

## סטודנטים יקרים

לפניכם ספר תרגילים בקורס סטטיסטיקה לניהול מלונאות ותיירות. הספר הוא חלק מקורס חדשני וראשון מסוגו בארץ בנושא זה, המועבר ברשת האינטרנט On-line.

הקורס באתר כולל פתרונות מלאים לספר התרגילים, וכן את התיאוריה הרלוונטית לכל נושא ונושא.

**הקורס כולו מוגש בסרטוני וידאו המלווים בהסבר קולי, כך שאתם רואים את התהליכים בצורה מובנית, שיטתית ופשוטה, ממש כפי שנעשה בשיעור פרטי, לדוגמה לחצו כאן.**

את הקורס בנה מר ברק קנדל, מרצה מבוקש במוסדות אקדמיים שונים ובעל ניסיון עתיר בהוראת המקצוע.

אז אם אתם עסוקים מידי בעבודה, סובלים מלקויות למידה, רוצים להצטיין או פשוט אוהבים ללמוד בשקט בבית, אנחנו מזמינים אתכם לחוויית לימודים יוצאת דופן וחדשה לחלוטין, היכנסו עכשיו לאתר [www.gool.co.il](http://www.gool.co.il).



אנו מאחלים לכם הצלחה מלאה בבחינות

צוות האתר GooL

**גול זה בול. בשבילך!**

לפתרון מלא בסרטון וידאו היכנסו ל- [www.GooL.co.il](http://www.GooL.co.il)

כתב ופתר - ברק קנדל ©

## תוכן

- פרק 1 - סטטיסטיקה תיאורית - סיווג משתנים וסולמות מדידה ..... 4
- פרק 2 - סטטיסטיקה תיאורית - הצגה של נתונים ..... 7
- פרק 3 - סטטיסטיקה תיאורית - סכימה ..... 15
- פרק 4 - סטטיסטיקה תיאורית- מדדי מיקום מרכזי ..... 18
- פרק 5 - סטטיסטיקה תיאורית - מדדי פיזור: הטווח, השונות וסטיית התקן ..... 23
- פרק 6 - סטטיסטיקה תיאורית - מדדי פיזור - ממוצע סטיות מוחלטות מהתציון ..... 28
- פרק 7 - סטטיסטיקה תיאורית - מדדי מיקום יחסי - ציון תקן ..... 31
- פרק 8 - סטטיסטיקה תיאורית - מדדי מיקום יחסי - אחוזונים בטבלת שכיתויות בדידה ..... 34
- פרק 9 - סטטיסטיקה תיאורית - שאלות מסכמות ..... 37
- פרק 10 - מדדי קשר - מדד הקשר של קרמר ..... 46
- פרק 11 - מדדי קשר - מדד הקשר של ספירמן ..... 49
- פרק 12 - מדדי קשר - מדד הקשר הלינארי (פירסון) ..... 53
- פרק 13 - מדדי קשר - השפעת טרנספורמציה לינאריות על מדד הקשר של פירסון ..... 61
- פרק 14 - מדדי קשר - רגרסיה ליניארית ..... 64
- פרק 15 - מדדי קשר - רגרסיה - שונות מוסברת ושונות לא מוסברת ..... 67
- פרק 16 - בעיות בסיסיות בהסתברות ..... 70
- פרק 17 - פעולות בין מאורעות (חיתוך ואיחוד), מאורעות זרים ומכילים ..... 75
- פרק 18 - הסתברות מותנית - במרחב מדגם אחיד ..... 86
- פרק 19 - הסתברות מותנית - מרחב לא אחיד ..... 89
- פרק 20 - דיאגרמת עצים, נוסחת בייס ונוסחת ההסתברות השלמה ..... 93
- פרק 21 - תלות ואי תלות בין מאורעות ..... 99
- פרק 22 - שאלות מסכמות בהסתברות ..... 103
- פרק 23 - התפלגויות בדידות מיוחדות - התפלגות בינומית ..... 107
- פרק 24 - התפלגות נורמלית ..... 112
- פרק 25 - התפלגות הדגימה ..... 116
- פרק 26 - מושגים בסיסיים באמידה ..... 125
- פרק 27 - רווח סמך לתוחלת (ממוצע האוכלוסייה) ..... 132
- פרק 28 - רווח סמך לפרופורציה ..... 144
- פרק 29 - רווח סמך להפרש תוחלות ממדגמים בלתי תלויים ..... 147
- פרק 30 - רווח סמך לתוחלת ההפרש במדגם מזווג ..... 149
- פרק 31 - תרגול מסכם ברווחי סמך ..... 152
- פרק 32 - בדיקת השערות כללית ..... 155
- פרק 33 - בדיקת השערות על פרמטרים ..... 163

168	פרק 34 - בדיקת השערות על תוחלת (ממוצע).....
205	פרק 35 - בדיקת השערות על פרופורציה.....
221	פרק 36 - בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים.....
232	פרק 37 - בדיקת השערות על תוחלת ההפרשים במדגמים מזווגים (תלויים).....
244	פרק 38 - הקשר בין רווח סמך לבדיקת השערות על הפרש תוחלות.....
248	פרק 39 - ניתוח שונות חד כיוונית.....

## פרק 1 - סטטיסטיקה תיאורית - סיווג משתנים וסולמות מדידה

### רקע:

סטטיסטיקה תיאורית הוא ענף בו לומדים כיצד לאסוף נתונים, להציג אותם ולנתח אותם. בסטטיסטיקה תיאורית אנו פונים לקבוצה מסוימת. באותה קבוצה אנו אוספים נתונים על הישויות באותה קבוצה.

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

חלוקה אחת של המשתנים הנמדדים היא לפי סולמות מדידה :

### מיון משתנים לפי סולמות המדידה:

1. **סולם שמי** (נומינאלי) – משתנה שלערכיו יש משמעות רק מבחינת הזהות ואין עניין של יותר או פחות לדוגמה: מצב משפחתי רווק/נשוי/אלמן/גרש; אזור מגורים. משתנה דיכוטומי (הינו מסולם שמי) אותם משתנים שיש להם רק שני ערכים אפשריות זכר/נקבה. מעשן/לא מעשן.
2. **סולם סדר** (אורדינאלי) – כאשר לערכים של המשתנה בנוסף לשם ישנה גם משמעות לסדר אבל אין משמעות לגודל ההפרש. למשל, דרגה בצבא.
3. **סולם רווחים** (אינטרוואלי) – משתנה שלערכים שלו בנוסף לשם ולסדר בניהם יש משמעות לרווחים בין הערכים אבל אין משמעות ליחס בין הערכים. למשל, קומה בבניין. סולם לא כל כך פופולרי.
4. **סולם מנה/יחס** – משתנה שלערכיו בנוסף לשם, לסדר ולרווח יש משמעות גם ליחס בין הערכים. למשל, מספר מכוניות למשפחה, משקל אדם בק"ג. הדרך הקלה ביותר כדי לזהות עם הסולם הוא סולם מנה היא על ידי מבחן האפס. בסולם מנה האפס הוא מוחלט, אבסולוטי, ומייצג אין.

### סוגי משתנים:

נבצע סיווג של המשתנים :

משתנה איכותי הוא משתנה שלערכיו אין משמעות של יותר או פחות , אין עניין כמותי לערכים המתקבלים.

כמו : מקום מגורים של אדם (רעננה, תל אביב, אשדוד..)

מין האדם (זכר, נקבה)

מצב משפחתי ( רווק, נשוי, גרוש, אלמן)

משתנה כמותי הוא משתנה שערכיו הם מספרים להם יש משמעות כמותית כמו : גובה אדם בס"מ, ציון בבחינה וכדומה.

את המשתנה הכמותי נסווג לשני סוגים :

משתנה בדיד : משתנה שערכיו מתקבלים מתוך סידרה של ערכים אפשריים. כמו : מספר ילדים למשפחה (1,2,3..)

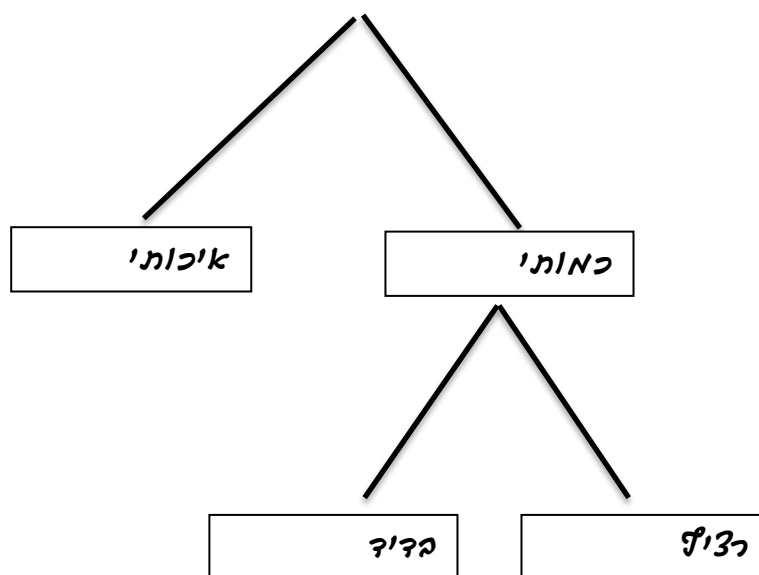
ציון בבחינה ( מ 0 ועד 100 בקפיצות של 1 )

משתנה רציף : משתנה שערכיו מתקבלים מתוך אינסוף ערכים בתחום מסוים , הערכים מתקבלים ברצף וללא קפיצות של ערכים .

כמו : גובה בס"מ – אם למשל, הגובה הנמוך ביותר הוא 150 ועד 190 ס"מ בקבוצה הגבהים הם

ברצף. גם בין 160 ל 161 ס"מ יש רצף אינסופי של ערכים אפשריים לגובה (16.233 ס"מ הוא גם גובה אפשרי)

משקל בק"ג , מהירות בקמ"ש וכולי.



## תרגילים:

1. לפניכם רשימה של משתנים:
- א. גובה אדם בס"מ.
  - ב. מספר ילדים למשפחה.
  - ג. מידת חרדה לפני מבחן.
  - ד. שביעות רצון משירות לקוחות בסקלה מ 1 עד 7 ( 1 כלל לא מרוצה עד 7 מרוצה מאד)
  - ה. השכלה.
  - ו. מספר אוטובוס.
  - ז. מקום מגורים.
  - ח. מין (1=גבר ו-2=אישה).
  - ט. מידת נעליים.

ציינו באיזה סולם מדידה המשתנה הנחקר ( שמי , סדר , רווחים או מנה)

2. להלן התפלגות מספר האיחורים לעבודה בחודש של העובדים בחברת "סטאר".  
בחברה 200 עובדים.

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

- א. מהו המשתנה הנחקר כאן?
- ב. האם מדובר במשתנה איכותי או כמותי? אם הוא כמותי האם הוא בדיד או רציף?  
באיזה סולם מדידה המשתנה?
3. לפניכם רשימה של משתנים כמותיים. ציין ליד כל משתנה אם הוא רציף או בדיד.
- א. שכר עובד בש"ח.
  - ב. ציון בחינת בגרות.
  - ג. תוצאה בהטלת קובייה.
  - ד. מהירות ריצה בתחרות.
  - ה. שיעור התמיכה בממשלה.

## פרק 2 - סטטיסטיקה תיאורית - הצגה של נתונים

### רקע:

דרכים להצגת נתונים שנאספו:

#### א. רשימה של תצפיות:

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

3 4 3 5 4

#### ב. טבלת שכיחויות בדידה:

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

רושמים את התצפיות בטבלה שבה עמודה אחת מבטאת את ערכי המשתנה והשנייה את השכיחות. יעיל עבור משתנה איכותי וכמותי בדיד וכשיש מספר רב של תצפיות. לא יעיל למשתנה כמותי רציף.

למשל, 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:  $F_i$  - השכיחות המצטברת נותנת כמה תצפיות קטנות או שוות לערך.

שכיחות יחסית (פרופורציה) – השכיחות מחולקת לכמות התצפיות הכללי:  $\frac{f_i}{n}$  - איזה חלק מהתצפיות בקבוצה שוות לערך.

### ג. טבלת שכיחויות במחלקות:

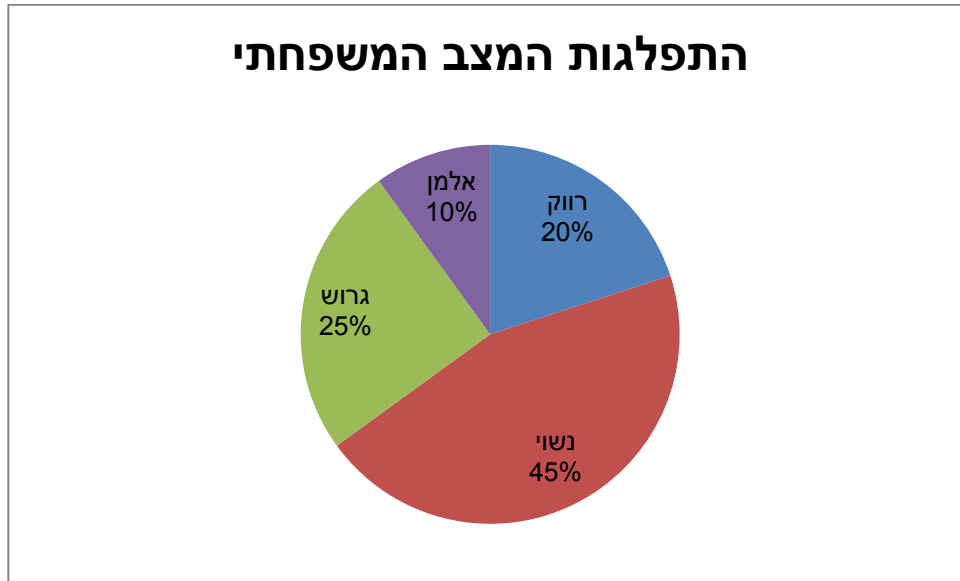
משתמשים שהמשתנה כמותי רציף או כאשר יש מספר ערכים רב במשתנה הבדיד וטבלת שכיחויות תהיה ארוכה מידי.  
למשל, נתנו לקבוצת ילדים לבצע משימה מסוימת ובדקו את התפלגות זמן ביצוע המשימה בדקות.  
להלן ההתפלגות שהתקבלה:

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



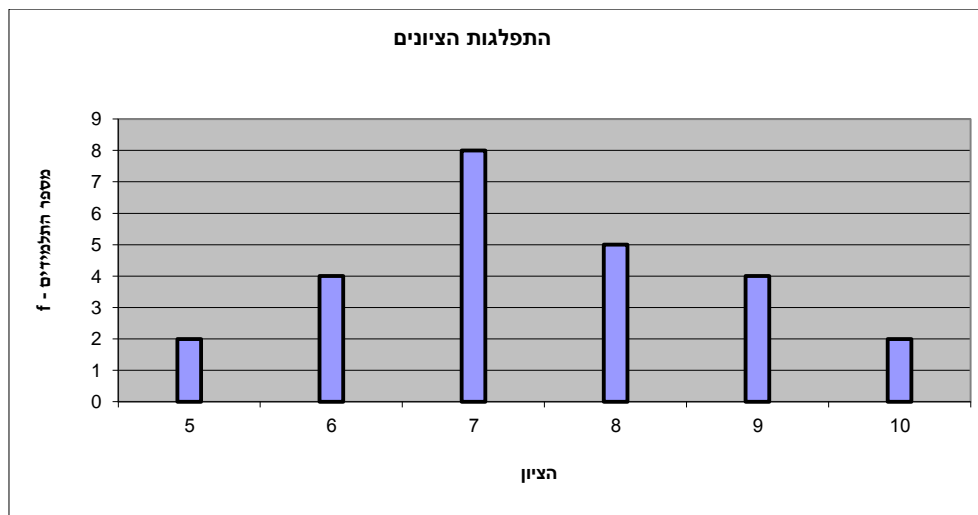
ד. **דיאגרמת עוגה :**

זהו התיאור הגרפי של משתנה איכותי. בדיאגרמת עוגה כל ערך במשתנה מקבל "נתח" יחסי מהעוגה. הנתח בעוגה פרופורציוני לשכיחות היחסית של ערך המשתנה בנתונים.



ה. **דיאגרמת מקלות :**

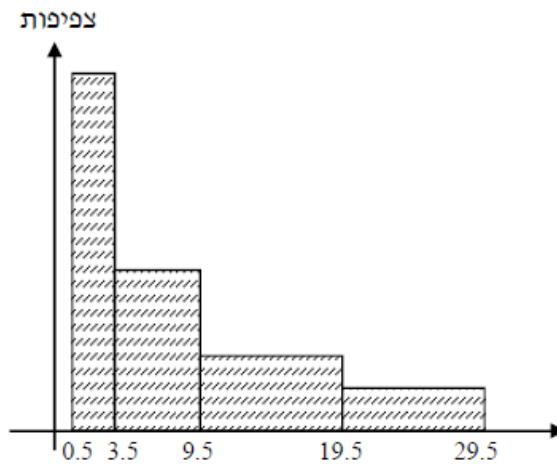
הציר האופקי הוא הציר של המשתנה הציר האנכי של השכיחות – הגובה של המקל מעיד על השכיחות .  
 רלבנטי למשתנה כמותי בדיד. לא נהוג להשתמש בתיאור למשתנה איכותי וכמו כן לא למשתנה כמותי רציף .  
 כמו כן בסולמות מדידה עבור משתנה מסולם סדר.



1. **היסטוגרמה:**

ההיסטוגרמה היא הדרך הגרפית כדי לתאר טבלת שכיחויות במחלקות. רלבנטית למשתנה כמותי רציף. בהיסטוגרמה ציר האופקי הוא הציר של המשתנה וציר האנכי הוא הציר של הצפיפות. הצפיפות מחושבת בכל מחלקה על ידי חלוקת השכיחות ברוחב של כל המחלקה והיא נותנת את מספר התצפיות הממוצע בכל מחלקה ליחידה. אם המחלקות הן שוות ברוחב, ניתן לשרטט את ההיסטוגרמה לפי השכיחות ואין צורך בצפיפות.

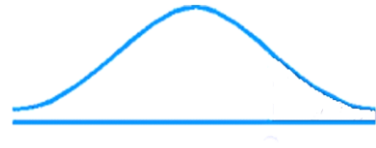
צפיפות	מצטברת	שכיחות	אמצע	רוחב	X
6.6667	20	20	2	3	0.5 - 3.5
3	38	18	6.5	6	3.5 - 9.5
1.4	52	14	14.5	10	9.5 - 19.5
0.8	60	8	24.5	10	19.5 - 29.5



**פוליגון-מצולעון:** אם נחבר את אמצע קצה כל מלבן בקווים ישרים. נותן מראה חזותי לצורה של התפלגות המשתנה.

### צורות התפלגות נפוצות

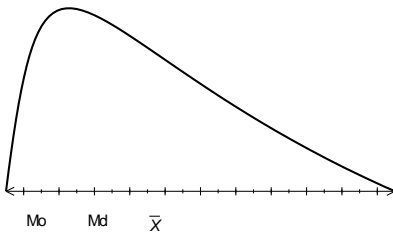
התפלגות סימטרית פעמונית- רוב התצפיות במרכז וככל שנתרחק מהמרכז יהיו פחות תצפיות באופן סימטרי. למשל, ציוני IQ.



ישנן התפלגויות סימטריות שאינן פעמוניות:

התפלגות אסימטרית ימנית (חיובית) – רוב התצפיות מקבלות ערכים נמוכים ויש מיעוט הולך וקטן של תצפיות שמקבלות ערכים גבוהים קיצוניים. למשל, שכר במשק.

#### התפלגות א-סימטרית ימנית או חיובית



התפלגות אסימטרית שמאלית (שלילית) רוב התצפיות מקבלות ערכים גבוהים ויש מיעוט הולך וקטן של תצפיות שמקבלות ערכים נמוכים קיצוניים. למשל, אורך חיים.

#### התפלגות א-סימטרית שמאלית או שלילית



**תרגילים:**

1. בסקר צפייה בטלוויזיה התקבלו התוצאות הבאות: 25 צפו בערוץ הראשון, 25 צפו בערוץ 10, 75 צפו בערוץ השני, 50 צפו באחד מערוצי הכבלים ו- 25 לא צפו בטלוויזיה בזמן הסקר.
- א. רשמו את טבלת השכיחות ואת השכיחות היחסית.
- ב. תארו את הנתונים באופן גרפי.

