0$ .
- ג. ברמת מובהקות של 0.05 אך לא של 0.01 נדחה את  $H_0$ .
- ד. ברמת מובהקות של 0.01 ושל 0.05 נדחה את  $H_0$ .

**(28)** רמת המובהקות במחקר הייתה 2% לכן.

- א. בסיכוי של 2% נדחה את השערת האפס.
- ב. בסיכוי של 2% לא נדחה את השערת האפס.
- ג. בסיכוי של 2% השערת האפס לא נכונה.
- ד. אף תשובה לא נכונה.

(29) נתון ש:  $(\sigma^2, \mu) \sim N \sim X$ . כמו כן נתונות ההשערות הבאות:  $\mu_0 < \mu$ :  $H_0$ ;  $\mu = \mu_0$ :  $H_1$ . חוקר בדק את ההשערות הללו על סמך מדגם שככל 10 תוצאות.

- $\sigma^2$  לא הייתה ידועה לחוקר. החוקר החליט לדוחות את השערת האפס ברמת מובהקות של 5%. אם הוא היה מגדיל את רמת המובהקות ל-10% אזי:
- cut-off הוא ידחה את השערת האפס.
  - cut-off הוא דוקא יקבל את השערת האפס.
  - cut-off לא ניתן לדעת מה תהיה מסקנתו.

(30) לצורך בדיקת השפעת היפנוזה על לימוד אנגלית, נבחרו 10 זוגות תאומים זהים. אחד התאומים למד אנגלית בהשפעת היפנוזה, והשני ללא היפנוזה. לאחר מכן נערכ לשניהם מבחן באנגלית. נניח שציוני המבחן מתפלגים נורמלית ללא ידיעת השונות האנטיתית. מספר דרגות החופש במבחן הוא:

- 9
- 19
- 18
- 8

(31) בתחום טיפת חלב מסוימת יש שני מכשירי שקילה. על מנת להשוות בין שני המשקלים נדגמו 4 תינוקות. כל תינוק בן חודשים נשקל בכל אחד מהמשקלים. להלן תוצאות השקילה (בק"ג):

	משקל במכשיר 1	4.5	9.6	0.7	2.5
	משקל במכשיר 2	3.5	6.9	1.7	0.5

נניח שהמשקלים מתפלגים נורמלית.  
המבחן שיש לבצע כאן הוא:  
א. מבחן Z למדגם יחיד.  
ב. מבחן T למדגם יחיד.  
ג. מבחן T למדגים בלתי תלויים.  
ד. מבחן T למדגים מזוגיים.

(32) כדי להשוות בין שני אצים נdagמו 5 תוצאות מריצת 100 מטר של כל אחד. זמני הריצה נרשמו ויש להניח שמתפלגים נורמלית. המטרה להשוות בין האצים. המבחן שיש לבצע כאן הוא:

- מבחן Z למדגם יחיד.
- מבחן T למדגם יחיד.
- מבחן T למדגים בלתי תלויים.
- מבחן T למדגים מזוגיים.

**(33)** סטטיסטיκאי ערך מבחן סטטיסטי. הוא חישב את עוצמת המבחן וקיבל 0. המשמעות של תוצאה זו היא :

- לעולם לא לדחות את השערת האפס כאשר היא לא נכונה.
- תמיד לדחות את השערת האפס כאשר היא נכונה.
- לעולם לא לדחות את השערת האפס כאשר היא נכונה.
- תמיד לדחות את השערת האפס כאשר היא לא נכונה.

**(34)** סטטיסטיκאי נתקבש לאמוד את הפרש הממוצעים של שני טיפולים לפי שני מדגמים מקרים בלתי תלויים. הוא חישב רוחסן ממוצע ברמת סמך 0.98, וקיבל את הרוחסן  $\mu_2 - \mu_1 = 4.5$ . אילו יתבקש החוקר לבדוק לפי אותם

נתונים את ההשערות :  $H_0: \mu_2 - \mu_1 = 0$  ;  $H_1: \mu_2 - \mu_1 \neq 0$ , ברמת מובהקות 0.05 מסקנתו תהיה :

- לדוחות את השערת האפס.
- לא לדוחות את השערת האפס.
- שלא ניתן לדעת את המסקנה עבור רמת מובהקות 0.05.
- שלא נתנות בשאלת סטיות התקן של האוכלוסיות, ולכן לא ניתן להסיק דבר.

**(35)** במטרה לבדוק האם קיים הבדל בין קווי זהב לבזק מבחינות ממוצע המחירים לשיחות בין"יל. נגדמו באקראי 7 מדינות ובעור כל מדינה נבדקה עלות דקט שיחה. בהנחה והמחקרים מתפללים נורמלית בנו רוחסן ממוצע ההפרשים וקיבלו :  $\mu_D = 0.2145 < 0.0293$  – רוחסן הסמך הוא ברמת סמך של 95%.

לכן מסקנת המחקר היא :

- ברמת מובהקות של 5% לא נוכל לקבוע שקיים הבדל בין החברות.
- ברמת מובהקות של 5% נקבע שקיים הבדל מובהק בין החברות.
- שלא ניתן לדעת מה המסקנה ברמת מובהקות של 5% כיון שלא נאמר מה ההגדרה של  $D$ .

**(36)** אם רמת מובהקות של מבחן סטטיסטי הינה 0, הכוונה היא :

- תמיד נדחה  $H_0$  כאשר היא נכונה, אך לא תמיד נדחה אותה כאשר היא לא נכונה.
- לא נדחה את  $H_0$  אף פעם.
- לא נדחה את  $H_0$  כאשר היא נכונה אך יתכן ונדחה אותה כאשר היא לא נכונה.
- כל התשובות לא נכונות.

(37) חוקר ביצע ניסוי. הוא ניסח את ההשערות הבאות:  $H_0: \mu = 10$ ,  $H_1: \mu \neq 10$

לצורך בדיקה הואלקח מוגן מקרי בגודל 5 מתוך אוכלוסייה המתפלגת נורמללית עם שונות לא ידועה. על סמך תוצאות המוגן הוא חישב וקבע:  $t_{\bar{x}} = -2.63$ . לכן המשקנה היא:

- א. הוא ידחה  $H_0$  ברמת מובהקות 0.1 אך לא כן ברמת מובהקות 0.05.
- ב. הוא ידחה  $H_0$  ברמת מובהקות 0.05 אך לא כן ברמת מובהקות 0.025
- ג. הוא ידחה  $H_0$  ברמת מובהקות 0.025 אך לא כן ברמת מובהקות 0.01
- ד. הוא לא ידחה  $H_0$  ברמת מובהקות 0.1

(38) האיגוד האמריקני לרפואת ילדים פרסם הנחיות חדשות הקובעות כי יש ליטול

תוספת יוד במהלך תקופת ההריון וההנקה. מושור במינרל זה עלול לגרום לפגיעה מוחית אצל העובר והתינוק. החלטה זו נקבעה על סמך מחקר בו השתתפו 1050 נשים שנלו יodium במהלך תקופת ההריון וההנקה. מתוך מותק הנשים שהשתתפו במחקר, רק 1-21 נמצאו ילדים בעלי פגיעה מוחית לעומת 3% באוכלוסייה הכללית. בנוסף, פורסם שהאיגוד האמריקאי מגיע למסקנותיו על סמך רמת מובהקות של 0.5%. מה הסיכוי לבצע טעות מסווג ראשון במחקר?

- א. 0.005
- ב. 0.03
- ג. 0.0287
- ד. 0.05

(39) חוקרת שיערה, כי משקלן של נשים כמנה לאחר החתונה גבוהה ממשקלן בעת החתונה.

החויקרת דגמה 15 נשים, ובדקתה את משקלן בשתי נקודות הזמן (בעת החתונה, ומשנה לאחריה), אך לא מצאה הבדל מובהק ברמת מובהקות 0.01. בהנחה, כי **במציאות** השערתה של החוקרת נכונה, סביר כי אם היא תגדיל את גודל המוגנים, אז:

- א. יקטן הסיכוי לטעות מסווג שני ( $\beta$ ).
- ב. תגדל רמת הביטחון ( $\alpha$ -1).
- ג. אף תשובה לא נכונה.
- ד. כל התשובות נכונות.

(40) איזה מהמשפטים הבאים נכון תמיד?

- א.  $\text{POWER} + \alpha + \beta = 1$
- ב.  $\text{POWER} = 0.5 - \beta$
- ג.  $\text{POWER} + \alpha = 1$
- ד.  $\beta + \alpha = 1$
- ה. הכל לא נכון.

**41)** מה נכון לגבי הנחת שיוויון השינויות ב מבחן D לדוגמים בלתי תלויים?

- א. היא אומרת שהשינויים המדגימות שוות.
- ב. בטעדייה אין שום דרך לבדוק השערה על הפרש בין תוצאות.
- ג. היא חשובה הן עבור דוגמים מזוגים והן עבור דוגמים בלתי תלויים.
- ד. אף תשובה אינה נכונה.

**42)** חוקר החליט לא לדוחות השערה ברמת מובהקות של  $\alpha$ . במידה וחוקר זה היה בודק השערה זו ברמת מובהקות של  $\alpha^2$  על סמך אותם נתונים, האם ההשערה תדחה?

- א. ההשערה תדחה.
- ב. ההשערה לא תדחה.
- ג. התשובה תלויות בעוצמת המבחן.
- ד. לא ניתן לדעת בוודאות אם ההשערה תדחה או לא.

**43)** חוקרת שיערה, כי בגילאי הגן בנות יותר תקשורתית מבנים. אם החוקרת תדגום אקרואית 30 בניים ו-30 בנות, ובמוגם יתקבל אותו ממוצע של ציון תקשורת. סטטיסטי המבחן יהיה :

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

**44)** עוצמה שווה ל-1 פרושה :

- א. לעולם לא לדוחות את השערת האפס כאשר היא נכונה.
- ב. תמיד לדוחות את השערת האפס כאשר היא נכונה.
- ג. לעולם לא לדוחות את השערת האפס כאשר היא לא נכונה.

**45)** מה מהבאים נבען לגבי מבחן D דוגמים מזוגים?

- א. כל התצפיות במחקר אינם תלויות זו בזו.
- ב. כל התצפיות במחקר תלויות זו בזו.
- ג. כל הצמידים של תצפיות במחקר אינם תלויים זה בזה.
- ד. התצפיות בתוך כל צמד אינם תלויות זו בזו.

- (46) לבדיקה ההשערה חד צדדית על התוחלת של התפלגות נורמלית  $\mu \geq 10$  ,  $H_0 : \mu = 10$  . נלקח מוגם והתקבלה רמת מובהקות מינימאלית לדחיתת השערת האפס 0.058 . לו רצינו לבדוק את ההשערה הדו צדדית  $\mu \neq 10$  ,  $H_1 : \mu = 10$  , אז על סמך תוצאה אותו המוגם ברמת מובהקות 0.05 :
- א. ניתן להכריע בין ההשערות רק אם שונות האוכלוסייה נתונה.
  - ב. מקבלים את השערת האפס.
  - ג. דוחים את השערת האפס.
  - ד. לא ניתן להכריע בין ההשערות שכן חסרים נתונים.

- (47) לבדיקה ההשערה חד צדדית ימנית  $\mu = 65$  ,  $H_1 : \mu > 65$  נלקח מוגם מקורי בגודל  $n$  מאוכלוסייה בעלת התפלגות נורמלית ושונות  $\sigma^2$  . רמת המובהקות היא 5% . נמצא שהעוצמה היא 0.9 . להלן 3 טענות :
- עבור מוגם בגודל  $n$  ורמות מובהקות 5% לבדיקה ההשערות :
    - .  $H_0 : \mu = 55$  ,  $H_1 : \mu = 60$  העוצמה תהיה גדולה מ-0.9.
    - .  $H_0 : \mu = 55$  ,  $H_1 : \mu = 65$  העוצמה תהיה גדולה מ-0.9.
    - .  $H_0 : \mu = 55$  ,  $H_1 : \mu = 65$  העוצמה תהיה קטנה מ-0.9.
- א. שלושת הטענות אינן נכונות.  
 ב. טענות 2 ו-3 אינן נכונות וטענה 1 נכונה.  
 ג. טענות 1 ו-2 נכונות וטענה 3 אינה נכונה.  
 ד. טענות 1 ו-3 אינן נכונות וטענה 2 נכונה.