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

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

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

3. להלן התפלגות ההשכלה במקום עבודה מסוים:

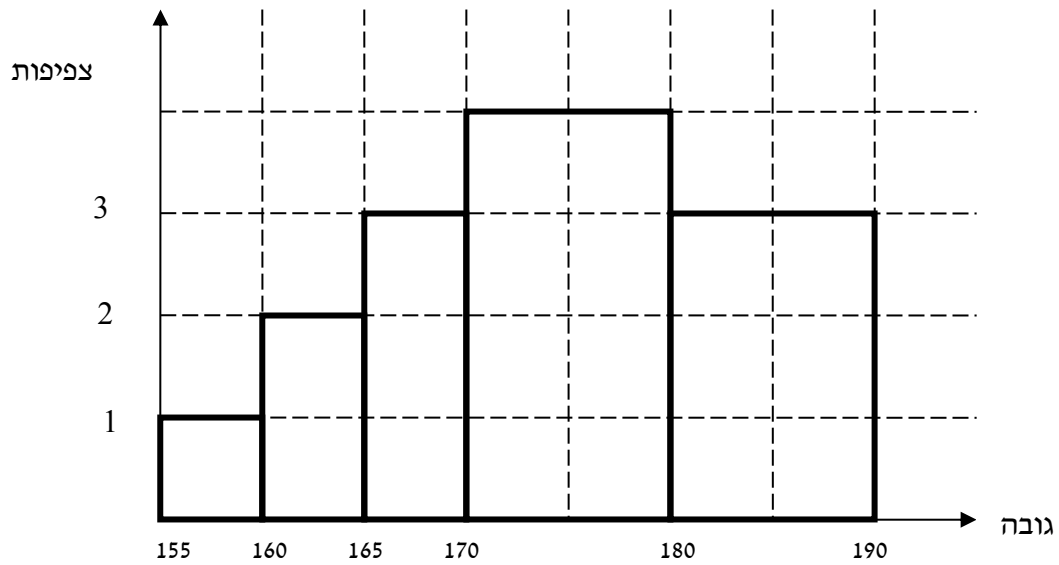
מספר העובדים	השכלה
60	נמוכה
120	תיכונית
20	אקדמאית

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

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

- א. מהו המשתנה? האם הוא בדיד או רציף?
- ב. תאר את הרשימה בטבלת שכיחויות.
- ג. הוסף שכיחויות יחסיות לטבלה.
- ד. תאר את הנתונים באופן גרפי.

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



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

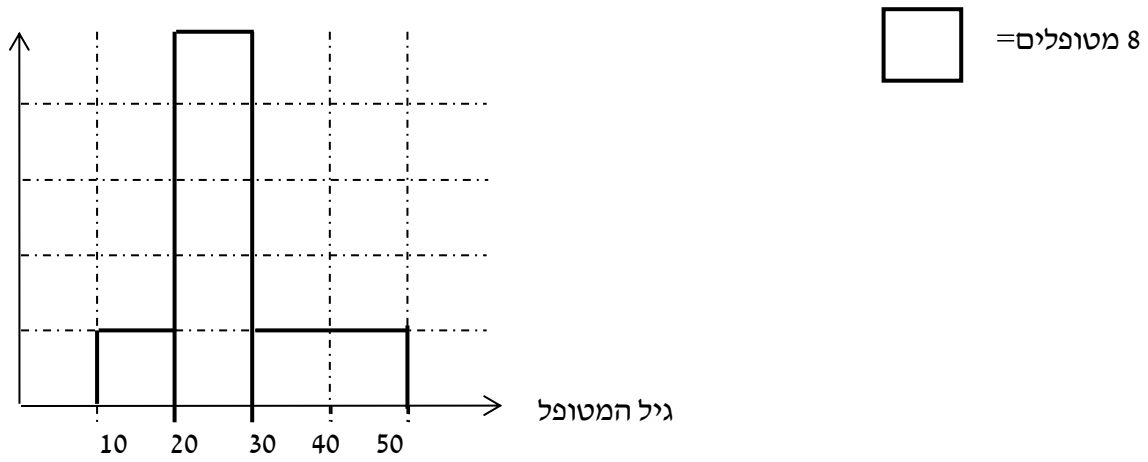
6. להלן התפלגות המשקל של קבוצה מסוימת בק"ג:

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

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

7. להלן גיל המטופלים של ד"ר שוורץ בשנים :

קנה מידה :



א. מה המשתנה הנחקר? האם הוא בדיד או רציף?

ב. מהי הקבוצה הנחקרת?

ג. תרגמו את ההסיטוגרמה לטבלת שכיחות.

ד. מהי הפרופורציה של המטופלים של ד"ר שוורץ בגילאים 20-30?

### פרק 3 - סטטיסטיקה תיאורית - סכימה

#### רקע:

בסטטיסטיקה ישנה צורת רישום מקובלת כדי לרשום סכום 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

תרגילים:

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

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

חשבו:

$$\sum_{i=1}^3 X_i$$

$$\sum_{i=1}^5 Y_i$$

$$\sum_{i=1}^4 X_i$$

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

$$\sum X_i$$

$$\sum X_i Y_i$$

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



2. נתון לוח ערכי המשתנים  $x_i$  ו-  $y_i$  כאשר:  $i=1,2,\dots,6$

$i$	1	2	3	4	5	6
$x_i$	3	2	4	-2	1	4
$y_i$	2	0	0	1	-5	2

ונתונים הקבועים:  $a=2$   $b=5$  חשבו את הנוסחאות הבאות:

$$\sum_{i=1}^4 y_i \quad \text{א.}$$

$$\sum_{i=1}^6 a \quad \text{ב.}$$

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

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

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

3. קבע לכל זהות אם היא נכונה:

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

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

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

## פרק 4 - סטטיסטיקה תיאורית - מדדי מיקום מרכזי

### רקע:

המטרה במדדי המיקום המרכזי למדוד את מרכז ההתפלגות של התצפיות.

### השכיח – MODE

השכיח הוא הערך הנפוץ ביותר בהתפלגות.

ברשימה : הערך החוזר על עצמו הכי הרבה פעמים .

בטבלת שכיחויות בדידה : הערך שהשכיחות שלו היא הגבוהה ביותר.

בדיאגרמת מקלות : שיעור ה-  $X$  של המקל הגבוה ביותר.

בעוגה : הערך של הפלח הגדול ביותר.

בטבלת שכיחויות במחלקות : המחלקה עם הצפיפות הגבוהה ביותר.

בהיסטוגרמה המחלקה הגבוהה ביותר.

יתכן שלהתפלגות יותר משכיח אחד.

השכיח הוא מדד הרלבנטי לכל סוגי המשתנים.

### MEDIAN – החציון

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

ברשימה : נסדר את התצפיות בסדר עולה.

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

אם יש מספר זוגי של איברים החציון יהיה הממוצע של האיבר ה-  $\frac{n}{2}$  והאיבר ה-  $\frac{n}{2} + 1$

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

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

בטבלת שכיחויות בדידה : נעשה תהליך דומה אך נעזר בשכיחות המצטברת.

בטבלת שכיחויות במחלקות : המחלקה החציונית היא המחלקה שמיקומה  $\frac{n}{2}$

החציון אינו רלבנטי למשתנה מסולם שמי ולא רלבנטי למשתנה איכותי.

**הממוצע:**

הנו מרכז הכובד של ההתפלגות.

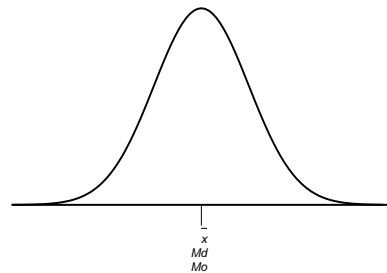
$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} : \text{ברשימה}$$

$$\bar{x} = \frac{\sum x \cdot f}{n} : \text{בטבלת שכיחויות}$$

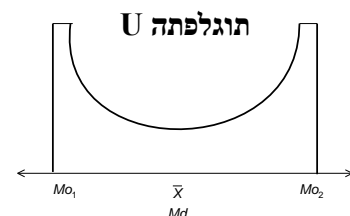
**במחלקות:** נשתמש באותה נוסחה רק נתייחס לאמצע המחלקה בתור ה  $X$ . הממוצע הזה יהיה ממוצע מקורב ולא אמיתי.  
הממוצע רלבנטי רק למשתנה כמותי.

**מדדי המיקום המרכזי בהתפלגויות המיוחדות:**

בהתפלגות סימטרית פעמונית כל מדדי המרכז שווים זה לזה:

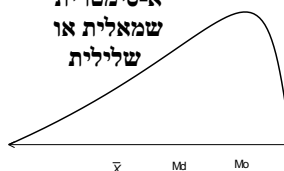
**התפלגות סימטרית**

בהתפלגות סימטרית השכיח לא חייב להיות במרכז:

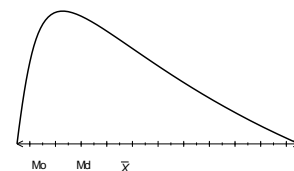


בהתפלגות אסימטרית

התפלגות  
א-סימטרית  
שמאלית או  
שלילית



התפלגות א-סימטרית  
ימנית או חיובית



**תרגילים:**

1. להלן רשימת הציונים של 20 תלמידים שנבחנו במבחן הבנת הנקרא:

7, 6, 8, 9, 10, 6, 4, 5, 8, 7, 6, 7, 6, 8, 9, 6, 7, 8, 5, 6

חשב את החציון, השכיח, והממוצע של הציונים.

2. בדקו את מספר החדרים לדירה בבניין בן 5 דירות והתקבל ממוצע 3.8

לגבי 4 דירות נמצא מספר חדרים: 4, 3, 4, 5.

א. כמה חדרים יש בדירה החמישית?

ב. מהו השכיח ומהו החציון?

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

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

א. חשב את הממוצע, החציון והשכיח של ההתפלגות.

ב. הסבר ללא חישוב כיצד כל מדד שחישבת בסעיף א' היה משתנה אם חלק מהמשפחות (לא

כולן) שלא היה להם עד היום טלוויזיה היו רוכשים מקלט אחד.

4. להלן התפלגות מספר המכוניות למשפחה בישוב "הגורן"

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

א. כמה משפחות יש בישוב?

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

ג. חשבו את הממוצע, החציון והשכיח.

הקפידו להסביר לגבי כל סעיף מה משמעות התוצאה שקיבלתם!

5. להלן התפלגות המשקל של קבוצה מסוימת בקייג:

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

המחלקה השכיחה והחציונית.

א. מהי

ב. חשב אומדן לממוצע.

ג. האם היה ניתן לדעת מהי התשובה לסעיף ב ללא חישוב? הסבר

**פתרונות:****שאלה 1:**

החציון: 7

השכיח: 6

הממוצע: 6.9

**שאלה 2:**

א. 3

ב. שכיח: 3,4 חציון: 4

**שאלה 3:**

א. הממוצע: 1.7

החציון: 1.5

השכיח: 1

ב. הממוצע יגדל ויתר המדדים לא ישתנו.

**שאלה 4:**

א. 630

ב. 34.13%

ג. שכיח וחציון: 3 ממוצע: 2.952

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

### רקע:

**המטרה :** למדוד את הפיזור של הנתונים כלומר כמה הם רחוקים זה מזה ושונים זה מזה.

### הטווחותחום RANGE:

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

### שונות וסטיית תקן:

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

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

דוגמה : נחשב את השונות של סדרת המספרים הבאה : 5,4,9

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

להלן התפלגות הציונים בכיתה מסוימת בה ממוצע הציונים הוא 7.44

$x^2 \cdot f$	השכיחות-f	הציון-X
50	2	5
144	4	6
392	8	7
320	5	8
324	4	9
200	2	10
1430		סה"כ

$$s_x^2 = \frac{\sum x^2 f(x)}{n} - \bar{x}^2 = \frac{1430}{25} - 7.44^2 = 1.8464$$

$$s = \sqrt{s_x^2} = \sqrt{1.8464} = 1.3588$$

כשיש מחלקות נעזר באמצע המחלקה כדי לחשב את השונות.

תרגילים:

1. להלן רשימת הציונים של 20 תלמידים שנבחנו במבחן הבנת הנקרא:

7, 6, 8, 9, 10, 6, 4, 5, 8, 7, 6, 7, 6, 8, 9, 6, 7, 8, 5, 6  
 חשבו את השונות, סטיית התקן והטווח של הציונים.

2. להלן התפלגות מספר המכוניות למשפחה בישוב "הגורן"

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

א. חשבו סטיית התקן.

ב. חשבו את הטווח של הנתונים.

הקפידו להסביר לגבי כל סעיף מה משמעות התוצאה שקיבלתם!

3. בחברה העוסקת בטלמרקטינג בדקו עבור כל עובד את מספר שנות הוותק שלו. התקבל שממוצע שנות הוותק הוא 4 שנים וסטיית התקן היא שנתיים.

א. האם הממוצע יגדל/יקטן/לא ישתנה וסטיית התקן תגדל/תקטן/לא תשנה כאשר יתווספו שני עובדים עם וותק של 4 שנים להתפלגות?

ב. האם הממוצע יגדל/יקטן/לא ישתנה וסטיית התקן תגדל/תקטן/לא תשנה כאשר יתווספו שני

עובדים אשר אחד עם וותק של 0 שנים והשני עם וותק של 8 שנים להתפלגות?

4. נתונה רשימה של 5 תצפיות, אך רק עבור 4 מהן נרשמו הסטיות שלהן מהממוצע:

2, 3, 2, -1. חשב את השונות של חמש התצפיות.



5. בשכונה בדקו בכל דירה את מספר החדרים לדירה. בשכונה 200 דירות.

פרופורציה	מספר חדרים
0.1	1
0.2	2
0.4	3
0.15	4
	5

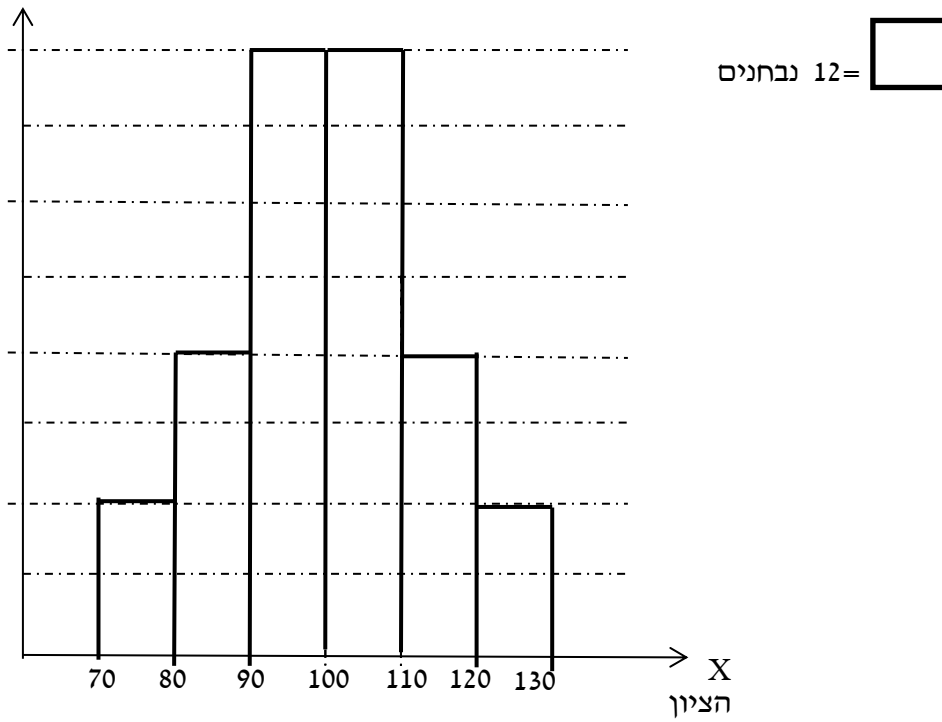
- א. מה הממוצע של מספר החדרים לשכונה בדירה?  
 ב. חשבו את סטיית התקן של מספר החדרים לדירה.  
 ג. חלק מבעלי הדירות בנות 2 החדרים הפכו את דירתם לדירת חדר. כיצד הדבר ישפיע (יקטין, יגדל, לא ישנה) כל מדד שחישבתם בסעיפים הקודמים.

6. להלן התפלגות המשקל של קבוצה מסוימת בק"ג:

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

מהי סטיית התקן של התפלגות המשקל?

7. להלן התפלגות הציונים במבחן אינטליגנציה:



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

**פתרונות :****שאלה 1:**

השונות : 2.19

סטיית תקן : 1.48

טווח : 6

**שאלה 2:**

א. סטיית תקן: 1.106

ב. טווח 4

**שאלה 3:**

א. ממוצע לא ישתנה, סטיית התקן תקטן.

ב. ממוצע לא ישתנה, סטיית התקן תגדל.

**שאלה 4:**

10.8

**שאלה 5:**

א. 3.05

ב. 1.16

**שאלה 6:**

7.73

**שאלה 7:**

א. 100

ב. 12.96

## פרק 6 - סטטיסטיקה תיאורית - מדדי פיזור - ממוצע סטיות מוחלטות מהחציון

### רקע:

מדד זה הוא מדד לפיזור בנוסף למדדים שנלמדו בפרקים הקודמים כמו סטיית התקן.

המדד בודק את הפיזור הממוצע סביב החציון.

הרעיון הוא למצוא בכמה בממוצע התצפיות סוטות בערך המוחלט מהחציון.

כדי לחשב את המדד יש לחשב קודם כל את החציון.

$$MAD = \frac{\sum_i |X_i - Md|}{n} : \text{אם מדובר ברשימה של תצפיות הנוסחה לחישוב המדד}$$

$$MAD = \frac{\sum |X_i - Md| \cdot f(X)}{n} : \text{אם מדובר בטבלת שכיחויות, הנוסחה לחישוב המדד}$$

כאשר מדובר על טבלת שכיחויות במחלקות ניקח בתור X את אמצע המחלקה.

### דוגמה: (הפתרון בהקלטה)

נתונה רשימת המספרים הבאה : 2 8 7 6 3

מה ממוצע הסטיות המוחלטות מהחציון?

**תרגילים:**

1. נתונה רשימת המספרים הבאה : 3 5 6 9 12 8 . מה ממוצע הסטיות המוחלטות מהחציון?

2. להלן התפלגות מספר מקלטי הטלויזיה שנספרו עבור כל משפחה בישוב מסוים :

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

א. חשב את החציון.

ב. חשב את ממוצע הסטיות המוחלטות מהחציון.

ג. הסבר ללא חישוב כיצד כל מדד שחישבת היה משתנה אם 5 ממשפחות שהיה להם מקלט יחיד היו מוכרים אותו.

**פתרונות:****שאלה 1 :**

2.5

**שאלה 2:**

א. 1.5

ב. 1.14

ג. חציון לא ישתנה וממוצע סטיות מוחלטות מהחציון יגדל.

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

### רקע:

המטרה למדוד איך תצפית ממוקמות יחסית לשאר התצפיות בהתפלגות.

### ציון תקן:

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

ציון התקן נותן כמה סטיות תקן סוטה התצפית מהממוצע.

כלומר, ציון התקן מעיד על כמה סטיות תקן התצפית מעל או מתחת לממוצע.

ציון תקן חיובי אומר שהתצפית מעל הממוצע.

ציון תקן שלילי אומר שהתצפית מתחת לממוצע.

ציון תקן אפס אומר שהתצפית בדיוק בממוצע.

### דוגמה: ( פתרון בהקלטה )

במקום עבודה מסוים ממוצע המשכורות 8 אלפי ₪ עם סטית תקן של 2 אלפי ₪ באותו מקום עבודה ההשכלה הממוצעת של העובדים הנה 14 שנים עם סטית תקן של 1.5 שנים. ערן מרוויח במקום עבודה זה 11 אלף ₪ והשכלתו 16 שנים. מה ערן יותר באופן יחסי משכיל או משתכר ?

### תרגילים

1. תלמידי כיתה ח' ניגשו למבחן בלשון ולמבחן במתמטיקה.  
להלן התוצאות שהתקבלו :

ממוצע	ממוצע	סטיית תקן
לשון	74	12
מתמטיקה	80	16

עודד קיבל: 68 בלשון ו70 במתמטיקה.

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

2. במפעל לייצור מצברים לרכב בדקו במשך 40 ימים את התפוקה היומית ( מספר מצברים במאות) ואת מספר הפועלים שעבדו באותו היום.

להלן טבלה המסכמת את האינפורמציה שנאספה על שני המשתנים :

מספר פועלים	תפוקה	מספר פועלים
15	48	ממוצע
2	10	סטיית תקן

באחד הימים מתוך כלל הימים שנבדקו התפוקה הייתה 50 מאות מצברים ובאותו היום עבדו 13 פועלים.

מה יותר חריג באותו היום יחסית לשאר הימים שנבדקו נתוני התפוקה או כמות הפועלים?  
בחר בתשובה הנכונה.

א. התפוקה.

ב. כמות הפועלים.

ג. חריגים באותה מידה.

ד. חסרים נתונים כדי לדעת זאת.

3. הגובה הממוצע של המתגייסים לצבא הוא 175 סנטימטר עם סטיית תקן 10 סנטימטר. המשקל

הממוצע 66 ק"ג עם סטיית תקן 8 ק"ג. ערך התגייס, גובהו 180 ס"מ ומשקלו 59 ק"ג.

א. במה ערך חריג יותר ביחס לשאר המתגייסים- גובהו או משקלו?