**תשובות סופיות:**

שאלה	תשובה	שאלה	תשובה	שאלה
א	25	א	א	1
א	26	ז	ז	2
ג	27	א	ג	3
ז	28	ג	ג	4
א	29	ג	ג	5
א	30	א	א	6
ז	31	ב	ב	7
ג	32	ג	ג	8
א	33	א	א	9
ג	34	א	א	10
א	35	א	א	11
ג	36	א	א	12
א	37	א	א	13
א	38	א	א	14
א	39	ג	ג	15
ה	40	א	א	16
ז	41	ג	ג	17
ז	42	ב	ב	18
א	43	ג	ג	19
ז	44	ז	ז	20
ג	45	ג	ג	21
ב	46	ג	ג	22
ז	47	א	א	23
		ב	ב	24

## סטטיסטיקה

### פרק 27 - מבחני חי בربוע

#### תוכן העניינים

150 .....	1. מבחן לאי תלות.....
154 .....	2. מקדם המתאים של קרמר.....
156 .....	3. ניתוח פלטימ בבחן אי תלות

## מבחן חי בריבוע לאי תלות בין משתנים – רקע

מבחן לאי תלות מטרתו לבדוק האם קיים קשר בין שני משתנים. שני המשתנים שנבדקים צריכים להיות מחולקים למספר קטגוריות.

**מבנה המבחן:**

**השערות:**

אין תלות בין המשתנים  $H_0$ .

יש תלות בין המשתנים  $H_1$ .

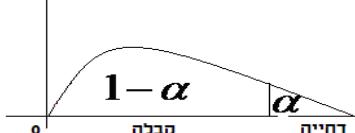
**כלל הכרעה:**

הערך הקритי נקבע על סמך התפלגות חי בריבוע. התפלגות זו היא אסימטרית חיובית ותלויה בדרגות החופש  $(r-1)(c-1)$ .  
 כאשר:  $r$  - מספר הקטגוריות של המשתנה שبشורות.  
 $c$  - מספר הקטגוריות של המשתנה שבעמודות.

הערך הקритי הוא:  $\chi^2_{1-\alpha,(r-1)(c-1)}$ , קלומר האחזוון ה-  $1 - \alpha$  בתפלגות חי בריבוע שדרגות החופש הן  $(r-1)(c-1)$ . אם  $\chi^2 > \chi^2_{1-\alpha,(r-1)(c-1)}$  או דוחים את השערת האפס.

$$\text{סטטיסטי המבחן: } \chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

כאשר:



$O_i$  - השכיחות נצפית במדגם בתא  $i$ .

$E_i$  - שכיחות צפואה במדגם בתא  $i$  בהנחה השערת האפס.

$$E_i = \frac{f(x) \cdot f(y)}{n}$$

**הערה:**

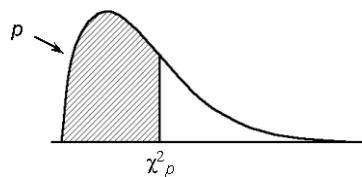
תנאי כדי לבצע את המבחן הוא  $5 \geq E_i$  לכל  $i$ . במידה ותנאי זה לא מתקיים יש אפשרות לאחד קטגוריות סמכות עד שהתנאי יתקיים.

תנאי חלופי: אין  $E$  קטן מ-1 וגם אין יותר מ 20% מהתאים  $E$  קטנים מ-5.

**דוגמה (הפתרון בהקלטה) :**

האם יש תלות בין המגדר לבין דעה מסוימת?  
יש לבדוק ברמת מובהקות של 5% על סמך תוצאות הסקר :

המגדר / דעתה	סה"כ	נמנע	נגד	بعد	גברים
נשים	20	60	20	50	10
סה"כ					

**טבלת התפלגות חי-בריבוע – ערכי החלוקה  $\chi^2_p$** 


df	$p$												
	.005	.01	.025	.05	.10	.25	.50	.75	.90	.95	.975	.99	.995
1	0.04393	0.03157	0.03982	0.0393	0.0158	0.102	0.455	1.32	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.103	0.211	0.575	1.39	2.77	4.61	5.99	7.38	9.21	10.6
3	0.0717	0.115	0.216	0.352	0.584	1.21	2.37	4.11	6.25	7.81	9.35	11.3	12.8
4	0.207	0.297	0.484	0.711	1.06	1.92	3.36	5.39	7.78	9.49	11.1	13.3	14.9
5	0.412	0.554	0.831	1.15	1.61	2.67	4.35	6.63	9.24	11.1	12.8	15.1	16.7
6	0.676	0.872	1.24	1.64	2.20	3.45	5.35	7.84	10.6	12.6	14.4	16.8	18.5
7	0.989	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.0	14.1	16.0	18.5	20.3
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.2	13.4	15.5	17.5	20.1	22.0
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.4	14.7	16.9	19.0	21.7	23.6
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.5	16.0	18.3	20.5	23.2	25.2
11	2.60	3.05	3.82	4.57	5.58	7.58	10.3	13.7	17.3	19.7	21.9	24.7	26.8
12	3.07	3.57	4.40	5.23	6.30	8.44	11.3	14.8	18.5	21.0	23.3	26.2	28.3
13	3.57	4.11	5.01	5.89	7.04	9.30	12.3	16.0	19.8	22.4	24.7	27.7	29.8
14	4.07	4.66	5.63	6.57	7.79	10.2	13.3	17.1	21.1	23.7	26.1	29.1	31.3
15	4.60	5.23	6.26	7.26	8.55	11.0	14.3	18.2	22.3	25.0	27.5	30.6	32.8
16	5.14	5.81	6.91	7.96	9.31	11.9	15.3	19.4	23.5	26.3	28.8	32.0	34.3
17	5.70	6.41	7.56	8.67	10.1	12.8	16.3	20.5	24.8	27.6	30.2	33.4	35.7
18	6.26	7.01	8.23	9.39	10.9	13.7	17.3	21.6	26.0	28.9	31.5	34.8	37.2
19	6.84	7.63	8.91	10.1	11.7	14.6	18.3	22.7	27.2	30.1	32.9	36.2	38.6
20	7.43	8.26	9.59	10.9	12.4	15.5	19.3	23.8	28.4	31.4	34.2	37.6	40.0
21	8.03	8.90	10.3	11.6	13.2	16.3	20.3	24.9	29.6	32.7	35.5	38.9	41.4
22	8.64	9.54	11.0	12.3	14.0	17.2	21.3	26.0	30.8	33.9	36.8	40.3	42.8
23	9.26	10.2	11.7	13.1	14.8	18.1	22.3	27.1	32.0	35.2	38.1	41.6	44.2
24	9.89	10.9	12.4	13.8	15.7	19.0	23.3	28.2	33.2	36.4	39.4	43.0	45.6
25	10.5	11.5	13.1	14.6	16.5	19.9	24.3	29.3	34.4	37.7	40.6	44.3	46.9
26	11.2	12.2	13.8	15.4	17.3	20.8	25.3	30.4	35.6	38.9	41.9	45.6	48.3
27	11.8	12.9	14.6	16.2	18.1	21.7	26.3	31.5	36.7	40.1	43.2	47.0	49.6
28	12.5	13.6	15.3	16.9	18.9	22.7	27.3	32.6	37.9	41.3	44.5	48.3	51.0
29	13.1	14.3	16.0	17.7	19.8	23.6	28.3	33.7	39.1	42.6	45.7	49.6	52.3
30	13.8	15.0	16.8	18.5	20.6	24.5	29.3	34.8	40.3	43.8	47.0	50.9	53.7

## שאלות

1) נבדקה ה תלות בין גודל הארגון לבין שבועות הרצון של העובדים.

להלן התוצאות:

סה"כ	גבואה	ביןונית	نمוכה	שבועות רצון	גודל המפעול
600	215	203	182		גודל
400	136	110	154		קטן
1000	351	313	336		סה"כ

מה המסקנה ברמת מובהקות של 2.5%?

2) מפעל עובד בשלוש משמרות. להלן מספר המוצרים הפוגומים והתקינים בכל

אחד מן המשמרות לפי מדגם שנעשה:

יום	ערב	לילה
פוגומים	60	50
תקינים	800	700

האם יש הבדל בין שיעורי הפוגומים במשמרות השונות?

הסיקו עבור רמת מובהקות של  $\alpha = 0.05$ .

3) נדגמו 50 מוצרים ממפעל מסוים מתוך 30 מוצרים שיוצרו ביום 17 נבחרו

לייצוא מתוך המוצרים שיוצרו בלילה 10 נבחרו לייצוא.

האם יש קשר בין להיות מוצר לייצוא למועד שבו הוא יוצר?

בדקו ברמת בטחון של 95%.

## תשובות סופיות

1) נסיק שיש קשר בין גודל הארגון לשבועות הרצון של העובדים.

2) נסיק שאין הבדל מובהק בין שיעור הפוגומים במשמרות השונות.

3) נסיק שאין קשר בין להיות מוצר לייצוא למועד שבו הוא יוצר.

## מדדי קשר-מדד הקשר של קרמר – רקע

מתי משתמשים במדד זהה? – כאשר אחד המשתנים הוא מסולםשמי 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 רוצחים לבדוק את הקשר בין מין לדעה בנושא מסוים, שאלו 100 גברים ו- 100 נשים האם הם בעד/נגד/נמנעים באיזשהו נושא. להלן טבלת שכיחויות המשותפת שהתקבלה.

בחזור של קרמר הטבלה נקראת טבלת O

(observed)

X - מין (גבר/אישה) – סולםשמי.

Y - דעה (בעד/נמנע/נגד) – סולםשמי/סדר.

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

### שלבים בחישוב $r_c$ :

#### שלב א' :

בנייה את טבלת E (Expected)

נעתק את המסגרת של טבלת O ואז כל  $E_i = \frac{f(x) \cdot f(y)}{n}$

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

#### שלב ב' :

$$\text{נחשב } \chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

#### שלב ג' :

$$\text{נחשב : } r_c = \sqrt{\frac{1}{n(L-1)} \chi^2}$$

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

## שאלות

1) להלן תוצאות מחקר שבודק את הקשר בין מין להשכלה. לגבי כל נחקר נבדק

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

המין שלו והשלתו.  
להלן התוצאות:

האם קיימים קשר בין מין להשכלה? נמקו!

2) נלקחו 200 אנשים שמתוכם 60 הוכיחו שהם עוסקים בפעילויות גופנית סדירה.

מתוך אלו שעוסקים בפעילויות גופנית סדירה 50 נמצאו במצב בריאותי תקין.

מתוך אלו שלא עוסקים בפעילויות גופנית סדירה 90 נמצאו במצב בריאותי תקין.

א. בנו טבלה שכיהות משותפת לנתחים שהוצעו בשאלת.

ב. האם קיימים קשר בין פעילות גופנית למצב בריאותי?  
חשבו לפי מודד הקשר של קרמר.

## תשובות סופיות

1) קיימים קשר בעוצמה בינונית בין המין להשכלה מקדם המתאים של קרמר הוא 0.595.

2) א. להלן טבלה:  
ב. מודד קרמר 0.19 מעיד על קשר בעוצמה נמוכה.

$f(x)$	לא נכון	נכון	$x/y$
60	10	50	כן
140	50	90	לא
200	60	140	$f(y)$

## פלטיטים על מבחן לאי תלות – רקע

מבחן לאי תלות מטרתו לבדוק האם קיים קשר בין שני משתנים. שני המשתנים שנבדקים צריכים להיות מחולקים למספר קטגוריות.

**מבנה המבחן:**

**השערות:**

אין תלות בין המשתנים :  $H_0$ .

יש תלות בין המשתנים :  $H_1$ .

**דרגות חופש :**  $d.f = (r-1)(c-1)$ .

כאשר :  $r$  - מספר הקטגוריות של המשתנה שبشורות.

$c$  - ספר הקטגוריות של המשתנה שבעמודות.

$$\text{סטטיסטי המבחן: } \chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

$O_i$  - שכיחות נצפית במדגם בתא  $i$ .

$E_i$  - שכיחות צפואה במדגם בתא  $i$  בהנחה השערת האפס.

$$E_i = \frac{f(x) \cdot f(y)}{n}$$

**הערה :**

תנאי כדי לבצע את המבחן הוא  $E_i \geq 5$  לכל  $i$ . במידה ותנאי זה לא מתקיים יש

אפשרויות לאחד קטגוריות סמכות עד שהתנאי יתקיים.

**תנאי חלופי:** אין  $E$  קטן מ-1 וגם אין יותר מ 20% מהתאים  $E$  קטן מ-5.

**דוגמה (פתרון בהקלטה) :**

- במחקר רצוי לבדוק את הקשר בין צבע שיער לבין צבע עיניים של אנשים. הפלטים שהתקבלו מצורפים.
- א. מהו המבחן הסטטיסטי המתאים?
  - ב. כמה קטגוריות יש לכל משתנה?
  - ג. רשמו את השערות המחקר.
  - ד. מה מספר דרגות החופש?
  - ה. כמה אנשים במדגם נמצאו עם שיער חום?
  - ו. כמה אנשים הייתה מזכה במדגם שהיה להם שיער חום ועיניים יroxות בהנחה ואין קשר בין צבע שיער לצבע עיניים?
  - ז. מתוך הבלונדיינים מה אחוז בעלי עיניים כחולות במדגם?
  - ח. מתוך בעלי עיניים יroxות מה אחוז הבלונדיינים במדגם?
  - ט. מה ערכו של סטטיסטי המבחן ומהי מובಹות התוצאה?
  - י. מה מסקנת המחקר?  $\alpha = 5\%$

להלן הפלטים שהתקבלו:

### Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
hair_color * eye_color	78	100.0%	0	0.0%	78	100.0%

### hair\_color \* eye\_color Crosstabulation

		eye_color			Total
		brown	green	Blue	
hair_color	black	Count	13	7	27
		Expected Count	10.7	8.3	27.0
		% within hair_color	48.1%	25.9%	25.9%
	brown	% within eye_color	41.9%	29.2%	30.4%
		Count	12	12	30
		Expected Count	11.9	9.2	30.0
	blond	% within hair_color	40.0%	40.0%	20.0%
		% within eye_color	38.7%	50.0%	26.1%
		Count	6	5	21
Total	blond	Expected Count	8.3	6.5	21.0
		% within hair_color	28.6%	23.8%	47.6%
		% within eye_color	19.4%	20.8%	43.5%
	Total	Count	31	24	78
		Expected Count	31.0	24.0	78.0
		% within hair_color	39.7%	30.8%	29.5%
		% within eye_color	100.0%	100.0%	100.0%

### Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	5.880 <sup>a</sup>	4	.208
Likelihood Ratio	5.641	4	.228
Linear-by-Linear Association	2.682	1	.101
N of Valid Cases	78		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 6.19.

## שאלות

- 1) בסקר שנעשה על ידי משרד ראש הממשלה נדגו 60 אזרחים. כל אזרח נשאל על מגדרו והאם הוא بعد הקמת מדינה פלסטינית.
- א. מה ההשערות הנבדקות ומהו סטטיסטי המבחן?
  - ב. אם סטטיסטי המבחן היה גדול כיצד הדבר היה משפייע על SIG שבפלט.
  - ג. האם קיים קשר בין מגדר ודעה ברמת מובהקות של 5%?
  - ד. מהו האומדן לאחוז התומכים במדינה פלסטינית מתוך הגברים?
  - ה. Aiיזה אחוז מהנשאלים שהיו بعد מדינה פלשתינית הם גברים?

להלן הפלטים:

### Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	1.973 <sup>a</sup>	2	.373
Likelihood Ratio	1.987	2	.370
Linear-by-Linear Association	1.882	1	.170
N of Valid Cases	60		

a. 0 cells (.0%) have expected count less than 5.

b. The minimum expected count is 7.25.

### gender \* opinion Crosstabulation

		opinion			Total	
		yes	now	no opinion		
gender	Male	Count	10	10	9	29
		Expected Count	12.6	9.2	7.3	29.0
		% within gender	34.5%	34.5%	31.0%	100.0%
		% within opinion	38.5%	52.6%	60.0%	48.3%
		% of Total	16.7%	16.7%	15.0%	48.3%
	female	Count	16	9	6	31
		Expected Count	13.4	9.8	7.8	31.0
		% within gender	51.6%	29.0%	19.4%	100.0%
		% within opinion	61.5%	47.4%	40.0%	51.7%
		% of Total	26.7%	15.0%	10.0%	51.7%
	Total	Count	26	19	15	60
		Expected Count	26.0	19.0	15.0	60.0
		% within gender	43.3%	31.7%	25.0%	100.0%
		% within opinion	100.0%	100.0%	100.0%	100.0%
		% of Total	43.3%	31.7%	25.0%	100.0%

2) להלן פلت על סמך סקר שנעשה בקרב סטודנטים, בסקר נשאלו הסטודנטים על המוזיקה אותה הם מעדיפים וצורת הבילוי המועדף עליהם.

#### Crosstab

Count

		בילוי			Total
		קריאה	ספורט	睦עדיון	
מוזיקה	רוק	0	0	11	11
	פופ	1	6	8	15
	קלסי	5	6	9	20
Total		6	12	28	46

#### Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	11.929 <sup>a</sup>	?	.018
N of Valid Cases	46		

a. 5 cells (55.6%) have expected count less than 5.

b. The minimum expected count is 1.43.

א. בין אלו משתנים נבדק הקשר? כמה קטגוריות לכל משתנה?

ב. האם התנאים של המודל מתקיים?

ג. מה מספר דרגות החופש בבדיקה הנ"ל?

ד. מה ההשערות של הבדיקה?

- (3) מחקר התענין לבדוק את הקשר בין רמת הכנסה של משפחה לבין צריכת עגבנייהות ארגניות. הפלטים מצורפים.
- השלימו את שלושת המספרים החסרים בטבלה (היכן שיש סימני שאלה).
  - מה ערכו של חי בריון הסטטיסטי.
  - תנו הערה למועדוקות התוצאות לבדוק הקשר בין רמת הכנסה של משפחה לבין צריכת עגבנייהות ארגניות.

### רמת\_הכנסה \* צרךן עגבנייהות Crosstabulation

		צרךן עגבנייהות		Total
		לא ארגני	ארוגני	
הרבה מתחת לממוצע	Count	17	42	59
הכנסה_רמת within	% within	28.8%	?	100.0%
עצביות צרךן חיקוי	% within	13.6%	33.6%	23.6%
מתחת לממוצע	Count	27	22	49
הכנסה_רמת within	% within	55.1%	44.9%	100.0%
עצביות צרךן חיקוי	% within	?	17.6%	19.6%
ממוצע	Count	31	29	60
הכנסה_רמת within	% within	51.7%	48.3%	100.0%
עצביות צרךן חיקוי	% within	24.8%	23.2%	24.0%
מעל הממוצע	Count	44	26	70
הכנסה_רמת within	% within	62.9%	37.1%	100.0%
עצביות צרךן חיקוי	% within	35.2%	20.8%	28.0%
הרבה מעל הממוצע	Count	?	6	12
הכנסה_רמת within	% within	50.0%	50.0%	100.0%
עצביות צרךן חיקוי	% within	4.8%	4.8%	4.8%
Total	Count	125	125	250

- (4) חוקר בדק את הקשר בין צבע השיער לבין צבע העיניים בעזרת מבחן Chi בריבוע בקרב 52 נבדקים. תוצאות המבחן מוצגות בטבלה. בנוסף ידוע כי סטטיסטי המבחן שהתקבל מעיבוד הנתונים הוא 8.08.
- מה תהיה מסקנת המבחן ברמתה מובהקות של 1%?
  - מה ערכו של E עבור עיניים כחולות וצבע שיער כהה.
  - מה יהיה בקירוב ערכו של מקדם המתאים של קרמר?
  - מהי פרופורציה בעלי צבע השיער הבاهיר מקרב בעלי העיניים הירוקות?

להלן הפלט:

### צבע עיניים \* צבע שיער Crosstabulation

		צבע שיער		Total
		כהה	בהיר	
צבע עיניים	כחול	Count		
		% within		50.0% 100.0%
		% within		21.6% 30.8%
		% of Total		15.4% 30.8%
חום	חום	Count		
		% within		83.3% 100.0%
		% within		27.0% 23.1%
		% of Total		19.2% 23.1%
ירוק	ירוק	Count		
		% within		79.2% 100.0%
		% within		51.4% 46.2%
		% of Total		36.5% 46.2%
Total		Count		
		% within		71.2% 100.0%
		% within		100.0% 100.0%
		% of Total		71.2% 100.0%

- 5) במחקר מסוים רצוי לבדוק האם יש קשר בין המגדר להוצאה על לבוש במשך שנה. דגמו באופן מקרי גברים ונשים ובדקו את רמת ההוצאה שלהם על לבוש השנה האחרונות. חוקר א' בדק האם קיימים הבדלים תוחלוטים בהוצאה בין גברים לנשים. חוקר ב' קיבץ את ההוצאה לקטגוריות ובאופן הזה בדק האם קיימים הבדלים בתפלגות ההוצאה בין גברים לנשים.
- הקטגוריות חולקו לשולש קבוצות הוצאה.
- איזה פלט מתאים לאיזה אחד מהחוקרים? נמקו.
  - מה מסקנתו של חוקר א'? בדקו ברמת מובהקות  $\alpha = 0.05$ . (רשמו השערות, נחו הנחות, ציינו כלל החלטה ותנו מסקנה במונחי המשתנים).
  - איזו טעות יכולה להיות במסקנתו של חוקר א'?  
נסחו את הטעות במונחי השאלה.
  - מהי מסקנתו של חוקר ב'? בדקו ברמת מובהקות  $\alpha = 0.05$ . (רשמו השערות, נחו הנחות, ציינו כלל החלטה ורשמו מסקנה במונחי המשתנים).
  - איזו טעות יכולה להיות במסקנתו של חוקר ב'?  
נסחו זאת במונחי השאלה.
  - כיצד ניתן ליישב את מסקנות שני החוקרים?

להלן פלט ראשון:

### T-Test

#### Group Statistics

gender	N	Mean	Std. Deviation	Std. Error Mean
female	40	2.9000	1.15025	.18187
male	40	2.6000	2.52982	.40000
dimension n1				

Independent Samples Test

	Levene's Test for Equality of Variances		t-test for Equality of Means						95% Confidence Interval of the Difference	
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference			
								Lower	Upper	
expese	Equal variances assumed	16.805	.000	.683	.78	.497	.30000	.43941	-.57479	1.17479
	Equal variances not assumed			.683	54.464	.498	.30000	.43941	-.58078	1.18078

להלן פلت שני:

**Crosstabs****Case Processing Summary**

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
gender * category	80	100.0%	0	.0%	80	100.0%

**gender \* category Crosstabulation**

		category			Total	
		a	b	c		
gender	Female	Count	2	30	8	40
		Expected Count	11.0	21.0	8.0	40.0
		% within gender	5.0%	75.0%	20.0%	100.0%
		% within category	9.1%	71.4%	50.0%	50.0%
	Male	Count	20	12	8	40
		Expected Count	11.0	21.0	8.0	40.0
		% within gender	50.0%	30.0%	20.0%	100.0%
		% within category	90.9%	28.6%	50.0%	50.0%
	Total	Count	22	42	16	80
		Expected Count	22.0	42.0	16.0	80.0
		% within gender	27.5%	52.5%	20.0%	100.0%
		% within category	100.0%	100.0%	100.0%	100.0%

**Chi-Square Tests**

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	22.442 <sup>a</sup>	2	.000
Likelihood Ratio	25.064	2	.000
N of Valid Cases	80		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 8.00.