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



**פתרונות:**

**שאלה 1:**

א. לשון

ב. 72

**שאלה 2:**

תשובה ב

**שאלה 3:**

א. משקל

ב. 70

**פרק 8 - סטטיסטיקה תיאורית - מדדי מיקום יחסי - אחוזונים בטבלת שכיחויות בדידה**

**רקע:**

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

מסמנים את האחוזון ה- $p$  ב- $X_p$ .

**חישוב האחוזון מתוך נתונים בטבלת שכיחויות בדידה:**

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

**דוגמה:** (פתרון בהקלטה)

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

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

מצא את האחוזון ה-25.

מצא את הערך ש-20% מהמקרים מעליו.

**תרגילים:**

1. להלן התפלגות של משתנה כלשהו.

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

מצא להתפלגות את:  
 האחוזון ה-60.  
 המאון ה-40.  
 העשירון העליון.  
 הטווח בין הרבעונים.

2. להלן התפלגות מספר המכוניות למשפחה בישוב "הגורן"

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

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

**פתרונות:**

**שאלה 2**

א. 1

ב. 2

ג. 4

ד. 4

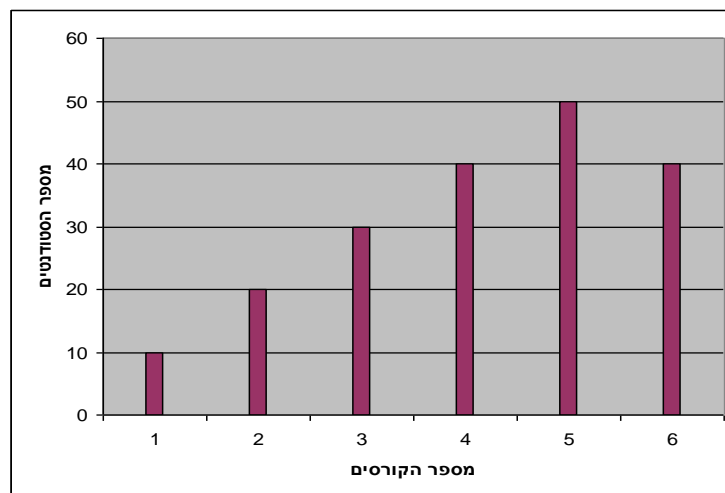
## פרק 9 - סטטיסטיקה תיאורית - שאלות מסכמות

1. בדקו עבור 5 תלמידים את המשקל שלהם :

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

- א. מהו המשתנה הנחקר? האם הוא בדיד או רציף?  
 ב. מהו המשקל החציוני, הממוצע והשכיח?  
 ג. מה הטווח וסטיית התקן של המשקל?  
 ד. לאותם תלמידים חישובו גם את הגובה בס"מ וקיבלו גובה ממוצע של 168 וסטיית תקן 6. במה תלמיד מספר 3 שגובהו 162 יותר חריג במשקל או בגובה?  
 ה. הוסיפו עוד תלמיד השוקל 52 ק"ג בדיוק. הסבירו ללא חישוב כיצד הדבר ישפיע על הממוצע וסטיית התקן? ( יגדילו יקטין או לא ישנה)

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



- א. מה המשתנה הנחקר? האם הוא בדיד או רציף?  
 ב. מהי צורת ההתפלגות?  
 ג. תאר את הנתונים בטבלת שכיחויות.  
 ד. חשב את השכיח, החציון והטווח .

3. להלן התפלגות הציונים בבחינה בלשון שנעשתה עבור תלמידי כיתות ד'. השתתפו במחקר 150 תלמידים.

$$\text{ממוצע הציונים שהתקבל: } \bar{X} = 7\frac{1}{15}$$

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

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

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

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

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

5. להלן התפלגות מספר שעות העבודה לשבוע של העובדים בחברת "סטאר".  
בחברה 200 עובדים.

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

- א. השלם את הטבלה.
- ב. חשב את החציון, השכיח, והממוצע של התפלגות מס' שעות העבודה בחברה.
- ג. מהי סטיית התקן של מס' שעות העבודה?
- ד. מה העשירון העליון של ההתפלגות?
- ה. איזה אחוז מהעובדים עובדים מעל 45 שעות בשבוע?
- ו. מה ציון התקן של רינה שעובדת 30 שעות בשבוע?
- ז. כיצד ישתנה החציון, הממוצע וסטיית התקן אם מספר שעות העבודה המינימאלי אינו 10 אלא 15? הסבר.



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

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

- א. מה אחוז האנשים ששלחו פחות מ-80 מסרונים בחודש?  
 ב. מה אחוז האנשים ששלחו בין 50 ל-120 מסרונים?  
 ג. הוחלט להעניק מתנה עבור  $\frac{1}{4}$  מהלקוחות שמשלמים במספר הרב ביותר של מסרונים בחודש. החל מאיזה כמות של מסרונים תחולק המתנה?  
 ד. ציינו איזה מדד ניתן לחשב ואיזה לא ניתן. אם ניתן חשב:
1. ממוצע
  2. שכיח
  3. חציון
  4. שונות

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

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

- א. שרטט היסטוגרמה לתיאור התפלגות זמן ביצוע המשימה.  
 ב. מתוך ההיסטוגרמה שבנית בסעיף א מהי צורת ההתפלגות?  
 ג. חשב את השכיח והחציון של ההתפלגות.  
 ד. הסבר, ללא חישובים, האם הזמן הממוצע לביצוע המשימה, קטן או גדול או שווה ביחס לשכיח ולחציון.

8. התפלגות ציוני מבחן אינטיליגנציה היא סימטרית .

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

נתון שהעשירון העליון הוא 130 והרבעון התחתון הוא 90.

נתון שלמבחן נגשו 500 מועמדים.

א. השלימו את הטבלה.

ב. מהו הממוצע והחציון של ההתפלגות?

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

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

9. להלן מספר טענות , עבור כל טענה ציין אם היא נכונה או לא נכונה ונמקד .

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

ב. ציון התקן של החציון תמיד יהיה 0.

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

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

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

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

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

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

**פתרונות:****שאלה 1:**

- א. המשתנה הנחקר כאן הוא משקל תלמיד בק"ג והוא משתנה כמותי רציף.  
ב.

$$\bar{X} = 52$$

$$Md = X_{\frac{n+1}{2}} = X_3 = 58$$

השכיח הוא 58

$$R = 28 \quad \text{ג.}$$

$$s = 10.12$$

- ד. הוא חריג יותר בגובה כי שם ציון התקן בערך מוחלט יותר גבוה.  
ה. הממוצע לא ישתנה אך סטיית התקן תקטן.

**שאלה 2:**

- א. מספר הקורסים. בדיד.  
ב. התפלגות אסימטרית שמאלית  
ד. השכיח: 5  
הטווח: 5

**שאלה 3:**

- א. 20 תלמידים קיבלו ציון 6 ו-40  
תלמידים קיבלו ציון 8.  
החציון: 7  
השכיח: 8  
ג. השונות: 2.533  
סטיית התקן: 1.592

**שאלה 4:**

- א. 20%  
ב. שביעות רצון (סדר)  
ג. 2  
ד. טווח: 4 שכיח: 2 חציון: 2.5  
ה. חציון: 4

**שאלה 5:**

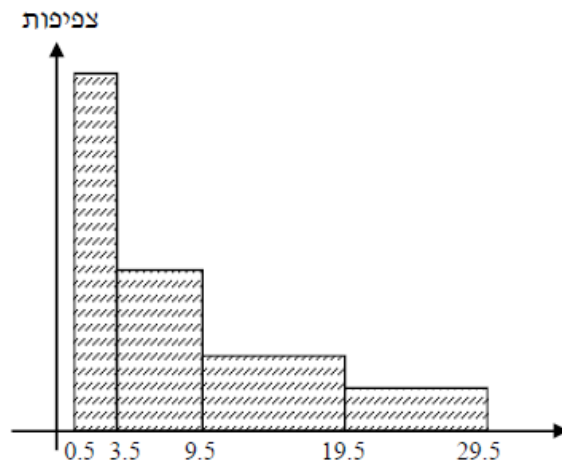
- ב. החציון : 35  
 השכיח : 35  
 הממוצע : 35  
 ג. סטיית תקן : 12.65  
 ד. 53.333  
 ה. 25%  
 ו. -0.395  
 ז. חציון לא ישתנה, ממוצע יגדל סטיית התקן תקטן.

**שאלה 6:**

- א. 38%  
 ב. 40%  
 ג. 150  
 ד. החציון : 100

**שאלה 7:**

א.



ב. ההתפלגות היא א-סימטרית ימנית .

ג. שכיח : 2 חציון : 6.83

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

**שאלה 8:**

א.

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

ב. 100

ג. 104

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

**שאלה 9:**

א. נכון

ב. לא נכון

ג. לא נכון

ד. לא נכון

ה. לא נכון

ו. נכון

ז. לא נכון

ח. לא נכון

## פרק 10 - מדדי קשר - מדד הקשר של קרמר

### רקע:

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

### דוגמה: (פתרון בהקלטה)

במחקר רוצים לבדוק את הקשר בין מין לדעה בנושא מסוים, שאלו 100 גברים ו-100 נשים האם הם בעד/נגד/נמנע/נשוא. להלן טבלת השכיחויות המשותפת שהתקבלה.

f(x)	נמנע	נגד	בעד	Y X
100	10	40	50	גבר
100	10	60	30	אישה
<b>n=200</b>	20	100	80	f(y)

בהקשר של קרמר הטבלה נקראת טבלת O (observed)

X - מין (גבר/אישה) – סולם שמי

Y - דעה (בעד/נמנע/נגד) – סולם שמי/סדר

שלבים בחישוב  $r_c$ :

שלב א': נבנה את טבלת E (Expected)

נעתיק את המסגרת של טבלת O ואז כל  $E_i = (f_{(x)} * f_{(y)}) / n$

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

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

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

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

**תרגילים:**

1. להלן תוצאות מחקר שבדק את הקשר בין מין להשכלה. לגבי כל נחקר נבדק המין שלו והשכלתו. להלן התוצאות:

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

האם קיים קשר בין מין להשכלה? נמק!

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

**פתרונות :**

**שאלה 1**

0.595

**שאלה 2**

ב. 0.19



## פרק 11 - מדדי קשר - מדד הקשר של ספירמן

### רקע:

מתי נשתמש במדד ספירמן ?

כאשר אחד המשתנים מסולם סדר והשני מסולם סדר ומעלה.

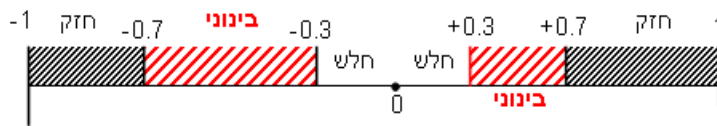
הקשר שהמדד בודק הוא קשר דירוגי.

מדד הקשר בודק :

1 כיוון של הקשר.

2 בודק את עצמת הקשר.

המדד מקבל ערכים בסקלה מ -1 ועד 1.



אם מדד הקשר של ספירמן יוצא 1 המשמעות היא שיש קשר דירוגי חיובי מלא : ככל המשתנה אחד עולה השני עולה ללא יוצא מן הכלל.

קשר דירוגי חיובי חלקי ( שמקדם המתאם בין 0 ל-1 ) אומר שככל שמשתנה אחד עולה לשני יש נטייה לעלות אך לא באופן מוחלט.

אם מדד הקשר של ספירמן יוצא -1 המשמעות היא שיש קשר דירוגי שלילי מלא : ככל שהמשתנה אחד עולה השני יורד ללא יוצא מן הכלל.

קשר דירוגי שלילי חלקי ( שמקדם המתאם הוא בין 0 ל- -1 ) אומר שככל שמשתנה אחד עולה לשני יש נטייה לרדת אך לא באופן מוחלט.

על מנת לחשב את הקשר יש לבצע פעולת דירוג ( RANK ) נלמד את פעולת הדירוג דרך הדוגמה הבאה ( פתרון בהקלטה )

שם התלמיד	הערכה	דירוג R
ערן	בינוני	
מיכל	מצוין	
עודד	חלש	
רוני	טוב	
יעל	טוב	

כאשר מדרגים אם יש כמה תצפיות שתופסות את אותו הערך אז הדירוג שלהם הוא הממוצע של המקומות שהן תופסות.

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

הנוסחה של מדד הקשר :

דוגמה : ( פתרון בהקלטה)

בתחרות רוקדים עם כוכבים השתתפו 7 זוגות, 2 שופטים נתנו את ציוניהם לריקוד של כל זוג.  
מהי מידת ההתאמה בין ציוני השופטים?

$X$ - ציון שופט א (סולם סדר)

$Y$ - ציון שופט ב (סולם סדר)

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			

**תרגילים:**

1. בתחרות יופי חילקו שני שופטים ציונים למועמדות:

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

האם קיים קשר בין שתי הערכות השופטים? נמק והסבר!

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

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

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

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

**פתרונות:****שאלה 1:**

0.973

**שאלה 2:**

-0.85

**שאלה 3:**

לא נכון

## פרק 12- מדדי קשר - מדד הקשר הלינארי (פירסון)

### רקע:

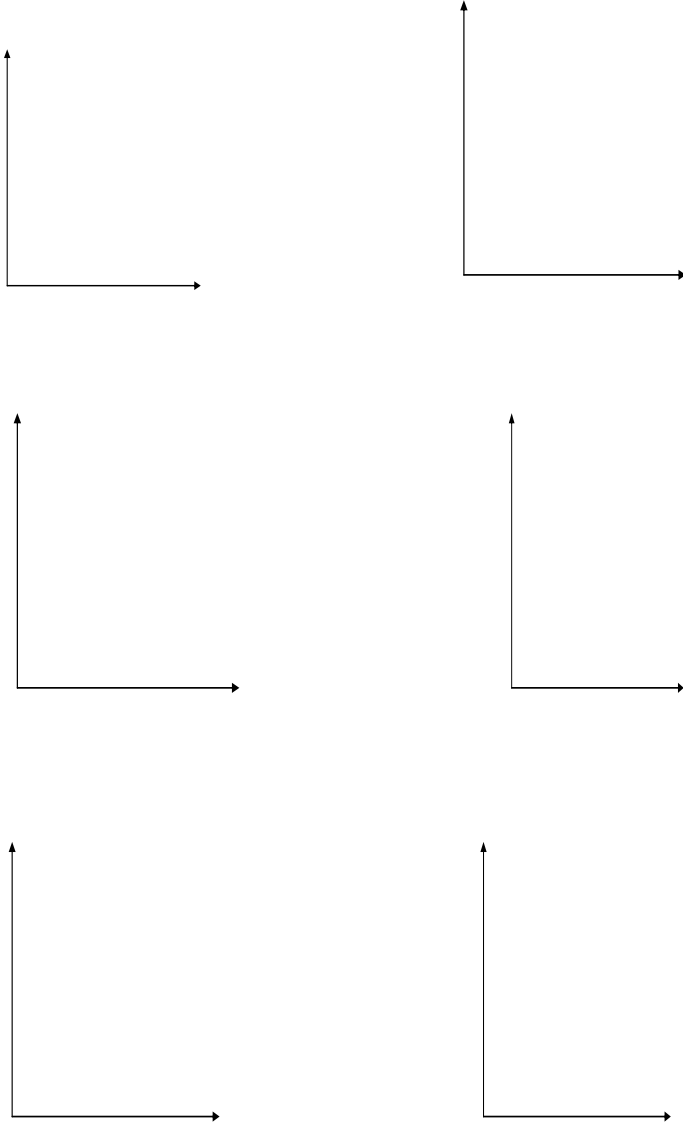
המטרה היא לבדוק האם קיים קשר (קורלציה, מתאם) של קו ישר בין שני משתנים כמותיים. מבחינת סולמות המדידה קשר בין סולמות רווחים ומנה. בדרך כלל, X הוא המשתנה המסביר (הבלתי תלוי) ו Y הוא המשתנה המוסבר (התלוי). למשל, נרצה להסביר כיצד השכלה של אדם הנמדדת בשנות לימוד X- מסבירה את ההכנסה שלו Y. במקרה זה שנות ההשכלה זהו המשתנה המסביר (או הבלתי תלוי) ואנחנו מעוניינים לבדוק כיצד שינויים בשנות ההשכלה של אדם יכולים להסביר את השינויים שלו בהכנסה, ולכן רמת ההכנסה זהו המשתנה המוסבר התלוי במשתנה המסביר אותו. בשלב הראשון, נהוג לשרטט דיאגרמת פיזור. זו דיאגרמה שנותנת אינדיקציה ויזואלית על טיב הקשר בין שני המשתנים. למשל, בבניין של 5 דירות בדקו את הנתונים הבאים: X - מס' חדרים בדירה. Y - מס' נפשות הגרות בדירה. להלן התוצאות שהתקבלו:

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

נשרטט מנתונים הללו דיאגרמת פיזור:



נתבונן בכמה מקרים של דיאגרמות פיזור וננתח אותן :



בשלב השני, מחשבים את מקדם המתאם (מדד הקשר) שבודק עד כמה קיים קשר לינארי בין שני המשתנים. המדד (ניקרא גם מדד הקשר של פירסון) מכמת את מה שנראה בשלב הראשון רק בעין.

המדד בודק את כיוון הקשר (חיובי או שלילי).

ואת עוצמת הקשר (חלש עד חזק).

מקדם מתאם זה מקבל ערכים בין -1 ל 1.

מקדם מתאם -1 או 1 אומר שקיים קשר לינארי מוחלט ומלא בין המשתנים שניתן לבטאו על ידי

$$y = bx + a$$

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

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

מתאם חיובי חלקי אומר שככל שמשנתה אחד עולה לשני יש נטייה לעלות בערכו אבל לא קיימת

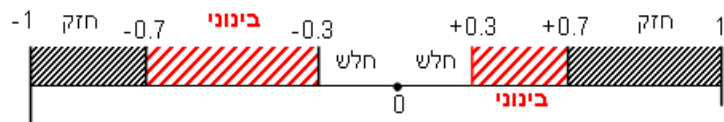
נוסחה לינארית שמקשרת את  $X$  ל- $Y$  באופן מוחלט ואילו מתאם שלילי חלקי אומר שככל

שמשנתה אחד עולה לשני יש נטייה לרדת אבל לא קיימת נוסחה לינארית שמקשרת את  $X$  ל- $Y$

באופן מוחלט.

ככל שערך מקדם המתאם קרוב לאפס נאמר שעוצמת הקשר חלשה יותר וככל שמקדם המתאם

רחוק מהאפס נאמר שעוצמת הקשר חזקה יותר.



מקדם המתאם יסומן באות  $r$ .

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

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

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

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

$$r_{xy} = \frac{cov(x, y)}{s_x \cdot s_y} : \text{מקדם המתאם הלינארי}$$



**תרגילים:**

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

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

- א. שרטט דיאגרמת פיזור לנתונים. מה ניתן להסיק מהדיאגרמה על טיב הקשר בין מספר החיסורים של תלמיד לציונו? מיהו המשתנה הבלתי תלוי ומיהו המשתנה התלוי?
- ב. חשב את מדד הקשר של פירסון. האם התוצאה מתיישבת עם תשובתך לסעיף א'?
- ג. הסבר ללא חישוב כיצד מקדם המתאם היה משתנה אם היה מתווסף תלמיד שהחסיר 4 פעמים וקיבל ציון 80?

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

x	y
10	12
14	15
15	15
18	17
20	21

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

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

$$\begin{aligned}\sum_{i=1}^{20} Y_i &= 200 & \sum_{i=1}^{20} X_i &= 240 \\ \sum_{i=1}^{20} (Y_i - \bar{Y})^2 &= 76 & \sum_{i=1}^{20} (X_i - \bar{X})^2 &= 76 \\ \sum_{i=1}^{20} (X_i - \bar{X})(Y_i - \bar{Y}) &= 60.8\end{aligned}$$

- א. חשב את מדד הקשר הלינארי בין  $X$  ל- $Y$ . מיהו המשתנה התלוי?  
 ב. מה המשמעות של התוצאה שקיבלת בסעיף א?

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

$$\begin{aligned}\sum_{i=1}^{20} Y_i &= 200 & \sum_{i=1}^{20} X_i &= 240 \\ \sum_{i=1}^{20} Y_i^2 &= 2080 & \sum_{i=1}^{20} X_i^2 &= 2960 \\ \sum_{i=1}^{20} X_i Y_i &= 2464\end{aligned}$$

חשב את מדד הקשר הלינארי בין  $X$  ל- $Y$ .

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

6. להלן רשימת טענות, לגבי כל טענה קבע נכון/לא נכון ונמק:

- א. מתווך דירות המיר מחירי דירות מדולר לשקל. נניח שדולר אחד הוא 3.5 ₪. אם מתווך הדירות יחשב את מדד הקשר של פירסון בין מחיר הדירה בשקלים למחיר הדירה בדולרים הוא יקבל 1.  
 ב. לסדרה של נתונים התקבל  $\bar{X} = \bar{Y} = 6$   $S_x = S_y = 1$  לכן מדד הקשר של פירסון יהיה 1.  
 ג. אם השונות המשותפת של  $X$  ושל  $Y$  הינה 0 אז בהכרח גם מקדם המתאם של פירסון יהיה 0.