## תשובות סופיות

- (1) א. 1.973      ב. קטן .  
 ג. לא נדחה  $H_0$  .  
 ד. 34.5%  
 ב. 38.5%
- (2) א. בילוי מועדף ומוזיקה מועדף עם 3 קטגוריות לכל משתנה.  
 ב. לא  
 ג. 4.  
 ד.  $H_0$  אין תלות בין בילוי מועדף למוזיקה מועדף.
- (3) א. 21.6%, 71.2%  
 ב. 15.8      ג. קטן מ-0.005.
- (4) א. לא נדחה  $H_0$  .  
 ב. 11.4      ג. 0.394  
 ד. 20.8%.
- (5) א. פלא אי- חוקר א', פלא ב'- חוקר ב'.  
 ב. קיבל את  $H_0$ .  
 ג. טעות מסוג שני- הכרענו שאין הבדל בין גברים לנשים למרות שיש במציאות הבדל.  
 ד. קיבל את  $H_1$ .
- ה. טעות מסוג ראשון- הכרענו שיש קשר בין מין להוצאה למרות שבמציאות אין קשר.  
 ו. כל חוקר פעל בשיטה סטטיסטית שונה ובמצב כזה יתכן מסקנות סותרות.

## סטטיסטיקה

פרק 28 - מדריך הקשר ספרמן

תוכן העניינים

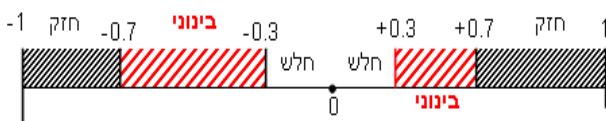
1. כללי .....

167 .....

## מדד 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 משתנה אחד עולה, השני עולה ללא יוצא מן הכלל.

**קשר דירוגי חיובי חלקי:**  
 מקדם המתאים בין 0 ל-1.  
 ככל משתנה אחד עולה, השני יש נטייה לעלות אך לא באופן מוחלט.

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

**קשר דירוגי שלילי חלקי:**  
 מקדם המתאים הוא בין 0 ל-(-1).  
 ככל משתנה אחד עולה לשני יש נטייה לרדת אך לא באופן מוחלט.  
 על מנת לחשב את הקשר יש לבצע פערות דירוג (RANK).  
 כאשר מדרגים, אם יש כמה תצפויות שתופסות את אותו הערך אז הדירוג שלהם הוא הממוצע של המקומות שהן תופסות.

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

הנוסחה של מדד הקשר :

**דוגמה (פתרון בהקלטה):**

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

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

מהי מידת ההתאמה בין ציוני השופטים?

X - ציון שופט א (סולם סדר).

Y - ציון שופט ב (סולם סדר).

**שאלות:**

1) בתרחות יופי חילקו שני שופטים ציוניים למועדות :

מספר מועדות	7	6	5	4	3	2	1
ציוון שופט א'	6	5	9	8	6	8	7
ציוון שופט ב'	7	5	9	8	7	8	8

האם קיים קשר בין שתי הערכות השופטים? נמקו והסבירו!

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

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

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 למספר החיסורים שלו.  
מדד הקשר : -0.85.

3) לא נכון.

## סטטיסטיקה

### פרק 29 - מקדם המתאים ( מחד קשור ) הלינארי ומובהקוותו

#### תוכן העניינים

170 .....	1. מקדם המתאים הלינארי ( פירסון ) .....
181 .....	2. חישוב מקדם המתאים הלינארי ( פירסון ) .....
186 .....	3. בדיקת השערות על מקדם המתאים הלינארי .....
190 .....	4. ניתוח פלטימס על מקדם המתאים הלינארי .....

## 막דם המתאים (מדד קשר) הליינארי ומובהקותו

### **מדד הקשר הליינארי (פירסון) – מבוא**

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

#### **דוגמה (פתרו בהקלטה) :**

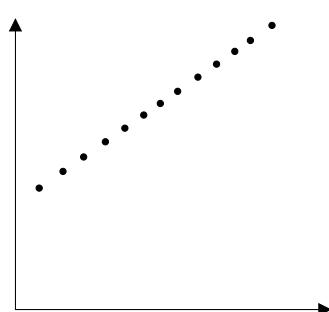
בבנייה 8 דירות בדקו לכל דירה את מספר החדרים שלה וכמו כן את מספר הנפשות הגורות בדירה. להלן התוצאות שהתקבלו :

מספר חדרים בדירה	מספר הנפשות בדירה
4	4
5	4

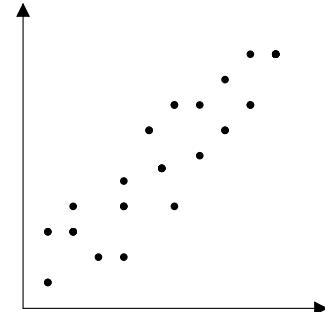
- 1) כמה תציפות ישנן בדוגמה?
- 2) כמה משתנים ישנים בדוגמה, מי הם?
- 3) שרטטו לנוטונים דיאגרמת פיזור.
- 4) מי המשתנה תלוי ומיהו המשתנה הבלתי תלוי?

### דיאגרמות פיזור לקשר בין משתנים וניתוחם

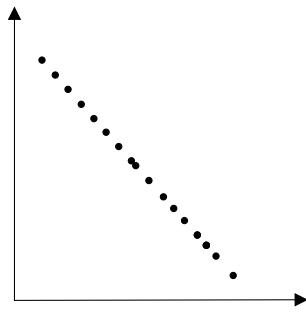
**קשר לנארוי חיובי מלא**



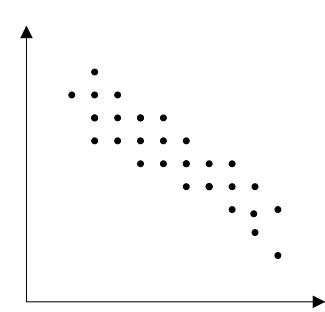
**קשר לנארוי חיובי חלק**



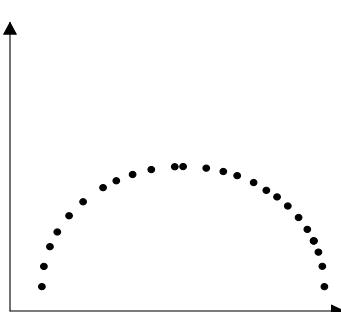
**קשר לנארוי שלילי מלא**



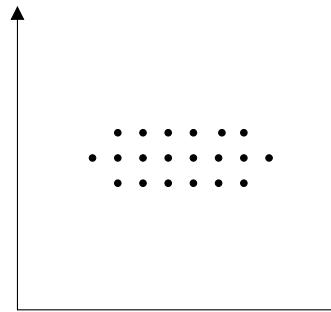
**קשר לנארוי שלילי חלק**



**אין קשר לנארוי**



**אין קשר**



#### משמעות מקדם המתאים:

כדי לבדוק עד כמה קיים קשר לנארוי בין שני המשתנים ישנו מדד קשר שנקרא גם מקדם המתאים הלינארי הידוע גם בשם מקדם המתאים של פירסון. מקדם מתאים זה מקבל ערכים בין 1 ל-1.

-1

0

1

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

### **מתאים חיובי מלא (מقدم מותאם 1):**

קיים קשר לנארי מלא בו השיפוע  $a$  יהיה חיובי ואילו מותאם שלילי (מقدم מותאם-1) מלא אומר שקיים קשר לנארי מלא בו השיפוע  $a$  שלילי.

### **מתאים חיובי חלק:**

כל משתנה אחד עולה לשני יש נטייה לעלות בערכו אבל לא קיימת נוסחה לינארית שמקשרת את  $X$  ל- $Y$  באופן מוחלט ואילו מותאם שלילי חלקי אומר שככל המשתנה אחד עולה לשני יש נטייה לרדת אבל לא קיימת נוסחה לינארית שמקשרת את  $X$  ל- $Y$  באופן מוחלט. ככל שמדובר המתאים הקרוב לאפס עצמת הקשר יותר חלשה וככל שהמדד רחוק יותר מהאפס העוצמה יותר חזקה. לsicום, מقدم המתאים בודק את עצמת הקשר הלינארי, ואת כיוון הקשר.

מقدم המתאים הלינארי אינו מושפע מייחדות המדידה. כל שינוי ביחסות המדידה של המשתנים, לא ישנה את מقدم המתאים.

מדד הקשר הלינארי באוכולוסייה, שנראה גם מقدم המתאים של פירסון או מדד הקשר של פירסון באוכולוסייה מסומן ב:  $r$  - פרמטר המאפיין את עצמת הקשר הלינארי באוכולוסייה וכיונו בין שני המשתנים הנחקרים. כאשר:

- מדד הקשר הלינארי במדגם שמהווה אומד לפרמטר  $r$ .

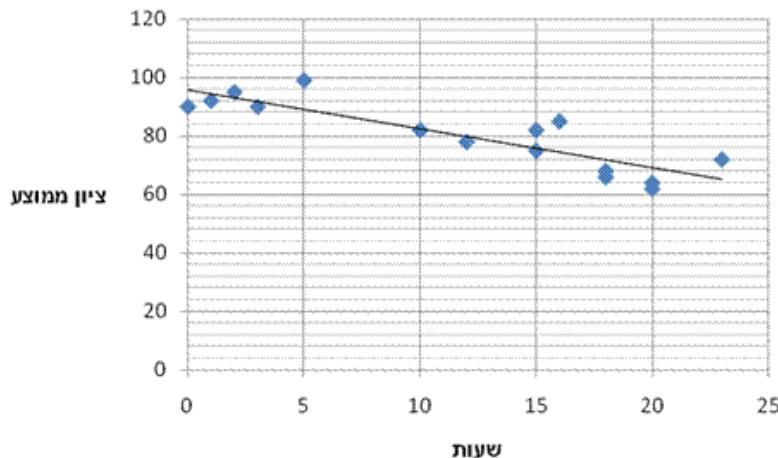
קיומו של מותאם בין שני משתנים אינו מצביע על סיבות בבחירה. למשל, אם נמצא מותאם חיובי בין כמות הסוכרזיות שאדם אוכל לבין משקל שלו אין זה אומר שהסיבה להשמנה היא הסוכרזית. מדד הקשר של פירסון הוא מדד קשר סימטרי,قولר אם נחליף את  $X$  ב- $Y$  התוצאה תהיה זהה.

### **דוגמה (פתרון בהקלטה):**

- מה ניתן להגיד על מועד המתאים של שני המשתנים על סמך דיאגרמת הפיזור שרטטנו?
- אם היינו משנים את הشرط כך שבציר האנכי היה המשתנה "מספר החדרים" ובציר האופקי היה "מספר הנפשות", האם הדבר היה משנה על מדד הקשר של פירסון?

## 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 השבועית לבין הציון הממוצע של הסמסטר? מה ניתן להגיד על עוצמת הקשר?

**2)** להלן טבלה המסכםת את מקדמי המתאים הליינארי בין ציוני מבחנים שונים שהתקבלו עבור תלמידים בכיתה מסוימת:

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

א. השלימו את מקדמי המתאים שמשמעותם בסימן שאלה בטבלה.

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

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

א. מיהו המשתנה התלו依 ומיהו המשתנה הבלתי תלוי בדוגמה זו?

ב. שרטטו דיאגרמת פיזור לנוטונים.

ג. מה ניתן לומר על הקשר בין המשתנים במדגם?

ד. מסתבר שבסוףו של דבר נתנו פקטור של 5

נקודות לציון הסופי. כיצד הדבר היה משנה את מקדם המתאים של המדגם?

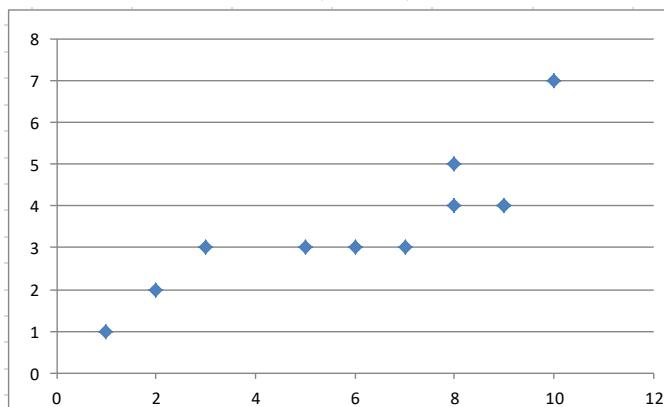
4) בتحقנה המטאורולוגית רצוי לבדוק את הקשר שבין הטמפרטורה במערכות כלזיות לכמות המשקעים במ"מ. הם אספו נתונים על 10 ימים במהלך חודש ינואר. המתאים שהתקבל היה 0.8.

א. השלימו את המשפט:

בחודש ינואר ככל שהטמפרטורה היומית נוטה לרדת, כך כמות המשקעים נוטה \_\_\_\_\_.

ב. הוחלט להעביר את הטמפרטורה למערכות פרנהייט על מנת שיוכלו להשוות אותה לנ נתונים מארה"ב. נוסחת המעבר היא  $F^0 = 32 + \frac{9}{5}C^0$ . כיצד הדבר ישפיע על מקדם המתאים בין הטמפרטורה במערכות פרניאיט לכמות המשקעים במ"מ?

5) להלן דיאגרמת פיזור המראה קשר בין שני משתנים:



א. השלימו: ניתן לראות קשר הוא לינארי \_\_\_\_\_ (מלאו חלקי) כיון שהקשר הוא (חיובי ושלילי).

ב. השלימו: אם היינו מושפעים תצפית שערך ה- X שלה הוא 4 וערך ה- Y שלה הוא 7, מקדם המתאים של פירסון היה \_\_\_\_\_ (גדלו קטו לא משתנה).

שאלות רב ברירה (יש לבחור את התשובה הנכונה):

6) חוקר אקלים דגם כמה ימים בשנה ומדד את הטמפרטורה בטורונטו שבקנדזה ואת הטמפרטורה בסידני שבאוסטרליה באותו היום. הוא חישב ומצא מקדם מתאים שלילי בין הטמפרטורה היומית בטורונטו לבין הטמפרטורה היומית בסידני. משמעות מקדם המתאים השלילי בדגם:

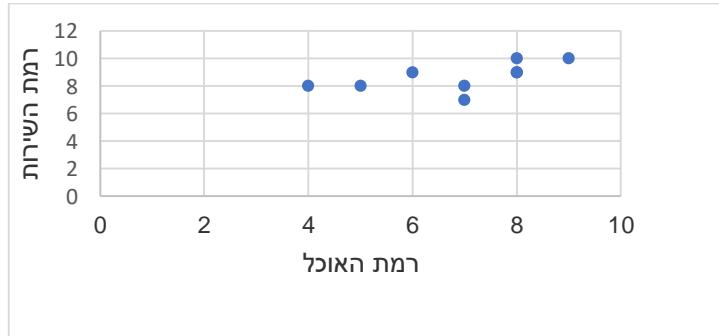
א. אין קשר בין הטמפרטורה בטורונטו לבין הטמפרטורה בסידני ביום שנדגמו.

ב. בדגם, רוב הטמפרטורות בטורונטו היו שליליות.

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

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

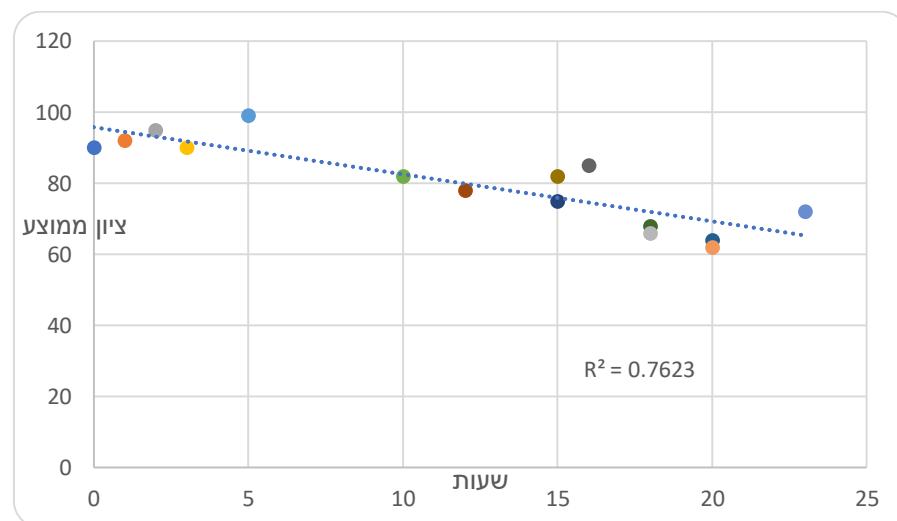
- 7) בסקר שביעות רצון שנערך בבית הקפה "fat لלחס" התבקוו הליקות לדרג את מידת שביעות הרצון שלהם (בסולם 1-10) בשני נושאים: רמת האוכל ורמת השירות.



מה יהיה ערכו של מקדם המתאים ( $r$ )?

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

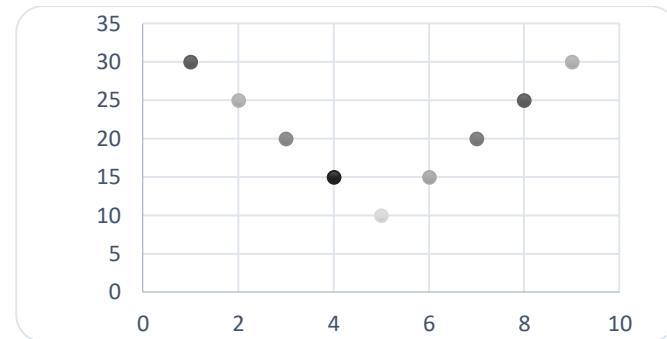
- 8) חוקר רצה לאפיין את הקשר בין מספר השעות בשבוע שסטודנט מקדיש לבילויים לבין הציון הממוצע שלו בסוף הסמסטר. לשם כך הוא אסף נתונים של 15 סטודנטים ויוצר דיאגרמת פיזור.



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

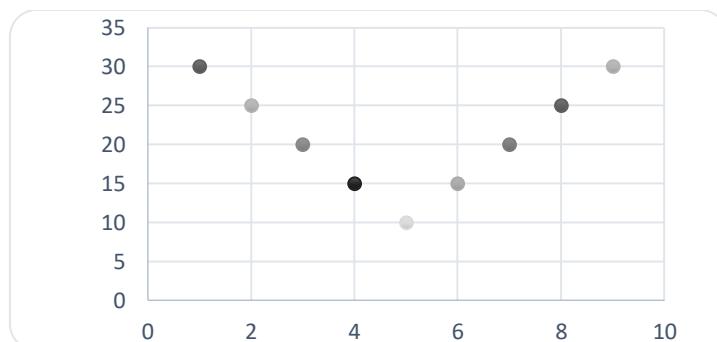
- א. ככל שմבלים יותר הציון נוטה לרדת.
- ב. אין קשר בין שעות הבילוי לציון.
- ג. ככל שմבלים פחות הציון נוטה לרדת.
- ד. ככל שהציון נוטה לרדת הסטודנט מבליה פחות.

**9)** התרשימים הבא מתאר קשר בין שני משתנים, איזה מהמתאים הבאים הוא המתאים ביותר לתיאור הקשר בין שני המשתנים?



- א.  $1 = z$  היות ושני המשתנים יוצרים קווים ישרים.
- ב.  $2 = z$  היות ויש שני קווים בעלי קשר מושלם.
- ג.  $0 = z$  היות והקו יורד ולאחר מכן עולה באותו האופן.
- ד.  $1 \pm 1 = z$  היות ויש קו עולה וגם קו יורד.

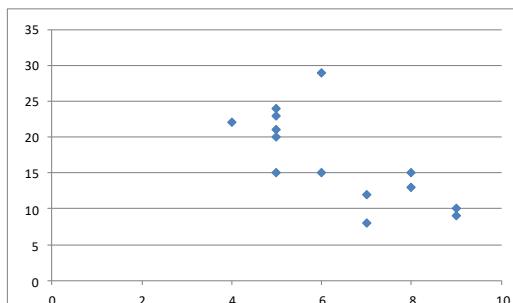
**10)** התרשימים הבא מתאר דיאגרמת פיזור.



איזה טענה נכונה?

- א. בתרשימים מוצג הקשר בין שני משתנים.
- ב. בתרשימים מוצג הקשר בין 9 משתנים.
- ג. בתרשימים מוצג הקשר בין 10 משתנים.
- ד. אין לדעת כמה משתנים מוצגים בתרשימים.

בגרף הבא מתוארת דיאגרמת פיזור של שני משתנים :



X - (משתנה בלתי תלוי בציר האופקי)  
ו- Y (משתנה תלוי).

במדגם התקבל  $r^2 = 0.52$ .

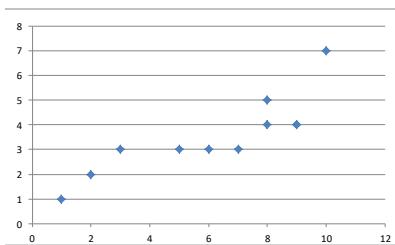
11) לאור הנתונים המופיעים בדיאגרמה, איזה מבחן הערכים הבאים מתאים להיות התוצאה של  $r$  ?

- א. -0.52
- ב. 0.72
- ג. -0.72
- ד. 0.52

12) אם מקדם המתאים בין שני משתנים הוא 1, אז :

- א. הערכים של המשתנים הם חיוביים.
- ב. עברו 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 המתאים בין שני המשתנים ?



- א. 1
- ב. 0.85
- ג. 0.15
- ד. 0

14) בבדיקה קשר בין שני משתנים התקבל :  $r = -1$ .  
א. קיימת נוסחה לינארית הקושרת בין כל התצפיות.  
ב. לא קיים קשר בין שני המשתנים.  
ג. ככל משתנה אחד נוטה לרדת גם לשני יש נטייה לרדת.  
ד. קיים קשר בין שני המשתנים, אך לא ניתן לדעת מאיזה סוג.

**15)** לפי הפטגס "רחוק מהעיר, רחוק מהלב", יש קשר \_\_\_\_ בין קרבה פיזית לקרבה נפשית.

- א. חיובי
- ב. שלילי
- ג. אפסי
- ד. לא ניתן לדעת.

**16)** מבחן אמייר הינו מבחן מיוון באנגלית של המרכז הארצי לבחינות והערכתה. הציון המינימלי בבחינה הינו 150 והמаксימלי הינו 250. בקורס הכנה לבחן השתתפו 19 תלמידים. להלן הציונים שלהם על פי פلت 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 עמודות הציונים שטראה לכל תלמיד כמה נקודות חסרות לו כדי להשלים לציוון המקסימלי בבחינה.

מה יהיה מקדם המתאים בין שתי העמודות (תלמיד, מקדם המתאים בין הציון לבין הנקודות החסרות)?

- א. -1
- ב. 1
- ג. -0.5
- ד. 0.5

**17)** מקדם המתאים בין שטחי דירה למחר שלחם חושב ונמצא 1.2. מה נובע לכך?

- 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 למחר שדייה.
- ד. מצב כזה שמתואר הנתונים לא אפשרי.

**18)** אם ניקח 10 אנשים וונרשום לכל אדם את הגובה במטר וכמה כו' את הגובה בס"מ. מה יהיה מקדם המתאים בין גובה האדם במטר לגובה האדם בס"מ?