### שאלות אמריקאיות:

לפתרון מלא בסרטון וידאו היכנסו ל- [www.GooL.co.il](http://www.GooL.co.il)

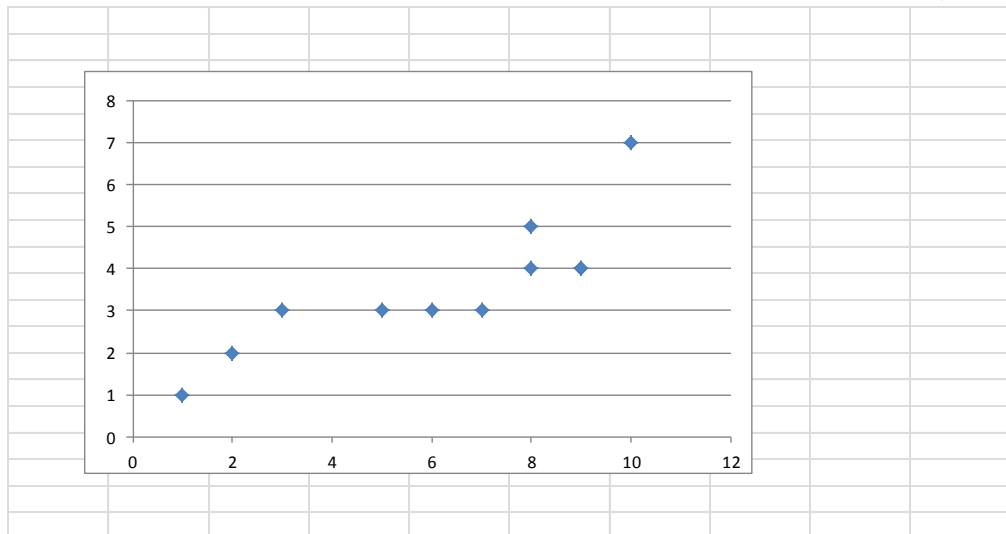
7. נמצא שקיים מקדם מתאם שלילי בין הציון בעברית לציון בחשבון בבחינה לכן :

- א. הדבר מעיד שהציונים בכתה היו שליליים.
- ב. ככל שהציון של תלמיד יורד בחשבון יש לו נטייה לרדת בעברית.
- ג. ככל שהציון של תלמיד עולה בחשבון יש לו נטייה לרדת בעברית.
- ד. אף אחת מהתשובות לא נכונה.

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

- א. 1
- ב. 0
- ג. 4.2
- ד. לא ניתן לדעת.

9. להלן דיאגרמת פיזור :



מה יהיה מקדם המתאם בין שני המשתנים?

- א. 1
- ב. 0.85
- ג. 0.15
- ד. 0

**פתרונות:****שאלה 1:**

א. בהקלטה

ב.  $-0.9325$ **שאלה 2:**א.  $\bar{y} = 16$  $\bar{x} = 15.4$ ב.  $r_{xy} = 0.96$ **שאלה 3:**

א : 0.8

**שאלה 4:**

0.8

**שאלה 5:**

1

**שאלה 6:**

א. נכון

ב. לא נכון

ג. נכון

**שאלה 7:**

התשובה : ג

**שאלה 8:**

התשובה : א

**שאלה 9:**

התשובה : ב

## פרק 13 - מדדי קשר - השפעת טרנספורמציה לינאריות על מדד הקשר של פירסון

### רקע:

טרנספורמציה לינארית בין אם נעשית על  $X$  ובין אם נעשית על  $y$ , או בין אם נעשית על שניהם, אינה משנה את עוצמת הקשר. היא עלולה רק לשנות את כיוונו אם השיפועים של שתי הטרנספורמציות שוני סימן.

$$r_{[(aX+b),(cY+d)]} = \begin{cases} r_{x,y} & \text{if } a \cdot c > 0 \\ -r_{x,y} & \text{if } a \cdot c < 0 \end{cases}$$

### תרגילים:

1. מבחן בנוי מחלק כמותי ומילולי.  
 מקדם המתאם בין שני הציונים של שני החלקים הוא 0.9.  
 א. אם יעלו את כל הציונים בחלק המילולי ב-20%, מה יהיה מקדם המתאם בין הציון המילולי החדש לציון הכמותי ובין הציון המילולי הישן לציון המילולי החדש?  
 ב. נגדיר משתנה חדש  $W$  להיות המרחק של הציון בחשיבה מילולית מהציון המקסימאלי בבחינה-150. מצא את מקדם המתאם בין הציון המילולי ל- $W$  ובין  $W$  ל-ציון הכמותי.
2. מקדם המתאם בין ההכנסה לבין ההוצאה של 10 משפחות חושב והתקבל 0.7. אם חל גידול של 5% בהכנסת האוכלוסייה כולה וגידול של 7% בהוצאה שלה, אז מה יהיה מקדם המתאם בין ההכנסה החדשה להוצאה החדשה?
3. חברת "לק" המייצרת גלידה החליטה לערוך מחקר לבדיקת הקשר בין מספר חבילות הגלידה הנמכרות ביום לבין הטמפרטורה באותו יום. נבדקו 10 ימים והתקבל מתאם לינארי 0.85. חברת "לק" דואגת להתחיל כל יום עם מלאי של 150 חבילות גלידה. בנוסף, מעוניינים כי הטמפרטורה תבוטא במעלות פרנהייט במקום במעלות צלסיוס. מה ערכו של מקדם המתאם בין מספר חבילות הגלידה שנשארות בסוף היום לבין הטמפרטורה במעלות פרנהייט?  
 הקשר בין מעלות צלסיוס ( $C^\circ$ ) למעלות פרנהייט ( $F^\circ$ ) נתון ע"י  $F = \frac{9}{5}C + 32$ .
- בחר בתשובה הנכונה:
- א. 0.85  
 ב. -0.85  
 ג. 1  
 ד. לא ניתן לדעת.
4. מקדם המתאם בין  $X$  ל- $Y$  הנו 0.4 כל ערכי ה- $X$  הוכפלו ב-2 לכן מקדם המתאם החדש בין שני המשתנים יהיה:  
בחר בתשובה הנכונה:
- א. 0.8  
 ב. 0.4  
 ג. -0.4  
 ד. לא ניתן לדעת.

**פתרונות :****שאלה 1:**

- א. בין הציון המילולי הישן לחדש 1:  
בין הציון המילולי החדש לכמותי 0.9:
- ב. בין  $W$  ל ציון המילולי : -1  
בין  $W$  לציון הכמותי : -0.9

**שאלה 2:**

0.7

**שאלה 3:**

התשובה : ב

**שאלה 4:**

התשובה : ב

## פרק 14 - מדדי קשר - רגרסיה ליניארית

### רקע:

במידה וקיים קשר חזק בין שני המשתנים הכמותיים נהוג לבצע ניבויי. לבנות קו ניבויים הנקרא גם קו רגרסיה המנבא משתנה אחד על סמך האחר.

מדובר בקו שמנבא את  $Y$  על סמך  $X$ . השיטה למציאת הקו הנ"ל נקראת שיטת הריבועים הפחותים והקו המתקבל נקרא קו הרגרסיה או קו הניבויים או קו הריבועים הפחותים.

a - בעצם נותן את ערך  $Y$  כאשר  $X$  הנו אפס על גבי קו הניבויים. הוא ניקרא החותך של הקו.

b - הוא שיפוע הקו נותן בכמה בעצם  $Y$  משתנה כאשר  $X$  גדל ביחידה אחת על גבי קו הניבויים.

להלן המשוואות למציאת הפרמטרים של קו הרגרסיה:

$$\tilde{Y} = bX + a$$

$$b = r \frac{S_y}{S_x}$$

$$a = \bar{Y} - b\bar{X}$$

אם נרצה לבנות קו ניבויים לניבוי  $X$  על סמך  $Y$  נצטרך לעדכן את הנוסחאות בהתאם.



תרגילים:

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

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

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

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

- א. חשב את מדד הקשר הלינארי בין  $X$  ל- $Y$ . מיהו המשתנה התלוי?  
 ב. מצא את קו הרגרסיה לניבוי ההוצאה של משפחה על סמך הכנסה שלה. הסבר את משמעות הפרמטרים של קו הרגרסיה.  
 ג. משפחת כהן הכניסה 15,000 ₪, מה ההוצאה הצפויה שלה?

1. נסמן ב- $X$  את ההשכלה של אדם בשנות למוד. נסמן ב- $Y$  את הכנסתו באלפי ₪. במחקר התקבלו התוצאות הבאות:

$$S_y = 5 \quad S_x = 2$$

$$\bar{Y} = 8 \quad \bar{X} = 14$$

$$COV(X, Y) = 7.5$$

- א. חשב את מדד הקשר של פירסון בין ההשכלה להכנסה.  
 ב. מה ההכנסה הצפויה לאדם שהשכלתו 12 שנים?  
 ג. מה ההשכלה הצפויה לאדם שהכנסתו 10,000 ₪?

3. חוקר רצה לחקור את הקשר הקווי שבין הציון המבחן בסטטיסטיקה לבין מספר שעות ההכנה של הסטודנטים למבחן. במדגם של 100 סטודנטים שנבחנו בקורס נרשמו התוצאות הבאות: הציון הממוצע של הסטודנטים היה 65 עם סטיית תקן של 27. מספר שעות ההכנה הממוצע היה 30 עם סטיית תקן של 18. מקדם המתאם בין הציון לשעות ההכנה היה 0.8.  
 א. על פי משוואת הרגרסיה שעת הכנה נוספת משפרת את ציון המבחן ב?  
 ב. על פי משוואת הרגרסיה תלמיד שייגש למבחן ללא שעות הכנה כלל יקבל ציון?  
 ג. מהו קו הרגרסייה לניבוי הציון לפי שעות ההכנה?

4. נתונים 2 משתנים  $Y, X$ . כמו כן נתון:  $X$  ממוצע = 1.5, שונות  $X$  = שונות  $Y$  = 4, וכן שקו הרגרסיה של  $Y$  על בסיס  $X$  הינו  $Y = -0.2X + 0.5$ . חשב מהו מקדם המתאם בין  $X$  ל- $Y$ ?

פתרונות:שאלה 1:

א. 0.8

ב.  $\tilde{Y} = 0.8X + 0.4$

ג. 12.4

שאלה 2:

א. 0.75

ב. 4.25 אלפי ש"ח

ג. 14.6 שנים

שאלה 3:

א. 1.2

ב. 29

ג.  $y = 1.2x + 29$

שאלה 4:

-0.2

## פרק 15 - מדדי קשר - רגרסיה - שונות מוסברת ושונות לא מוסברת

### רקע:

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

$r^2$  - נותן בעצם איזה חלק מהשונות של המשתנה התלוי מוסבר. השונות המוסברת נקראת גם שונות ניבויים. השונות הלא מוסברת נקראת גם שונות טעויות.

### תרגילים :

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

- א. איזה אחוז מהשונות של מחירי הדירות מוסבר על ידי שטח הדירה?
- ב. איזה אחוז מהשונות של מחירי הדירות לא מוסבר על ידי שטח הדירה?
- ג. מהי השונות המוסברות ומהי השונות הלא מוסברת של מחירי הדירות?

2. להלן רשימת טענות, לגבי כל טענה קבע נכון/לא נכון ונמק!

- א. אם שונות הטעויות שווה ל-0 (השונות הלא מוסברת) אז מקדם המתאם של פירסון יהיה 1.
- ב. אם מקדם המתאם של פירסון בין שני משתנים הוא 1 אזי שונות הטעויות (השונות הלא מוסברת) תהיה 0.
- ג. אם השונות המשותפת של X ושל Y הינה 0 אז בהכרח גם מקדם המתאם של פירסון יהיה 0.

### שאלות אמריקאיות:

בשאלות הבאות יש לבחור בתשובה הנכונה.

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

- א. ללא יוצא מן הכלל ככל שערכי משתנה אחד עולה השני יעלה.
- ב. 64% מהשונות של משתנה אחד מוסבר על ידי המשתנה השני.
- ג. הקשר בין שני המשתנים הוא בעוצמה של 0.64.
- ד. כל התשובות נכונות.

4. אם מגדילים את  $r^2$  מה ניתן לומר?

- א. אחוז השונות המוסברת יקטן
- ב. אחוז השונות המוסברת יגדל
- ג. אחוז השונות המוסברת יישאר ללא שינוי.
- ד. סטיית התקן משתנה
- ה. לא ניתן לדעת

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

א. 0.44 .

ב. - 0.44 .

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

ד. אין אפשרות לחשב את מקדם המתאם.

ה. 0.35

## פרק 16 - בעיות בסיסיות בהסתברות

### רקע :

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

**מרחב מדגם :** כלל התוצאות האפשריות בניסוי המקרי :

בהטלת קובייה :  $\{1,2,3,4,5,6\}$ .  
 מזג האוויר בעוד שבועיים :  $\{ \text{נאה, שרבי, מושלג, גשום, מעונן חלקית, אביך} \}$

**מאורע :** תת קבוצה מתוך מרחב במדגם. מסומן באותיות :  $A, B, C, \dots$

בהטלת קובייה, למשל, לקבל לפחות 5 :  $A = \{5, 6\}$   
 לקבל תוצאה זוגית :  $B = \{2, 4, 6\}$

**גודל מרחב המדגם :** מספר התוצאות האפשריות במרחב המדגם :

בהטלת הקובייה :  $|\Omega| = 6$

**גודל המאורע :** מספר התוצאות האפשריות במאורע עצמו.

בהטלת הקובייה :  $|A| = 2$        $|B| = 3$

**מאורע משלים :** מאורע המכיל את כל התוצאות האפשריות במרחב המדגם פרט לתוצאות במאורע אותו הוא משלים :

בהטלת הקובייה :  $\bar{A} = \{1, 2, 3, 4\}$        $\bar{B} = \{1, 3, 5\}$

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

**הסתברות במרחב מדגם אחיד :**

במרחב מדגם אחיד הסיכוי למאורע יהיה :  $p(A) = \frac{|A|}{|\Omega|}$

למשל, מה הסיכוי בהטלת קובייה לקבל לפחות 5 ?  $p(A) = \frac{|A|}{|\Omega|} = \frac{2}{6}$

מה הסיכוי בהטלת קובייה לקבל תוצאה זוגית ?  $p(B) = \frac{|B|}{|\Omega|} = \frac{3}{6}$

**הסתברות במרחב לא אחיד :**

יחושב לפי השכיחות היחסית :  $\frac{f}{n}$

להלן התפלגות הציונים בכיתה מסוימת :

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

א. מה ההסתברות שתלמיד אקראי שניבחר בכיתה קיבל את הציון 8 ?  $\frac{f}{n} = \frac{5}{25} = 0.2$

ב. מה ההסתברות שתלמיד אקראי שניבחר בכיתה יכשל?

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

**הסתברות למאורע משלים :**

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

למשל, בדוגמה הקודמת הסיכוי לעבור את הבחינה יכול להיות מחושב לפי הסיכוי להיכשל :

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

**תרגילים:**

1. מהאותיות E, F ו-G יוצרים מילה בת 2 אותיות לא בהכרח בת משמעות.
- הרכב את כל המילים האפשריות.
  - רשום את המקרים למאורע:  
A - במילה נמצאת האות E.  
B - במילה האותיות שונות.  
ג. רשום את המקרים למאורע  $\bar{A}$ .
2. מטילים זוג קוביות.
- רשום את מרחב המדגם של הניסוי. האם המרחב מדגם הוא אחיד?
  - רשום את כל האפשרויות למאורעות הבאים:  
A - סכום התוצאות 7.  
C - מכפלת התוצאות 12.  
ג. חשב את הסיכויים למאורעות שהוגדרו בסעיף ב.
3. בוחרים באקראי ספרה מבין הספרות 0-9.
- מה ההסתברות שהספרה שנבחרה גדולה מ-5?
  - מה ההסתברות שהספרה שנבחרה היא לכל היותר 3?
  - מה ההסתברות שהספרה שנבחרה היא אי זוגית?
4. להלן התפלגות מספר מקלטי הטלוויזיה שנספרו עבור כל משפחה בישוב מסוים:

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

- נבחרה משפחה באקראי מהישוב.
- מה ההסתברות שאין מקלטים למשפחה?
  - מה ההסתברות שיש מקלטים למשפחה?
  - מה ההסתברות שיש לפחות 3 מקלטים למשפחה?



5. להלן התפלגות מספר המכוניות למשפחה ביישוב "עדן":

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

נבחרה משפחה אקראית מן הישוב.

א. מה ההסתברות שאין לה מכוניות?

ב. מה ההסתברות שבבעלות המשפחה לפחות 3 מכוניות?

ג. מה הסיכוי שבבעלותה פחות מ-3 מכוניות?

6. מטילים מטבע רגיל 3 פעמים. בצד אחד של המטבע מוטבע עץ ובצד השני פלי.

א. רשום את מרחב המדגם של הניסוי. האם המרחב מדגם הוא אחיד?

ב. רשום את כל האפשרויות למאורעות הבאים:

A- התקבל פעם אחת עץ.

D- התקבל לפחות פלי אחד.

ג. מהו המאורע המשלים ל-D.

ד. חשבו את הסיכויים למאורעות שהוגדרו בסעיפים ב- ג.

**פתרונות:****שאלה 2**

ג. הסיכוי ל-A:  $\frac{1}{6}$

הסיכוי ל-B:  $\frac{1}{9}$

**שאלה 3**

א. 0.4

ב. 0.4

ג. 0.5

**שאלה 4**

א. 0.22

ב. 0.78

ג. 0.32

## פרק 17 - פעולות בין מאורעות (חיתוך ואיחוד), מאורעות זרים ומכילים

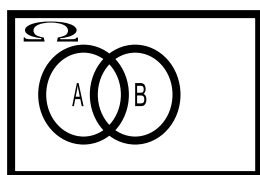
רקע:

פעולת חיתוך:

נותנת את המשותף בין המאורעות הנחתכים, חיתוך בין המאורע A למאורע B יסומן כך:

$$A \cap B$$

מדובר בתוצאות שנמצאות ב-A וגם ב-B.



בהטלת קובייה, למשל, לקבל לפחות 5 :  $A = \{5, 6\}$

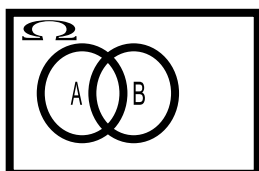
לקבל תוצאה זוגית :  $B = \{2, 4, 6\}$

$$A \cap B = \{6\}$$

פעולת איחוד:

נותנת את כל האפשרויות שנמצאות לפחות באחת מהמאורעות. הסימון הוא:  $A \cup B$  נותנת את

אשר נימצא ב-A או ב-B. כלומר, לפחות אחד מהמאורעות קורה.



בהטלת קובייה, למשל, לקבל לפחות 5 :  $A = \{5, 6\}$

לקבל תוצאה זוגית :  $B = \{2, 4, 6\}$

$$A \cup B = \{2, 4, 5, 6\}$$

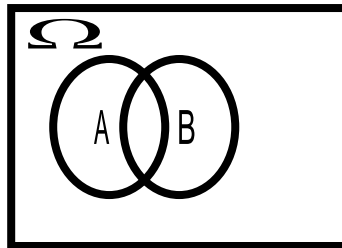
דוגמה ( הפתרון נמצא בהקלטה )

סטודנט ניגש בסמסטר לשני מבחנים. מבחן בסטטיסטיקה ומבחן בכלכלה. ההסתברות שלו לעבור את המבחן בסטטיסטיקה הוא 0.9. ההסתברות שלו לעבור את המבחן בכלכלה הוא 0.8. ההסתברות לעבור את המבחן בסטטיסטיקה ובכלכלה היא 0.75.

- א. מה ההסתברות שלו לעבור את המבחן בסטטיסטיקה בלבד?  
 ב. מה ההסתברות שלו להיכשל בשני המבחנים?  
 ג. מה ההסתברות לעבור לפחות מבחן אחד?

נוסחת החיבור לשני מאורעות :

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

חוקי דה מורגן לשני מאורעות:

$$\overline{A \cup B} = \bar{A} \cap \bar{B}$$

$$\overline{A \cap B} = \bar{A} \cup \bar{B}$$

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

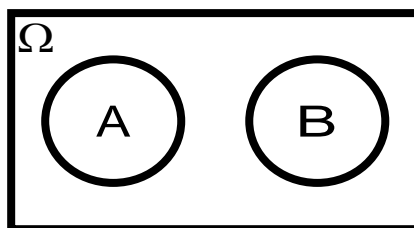
$$P(A \cup B) = 1 - P(\bar{A} \cap \bar{B})$$

שיטת ריבוע הקסם :

השיטה רלבנטית רק אם יש שני מאורעות במקביל בדומה לתרגיל הקודם :

	$\bar{A}$	$A$	
$B$	$P(\bar{A} \cap B)$	$P(A \cap B)$	$P(B)$
$\bar{B}$	$P(\bar{A} \cap \bar{B})$	$P(A \cap \bar{B})$	$P(\bar{B})$
	$P(\bar{A})$	$P(A)$	1

מאורעות זרים : מאורעות שאין להם מהמשותף : לא יכולים להתרחש בו זמנית.



$$A \cap B = \{\}$$

$$P(A \cap B) = 0$$

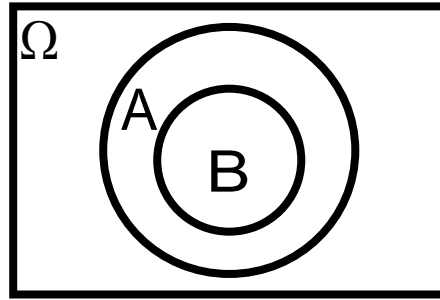
$$P(A \cup B) = P(A) + P(B)$$

למשל , בהטלת קובייה

$$A = \{5, 6\} \quad : \text{ לקבל לפחות 5}$$

$$B = \{3\} \quad : \text{ לקבל 3}$$

$$A \cap B = \{\}$$

מאורעות מכילים :

מאורע A מכיל את מאורע B כל התוצאות שנמצאות ב-B מוכלות בתוך המאורע-A.