- א. 1
- ב. 0
- ג. -1
- ד. לא ניתן לדעת.

- 19)** נמצא מתאים חיובי בעוצמה גבוהה בין  $X$  – ציון בගירות בלשון ל  $Y$  – ציון בගירות במתמטיקה. אילו מהמשפטים הבאים נכון?
- ניתן לומר שאחת מהסיבות להבדלים שיש לסטודנטים במתמטיקה נובעים מההבדלים שיש להם בלשון.
  - קיימת נוסחה של קו ישר שקשורה בין ציון בගירות במתמטיקה לציון בගירות בלשון.
  - לא יוצא מן הכלל, ניתן להגיד שככל תלמיד שמציל יותר מטלמיד אחר בלשונו גם יצליח יותר מאותו תלמיד במתמטיקה.
  - אף אחד מהטענות שהוצעו אינה בהכרח נכונה.
- 20)** עברו 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. מוקדם המתאים של פירסון יהיה בהכרח:
- 1
  - 1
  - 0
  - אף אחת מהתשובות.

## תשובות סופיות

- ב. הקשר חלקי, כיוןו הקשר שלילי.  
ב. ספורט ולשון.

- (1) א. שעות בילוי.  
(2) א. להלן טבלה:

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

- ב. ראה גרף בפתרון וידאו.  
ד. מקדם המתאים לא היה משתנה.  
ב. לא ישפיע על מקדם המתאים.  
ב. קטן.

- (3) א. ב'ית- מס' שעות התרגול, תלוי- ציון.  
ג. קשר לינארי חיובי חלקי.

- (4) א. עלות.  
(5) א. חלקי, חיובי.  
(6) ד'. ד'. א'. א'. ג'. (10)  
(11) ג'. א'. א'. א'. א'. (15)  
(16) א'. א'. ד'. ד'. א'. א'. (19)  
(17) א'. א'. א'. א'. א'. (20)

## מדדי קשר – מדד הקשר הلينארי (פירסון) – רקע

המטרה היא לבדוק האם קיים קשר (קורלציה, מתאים) של קו ישר בין שני משתנים כמותיים. מבחינת סולמות המדיידה קשר בין סולמות רוחניים ומנה. בדרך כלל,  $X$  הוא המשתנה המסביר (הבלתי תלוי) ו- $Y$  הוא המשתנה המוסבר (התלויה).

**דוגמה:**

נרצה להסביר כיצד השכלה של אדם הנ마다 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 להשביר את השינויים שלו בהכנסה, וכך רמת ההכנסה זהו המסביר התלויה במשתנה המסביר אותו.

**שלב ראשון:** נהוג לשרטט דיאגרמת פיזור. זו דיאגרמה שנוננת אינדיקטיבית ויזואלית על טיב הקשר בין שני המשתנים.

**דוגמה:**

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

בבנייה של 5 דירות בדקנו את הנתונים הבאים :

$X$  - מספר חדרים בדירה.  $Y$  - מס' נפשות הגרות בדירה.

להלן התוצאות שהתקבלו :

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

**שלב שני:** מחשבים את מקדם המתאים (מדד הקשר) שבזוק עד כמה קיים קשר לינארי בין שני המשתנים. המדד (נקרא גם מדד הקשר של פירסון) מכמת את מה שנראה בשלב הראשון רק בעין.

המדד בודק את כיוון הקשר (חיובי או שלילי) ואת עוצמת הקשר (חלש עד חזק).

מקדם מתאים זה מקבל ערכאים בין 1- -1.

מקדם מתאים 1- או 1 אומר שקיים קשר לינארי מוחלט ומלא בין המשתנים שניינו לבטא על ידי הנוסחה :  $y = bx + a$ .

### מתאים חיובי מלא (מקדם מתאים 1):

קיים קשר לנاري מלא בו השיפוע  $b$  יהיה חיובי ואילו מתאים שלילי מלא אומר שקיים קשר לנاري מלא בו השיפוע  $b$  שלילי (מקדם מתאים 1-).

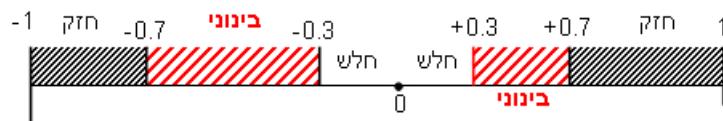
**מתאים חיובי חלקי:**

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

**מתאים שלילי חלקי:**

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

ככל שערך מקדם המתאים קרוב לאפס נאמר שעוצמת הקשר חלה יותר וככל שמקדם המתאים רחוק מהאפס נאמר שעוצמת הקשר חזקה יותר :



מקדם המתאים יסומן באות  $r$ .

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

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

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

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

$$\text{מקדם המתאים הלינאי: } r_{xy} = \frac{COV(x, y)}{S_x \cdot S_y}$$

## שאלות

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

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

א. שרטטו דיאגרמת פיזור לנ נתונים. מה ניתן להסיק מהדיאגרמה על טיב הקשר בין מספר החיסורים של תלמיד לציונו? מיהו המשתנה הבלטי תלוי ומיהו המשתנה התלויה?

ב. חשבו את מדד הקשר של פירסון. האם התוצאה מתוישבת עם תשובה בסעיף א'?

ג. הסבירו, ללא חישוב, כיצד מוקדם המתאים היה משתנה אם היה מתווסף תלמיד שהיחסיר 4 פעמים וקיבל ציון 80?

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

**2)** במחקר רפואי רצוי לבדוק האם קיים קשר בין רמת ההורמון X בدم החולים לרמת ההורמון Y שלו. לצורך כך מדדו את רמת ההורמוניים ההלו עבור חמישה חולים. להלן התוצאות שהתקבלו:

א. מה הממוצע של כל רמת ההורמו?

ב. מהו מוקדם המתאים בין ההורמוניים? ומה המשמעות ההתואאה?

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

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

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

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

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

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

א. חשב את מדד הקשר הילינארי בין X ל- Y. מיהו המשתנה התלויה?

ב. מה המשמעות של התוצאה שקיבלת בסעיף א'?

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

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

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

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

חשבו את ממד הקשרlienاري בין  $X$  ל-  $Y$ .

5) במוסד אקדמי ציון ההתאמה מחושב כך: מכפילים את הציון הממוצע בוגרות ב-3 ומחיתנים 2 נקודות. ידוע שעבור 40 מועמדים סטיטית התקן של ממוצע הציון בוגרות הייתה 2. מה מגדם המתאים בין ציון ההתאמה לציון הממוצע בוגרות שלהם?

- 6)
- הלו רשימה טענות, לגבי כל טענה קבעו נכון/לא נכון ונמקו.
  - א. מתוויך דירות המיר מחירי דירות מדולר לשקל. נניח שдолר אחד הוא 3.5 נק. אם מתוויך הדירות יחשב את ממד הקשר של פירסון בין מחיר הדירה בשקלים למחיר הדירה בדולרים הוא יקבל 1.
  - ב. לסדרה של נתונים התקבל  $S_x = S_y = 1$ ,  $\bar{X} = \bar{Y}$ . לכן, ממד הקשר של פירסון יהיה 1.
  - ג. אם השונות המשותפת של  $X$  ושל  $Y$  הינה 0 אז בהכרח גם מגדם המתאים של פירסון יהיה 0.

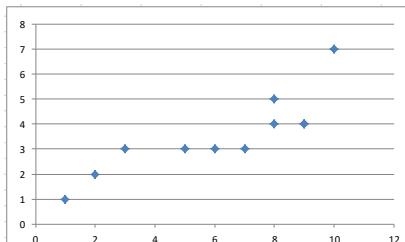
#### שאלות רב-ברירה:

- 7) נמצא שקיים מגדם מתאים שלילי בין הציון בעברית לחשבון בבחינה לכן:
- א. הדבר מעיד שהציונים בכתה היו שליליים.
  - ב. ככל שהציון של תלמיד יורך בחשבון יש לו נטייה לרדת בעברית.
  - ג. ככל שהציון של תלמיד עולה בחשבון יש לו נטייה לרדת בעברית.
  - ד. אף אחת מהתשובות לא נכונה.

**8)** נלקחו 20 מוצרים ונבדק ביום מסוים המחיר שלהם בדולרים והמחיר שלהם בש"ח (באותו היום ערך הדולר היה-2.4₪). מהו מקדם המתאים בין המחיר בדולר למחיר בש"ח?

- א. 1
- ב. 0
- ג. 4.2
- ד. לא ניתן לדעת.

**9)** להלן דיאגרמת פיזור:  
מה יהיה מקדם המתאים 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 וחליqi בין מספר החיסורים לציון התלמיד.  
 ב.  $r_{xy} = -0.9325$ .  
 ג. הקשר יישאר לינארי שלילי חליqi אך עוצמתו תחלש.
- 2)** א.  $r_{xy} = 0.96$       ב.  $\bar{x} = 15.4$  ,  $\bar{y} = 16$  .  
 ג. נכון.      ד. נכון.
- 3)** א. 0.8  
 ב. 0.8  
 ג. 1.
- 4)** א. נכון.  
 ב. נכון.  
 ג. נכון.
- 5)** א. נכון.  
 ב. נכון.  
 ג. נכון.
- 6)** א. נכון.  
 ב. נכון.  
 ג. נכון.
- 7)** א. נכון.  
 ב. נכון.  
 ג. נכון.
- 8)** א. נכון.  
 ב. נכון.
- 9)** א. נכון.  
 ב. נכון.

### בדיקות השערות על מקדם המתאיםlienاري – רקע

מדד הקשרlienاري באוכולוסייה, שנראה גם מקדם המתאים של פירסון או מדד הקשר של פירסון באוכולוסייה מסומן ב:  $r$  - פרמטר המאפיין את עצמת הקשרlienاري וכיוונו בין שני המשתנים הנחקרים באוכולוסייה. כאשר:

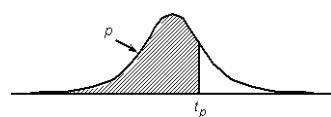
- מדד הקשרlienاري במדגם שמהווה אומד לפרמטר  $r$ .

**השערת האפס:** תהיה שבאוכולוסייה לא קיים כלל קשרlienاري בין שני המשתנים  $\rho = 0$ .  
ההנחה שעלייה אנו מtabסים בתחילת היא שני המשתנים הנחקרים מתפלגים דו נורמלית.

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

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

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

**טבלת ערכים קרייטיים של  $\zeta$  - 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
$\infty$	0.674	1.282	1.645	1.960	2.326	2.576	3.291

## שאלות

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

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

מقدم המתאים חושב והתקבל : 0.31 --.

א. האם קיימים מתאים בין וווטק העובד להשכלה? בדקו ברמת מובהקות של 5%?

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

**2)** מחקר התעניין לבדוק את הקשר בין גיל נשים בהריאן לרמת ההמוגולובי שלחן בדם בזמן הריאן. נדרגו 7 נשים והתקבלו התוצאות הבאות :

גיל	1	2	3	4	5	6	7	נבדק
המוגולובי	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 איש היא יותר מבוגרת אזי בהכרח יש לה יותר המוגולובי בדם?

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

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

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

ב. כיצד הייתה המשתנה התשובה לסעיף א' אם היו מוסיפים עוד תצפיות למדגם?

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

**4)** מtower דירות חישב את מועד המתאים בין שטח דירה במרכז תל אביב לבין המחיר של הדירה עבור 17 דירות. מועד המתאים שקיבל היה 0.6.

א. בדקו 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 השערת שקיים קשר ישר עולה בין שטח הדירה לבין מחיר הדירה בתל אביב.

**תשובות סופיות**

- ב. לא תשתנה.      1) א. לא נדחה את  $H_0$ .
- .  $H_0$       2) א. לא
- ב. לא ניתן לדעת.      3) א. נדחה את  $H_0$ .
- .  $0.005 < P_v < 0.01$       4) א. נדחה את  $H_0$ .

## מדד הקשר הלינארי – ניתוח פלטיטים – רקע

מדד הקשר הלינארי באוכלוסייה, שנקרה גם מקדם המתאים של פירסון או מדד הקשר של פירסון באוכלוסייה מסומן ב:  $\rho$  - פרמטר המאפיין את עוצמת הקשר הלינארי באוכלוסייה וכיונו בין שני המשתנים הנחקרים. כאשר :

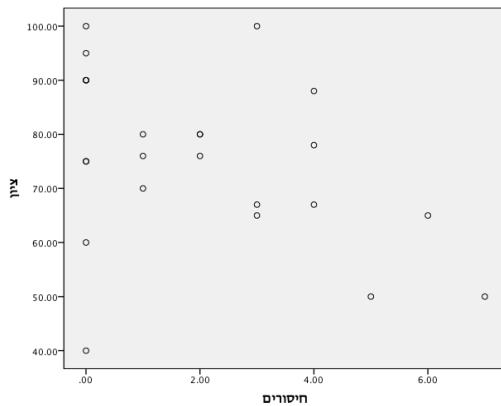
- ג' - מדד הקשר הלינארי במדגם שמהווה אומד לפרמטר  $\rho$ .

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

ההנחה שעלייה אנו מtabססים בתחילת הימ"ש היא שני המשתנים הנחקרים מתפלגים דו נורמלית.

**דוגמה (פתרון בהקלטה) :**

הזדיקן ביקש לדגום סטודנטים כדי לבדוק את הקשר בין ציון הסטודנט בקורס במספר הפעמים שהוא החסיר שיעור בקורס. דיאגרמת הפיזור שהתקבל במדגם שבוצע :



- א. מיהו המשתנה תלוי  
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			
ציון	Pearson Correlation	-.389	1
	Sig. (2-tailed)	.060	
	N	24	24

- ג. מהו מקדם המתאים שהתקבל במדגם? מה המשמעות שלו?
- ד. האם ניתן להגיד ברמת מובהקות של 5% שיש קיים מתאם לינארי שלילי בין מספר החיסורים של הסטודנטים מהקורס לבין הציון של הסטודנטים בקורס?

## שאלות

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

**Correlations**

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

- א. מהי אוכלוסיית המחקר?
- ב. מבחן השערות המחקר?
- ג. מהו המשתנה הבלטי תלוי ומהו המשתנה התלו依 במחקר?
- ד. מהי מסקנת המחקר ברמת מובהקות של 5%?

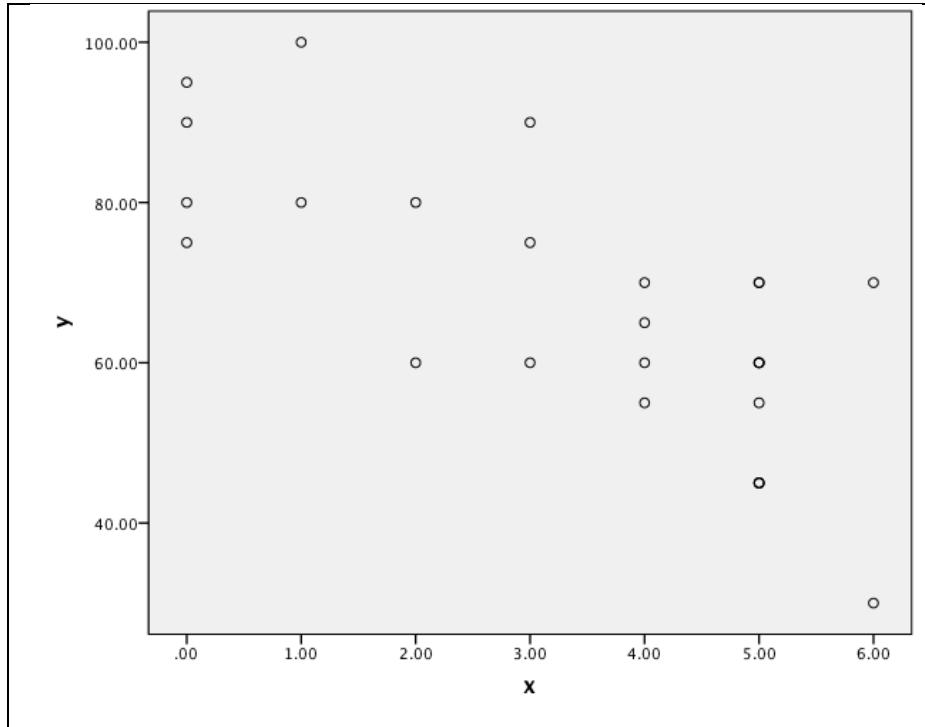
2) במדגם שנעשה נבדקו מספר משתנים על התכפיות שנדגמו. להלן פלט שהופק על המדגם?

### Correlations

		x	y	z	w
x	Pearson Correlation	???	-.682	.134	.176
y	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 בעלי קשר לינארי מובהק?  
 ג. השלימו את המספרים המסומנים בפלט בסימני שאלה.

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

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

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

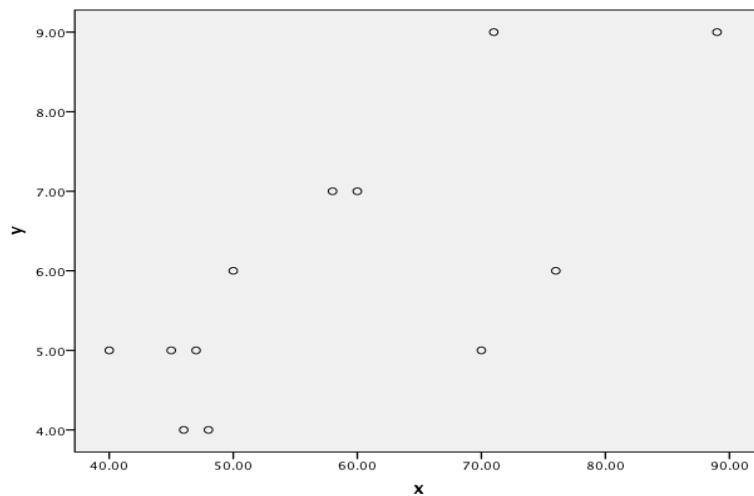
- 4) נדגמו ילדים בגיל 8 ונבדק עבור כל ילד גובהו בס"מ ומשקלו בק"ג. להלן הפלט שהתקבל עבור תוצאות המדגם:

### Correlations

	גובה	משקל	
Pearson Correlation	1	.552	גובה
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 לנארី חיובי בין המשקל והגובה.  
 ב. באילו רמות מובהקות ניתן לקבוע שקיים קשר לנארី חיובי בין במשקל והגובה?  
 ג. כיצד התשובה לסעיף הקודם הייתה משתנה אם היו מתווספות עוד 3  
 תכפיות למדגם?

- 5) בתקלิก כימי מסוים חוקר בדק את הקשר בין הטמפרטורה בתקליך (X) לבין אחוז החומר (Y) בתקליך. דיאגרמת הפיזור שהתקבל 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 שנמצא במדגם בין הטמפרטורה בתקליך לאחוז החומר?
- ב. האם הקשר בין הטמפרטורה בתקליך לבין אחוז החומר הוא קויי חיובי מובהק? בדקו ברמת מובהקות של 5%.
- ג. מה הייתה קורה למقدم המתאם במדגם ומובהקות התוצאה אם הייתה מתווספת תצפויות שבה הטמפרטורה היא 40 ואחוז החומר 9?

### תשובות סופיות

$$\begin{aligned} H_0 : p &= 0 \\ H_1 : p &\neq 0 \end{aligned}$$

**1)** א. נשים בהריון.

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

ד. קיימים 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 בברגות 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.

**5)** א. קיימים קשר לינארי חיובי וחלוקת שעוצמתו: 0.732.      ב. נדחה את  $H_0$ .

ג. מקדם המתאים קטן ומובהקות התוצאה גדולה.

## סטטיסטיקה

פרק 30 - מדדי קשר- השפעת טרנספורמציה לינארית על פירסון

תוכן העניינים

1. כללי .....

197 .....

## מדדי קשר – השפעת טרנספורמציה לינארית על פירסום:

---

**רקע:**

טרנספורמציה לינארית, בין אם נעשית על  $X$ , בין אם נעשית על  $Y$ , ובין אם נעשית על שניהם, אינה משנה את עוצמת הקשר. היא עלולה רק לשנות את כיוונו אם השיפועים של שתי הטרנספורמציות שונות סימן.

$$\cdot r_{[(aX+b),(cY+d)]} = \begin{cases} r_{x,y} & \text{if } a \cdot c > 0 \\ -r_{x,y} & \text{if } a \cdot c < 0 \end{cases}$$

**שאלות:**

- 1)** מבחן בניו שני חלקים: חלק כמותי וחלק מילולי. מקדם המתאים בין שני הציונים של שני החלקים הוא 0.9.  
 א. אם יعلו את כל הציונים בחלק המילולי ב-20%, מה יהיה מקדם המתאים בין הציון המילולי החדש לציון הכמותי ובין הציון המילולי היישן לציון המילולי החדש?  
 ב. נגידר משתנה חדש  $W$  להיות המרחק של הציון בחשיבות מילולית מהציון המקורי בבחינה-150. מצאו את מקדם המתאים בין הציון המילולי ל- $-W$  ובין  $W$  לציון הכמותי.
- 2)** מקדם המתאים בין הכניסה לבני ההוצאה של 10 משפחות חושב והתקבל 0.7. אם חל גידול של 5% בהכנסת האוכלוסייה כולה וגידול של 7% בהוצאה שלה, מה יהיה מקדם המתאים בין הכניסה החדש להוצאה החדש?
- 3)** חברת "לק" המייצרת גליידה החלילית לערוֹץ מחקר לבדיקת הקשר בין מספר חבילות הגלידה הנמכרות ביום לבין הטמפרטורה באותו יום. נבדקו 10 ימים והתקבל מוגן מתאים לינארי 0.85. חברת "לק" דואגת להתחילה כל יום עם מלאי של 150 חבילות גליידה. בנוסף, מעוניינים כי הטמפרטורה תboveoa במלוא פרנהייט במקומות במעלה צלסיוס. מה ערכו של מקדם המתאים בין מספר חבילות הגלידה שנשארות בסוף היום לבין הטמפרטורה במעלה פרנהייט?  
 הקשר בין מעלות צלסיוס ( $C$ ) למעלות פרנהייט ( $F$ ) נתון ע"י:  

$$F = \frac{9}{5}C + 32$$
 בחרו בתשובה הנכונה:  
 א. 0.85.  
 ב. 0.-85.  
 ג. 1.  
 ד. לא ניתן לדעת.
- 4)** מקדם המתאים בין  $X$  ל- $Y$  הנו 0.4. כל ערכי ה-  $X$  הוכפלו ב-2.  
 מה יהיה מקדם המתאים החדש בין שני המשתנים?  
 א. 0.8.  
 ב. 0.4.  
 ג. 0.-4.  
 ד. לא ניתן לדעת.

**תשובות סופיות:**

- 1) א. בין הציון המילולי הישן לחידש : 1. בין הציון המילולי החדש לכמותי : 0.9.  
ב. בין  $W$  לציון המילולי : -1, בין  $W$  לציון הכמותי : -0.9.
- (2) 0.7  
(3) ב'.  
(4) ב'.

## סטטיסטיקה

פרק 31 - רגרסיה

תוכן העניינים

200 .....

1. כללי .....

## מדדי קשר – רגרסיה ליניארית:

**רקע:**

במידה וקיים קשר חזק בין שני המשתנים הבלתיים נוהג לבצע ניבוי. לבנות קו ניבויים הנקרא גם קו רגרסיה המנeba משתנה אחד על סמך الآخر. מדובר בקו שמנבא את  $Y$  על סמך  $X$ .