קשר זה מסומן באופן הבא:  $B \subset A$

$$A \cap B = B$$

$$A \cup B = A$$

$$P(A \cap B) = P(B)$$

$$P(A \cup B) = P(A)$$

למשל:

$$A = \{2, 4, 6\}$$

$$B = \{2, 4\}$$

תרגילים:

לפתרון מלא בסרטון וידאו היכנסו ל- [www.GooL.co.il](http://www.GooL.co.il)

1. מהאותיות E, F ו-G יוצרים מילה בת 2 אותיות לא בהכרח בת משמעות. נגדיר את המאורעות הבאים :

-E במילה נמצאת האות E.

-F במילה אותיות שונות.

א. רשום את כל האפשרויות לחיתוך A עם B.

ב. רשום את כל האפשרויות לאיחוד של A עם B.

2. תלמיד ניגש בסמסטר לשני מבחנים מבחן בכלכלה ומבחן בסטטיסטיקה. נגדיר את המאורעות הבאים :

-A לעבור את המבחן בסטטיסטיקה.

-B לעבור את המבחן בכלכלה.

העזר בפעולות חיתוך, איחוד ומשלים בלבד כדי להגדיר את המאורעות הבאים וסמן בדיאגרמת וון את השטח המתאים :

א. התלמיד עבר רק את המבחן בכלכלה.

ב. התלמיד עבר רק את המבחן בסטטיסטיקה.

ג. התלמיד עבר את שני המבחנים.

ד. התלמיד עבר לפחות מבחן אחד.

ה. התלמיד נכשל בשני המבחנים.

ו. התלמיד נכשל בכלכלה.

3. נתבקשתם לבחור ספרה באקראי. נגדיר את A להיות הספרה שנבחרה היא זוגית. נגדיר את B להיות הספרה שנבחרה קטנה מ-5.

א. רשמו את כל התוצאות למאורעות הבאים :

$$A =$$

$$B =$$

$$\bar{B} =$$

$$A \cap B =$$

$$A \cup B =$$

ב. חשבו את ההסתברויות לכל המאורעות מהסעיף הקודם.

4. נסמן ב- $\Omega$  את מרחב המדגם וב- $\phi$  קבוצה ריקה.

נתון כי  $A$  הינו מאורע בתוך מרחב המדגם.  
להלן מוגדרים מאורעות שפתרונם הוא  $\Omega$  או  $\phi$  או  $A$ .  
קבע עבור כל מאורע מה הפתרון שלו.

$$\begin{aligned}
 &= \\
 &\bar{A} \\
 &A \cap \phi \\
 &A \cup \phi \\
 &A \cap \Omega \\
 &A \cup \Omega \\
 &A \cap \bar{A} \\
 &\bar{\phi} \\
 &A \cup \bar{A}
 \end{aligned}$$

5. הוגדרו המאורעות הבאים :

$A =$  אדם שגובהו מעל 1.7 מטר

$B =$  אדם גובהו מתחת ל-1.8 מטר

קבע את גובהם של האנשים הבאים :

א.  $A \cap B$

ב.  $A \cup B$

ג.  $\bar{A} \cap B$

ד.  $\bar{A} \cup \bar{B}$

$=$

ה.  $\bar{A}$



6. נגדיר את המאורעות הבאים :

A - אדם דובר עברית.

B - אדם דובר ערבית.

C - אדם דובר אנגלית.

השתמש בפעולות איחוד, חיתוך והשלמה לתיאור המאורעות הבאים :

א. אדם דובר את כל שלוש השפות.

ב. אדם דובר רק עברית.

ג. אדם דובר לפחות שפה אחת מתוך השפות הללו.

ד. אדם אינו דובר אנגלית.

ה. קבוצת התלמידים דוברי 2 שפות בדיוק (מהשפות הנ"ל).

7. שתי מפלגות רצות לכנסת הבאה. מפלגת "גדר" תעבור את אחוז החסימה בהסתברות של 0.08. מפלגת עתיד תעבור את אחוז החסימה בהסתברות של 0.20. בהסתברות של 76% שתי המפלגות לא תעבורנה את אחוז החסימה.

א. מה ההסתברות שלפחות אחת מהמפלגות תעבור את אחוז החסימה?

ב. מה ההסתברות ששתי המפלגות תעבורנה את אחוז החסימה?

ג. מה ההסתברות שרק מפלגות "עתיד" תעבור את אחוז החסימה?

8. במקום עבודה מסוים 40% מהעובדים הם גברים. כמו כן 20% מהעובדים הם אקדמאים. 10% מהעובדים הינן נשים אקדמאיות.

א. איזה אחוז מהעובדים הם גברים אקדמאיים?

ב. איזה אחוז מהעובדים הם גברים או אקדמאיים?

ג. איזה אחוז מהעובדים הם נשים לא אקדמאיות?

9. הסיכוי של מניה A לעלות הנו 0.5 ביום מסוים והסיכוי של מניה B לעלות ביום מסוים הנו 0.4. בסיכוי של 0.7 לפחות אחת מהמניות תעלה ביום מסוים. חשב את ההסתברויות הבאות לגבי שתי המניות הללו ביום מסוים :

א. ששתי המניות תעלנה.

ב. שאף אחת מהמניות לא תעלנה.

ג. שמניה A בלבד תעלה.

10. מטילים זוג קוביות אדומה ושחורה. נגדיר את המאורעות הבאים :

A- בקובייה האדומה התקבלה התוצאה 4 ובשחורה 2.

B- סכום התוצאות משתי הקוביות 6.

C- מכפלת התוצאות בשתי הקוביות 10.

א. האם A ו-B מאורעות זרים?

ב. האם המאורע B מכיל את המאורע A?

ג. האם A ו-C מאורעות זרים?

ד. האם A ו-C מאורעות משלימים?

11. עבור המאורעות A ו-B ידועות ההסתברויות הבאות :

$$p(\bar{A} \cap \bar{B}) = 0.1 \quad p(B) = 0.3 \quad p(A) = 0.6$$

א. האם A ו-B מאורעות זרים?

ב. חשב את  $p(\bar{A} \cap B)$

12. מטבע הוטל פעמיים. נגדיר את המאורעות הבאים :

A- קיבלנו עץ בהטלה הראשונה.

B- קיבלנו לפחות עץ אחד בשתי ההטלות.

איזו טענה נכונה?

א. A ו-B מאורעות זרים.

ב. A ו-B מאורעות משלימים.

ג. B מכיל את A.

ד. A מכיל את B.

13. בהגרלה חולקו 100 כרטיסים על 3 מהם רשום חופשה ועל 2 מהם רשום מחשב שאר

הכרטיסים ריקים. אדם קיבל כרטיס אקראי.

א. מה הסיכוי לזכות בחופשה או במחשב? האם המאורעות הללו זרים?

ב. מה ההסתברות לא לזכות בפרס?

.14

$$P(A) = 0.3$$

$$P(B) = 0.25$$

$$P(A \cup B) = 0.49$$

א. חשב את הסיכוי ל-  $P(A \cap B)$

ב. האם A ו- B מאורעות זרים?

ג. מה ההסתברות שרק A יקרה או רק B יקרה?

15. A ו- B מאורעות זרים. נתון ש:  $2 \cdot P(B \cap \bar{A}) = P(A \cap \bar{B}) = P(\bar{A} \cap \bar{B})$

מה הסיכוי למאורע A ומה ההסתברות למאורע B?

16. קבע אילו מהטענות הבאות נכונות:

א.  $A \cap B = B \cap A$

ב.  $\overline{A \cup B} = A \cap B$

ג.  $A \cap B \cap C = A \cap B \cap (C \cup B)$

ד.  $\overline{A \cap B \cap C} = \bar{A} \cup \bar{B} \cup \bar{C}$

17. נתון ש A ו- B מאורעות במרחב מדגם. נתון ש  $P(A) = 0.3$  ו-  $P(B) = 0.2$

א. האם יתכן ש-  $p(A \cup B) = 0.4$  ?

ב. האם יתכן ש-  $p(A \cup B) = 0.6$  ?

ג. אם A ו- B זרים מה הסיכוי  $p(A \cup B)$  ?

ד. אם A מכיל את B מה הסיכוי  $p(A \cup B)$  ?

18. מתוך אזרחי המדינה הבוגרים ל- 30% חשבון בבנק הפועלים. ל-28% חשבון בבנק לאומי ול-15% חשבון בבנק מזרחי. כמו כן נתון כי 6% מחזיקים חשבון בבנק לאומי ובבנק הפועלים. ל-5% חשבון בבנק פועלים ומזרחי. ול-4% חשבון בבנק לאומי ומזרחי. כמו כן ל-1% מהאוכלוסייה הבוגרת חשבון בנק בשלושת הבנקים יחד.
- א. מה אחוז האזרחים להם חשבון בבנק לאומי בלבד?  
 ב. מה ההסתברות שאזרח כלשהו יחזיק חשבון בבנק פועלים ולאומי אבל לא בבנק מזרחי?  
 ג. מה ההסתברות שלאזרח יהיה חשבון בפועלים או במזרחי אבל לא בבנק לאומי?  
 ד. מה אחוז האזרחים שיש להם חשבון בנק אחד בלבד?  
 ה. מה אחוז האזרחים שיש להם בדיוק חשבון בשני בנקים בלבד?  
 ו. מה ההסתברות שלאזרח בוגר אין חשבון בנק באף אחד מהבנקים הללו?  
 ז. לאיזה אחוז מהאזרחים יש חשבון בנק בלפחות אחד מהבנקים הללו?

19. חברה מסוימת פרסמה את הנתונים הבאים לגבי האזרחים מעל גיל 21. הנתונים שהתקבלו היו: 40% מהאנשים מחזיקים כרטיס "ויזה", 52% מחזיקים כרטיס "ישראכרט", 20% מחזיקים כרטיס "אמריקן אקספרס", 15% מחזיקים כרטיס ויזה וגם ישראכרט, 8% מחזיקים כרטיס ישראכרט וגם אמריקן אקספרס ו-7% מחזיקים כרטיס ויזה וגם אמריקן אקספרס. כמו כן, 13% לא מחזיקים באף אחד משלושת הכרטיסים הנ"ל.
- א. מה אחוז מחזיקי שלושת כרטיס האשראי גם יחד?  
 ב. מה אחוז מחזיקי ישראכרט וויזה אך לא את אמריקן אקספרס?  
 ג. מה אחוז מחזיקי כרטיס אחד בלבד?

20. הוכח:  $P(\bar{A} \cap \bar{B}) = 1 - P(A) - P(B) + P(A \cap B)$

21. A ו-B מאורעות במרחב המדגם האם נכון לומר שהסיכוי שיתרחש בדיוק מאורע אחד הוא:  $P(A) + P(B) - 2P(A \cap B)$

**פתרונות:****שאלה 7**

א. 0.24

ב. 0.04

ג. 0.16

**שאלה 8**

א. 10%

ב. 50%

ג. 50%

**שאלה 9**

א. 0.2

ב. 0.3

ג. 0.3

**שאלה 10**

א. לא.

ב. כן.

ג. כן.

ד. לא.

**שאלה 11**

א. כן

ב. 0.3

**שאלה 12**

התשובה הנכונה ג

**שאלה 13**

א. 0.05

ב. 0.95

**שאלה 14**

א. 0.06

ב. לא זרים

ג. 0.43

**שאלה 18**

א. 0.19

ב. 0.05

ג. 0.31

ד. 0.46

ה. 0.12

ו. 0.41

ז. 0.59

## פרק 18 - הסתברות מותנית - במרחב מדגם אחיד

### רקע:

לעיתים אנו נדרשים לחשב הסתברות למאורע כלשהו כאשר ברשותנו אינפורמציה לגבי מאורע אחר. הסתברות מותנית הינה סיכוי להתרחשות מאורע כלשהו אשר ידוע שמאורע אחר התרחש/ לא התרחש.

ההסתברות של A בהינתן ש-B כבר קרה:

$$P(A|B)$$

$$P(A|B) = \frac{|A \cap B|}{|B|} \quad \text{כשמרחב המדגם אחיד:}$$

למשל, (פתרון בהקלטה)

מטילים קובייה.

נגדיר:

A – התוצאה זוגית.

B – התוצאה גדולה מ-3.

נרצה לחשב את:

$$P(A|B)$$

**תרגילים:**

1. נבחרה ספרה זוגית באקראי. מה הסיכוי שהספרה גדולה מ-6?
2. יוסי הטיל קובייה. מה הסיכוי שקיבל את התוצאה 4 אם ידוע שהתוצאה שהתקבלה זוגית?
3. מטילים צמד קוביות.  
נגדיר:  
A – סכום התוצאות בשתי ההטלות הינו 7  
B – מכפלת התוצאות 12  
חשבו את  $P(A|B)$ .
4. הוטל מטבע פעמיים. ידוע שהתקבל לכל היותר ראש אחד, מה הסיכוי שהתקבלו שני ראשים?
5. אדם הטיל זוג קוביות והתקבל שהתוצאות זהות. מה הסיכוי שלפחות אחת התוצאות 5?
6. אדם הטיל זוג קוביות והתקבל לפחות פעם אחת 4. מה הסיכוי שאחת התוצאות 5?
7. נבחרה משפחה בת שני ילדים. ידוע שאחד הילדים בן. מה ההסתברות שבמשפחה שני בנים בקרב הילדים?
8. נבחרה משפחה בת שלושה ילדים. נתון שהילד האמצעי בן. מה הסיכוי שיש בנות בקרב הילדים?

**השאלות הבאות משלבות קומבינטוריקה:**

9. בכיתה 6 בנים ו-7 בנות. נבחרו ארבעה ילדים מהכיתה.  
אם ידוע שנבחרו 2 בנים ושתי בנות, מה הסיכוי שאלעד לא נבחר?
10. חמישה חברים יצאו לבית קולנוע והתיישבו זה ליד זה באקראי בכיסאות מספר 5 עד 9.  
אם ידוע שערך ודין התיישבו זה ליד זה. מה ההסתברות שהם יושבים בכיסאות מספר 6 ו 7?

פתרונות:שאלה 1

0.2

שאלה 2

1/3

שאלה 3

0.5

שאלה 4

0

שאלה 5

1/6

שאלה 6

2/11

שאלה 7

1/3

שאלה 8

3/4

שאלה 9

2/3

שאלה 10

1/4



## פרק 19 - הסתברות מותנית - מרחב לא אחיד

### רקע:

הסיכוי שמאורע A יתרחש בהינתן ש – מאורע B כבר קרה :

$$P(A | B) = \frac{P(A \cap B)}{P(B)}$$

במונה : הסיכוי לחיתוך של שני המאורעות זה הנשאל וזה הנתון שהתרחש.  
במכנה : הסיכוי למאורע שנתון שהתרחש :

למשל,

נלקחו משפחות שיש להם שתי מכוניות. ל- 30% מהמשפחות הללו המכונית הישנה יותר היא מתוצרת אירופה ואצל 60% מהמשפחות הללו המכונית החדשה יותר מתוצרת אירופה. כמו כן 15% מהמשפחות הללו שתי המכוניות הן מתוצרת אירופאית.  
אם המכונית הישנה של המשפחה היא אירופאית, מה ההסתברות שגם החדשה אירופאית? (פתרון בהקלטה)

## תרגילים:

1. תלמיד ניגש בסמסטר לשני מבחנים מבחן בכלכלה ומבחן בסטטיסטיקה :  
 נגדיר את המאורעות הבאים : A- לעבור את המבחן בסטטיסטיקה. B- לעבור את המבחן בכלכלה.  
 כמו כן נתון שהסיכוי לעבור את המבחן בכלכלה הנו 0.8 והסיכוי לעבור את המבחן בסטטיסטיקה הנו 0.9.  
 הסיכוי לעבור את שני המבחנים הנו 0.75. חשבו את הסיכויים למאורעות הבאים :
- א. התלמיד עבר בסטטיסטיקה , מה ההסתברות שהוא עבר בכלכלה?  
 ב. התלמיד עבר בכלכלה , מה ההסתברות שהוא עבר בסטטיסטיקה?  
 ג. התלמיד עבר בכלכלה , מה ההסתברות שהוא נכשל בסטטיסטיקה?  
 ד. התלמיד נכשל בסטטיסטיקה מה ההסתברות שהוא נכשל בכלכלה?  
 ה. התלמיד עבר לפחות מבחן אחד מה ההסתברות שהוא יעבור את שני המבחנים?
2. במדינה שתי חברות טלפון סלולארי "סופט" ו"בל". 30% מהתושבים הבוגרים רשומים אצל חברת "בל". 60% מהתושבים הבוגרים רשומים אצל חברת "סופט".  
 ל-15% מהתושבים הבוגרים אין טלפון סלולארי בכלל.  
 א. איזה אחוז מהתושבים הבוגרים רשומים אצל שתי החברות?  
 ב. נבחר אדם שרשום אצל חברת "סופט", מה ההסתברות שהוא רשום גם אצל חברת "בל"?  
 ג. אם אדם לא רשום אצל חברת "בל", מה ההסתברות שהוא כן רשום בחברת "סופט"?  
 ד. אם אדם רשום אצל חברה אחת בלבד, מה ההסתברות שהוא רשום בחברת "סופט"?
3. במכללה שני חניונים : חניון קטן וחניון גדול. בשעה 08:00 יש סיכוי של 60% שבחניון הגדול יש מקום, סיכוי של 30% שבחניון הקטן יש מקום וסיכוי של 20% שבשני החניונים יש מקום.  
 א. מה ההסתברות שיש מקום בשעה 08:00 רק בחניון הגדול של המכללה?  
 ב. ידוע שבחניון הקטן יש מקום בשעה 08:00, מה הסיכוי שבחניון הגדול יש מקום?  
 ג. אם בשעה 08:00 בחניון הגדול אין מקום, מה ההסתברות שבחניון הקטן יהיה מקום?  
 ד. נתון שלפחות באחד מהחניונים יש מקום בשעה 08:00, מה ההסתברות שבחניון הגדול יש מקום?
4. נלקחו 200 שכירים ו-100 עצמאים, מתוך השכירים 20 הם אקדמאיים, מתוך העצמאיים 30 הם אקדמאיים.  
 א. בנו טבלת שכיחות משותפת לנתונים.  
 ב. נבחר אדם אקראי מהי ההסתברות שהוא שכיר?  
 ג. מה ההסתברות שהוא שכיר ולא אקדמאי?  
 ד. מה ההסתברות שהוא שכיר או אקדמאי?  
 ה. אם האדם שנבחר הוא עצמאי מהי ההסתברות שהוא אקדמאי?  
 ו. אם הבן אדם שנבחר הוא לא אקדמאי, מה ההסתברות שהוא שכיר?

5. חברה מסוימת פרסמה את הנתונים הבאים לגבי האזרחים מעל גיל 21 :  
הנתונים שהתקבלו היו : 40% מהאנשים מחזיקים כרטיס "ויזה", 52% מחזיקים כרטיס "ישראלכרט",  
20% מחזיקים כרטיס "אמריקן אקספרס", 15% מחזיקים כרטיס ויזה וגם ישראלכרט, 8% מחזיקים  
כרטיס ישראלכרט וגם אמריקן אקספרס ו-7% מחזיקים כרטיס ויזה וגם אמריקן אקספרס. כמו כן,  
5% מחזיקים בכל שלושת הכרטיסים הנ"ל.

א. אם לאדם יש ויזה, מה הסיכוי שאין לו כרטיס ישראלכרט?

ב. אם לאדם שני כרטיסי אשראי, מה הסיכוי שאין לו כרטיס ישראלכרט?

ג. אם לאדם לפחות כרטיס אשראי אחד, מה הסיכוי שאין לו כרטיס ישראלכרט?