השיטה למציאת הקו הניל נקראת שיטת הריבועים הפחותים והקו המתתקבל נקרא קו הרגרסיה או קו הניבויים או קו הריבועים הפחותים.

- a - נותן את ערך  $Y$  כאשר  $X$  הינו אפס על גבי קו הניבויים. הוא נקרא החותך של הקו.
- b - הוא שיפוע הקו נותן בכמה בפעם עצם  $Y$  משתנה כאשר  $X$  גדל ביחידת אחת על גבי קו הניבויים.

להלן המשוואות למציאת הפרמטרים של קו הרגרסיה: 
$$Y = bX + a \quad , \quad b = r \frac{S_r}{S_x}$$

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

**שאלות:**

**1)** נסמן ב-  $X$  את הכנסה של משפחה באלפי נק. נסמן ב-  $Y$  את ההוצאות של משפחה באלפי נק. נלקחו 20 משפחות והתקבלו התוצאות הבאות:

$$\begin{aligned}\sum_{i=1}^{20} Y_i &= 200, \quad \sum_{i=1}^{20} X_i = 240 \\ \cdot \sum_{i=1}^{20} (X_i - \bar{X})^2 &= 76, \quad \sum_{i=1}^{20} (Y_i - \bar{Y})^2 = 76 \\ \sum_{i=1}^{20} (X_i - \bar{X})(Y_i - \bar{Y}) &= 60.8\end{aligned}$$

א. חשבו את מודד הקשר הlienاري בין  $X$  ל-  $Y$ . מיהו המשתנה תלוי?

ב. מצאו את קו הרוגסיה לניבוי ההוצאה של משפחה על סמך הכנסה שלה. הסבירו את משמעות הפרמטרים של קו הרוגסיה.

ג. משפחת כהן הכנסה 15,000 נק. מה ההוצאה הצפואה שלה?

**2)** נסמן ב-  $X$  את ההשכלה של אדם בשנות לימוד. נסמן ב-  $Y$  את הכנסתו באלפי נק. במחקר התקבלו התוצאות הבאות:

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

א. חשבו את מודד הקשר של פירסונו בין ההשכלה להכנסה.

ב. מה הכנסה הצפואה לאדם שהשכלתו 12 שנים?

ג. מה ההשכלה הצפואה לאדם שהכנסתו 10,000 נק?

**3)** חוקר רצה לחקור את הקשר הקוווי 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 המתאים בין הציון לשעות ההכנה היה 0.8.

א. על פי משווהת הרוגסיה, שעת הכנה נספתח משפרת את ציון המבחן ב-?

ב. על פי משווהת הרוגסיה, תלמיד שייגש למבחן ללא שעות הכנה כל יקבל ציון?

ג. מהו קו הרוגסיה לניבוי הציון לפי שעות ההכנה?

**4)** נתונים 2 משתנים  $X$  ו-  $Y$ . כמו כן נתון:  $S_x = S_y = 4$ ,  $\bar{X} = 1.5$ .

וכן שקו הרוגסיה של  $Y$  על בסיס  $X$  הינו:  $Y = -0.2X + 0.5$ .  
חשבו מהו מקדם המתאים בין  $X$  ל-  $Y$ .

### תשובות סופיות:

- |         |                    |         |                    |
|---------|--------------------|---------|--------------------|
| ג. 12.4 | . $Y = 0.8X + 0.4$ | ב. 0.8  | א. 0.8 <b>(1)</b>  |
| ג. 14.6 | . $Y = 1.2X + 29$  | ב. 4.25 | א. 0.75 <b>(2)</b> |
| .       | .                  | ב. 29   | א. 1.2 <b>(3)</b>  |
| .       | .                  | .       | -0.2 <b>(4)</b>    |

## סטטיסטיקה

פרק 32 - מדדי קשר-רגסיבית - שונות מוסברת ושונות לא מוסברת

תוכן העניינים

203.....  
1. כללי.....

## מדדי קשר – רgresיה – שונות מוסברת ושונות לא מוסברת:

**רקע:**

המטרה ברגרסיה היא להסביר את השונות של המשתנה תלוי. למשל, להסביר את השונות של המשכורת באמצעות הוותק או להסביר את השוני בציונים באמצעות כמה חחיסורים.  
<sup>2</sup> r<sup>2</sup> - החלק מהשונות של המשתנה תלוי מוסבר. השונות המוסברת נקראת גם שונות ניבויים. השונות הלא מוסברת נקראת גם שונות טעויות.

**שאלות:**

- 1)** נמצא קשר חיובי בעוצמה של 0.7 בין שטח דירה למחירה. כמו כן, נתון שטיתת התקן של מחירי הדירות הינה 200.
- איזה אחוז מהשונות של מחירי הדירות מוסבר על ידי שטח הדירה?
  - איזה אחוז מהשונות של מחירי הדירות לא מוסבר על ידי שטח הדירה?
  - מהי השונות המוסברות ומהי השונות הלא מוסברת של מחירי הדירות?
- 2)** להלן רשימת טענות, לגבי כל טענה קבעו נכון/לא נכון וכוכן וنمוקו!
- אם שונות הטעויות שווה ל-0 (השונות הלא מוסברת) אז מקדם המתאים של פירסון יהיה 1.
  - אם מקדם המתאים של פירסון בין שני משתנים הוא 1 אז שונות הטעויות (השונות הלא מוסברת) תהיה 0.
  - אם השונות המשותפת של  $X$  ושל  $Y$  היא 0 אז בהכרח גם מקדם המתאים של פירסון יהיה 0.

**שאלות רב-ברירה:**

- 3)** הקשר בין שני משתנים התקבל:  $r^2 = 0.64$ , לכן:
- לא יוצא מן הכלל ככל שערכי משתנה אחד עולה המשתנה השני עולה.
  - 64% מהשונות של משתנה אחד מוסבר על ידי המשתנה השני.
  - הקשר בין שני המשתנים הוא בעוצמה של 0.64.
  - כל התשובות נכונות.
- 4)** אם מגדילים את  $r^2$ , ניתן לומר כי:
- אחוז השונות המוסברת יקטן.
  - אחוז השונות המוסברת יגדל.
  - אחוז השונות המוסברת ישאר ללא שינוי.
  - סטודנטית התקן משתנה.
  - לא ניתן לדעת.

- 5) בקורס מבוא לכלכלה ניתנו במשך השנה שני מבחנים: מבחן בסוף סמסטר א' X ו מבחן בסוף סמסטר ב' Y. כאשר בנו את קו הרגרסיה של הציון במבחן סוף סמסטר ב' לפי הציון במבחן סוף סמסטר A התקבלה שונות טעויות של 80, ושונות ניבויים של 20.
- לפי נתוניים אלו, מקדם המתאים בין הציון במבחן סוף סמסטר A' לבין הציון במבחן סוף סמסטר B' הוא:
- .0.44.
  - .-0.44.
  - 0.44.
  - אין אפשרות לחשב את מקדם המתאים.
  - .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'.

## סטטיסטיקה

פרק 33 - מדי קשר - בחירת מדי מתאים

תוכן העניינים

206 ..... 1. בחירת מדי מתאים

## **מדדי קשר – בחירת מדד מתאים:**

**רקע:**

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

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

בחירת מדד הקשר נעשו לפי סולמות המדידה של שני המשתנים שאחנו רוצים לבדוק את הקשר ביניהם. 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 רוחניים ומנה יחד.

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

להלן טבלה שמסכמת את בחירת המדד המתאים:

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

**דוגמה (פתרון בהקלטה):**

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

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

**שאלות:**

- 1)** בקרב תלמידי כיתות א' בבית הספר גבריאלי אשר בתל אביב בדקו לכל תלמיד את גובהו בס"מ ואת משקלו בק"ג. מהו מודד הקשר המתאים כדי לבדוק האם קיים קשר בין גובה התלמיד למשקלו?
- פירסון.
  - ספירמן.
  - קרמר.
  - אין מספיק נתונים כדי לדעת.
- 2)** בסקר שנעשה על אוזחים במדינה בדקו לכל אוזח את השכלתו ואת שכרו. מהו מודד הקשר המתאים כדי לבדוק האם קיים קשר בין השכלה לשכר?
- פירסון.
  - ספירמן.
  - קרמר.
  - אין מספיק נתונים כדי לדעת.
- 3)** בגין הילדים של שולח אספו נתונים על 25 הילדים שבוגן. על כל ילד בדקו את רמת הביטחון העצמי שלו ( $X$ ) = מודד 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 שזכה בילדים ובחון אותן. מהו מודד המתאים לבדיקת התלות בין  $X$  לבין  $Y$ ?
- פירסון.
  - ספירמן.
  - קרמר.
  - 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 שלמדו סטטיסטיקה אצל איזה מרצה הוא למד (היו 3 מרצים שונים) והאם הוא עבר את הבחינה. מהו מודד הקשר המתאים במקרה זה?
- פירסון.
  - ספירמן.
  - קרמר.
  - אין מספיק נתונים כדי לדעת.

5) בבורסה בתל אביב רצוי לבדוק את הקשר בין גובה הריבית במשק בסוף החודש (ב אחוזים), לבין תשואת מניאת אקטר (ב אחוזים) 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 הלוקח מאדיבות השירות של המלצר בסקלה של 1 עד 5.
- מה גילו של הלוקח בשנים.
- מה גובה התשר (טיפ) ב-₪ אשר נתן הלוקח למלאר ברכתו מהמסעדה.

מהו המדד המתאים כדי לבדוק האם קיימים מתחים חיוביים בין מידת שביעות הרצון של הלוקח מאדיבות השירות לבין גובה התשר שהוא נתן למלאר?

- 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 הלוקח מאדיבות השירות של המלצר בסקלה של 1 עד 5.
- מה גילו של הלוקח בשנים.
- מה גובה התשר (טיפ) ב-₪ אשר נתן הלוקח למלאר ברכתו מהמסעדה.

מהו המדד המתאים כדי לבדוק האם קיימים מתחים בין גיל הלוקח לגובה התשר שהעניק לשירות?

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

### הנתונים הבאים מותאים ל-3 השאלות הבאות:

חוקרים ערכו מדגם של ילדים מכיתות ב' ו-ג' מ-4 בתים ספר שונים. הועבר לילדים שאלון בו תואר מצב מסוים והילדים התבקו ציינו את רמת החרצה 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 חרצה (ציון שהילד היה צריך לתת בסקלה של 1 עד 10).
- גיל התלמיד בחודשים.

(8) מהו מודד הקשר המתאים כדי לבדוק את הקשר בין גיל התלמיד לבין רמת החרצה שלו מהמצבי?

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

(9) מהו מודד הקשר המתאים כדי לבדוק את הקשר בין המגדר לבין רמת החרצה שלו מהמצבי?

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

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

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

**11)** בטבלה שלහן נתונות שכיחויות ההצלחה והכשלון של 150 חוליים :

C	B	A	תוצאה/ הת蘗פה
45	13	35	נרא
5	37	15	לא נראה

החולים קיבלו 3 תרופות שונות ובדקו עבור כל חוליה אם הת蘗פה הצליחה בריפוי. מהו מדד הקשור המתאים?

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

**12)** שני מוסיקאים מפורסמים נתנו ציון בסולם של 1-10 ל��ולם של 8 מתמודדים בתוכנית ראליטי ידועה. ציון 10 ניתן לקול שמצא חן ביוטר בעני המוסיקאי. מפיק התוכנית רצה לבדוק האם יש קוורלציה בין המוסיקאים מבחינת הטעם. בטבלה הבאה נתונים הציונים של כל אחד מהמוסיקאים את שמות המתמודדים:

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

מהו מדד הקשור המתאים?

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

**13)** להלן טבלה המסכםת את השכר באלפי ש' של עובדים בחברה ואת רמת המוטיבציה שלהם מ-1 עד 5 :

30	15	20	18	12	שכר
5	3	5	4	4	מוטיבציה

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

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

14) להלן טבלה על נתונים שנאספו על מספר תצפיות:

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

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

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

#### תשובות סופיות:

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