**פתרונות:****שאלה 1**

- א. 0.833
- ב. 0.9375
- ג. 0.0625
- ד. 0.5
- ה. 0.789

**שאלה 2**

- א. 5%
- ב. 0.0833
- ג. 0.786
- ד. 0.6875

**שאלה 3**

- א. 0.4
- ב.  $\frac{2}{3}$
- ג. 0.25
- ד.  $\frac{6}{7}$

## פרק 20 - דיאגרמת עצים, נוסחת בייס ונוסחת ההסתברות השלמה

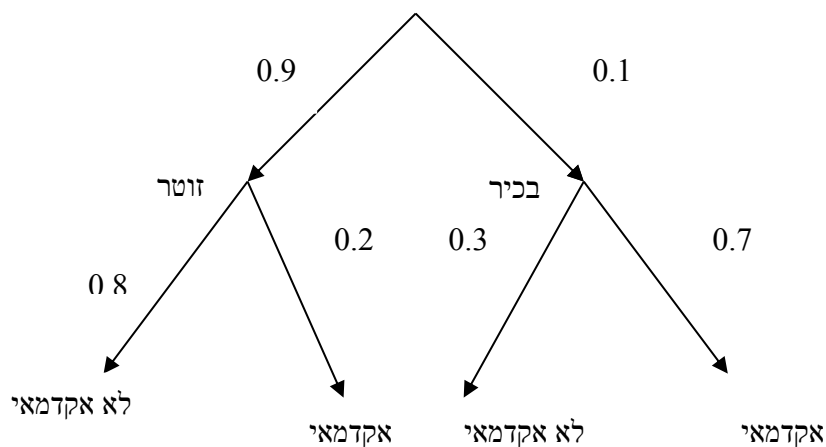
### רקע:

נשתמש בשיטה זו כאשר יש תרגיל שבו התרחשות המאורעות היא בשלבים, כך שכל תוצאה של כל שלב תלויה בשלב הקודם, פרט לשלב הראשון:

למשל,

בחברה מסוימת 10% מוגדרים בכירים והיתר מוגדרים זוטרים.  
מבין הבכירים 70% הם אקדמאים ומבין הזוטרים 20% הם אקדמאים.

נשרטט עץ שיתאר את הנתונים, השלב הראשון של העץ אינו מותנה בכלום ואילו השלב השני מותנה בשלב הראשון.



כדי לקבל את הסיכוי לענף מסוים נכפיל את כל ההסתברויות על אותו ענף.  
נבחר אדם באקראי מאותה חברה.

א. מה הסיכוי שהוא בכיר אקדמאי ?

$$0.1 * 0.7 = 0.07$$

ב. מה הסיכוי שהוא זוטר לא אקדמאי ?

$$0.9 * 0.8 = 0.72$$

כדי לקבל את הסיכוי לכמה ענפים נחבר את הסיכויים של כל ענף ( רק אחרי שבתוך הענף הכפלנו את ההסתברויות )

ג. מה הסיכוי שהוא אקדמאי ?

$$0.1*0.7+0.9*0.2=0.25$$

ד. נבחר אקדמאי מה ההסתברות שהוא עובד זוטרי?

מדובר כאן על שאלה בהסתברות מותנה ולכן נשתמש בעיקרון של הסתברות מותנה

$$P(zutar | academay) = \frac{0.9*0.2}{0.25} = \frac{0.18}{0.25} = 0.72$$

### נוסחת ההסתברות השלמה

B מאורע כלשהו,  $A_1, \dots, A_n$  חלוקה ממצה של  $\Omega$ .

$$P(B) = \sum_{i=1}^n P(A_i) \cdot P(B/A_i) \quad \text{אזי:}$$

### נוסחת בייס

$$P(A_j/B) = \frac{P(A_j)P(B/A_j)}{\sum_{i=1}^n P(A_i) \cdot P(B/A_i)}$$

### תרגילים:

1. בשקית סוכריות 4 סוכריות תות ו-3 לימון . מוציאים באקראי סוכרייה אם היא בטעם תות אוכלים אותה ומוציאים סוכרייה נוספת , אך אם היא בטעם לימון מחזירים אותה לשקית ומוציאים סוכרייה נוספת.
  - א. מה ההסתברות שהסוכרייה הראשונה שהוצאה בטעם תות והשנייה בטעם לימון ?
  - ב. מה ההסתברות שהסוכרייה השנייה בטעם לימון?
  
2. באוכלוסיה מסוימת 30% הם ילדים, 50% בוגרים והיתר קשישים. לפי נתוני משרד הבריאות הסיכוי שילד יחלה בשפעת במשך החורף הוא 80%, הסיכוי שמבוגר יחלה בשפעת במשך החורף הוא 40% והסיכוי שקשיש יחלה בשפעת במשך החורף הוא 70%.
  - א. איזה אחוז מהאוכלוסייה הינו קשישים שלא יחלו בשפעת במשך החורף?
  - ב. מה אחוז האנשים שיחלו בשפעת במשך החורף?
  - ג. נבחר אדם שחלה במשך החורף בשפעת, מה ההסתברות שהוא קשיש?
  - ד. נבחר ילד, מה ההסתברות שהוא לא יחלה בשפעת במשך החורף?
  
3. בכד א' 5 כדורים כחולים ו-5 כדורים אדומים. בכד ב' 6 כדורים כחולים ו-4 כדורים אדומים. בוחרים באקראי כד, מוציאים ממנו כדור ומבלי להחזירו מוציאים כדור נוסף.
  - א. מה ההסתברות ששני הכדורים שיוצאו יהיו בצבעים שונים?
  - ב. אם הכדורים שהוצאו הם בצבעים שונים, מה ההסתברות שהכדור השני שהוצא יהיה בצבע אדום?
  
4. חברת סלולר מסווגת את לקוחותיה לפי 3 קבוצות גיל: נוער, בוגרים ופנסיונרים. נתון כי :
  - 10% מהלקוחות בני נוער, 70% מהלקוחות בוגרים והיתר פנסיונרים. מתוך בני הנוער 90% מחזיקים בסמארט-פון, מתוך האוכלוסייה הבוגרת ל 70% יש סמארט-פון ומתוך אוכלוסיית הפנסיונרים 30% מחזיקים בסמארט-פון.
  - א. איזה אחוז מלקוחות החברה הם בני נוער עם סמארט-פון?
  - ב. נבחר לקוח אקראי ונתון שיש לו סמארט-פון. מה ההסתברות שהוא פנסיונר?
  - ג. אם ללקוח אין סמארט-פון, מה ההסתברות שהוא לא בן נוער?

5. כדי להתקבל למקום עבודה יש לעבור שלושה מבחנים. המבחנים הם בשלבים, כלומר אם נכשלתם במבחן מסוים אינכם ניגשים למבחן הבא אחריו.

70% מהמועמדים עוברים את המבחן הראשון.

מתוכם 50% עוברים את המבחן השני.

מבין אלה שעוברים את המבחן השני 40% עוברים את המבחן השלישי.

א. מה ההסתברות להתקבל לעבודה?

ב. מועמד לא התקבל לעבודה. מה ההסתברות שהוא נכשל במבחן הראשון?

ג. מועמד לא התקבל לעבודה. מה ההסתברות שהוא עבר את המבחן השני?

6. משרד הבריאות פרסם את הנתונים הבאים:

מתוך אוכלוסיית הילדים והנוער 80% חולים בשפעת בזמן החורף.

מתוך אוכלוסיית המבוגרים (עד גיל 65) 60% חולים בשפעת בזמן החורף.

30% מהתושבים הם ילדים ונוער.

50% הם מבוגרים.

היתר קשישים.

כמו כן נתון ש 68% מהאוכלוסייה תחלה בשפעת בחורף.

א. מה אחוז החולים בשפעת בקרב האוכלוסייה הקשישה?

ב. נבחר אדם שלא חלה בשפעת, מה ההסתברות שהוא לא קשיש?

7. רדאר שנמצא על החוף צריך לקלוט אנייה הנמצאת ב-1 מ-4 האזורים: A B C D.

אם האנייה נמצאת באזור A הרדאר מזהה אותה בסיכוי 0.8, סיכוי זה פוחת ב-0.1 ככל שהאנייה מתקדמת באזור.

כמו כן נתון שבהסתברות חצי האנייה נמצאת באזור D, בהסתברות 0.3 באזור C, באזור B היא נמצאת בסיכוי 0.2, אחרת היא נמצאת באזור A.

א. מה הסיכוי ש האנייה תתגלה ע"י הרדאר?

ב. אם האנייה התגלתה ע"י הרדאר, מה ההסתברות שהיא נמצאת באזור C?

ג. אם האנייה התגלתה ע"י הרדאר, מה הסיכוי שהיא לא נמצאת באזור B?



8. סימפטום X מופיע בהסתברות של 0.4 במחלה A, בהסתברות של 0.6 במחלה B ובהסתברות של 0.5 במחלה C.

סימפטום X מופיע אך ורק במחלות הללו, אדם לא יכול לחלות ביותר ממחלה אחת מבין המחלות הללו.

לקליניקה מגיעים אנשים כדלקמן:

8% חולים במחלה A, 10% במחלה B, 2% במחלה C והיתר בריאים. כמו כן נתון שבמחלה A, סימפטום X מתגלה בסיכוי של 80%. במחלות B, C הסימפטום מתגלה בסיכוי של 90% בכל מחלה.

א. מה ההסתברות שאדם הגיע לקליניקה וגילו אצלו את סימפטום X?

ב. אם התגלה אצל אדם סימפטום X, מה ההסתברות שהוא חולה במחלה A?

ג. אם לאדם יש את סימפטום X, מה ההסתברות שהוא חולה במחלה A?

ד. אם לא גילו אצל אדם את סימפטום X, מה ההסתברות שהוא בריא?

9. סטודנט ניגש למבחן אמריקאי. הסיכוי שהוא יודע לשאלה מסוימת את התשובה הוא  $p$ , אם הוא לא יודע את התשובה הוא מנחש. בכל מקרה הוא עונה על השאלה.

נתון שלשאלה יש  $k$  תשובות אפשריות.

אם הסטודנט ענה נכון על השאלה, מה הסיכוי שהוא ידע אותה?

10. אדם משחק נגד שני מתמודדים, רונית ודולב. האדם צריך לשחק שלושה משחקים ויש לו לבחור איזה סדר משחקים עדיף לו:

א. דולב, רונית, דולב

ב. רונית, דולב, רונית

בכל משחק מישהו חייב לנצח (אין תיקו). האדם ינצח בטורניר רק אם ינצח בשני משחקים ברציפות. נתון שדולב שחקן טוב יותר מאשר רונית. איזו אפשרות עדיפה יותר על האדם כדי לנצח בטורניר?

פתרונות:שאלה 1א.  $2/7$ ב.  $23/49$ שאלה 2

א. 6%

ב. 58%

ג. 0.241

ד. 0.2

שאלה 3

א. 0.544

ב. 0.5

שאלה 4

א. 9%

ב. 0.09375

ג. 0.9722

שאלה 8

א. 0.0886

ב. 0.2889

ג. 0.3137

ד. 0.8778

שאלה 9

$$\frac{kp}{1 + (k-1)p}$$

שאלה 10

אפשרות א עדיפה

## פרק 21 - תלות ואי תלות בין מאורעות

### רקע:

אם מתקיים ש:  $P(B|A) = p(B)$  נגיד שמאורע B בלתי תלוי ב-A.

הדבר גורר גם ההפך:  $P(A|B) = p(A)$  כלומר A אינו תלוי גם ב-B.

כשהמאורעות בלתי תלויים מתקיים ש:  $P(A \cap B) = P(A) \cdot P(B)$ .

הוכחה לכך:

$$P(A|B) = P(A) = \frac{P(A \cap B)}{P(B)} \Rightarrow P(A \cap B) = P(A) \cdot P(B)$$

נשתמש בנוסחאות של מאורעות בלתי תלויים רק אם נאמר במפורש שהמאורעות בלתי תלויים

בתרגיל או שמההקשר אפשר להבין ללא צל של ספק שהמאורעות בלתי תלויים.

למשל,

חוקר מבצע שני ניסויים בלתי תלויים הסיכוי להצליח בניסוי הראשון הנו 0.7 והסיכוי להצליח

בניסוי השני הוא 0.4.

א. מה הסיכוי להצליח בשני הניסויים יחדו?

כיוון שהמאורעות הללו בלתי תלויים:

$$p(A \cap B) = P(A) \cdot P(B) = 0.7 \cdot 0.4 = 0.28$$

ב. מה הסיכוי להיכשל בשני הניסויים?

$$p(\bar{A} \cap \bar{B}) = P(\bar{A}) \cdot P(\bar{B}) = (1-0.7)(1-0.4) = 0.18$$

באופן דומה:

### הרחבה: אי תלות בין n מאורעות

n מאורעות  $A_1, \dots, A_n$  הם בלתי תלויים אם ורק אם:

$$P\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n P(A_i)$$

**תרגילים:**

1. נתון:

$$p(A) = 0.2$$

$$P(B) = 0.5$$

$$P(A \cup B) = 0.6$$

האם המאורעות הללו בלתי תלויים?

2. תלמיד ניגש לשני מבחנים שהצלחתם לא תלויה זו בזו. הסיכוי שלו להצליח במבחן הראשון הוא 0.7 והשני 0.4 .

א. מה הסיכוי להצליח בשני המבחנים יחדו?

ב. מה הסיכוי שניכשל בשני המבחנים ?

3. במדינה מסוימת 8% אבטלה, נבחרו באקראי שני אנשים מהמדינה.

א. מה ההסתברות ששניהם מובטלים?

ב. מה ההסתברות שלפחות אחד מהם מובטל?

4. מוצר צריך לעבור בהצלחה ארבע בדיקות בלתי תלויות לפני שיווקו, אחרת הוא נפסל ולא יוצא לשוק. הסיכוי לעבור בהצלחה כל אחת מהבדיקות הוא 0.8. בכל מקרה מבוצעות כל 4 הבדיקות.

א. מה הסיכוי שהמוצר יפסל?

ב. מה ההסתברות שהמוצר יעבור בהצלחה לפחות בדיקה אחת?

5. מדינה מסוימת 8% אבטלה, נבחרו באקראי חמישה אנשים מהמדינה.

א. מה ההסתברות שכולם מובטלים במדגם?

ב. מה ההסתברות שלפחות אחד מהם מובטל?

6. עבור שני מאורעות A ו-B המוגדרים על אותו מרחב מדגם נתון ש:  $P(A \cup B) = 0.9$ ,  $P(A|B) = 0.6$ ,  $P(A \cap \bar{B}) = 0.3$ . האם A ו-B מאורעות בלתי תלויים?

7. הוכח אם

$$P(A/B) = P(B/A)$$

אז:

$$P(A) = P(B)$$

8. קבע אילו מהטענות הבאות נכונות. נמק!

- א. אם  $p(A \cup B) = p(A) + p(B)$  אזי המאורעות בלתי תלויים.
- ב. מאורע A כלול במאורע B.  $P(A) > 0$ ,  $0 < p(B) < 1$ , לכן  $p(A/B) < p(A)$ .
- ג. A ו-B מאורעות זרים שסיכוייהם חיובים לכן הם מאורעות תלויים.
- ד. A ו-B מאורעות תלויים שסיכוייהם חיובים לכן A ו-B מאורעות זרים.
- ה.  $P(\bar{A} \cap \bar{B}) = 1 - P(A) - P(B)$  לכן A ו-B מאורעות זרים.

פתרונות :שאלה 1

כן

שאלה 2

א. 0.28

ב. 0.18

שאלה 3

א. 0.0064

ב. 0.1536

שאלה 4

א. 0.5904

ב. 0.9984

שאלה 8

א. לא נכון

ב. לא נכון

ג. נכון

ד. לא נכון

ה. נכון

## פרק 22 - שאלות מסכמות בהסתברות

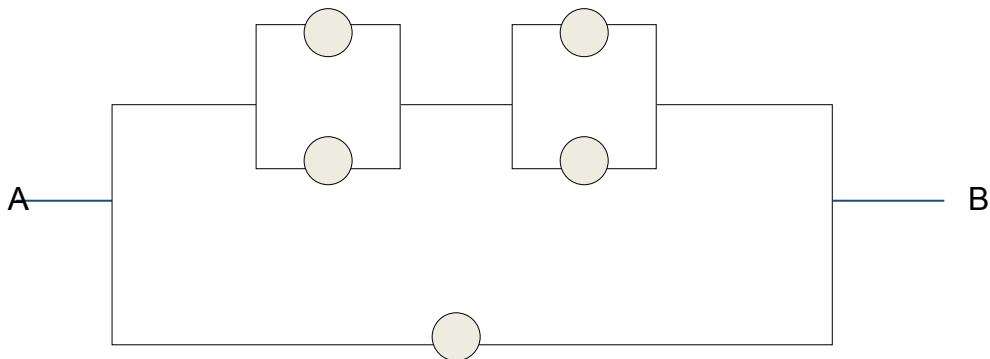
1. נלקחו משפחות שיש להם שתי מכוניות. ל- 30% מהמשפחות הללו המכונית הישנה יותר היא מתוצרת אירופה ואצל 60% מהמשפחות הללו המכונית החדשה יותר מתוצרת אירופה. כמו כן 15% מהמשפחות הללו שתי המכוניות הן מתוצרת אירופאית.
- א. מה ההסתברות שמשפחה אקראית בת שתי מכוניות תהיה ללא מכוניות מתוצרת אירופה?  
 ב. מה ההסתברות שלפחות מכונית אחת תהיה אירופאית?  
 ג. ידוע שלמשפחה יש מכונית אירופאית. מה ההסתברות שרק המכונית החדשה שלה היא מתוצרת אירופאית?  
 ד. אם המכונית הישנה של המשפחה היא אירופאית, מה ההסתברות שגם החדשה אירופאית?
2. במדינת "שומקום" 50% מהחלב במרכולים מיוצר במחלבה א' 40% במחלבה ב' והיתר במחלבה ג'. 3% מתוצרת מחלבה א' מגיעה חמוצה למרכולים ואילו במחלבה ב' 10%. כמו כן ידוע שבמדינת "שומקום" בסך הכול 7.5% מהחלב חמוץ.
- א. איזה אחוז מהחלב שמגיע למרכול ממחלבה ג' חמוץ?  
 ב. אם נרכש חלב חמוץ במרכול. מה הסיכוי שהוא יוצר במחלבה ג?  
 ג. ברכישת חלב נימצא שהוא אינו חמוץ. מה הסיכוי שהוא יוצר במחלבה א?  
 ד. האם המאורעות: "חלב חמוץ" ו-"יוצר במחלבה א" בלתי תלויים?
3. רוני ורונה יצאו לבלות במרכז בילויים עם מספר אפשרויות בילוי:  
 בהסתברות של 0.3 הם ייצאו לבאולינג  
 בהסתברות של 0.5 הם ייצאו לבית קפה  
 בהסתברות של 0.7 הם יצאו לפחות לאחד מהם, באולינג/קפה.
- א. מה ההסתברות שהם יצאו רק לבאולינג?  
 ב. האם המאורעות "לצאת לבאולינג" לצאת לבית קפה" זרים?  
 ג. האם המאורעות "לצאת לבאולינג" לצאת לבית קפה" תלויים?  
 ד. מה ההסתברות שיום אחד הם יצאו רק לבאולינג וביום למחרת לא יצאו לאף אחד מהמקומות?

4. 70% מהנבחנים בסטטיסטיקה עוברים את מועד א'. כל מי שלא עובר את מועד א' ניגש לעשות מועד ב', מתוכם 80% עוברים אותו. מבין אלה שנכשלים בשני המועדים 50% נרשמים לקורס מחדש, והיתר פורשים מהתואר.
- א. מה הסיכוי שסטודנט אקראי עבר את הקורס?
- ב. אם סטודנט אקראי עבר הקורס, מה הסיכוי שעבר במועד ב'?
- ג. מה אחוז הסטודנטים שפורשים מהתואר?
- ד. נבחרו 2 סטודנטים אקראיים רונית וינאי, מה ההסתברות שרונית עברה במועד א' ושינאי עבר במועד ב'?
5. באוכלוסייה מסוימת 40% הם גברים והיתר הן נשים. מבין הגברים 10% מובטלים. בסך הכול 13% מהאוכלוסייה מובטלת.
- א. מה אחוז האבטלה בקרב הנשים?
- ב. נבחר אדם מובטל, מה ההסתברות שזו אישה?
- ג. נגדיר את המאורעות הבאים:
- A - נבחר אדם מובטל
- B - נבחר גבר
- האם המאורעות הללו זרים? והאם הם בלתי תלויים?
6. בתיבה 10 מטבעות, מתוכם 7 מטבעות רגילים (ראש, זנב) ו-3 מטבעות שבשני צדדיהם טבוע ראש. אדם בוחר באקראי מטבע ומטיל אותו פעמיים. נסמן ב-A את ההטלה הראשונה ראש, ב-B את ההטלה השנייה ראש.
- א. חשבו את הסיכויים למאורעות A ו-B.
- ב. האם המאורע A ו-B בלתי תלויים?
- ג. ידוע שבהטלה הראשונה התקבל ראש, מה ההסתברות שהמטבע שהוטל הוא מטבע הוגן?



7. ערן מעוניין למכור את רכבו, הוא מפרסם מודעה באינטרנט ומודעה בעיתון. מבין אלה שמעוניינים לרכוש רכב משומש 30% יראו את המודעה באינטרנט, 50% יראו את המודעה בעיתון ו-72% יראו את המודעה בלפחות אחת מהמדיות.
- א. מה אחוז האנשים מאלה שמעוניינים לרכוש רכב משומש יראו את 2 המודעות?  
 ב. אם אדם ראה את המודעה באינטרנט, מה ההסתברות שהוא לא ראה את המודעה בעיתון?  
 ג. האם המאורעות: "לראות את המודעה באינטרנט" ו"לראות את המודעה בעיתון" בלתי תלויים?  
 ד. אדם שראה את המודעה באינטרנט בלבד יתקשר לערן בהסתברות של 0.7, אם הוא ראה את המודעה בעיתון בלבד הוא יתקשר לערן בהסתברות של 0.6. ואם הוא ראה את שתי המודעות הוא יתקשר לערן בהסתברות של 0.9.
1. מה ההסתברות שאדם המעוניין לרכוש רכב משומש יתקשר לערן?  
 2. אדם המעוניין לרכוש רכב משומש התקשר לערן. מה ההסתברות שהוא ראה את שתי המודעות?

8. נתונה המערכת החשמלית הבאה :



- כל יחידה עובדת באופן בלתי תלוי ובהסתברות  $p$ .  
 כדי שהמערכת תפעל צריך לעבור זרם מהנקודה A לנקודה B.  
 הוכח שהסיכוי שהמערכת תפעל:

$$P + (1 - P)(2P - P^2)^2$$

**פתרונות:****שאלה 1**

א. 0.25

ב. 0.75

ג. 0.6

ד. 0.5

**שאלה 2**

א. 0.2

ב. 0.267

ג. 0.524

ד. המאורעות תלויים.

**שאלה 3**

א. 0.2

ב. המאורעות אינם זרים.

ג. המאורעות הללו תלויים.

ד. 0.06

**שאלה 4**

א. 0.94

ב. 0.255

ג. 0.03

ד. 0.168

**שאלה 5**

א. 15%

ב. 0.692

ג. לא זרים ותלויים.

**שאלה 6**

א. 0.65

ב. A ו-B תלויים.

ג. 0.5384

**שאלה 7**

א. 8%

ב. 0.733

ג. תלויים.

ד. 1. 0.478

2. 0.15

## פרק 23 - התפלגויות בדידות מיוחדות - התפלגות בינומית

### רקע:

נגדיר את המושג ניסוי ברנולי:  
 ניסוי ברנולי הנו ניסוי שיש לו שתי תוצאות אפשריות: "הצלחה" ו"כישלון" כמו: מוצר פגום או תקין אדם עובד או מובטל עץ או פלי בהטלת מטבע וכדומה.

בהתפלגות בינומית חוזרים על אותו ניסוי ברנולי  $n$  פעמים באופן בלתי תלוי זה בזה. מגדירים את  $X$  להיות מספר ההצלחות שהתקבלו בסך הכול. נסמן ב  $p$  את הסיכוי להצלחה בניסוי בודד וב  $q$  את הסיכוי לכישלון בניסוי בודד.

ואז נגיד ש:  $X \sim B(n, p)$ .

**פונקציית ההסתברות של  $X$** :

$$k = 0, 1, 2, \dots, n; P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{לכל}$$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}; \quad n! = n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot 1; \quad 0! = 1; \quad \text{כאשר}$$

לגודל  $\binom{n}{k}$ : ניתן לחשב באמצעות המחשבון.

$$E(X) = np \quad \text{תוחלת}$$

$$V(X) = npq \quad \text{שונות}$$

שימו לב כדי לזהות שמדובר בהתפלגות בינומית צריכים להתקיים כל התנאים הבאים:

- (1) חוזרים על אותו ניסוי ברנולי באופן בלתי תלוי זה בזה.
- (2) חוזרים על הניסוי  $n$  פעמים.
- (3)  $X$  – מוגדר כמספר ההצלחות המתקבלות בסך הכול.

**דוגמה**: (פתרון בהקלטה)

- במדינה מסוימת ל-80% מהתושבים יש רישיון נהיגה. נבחרו 10 תושבים אקראיים מהמדינה.
- א. מהי ההסתברות שבדיוק ל-9 מהם יש רישיון נהיגה?
- ב. מה ההסתברות שלפחות ל-9 מהם יש רישיון נהיגה?
- ג. מהי התוחלת ומהי סטיית התקן של מספר התושבים שנדגמו ושיש להם רישיון נהיגה?

**תרגילים:**

1. במדינה 10% מהאוכלוסייה מובטלת. נבחרו 5 אנשים באקראי מאותה אוכלוסייה.

נגדיר את  $X$  להיות מספר המובטלים שהתקבלו במדגם.

א. מהי ההתפלגות של  $X$ ?

ב. מה ההסתברות שיהיה בדיוק מובטל אחד?

ג. מה ההסתברות שכולם יעבדו במדגם?

ד. מה ההסתברות ששלושה יעבדו במדגם?

ה. מה ההסתברות שלפחות אחד יהיה מובטל?

ו. מה תוחלת ומהי השונות של מספר המובטלים במדגם?

2. על פי נתוני משרד התקשורת ל-70% מהאוכלוסייה יש סמארט-פון. נבחרו 10 אנשים

באקראי. נגדיר את  $X$  כמספר האנשים שנדגמו עם סמארט-פון.

א. מהי ההתפלגות של  $X$ ? הסבירו.

ב. מה ההסתברות שבמדגם ל-8 אנשים יש סמארט-פון?

ג. מה ההסתברות שבמדגם לפחות ל-9 יהיו סמארט-פון?

ד. מה התוחלת ומה סטיית התקן של מספר האנשים שנדגמו ולהם סמארט-פון?

3. בבית הימורים יש שורה של 6 מכונות מזל מאותו סוג. משחק במכונת מזל כזו עולה 5 ₪.

ההסתברות לזכות ב-20 ₪, בכל אחת מהמכונות היא 0.1 וההסתברות להפסיד את ההשקעה

היא 0.9 בכל מכונה. מהמר נכנס לבית ההימורים ומכניס 5 ₪ לכל אחת מ-6 המכונות.

א. מה ההסתברות שיפסיד בכל המכונות?

ב. מה ההסתברות שיזכה בדיוק בשתי מכונות?

ג. מה ההסתברות שיזכה ביותר כסף מה-30 ₪ שהשקיע?

ד. מהן התוחלת וסטיית התקן של הרווח נטו של המהמר (הזכיות בניכוי ההשקעה)?

4. במדינה מסוימת התפלגות ההשכלה בקרב האוכלוסייה מעל גיל 30 היא כזו:

השכלה	נמוכה	תיכונית	תואר I	תואר II ומעלה
פרופורציה	0.1	0.6	0.2	0.1

נבחרו 20 אנשים אקראיים מעל גיל 30 מהמדינה הנ"ל.

א. מה ההסתברות ש-5 מהם אקדמאים?

ב. מה התוחלת של מס' בעלי ההשכלה הנמוכה?

5. במכללה מסוימת 20% מהסטודנטים גרים בת"א. מבין הסטודנטים שגרים בת"א 30% מגיעים ברכבם ומבין הסטודנטים שלא גרים בת"א 50% מגיעים ברכבם למכללה.
- א. השומר בשער המכללה בודק לכל סטודנט את תיקו בהיכנסו למכללה. מה ההסתברות שבקרב 5 סטודנטים שנבדקו ע"י השומר רק 1 מתוכם הגיע למכללה ברכבו?
- ב. בהמשך לסעיף הקודם מה ההסתברות שרוב הסטודנטים בקרב ה-5 הגיעו למכללה ברכבם?
6. במבחן אמריקאי 20 שאלות. סטודנט ניגש למבחן והסיכוי שהוא יודע שאלה היא 0.8. אם הוא לא יודע הוא מנחש את התשובה. לכל שאלה 4 תשובות אפשריות שרק אחת מהן נכונה.
- א. מה הסיכוי לענות על שאלה מסוימת נכון?
- ב. מה הסיכוי שיענה נכונה על בדיוק 16 שאלות?
- ג. על כל שאלה שענה נכון התלמיד מקבל 5 נקודות, על כל שאלה ששגה מופחתת נקודה, מה התוחלת ומהי השונות של ציון התלמיד?
7. 5% מקו היצור פגום. המוצרים נארזים בתוך קופסת קרטון. בכל קופסא 10 מוצרים שונים. הקופסאות נארזות בתוך מכולה. בכל מכולה 20 קופסאות.
- א. מה ההסתברות שבקופסא אקראית לפחות מוצר פגום אחד?
- ב. מה התוחלת ומהי סטיית התקן של מספר הקופסאות במכולה בהן לפחות מוצר פגום אחד?
8. מטילים מטבע הוגן 5 פעמים. נגדיר את  $X$  – מספר הפעמים שהתקבל עץ. חשבו את  $E(x^2)$ .

**פתרונות :****שאלה 7 :**

- א. 0.401  
 ב. תוחלת : 8.025  
 סטיית תקן : 2.193

**שאלה 2 :**

- ב. 0.2335  
 ג. 0.1493  
 ד. תוחלת : 7  
 סטיית תקן : 1.449

**שאלה 8 :**

7.5

**שאלה 3 :**

- א. 0.5314  
 ב. 0.0984  
 ג. 0.1143  
 ד. תוחלת : -18  
 סטיית תקן : 14.697

**שאלה 4 :**

- א. 0.1789  
 ב. 2

**שאלה 5 :**

- א. 0.1956  
 ב. 0.4253

**שאלה 6 :**

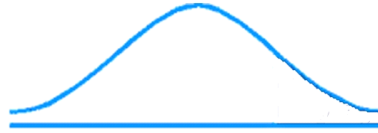
- א. 0.85  
 ב. 0.182  
 ג. תוחלת : 82 נקודות  
 שונות : 91.8 נקודות

## פרק 24 - התפלגות נורמלית

### רקע

התפלגות זו מאפיינת משתנה רציף כמו גובה של אוכלוסייה, משקל של אוכלוסייה, זמן ייצור וכדומה.

בהתפלגות זו רוב האוכלוסייה נמצאת במרכז של ההתפלגות וככל שמתרחקים מהמרכז אחוז המקרים הולך וקטן באופן סימטרי. צורת העקומה של ההתפלגות הנורמלית היא זו:



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

אם נתון, למשל, שציוני פסיכומטרי מתפלגים נורמלית עם ממוצע 500 וסטיית תקן 100, נרשום

$$\mu = 500$$

$$\sigma = 100$$

נתון זה בצורה הבאה:

$$X \sim N(500, 100^2)$$

ובהכללה, נרשום זאת כך:  $X \sim N(\mu, \sigma^2)$

השטח שמתחת לעקומה יבטא את ההסתברות או הפרופורציה של המקרים המבוקשים.

כדי לחשב שטחים אלה נמיר כל התפלגות נורמלית כלשהי להתפלגות נורמלית סטנדרטית.

התפלגות נורמלית סטנדרטית היא התפלגות נורמלית שהממוצע שלה הוא אפס וסטיית התקן

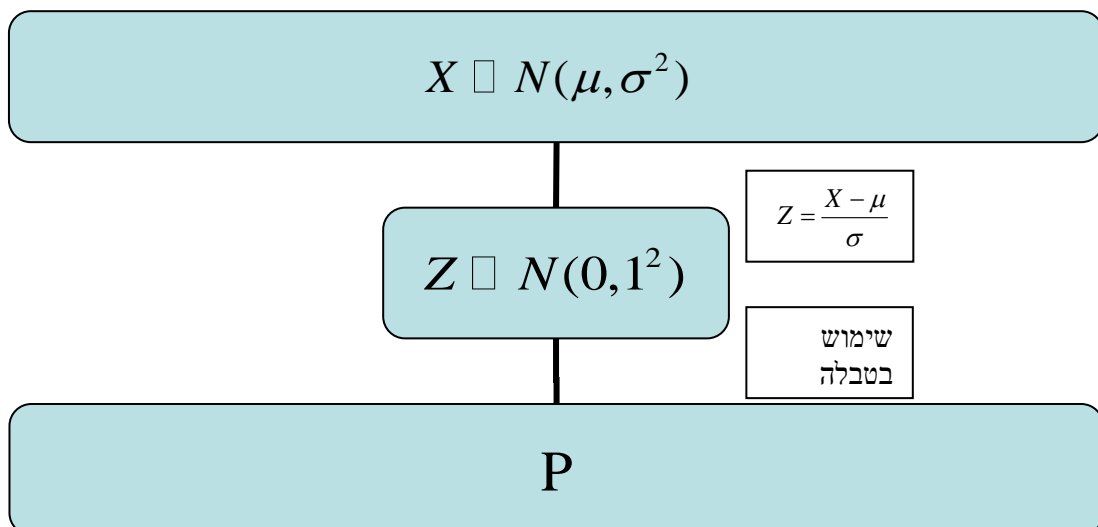
היא אחת והיא תסומן באות  $Z$ .  $Z \sim N(0, 1^2)$

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

הערך המתקבל הוא ציון התקן הנותן בכמה סטיות תקן אנו סוטים מהממוצע.

לאחר שמצאנו את ציון התקן נעזר בטבלה שבסופו של דבר תיתן לנו את ההסתברות הרצויה.

באופן כללי נתאר את הסכמה הבאה לתהליך:





**תרגילים:**

1. הגובה של אנשים באוכלוסייה מסוימת מתפלג נורמלית עם ממוצע של 170 ס"מ וסטית תקן של 10 ס"מ .
- מה אחוז האנשים שגובהם מתחת ל- 182.4 ס"מ?
  - מה אחוז האנשים שגובהם מעל 190 ס"מ?
  - מה אחוז האנשים שגובהם בדיוק 173.6 ס"מ?
  - מה אחוז האנשים שגובהם מתחת ל- 170 ס"מ?
  - מה אחוז האנשים שגובהם לכל היותר 170 ס"מ?
2. 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 פרופורציית האנשים באוכלוסייה שמשקלם לפחות 50 ק"ג?
  - מהי השכיחות היחסית של האנשים באוכלוסייה שמשקלם בין 60 ל- 70 ק"ג?
  - לאיזה חלק מהאוכלוסייה משקל הסוטה מהמשקל הממוצע בלא יותר מ- 4 ק"ג?
  - מה אחוז האנשים באוכלוסייה הזו ששוקלים מתחת ל – 140 ק"ג?
3. ציוני מבחן אינטליגנציה מתפלג נורמלית עם ממוצע 100 ושוונות 225 .
- מה העשירון העליון של הציונים במבחן האינטליגנציה?
  - מה העשירון התחתון של ההתפלגות?
  - מהו הציון ש- 20% מהנבחנים מקבלים מעליו?
  - מהו האחוזון ה- 20?
  - מה הרבעון התחתון?
4. אורך חים של מכשיר מתפלג נורמלית . ידוע שמחצית מהמכשירים חים פחות מ- 500 שעות , כמו כן ידוע ש- 67% מהמכשירים חים פחות מ- 544 שעות.
- מהו ממוצע אורך חיי מכשיר?
  - מהי סטיית בתקן של אורך חיי מכשיר?
  - מה הסיכוי שמכשיר אקראי יחיה פחות מ- 460 שעות?
  - מהו המאיון העליון של אורח חיי מכשיר?
  - 1% מהמכשירים בעלי אורך החיים הקצר ביותר נשלח למעבדה לבדיקה מעמיקה. מהו אורך החיים המקסימלי לשליחת מכשיר למעבדה?

5. הזמן שלוקח לאדם להגיע לעבודתו מתפלג נורמלית עם ממוצע של 40 דקות וסטיית תקן של 5 דקות.
- א. מה ההסתברות שמשך הנסיעה של האדם לעבודתו יהיה לפחות שלושת רבעי השעה?
- ב. אדם יצא לעבודתו בשעה 08:10 מביתו. הוא צריך להגיע לעבודתו בשעה 09:00. מה הסיכוי שיאחר לעבודתו?
6. אורך שיר אקראי המשודר ברדיו מתפלג נורמלית עם תוחלת של 3.5 דקות וסטיית תקן של שלושים שניות.
- ד. מה ההסתברות שאורך של שיר אקראי המנוגן ברדיו יהיה בין 3 ל 2.5 דקות?
- ה. מהו הטווח הבין רבעוני של אורך שיר המשודר ברדיו?
-

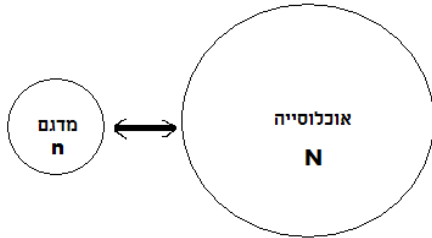
**פתרונות:**

<u>שאלה 1</u>		<u>שאלה 2</u>	
א. 89.25%		א. 26.43%	
ב. 2.28%		ב. 89.44%	
ג. 0		ג. 39.44%	
		ד. 0.383	
ד. 50%		ה. 100%	
<u>שאלה 3</u>		<u>שאלה 4</u>	
א. 119.2		א. 500	
ב. 80.8		ב. 100	
ג. 112.6		ג. 0.3446	
ד. 87.4		ד. 733	
ה. 89.95		ה. 267	
<u>שאלה 5</u>		<u>שאלה 6</u>	
א. 0.1587		א. 0.1359	
ב. 0.0228		ב. 0.67	
ג.			

## פרק 25 - התפלגות הדגימה

### התפלגות ממוצע המדגם ומשפט הגבול המרכזי

#### הקדמה כללית :



**אוכלוסייה** – קבוצה שאליה מפנים שאלה מחקרית. למשל, חברת תרופות שמעוניינת לפתח תרופה למחלת הסכרת מתעניינת באוכלוסיית חולי הסכרת בעולם. **מדגם** – חלק מתוך האוכלוסייה. למשל, אם נדגום באקראי 10 אנשים מתוך חולי הסכרת אז זהו מדגם מתוך אוכלוסיית חולי הסכרת.

הקורס עוסק ביחסי הגומלין בין המדגם לבין האוכלוסייה : במקרים רבים אין אפשרות לחקור את כל האוכלוסייה כיוון שאין גישה לכולה, היא גדולה מידי, אנו מוגבלים בזמן ובאמצעים טכניים ולכן מבצעים מדגם במטרה לבצע הסקה סטטיסטית מהמדגם לאוכלוסייה.

הדגימה בקורס תהייה דגימה מקרית הכוונה לדגימה שבה לכל תצפית באוכלוסייה יש את אותו סיכויי להיכלל במדגם.

**סטטיסטי** – גודל המחושב על המדגם.  
**פרמטר** – גודל המתאר את האוכלוסייה.

הסימונים לפרמטר וסטטיסטי בקורס הם שונים והדבר מאד משמעותי למשל,

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

הערה : פרמטר הוא גודל קבוע גם אם אנו לא יודעים אותו סטטיסטי הוא משתנה ממדגם למדגם ולכן נדון בהתפלגות שלו שזה נושא המפגש.

תרגילים :

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

א. מי האוכלוסייה?

ב. מה המשתנה?

ג. מהם הפרמטרים?

ד. מהו גודל המדגם?

2. 58% מאזרחי המדינה תומך בהצעת החוק של חבר כנסת מסוים. הוחלט לדגום 200 אזרחים ומתוכם לבדוק מהו אחוז התומכים בהצעת החוק.

א. מי האוכלוסייה?

ב. מה המשתנה?

ג. מה הפרמטרים?

ד. מהו גודל המדגם?

ה. מהו הסטטיסטי שמתכננים להוציא מהמדגם?

ו. האם הפרמטר או הסטטיסטי הוא משתנה מקרי?

**א. התפלגות ממוצע המדגם ומשפט הגבול המרכזי**

$$\bar{x} = \frac{\sum x_i}{n} \quad \text{התפלגות ממוצע הדגימה}$$

ממוצע האוכלוסייה:  $\mu$

שונות אוכלוסייה:  $\sigma^2$

סטיית תקן של אוכלוסייה:  $\sigma$

**א. תכונות התפלגות  $\bar{x}$** 

**תכונה 1:**

ממוצע כל ממוצעי המדגם האפשריים שווה לממוצע האוכלוסייה.

$$\mu_{\bar{x}} = \mu$$

**תכונה 2:**

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

שונות כל ממוצעי המדגם האפשריים שווה לשונות האוכלוסייה מחולק ב- n - תכונה זו נכונה רק במדגם מקרי.

יש יחס הפוך בין גודל המדגם לבין שונות ממוצעי המדגם.

אם נוציא שורש לשונות נקבל סטיית תקן:

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

**ב. דגימה מהתפלגות נורמלית**

אם נדגום מתוך אוכלוסייה שהמשתנה בה מתפלג נורמלית עם ממוצע  $\mu$  ושונות  $\sigma^2$  ממוצע המדגם גם יתפלג נורמלית:

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$Z_{\bar{x}} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

**ג. משפט הגבול המרכזי**

אם אוכלוסייה מתפלגת כלשהו עם ממוצע  $\mu$  ושונות  $\sigma^2$  אזי עבור מדגם מספיק גדול ( $n \geq 30$ )

ממוצע המדגם גם מתפלג נורמלית  $\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ .

1. משקל תינוק ביום היוולדו מתפלג נורמאלית עם ממוצע 3400 גרם וסטיית תקן של 400 גרם.
- א. מה ההסתברות שתינוק אקראי בעת הלידה ישקול פחות מ-3800 גרם?
- נתון שביום מסוים נולדו 4 תינוקות.
- ב. מה ההסתברות שהמשקל הממוצע שלהם יעלה על 4 ק"ג ?
- ג. מה ההסתברות שהמשקל הממוצע של התינוקות יהיה מתחת ל-2.5 ק"ג?
- ד. מה ההסתברות שהמשקל הממוצע של התינוקות יהיה רחוק מהתוחלת בלא יותר מ-50 גרם?
2. הזמן הממוצע שלוקח לאדם להגיע לעבודתו 30 דקות עם שונות של 16 דקות רבועיות. האדם נוסע לעבודה במשך שבוע 5 פעמים. לצורך פתרון הניחו שזמן הנסיעה לעבודה מתפלג נורמאלית.
- א. מה ההסתברות שבמשך שבוע משך הנסיעה הממוצע יהיה מעל 33 דקות?
- ב. מהו הזמן שבהסתברות של 90% ממוצע משך הנסיעה השבועי יהיה פחות ממנו?
- ג. מה ההסתברות שממוצע משך הנסיעה השבועי יהיה מרוחק מ-30 דקות בלפחות 2 דקות?
3. נפח היין בבקבוק מתפלג נורמאלית עם תוחלת של 750 סמ"ק וסטיית תקן של 10 סמ"ק.
- א. בארגז 4 בקבוקי יין. מה ההסתברות שהנפח הממוצע של הבקבוקים בארגז יהיה בדיוק 755 סמ"ק?
- ב. בארגז 4 בקבוקי יין. מה ההסתברות שהנפח הממוצע של הבקבוקים בארגז יהיה יותר מ-755 סמ"ק?
- ג. בארגז 4 בקבוקי יין. מה ההסתברות שהנפח הממוצע של הבקבוקים בארגז יהיה לפחות 755 סמ"ק?
- ד. בקבוקי היין שבארגז נמזגים לקערה עם קיבולת של שלושה ליטר. מה ההסתברות שהיין יגלוש מהקערה?

4. משתנה מתפלג נורמאלית עם תוחלת 80 וסטיית תקן 4 .
- א. מה ההסתברות שממוצע המדגם יסטה מתוחלתו בלא יותר מיחידה כאשר גודל המדגם הוא 9?
- ב. מה ההסתברות שממוצע המדגם יסטה מתוחלתו בלא יותר מיחידה שגודל המדגם הוא 16?
- ג. הסבר את ההבדל בתשובות של שני הסעיפים.
5. 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 ₪?
6. מטילים קובייה 50 פעמים בכל פעם מתבוננים בתוצאה של הקובייה. מה ההסתברות שהממוצע של התוצאות יהיה לפחות 3.72 ב- 50 ההטלות?
7. אורך צינור שמפעל מייצר הינו עם ממוצע של 70 ס"מ וסטיית תקן של 10 ס"מ
- א. נלקחו באקראי 100 מוטות, מה ההסתברות שממוצע 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 למלאכה?
- ג. מה צריך להיות גודל המדגם המינימאלי, כדי שבהסתברות של 5% ממוצע המדגם יהיה קטן מ-69 ס"מ. העזר במשפט הגבול המרכזי.

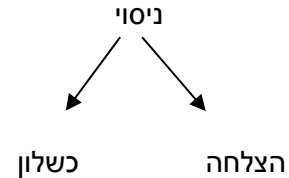


## ב. התפלגות מספר הצלחות במדגם ופרופורציית הצלחות במדגם

### התפלגות מספר הצלחות במדגם

כעת אנו רוצים לדון לא על ממוצע של משקלים או על ממוצע של ציונים במדגם כלשהו, אלא על סטטיסטי אחר:

מספר הצלחות במדגם שנסמן אותו ב-  $Y$ . מחלקים כל אפשרות במדגם להצלחה וכישלון. כעת מה שמתנה מתצפית לתצפית הוא משתנה דיכוטומי (משתנה שיש לו שני ערכים)



הסיכוי להצלחה יסומן עם הפרמטר  $p$  וכישלון יסומן ע"י הפרמטר  $q = 1 - p$ . מבצעים מדגם אקראי בגודל  $n$  ו-  $y$  מספר הצלחות במדגם (סטטיסטי - גודל המתאר את המדגם).

$y$  מתפלג בינומית (מספר הצלחות במדגם)

פונקציית ההסתברות של ההתפלגות הבינומית היא:  $p(y = k) = \binom{n}{k} p^k q^{n-k}$

תוחלת של ההתפלגות הבינומית  $E(y) = np$

שוונות  $V(y) = npq$

### התפלגות פרופורציית הצלחות במדגם

$y$  מספר הצלחות במדגם (למשל, מספר האקדמאים במדגם)

פרופורציית הצלחות במדגם (למשל, אחוז האקדמאים במדגם)  $\hat{p} = \frac{y}{n}$

$n = 200$

מספר המובטלים: 20

פרופורציית המובטלים  $\hat{p} = \frac{20}{200} = 0.1$

$$E(\hat{P}) = p$$

$$V(\hat{P}) = \frac{pq}{n}$$

$$\sigma(\hat{p}) = \sqrt{\frac{pq}{n}}$$

משפט הגבול המרכזי עבור הפרופורצייה המדגמית:

אם  $np \geq 10$  &  $nq \geq 10$  אזי  $\hat{p} \sim N(p, \frac{pq}{n})$

$$Z_{\hat{p}} = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

תרגילים:

1. נתון ש-20% מאוכלוסייה מסוימת אקדמאית. נבחרו באקראי 10 אנשים באותה אוכלוסייה.
  - א. מה ההסתברות ששלושה מהם אקדמאים?
  - ב. מה ההסתברות שלכל היותר אחד מהם אקדמאי?
  - ג. מה התוחלת ומהי סטיית התקן של מספר האקדמאים במדגם?
  
2. במפעל 10% מהמוצרים פגומים. נלקחו 100 מוצרים באקראי מקו הייצור.
  - א. מה ההסתברות שנדגמו לפחות 6 מוצרים פגומים?
  - ב. מה ההסתברות שמספר המוצרים הפגומים יהיה לכל היותר 11 במדגם?
  
3. ציוני פסיכומטרי בקרב הנרשמים למוסד מסוים מתפלגים נורמאלית עם ממוצע 500 וסטיית תקן 100. למוסד מסוים הוחלט לקבל אך ורק סטודנטים שקיבלנו מעל 600 בפסיכומטרי. 100 סטודנטים אקראיים נרשמו למוסד. מה ההסתברות שלפחות 20 יתקבלו?
  
4. במדינה יש 10% של אבטלה. נדגמו באקראי 140 אנשים מהמדינה.
  - א. מה התוחלת ומהי השונות של פרופורציות המובטלים שנדגמו?
  - ב. מה ההסתברות שבמדגם לפחות 10% יהיו מובטלים?
  - ג. מה ההסתברות שלכל היותר 9% מהמדגם יהיו מובטלים?
  
5. נניח ש-30% מהאוכלוסייה תומכת בהצעת חוק מסוימת. אם נדגום מהאוכלוסייה 200 איש. חשבו את ההסתברויות הבאות:
  - א. לפחות 35% יתמכו בהצעת החוק במדגם.
  - ב. לכל היותר 25% יתמכו בהצעת החוק במדגם.
  - ג. יותר מ – 27% יתמכו בהצעת החוק במדגם.

## ג. תשובות סופיות להתפלגות הדגימה

פרק א' - התפלגות ממוצע מדגם ומשפט הגבול המרכזי

<u>שאלה 3</u>	<u>שאלה 4</u>
א. 0.8413	א. 0.0465
ב. 0.0013	ב. 32.29
ג. 0	ג. 0.2628
ד. 0.1974	
<u>שאלה 5</u>	<u>שאלה 6</u>
א. 0	א. 0.5468
ב. 0.1587	ב. 0.6826
ג. 0.1587	
ד. 0.5	
<u>שאלה 8</u>	
	0.0475
<u>שאלה 9</u>	<u>שאלה 10</u>
0.1814	א. 0.9772
	ב. 0.0228
	ג. 271

**פרק ב' - התפלגות מספר ההצלחות במדגם ופרופורציות ההצלחות במדגם****שאלה 2**

א. 0.9082

ב. 0.6293

**שאלה 1**

א. 0.201

ב. 0.3758

ג. התוחלת: 2

סטיית התקן: 1.2649

**שאלה 4**

א. התוחלת: 0.1

השונות: 0.00064

ב. 0.5

ג. 0.3483

**שאלה 3**

0.1292

**שאלה 5**

א. 0.0618

ב. 0.0618

ג. 0.8238

## פרק 26 - מושגים בסיסיים באמידה

### רקע:

כזכור מהמפגש הקודם פרמטר הוא גודל המתאר את האוכלוסייה או התפלגות מסוימת.

כמו ממוצע הגבהים בקרב מתגייעים לצה"ל- $\mu$ .

כמו פרופורציית התומכים בממשלה בקרב אזרחי המדינה -  $p$ .

בדרך כלל הפרמטרים הם גדלים שאינם ידועים באמת, ולכן מבצעים מדגמים במטרה לאמוד אותם. אין אפשרות לחשב אותם הניסיון הוא בלהעריך כמה הם שווים ככל שניתן.

• נסמן באופן כללי פרמטר באות  $\theta$  ואומד ב- $\hat{\theta}$ .  $\hat{\theta}$  הוא סטטיסטי המחושב על המדגם ובאמצעותו נאמוד את  $\theta$ .

• שגיאת אמידה:  $|\hat{\theta} - \theta|$  - ההפרש בין האומד לאמת(הפרמטר).

### דוגמה: (פתרון בהקלטה)

בכנסת ה-19 קיבלה מפלגת העבודה 15 מנדטים. בערוץ 10 ברגע סגירת הקלפיות העריכו את מספר המנדטים של המפלגה להיות 17 מנדטים וזאת על סמך תוצאות מדגם של הערוץ.

מה הפרמטר בדוגמה זו?

מהי טעות האמידה של ערוץ 10?

•  $E(\hat{\theta}) = \theta$  יהיה אומד חסר הטיה ל  $\theta$  אם התוחלת של  $\hat{\theta}$  תהיה שווה ל  $\theta$  :

• טעות התקן של אומד היא סטיית התקן שלו, כלומר:  $\sigma(\hat{\theta}) = S.E$

להלן פרמטרים מרכזיים והאומדים שלהם:

ממוצע האוכלוסייה:  $\mu$

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

$E(\bar{x}) = \mu$  לכן  $\bar{x}$  הינו אומר חסר הטויה ל  $\mu$ .

כמו כן טעות תקן:  $SE = \frac{\sigma}{\sqrt{n}} = \sigma(\bar{x})$

פרופורציה באוכלוסייה:  $p$

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

$E(\hat{p}) = p$  לכן  $\hat{p}$  הינו אומר חסר הטויה ל  $p$ .

כמו כן טעות התקן:  $\sigma(\hat{p}) = \sqrt{\frac{p \cdot (1-p)}{n}}$

שונות האוכלוסייה:  $\sigma^2$

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

$E(S^2) = \sigma^2$  ולכן  $S^2$  הינו אומד חסר הטויה ל  $\sigma^2$ .

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{\sum x_i^2 - n\bar{x}^2}{n-1}$$

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:

2,1,3,2,1,4,5,2,1,3

אמדו באמצעות אומדים 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.

**תרגילים:**

1. מתוך 500 טירונים נמצאו 120 בעלי שברי הליכה. נתון שהסיכוי שטירון יהיה עם שבר הליכה הוא 0.25.

א. מהי האוכלוסייה המוצגת בשאלה? מהם הפרמטרים שלה?

ב. מהי טעות התקן של האומדן כשהמדגם בגודל 500?

ג. מהו האומדן לפרמטר?

ד. מהי טעות האמידה?

2. לפי נתוני היצרן מקרר צורך בממוצע 2400 וואט לשעה עם סטיית תקן של 500 וואט לשעה.

במדגם של 25 מקררים של היצרן התקבל ממוצע של 2342 וואט לשעה.

א. מהי האוכלוסייה המוצגת בשאלה? מהם הפרמטרים שלה?

ב. מהי טעות התקן של האומדן?

ג. מהו האומדן לפרמטר?

ד. מהי טעות האמידה?

3. נדגמו עשרה מתגייסים לצה"ל. גובהם נמדד בס"מ. להלן התוצאות שהתקבלו:

175, 168, 187, 177, 180, 171, 192, 184, 168, 175.

א. מצא אומדן חסר הטיה לגובה הממוצע של מתגייסי צה"ל.

ב. מצא אומדן חסר הטיה לשונות הגבהים של מתגייסי צה"ל.

ג. מצא אומדן חסר הטיה לפרופורציות המתגייסים בגובה של לפחות 180 ס"מ.

4. נדגמו 20 שכירים באקראי. עבור כל שכיר נמדד השכר באלפי שקלים. להלן התוצאות שהתקבלו:

$$\sum_{i=1}^{20} X_i^2 = 1502.2 \quad \sum_{i=1}^{20} X_i = 162$$

א. אמדו את השכר הממוצע של השכירים במשק.

ב. אמדו את סטיית התקן של שכר השכירים במשק.



5. במטרה לאמוד את ממוצע האוכלוסייה. דגמו תצפיות בלתי תלויות מהאוכלוסייה וחישובו את הממוצע שלהם. מהי טעות התקן?  
 א. סטיית התקן של האוכלוסייה.  
 ב. סטיית התקן של ממוצע האוכלוסייה.  
 ג. סטיית התקן של המדגם.  
 ד. 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 תהייה:

א. 3

ב. 2.5

ג. 1.581

ד. אין מספיק נתונים לדעת.

7. במדגם מקרי, מתי 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 נורמאלית.

ד. כאשר מעוניינים באומדן חסר הטיה לשונות האוכלוסייה ממנה הוצא המדגם.

ה. כאשר מעוניינים לחשב את שונות התפלגות הדגימה של ממוצע המדגם.

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

$\sigma^2 = 64$ . טעות התקן של האומדן 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 הפרמטר באוכלוסייה.
- ג. אומד שממוצע התפלגות הדגימה שלו שווה לערך הפרמטר באוכלוסייה.
- ד. אומד שהסיכוי שערכו יהיה גבוה מערך הפרמטר באוכלוסייה שווה לסיכוי שיהיה נמוך ממנו.

**פתרונות:**

**שאלה 3**

א. 177.9

ב. 64.1

ג. 0.4

**שאלה 4**

א. 8.1

ב. 3.16

**שאלה 5**

התשובה היא ד.

**שאלה 6**

התשובה היא ג.

**שאלה 7**

התשובה היא ד.

**שאלה 8**

התשובה היא ד.

**שאלה 9**

התשובה היא ג.

## פרק 27 - רווח סמך לתוחלת (ממוצע האוכלוסייה)

### רווח סמך כששונות האוכלוסייה ידועה

#### רקע:

ממוצע המדגם הוא אומדן לממוצע האוכלוסייה, אך לא באמת ניתן להבין ממנו על גודלו של ממוצע האוכלוסייה. ההסתברות שממוצע המדגם יהיה בדיוק כמו הממוצע האמתי הוא אפסי. מה שנהוג לעשות כדי לאמוד את ממוצע האוכלוסייה זה לבנות רווח סמך. נבנה מרווח בטחון שהסיכוי שהפרמטר  $\mu$  ייכלל בתוכו הוא  $1-\alpha$ .

$1-\alpha$  : נקרא רמת בטחון או רמת סמך.

כך ש:  $P(A \leq \mu \leq B) = 1 - \alpha$

A - גבול התחתון של רווח הסמך

B - הגבול העליון של רווח הסמך

$L = B - A$  - אורך רווח הסמך

#### דוגמה: (פתרון בהקלטה)

חוקר דגם 25 חיילים שנבחנו במבחן הפסיכומטרי. הוא בנה רווח סמך לממוצע הציונים במבחן הפסיכומטרי בקרב אוכלוסיית החיילים וקיבל בין 510 ל-590. רווח הסמך נבנה ברמת סמך של 95%.

מהי אוכלוסיית המחקר?

מה המשתנה באוכלוסייה?

מה הפרמטר שהחוקר רצה לאמוד?

מהו רווח הסמך?

מה אורך רווח הסמך?

מהי רמת הביטחון של רווח הסמך?

בפרק זה נרצה לבנות רווח סמך לתוחלת (  $\mu$  ) במקרה ש  $\sigma^2$  (שונות האוכלוסייה) ידועה

הפרמטר שנרצה לאמוד :  $\mu$

האומד נקודתי :  $\bar{x}$

התנאים לבניית רווח הסמך :

1  $X \sim N$  או  $n \geq 30$

2  $\sigma^2$  (שונות האוכלוסייה) ידועה

הנוסחה לרווח הסמך :

$$\bar{x} \pm Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

דוגמה : (פתרון בהקלטה)

על פי נתוני היצרן אורך חיי סוללה מתפלג נורמאלית עם סטיית תקן של 1 שעה.

מעוניינים לאמוד את תוחלת חיי סוללה.

נדגמו באקראי 4 סוללות, אורך החיים הממוצע שהתקבל הוא 13.5 שעות.

בנו רווח סמך ברמת סמך של 95% לתוחלת אורך חיי סוללה.

שגיאת האמידה המקסימלית:

$$\varepsilon = Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

$\varepsilon$ -נותן את שגיאת האמידה המקסימלית, דבר שנקרא גם טעות סטטיסטית, טעות דגימה.

דוגמה: (פתרון בהקלטה)

בהמשך לשאלה עם הסוללות. מה ניתן להגיד בביטחון של 95% על שגיאת האמידה?

קשרים מתמטיים ברווח הסמך:

- אורך רווח הסמך הוא פעמיים שגיאת האמידה המקסימלית:  $L = 2\varepsilon$ .
- ממוצע המדגם נופל תמיד באמצע רווח הסמך:  $\bar{X} = \frac{A+B}{2}$
- ככל שמספר התצפיות ( $n$ ) גבוה יותר, כך יש יותר אינפורמציה ולכן האומדן יותר מדויק, ולכן נקבל רווח סמך יותר קצר.
- ככל שרמת הביטחון  $(1 - \alpha)$  גבוהה יותר כך  $z_{1-\alpha/2}$  יותר גבוה, ורווח הסמך יותר ארוך.

### תרגילים :

1. חוקר התעניין לאמוד את השכר הממוצע במשק. על סמך מדגם הוא קבע שבביטחון של 95% כי השכר הממוצע במשק נע בין 9200 ל-9800.
  - א. מי האוכלוסייה במחקר?
  - ב. מה המשתנה הנחקר?
  - ג. מה הפרמטר שאותו רוצים לאמוד?
  - ד. מה רווח הסמך לפרמטר?
  - ה. מהי רמת הסמך לפרמטר?
  - ו. מה אורך רווח הסמך?
  - ז. מה הסיכוי שטעות הדגימה תעלה על 300 ₪?
  
2. מעוניינים לאמוד את התפוקה היומית הממוצעת של מפעל מסוים ברמת סמך של 95%. במדגם אקראי של 100 ימים התקבלה תפוקה ממוצעת 4950 מוצרים ביום. לצורך פתרון הנח שסטיית התקן האמתית ידועה ושווה 150 מוצרים ביום. בנה את רווח הסמך.
  
3. מעוניינים לאמוד את ממוצע אורך החיים של מכשיר. מנתוני היצרן ידוע שאורך החיים מתפלג נורמאלי עם סטיית תקן של 20 שעות. נדגמו 25 מכשירים ונמצא כי ממוצע אורך החיים שלהם היה 230 שעות.
  - א. בנו רווח סמך ברמת סמך של 90% לאורך החיים הממוצע של מכשיר.
  - ב. בנו רווח סמך ברמת סמך של 95% לאורך החיים הממוצע של מכשיר.
  - ג. הסבר כיצד ומדוע השתנה רווח הסמך.
  
4. דגמו 200 עובדים מהמשק הישראלי. השכר הממוצע שלהם היה 9700 ₪. נניח שסטיית התקן של השכר במשק היא 3000 ₪.
  - א. בנו רווח סמך ברמת סמך של 95% לתוחלת השכר במשק.
  - ב. מה ניתן לומר בביטחון של 95% על הסטייה המרבית בין ממוצע המדגם לתוחלת השכר?
  - ג. מה היה צריך להיות גודל המדגם אם הינו רוצים להקטין את רווח הסמך ב-50%?
  - ד. אם היינו מגדילים את גודל המדגם ובונים רווח סמך באותה רמת סמך האם היה ניתן לטעון בביטחון רב יותר שרווח הסמך מכיל את הפרמטר?
  
5. בנו רווח סמך לממוצע הציונים של מבחן אינטליגנציה. ידוע שסטיית התקן היא 15 והמדגם מתבסס על 100 תצפיות. רווח הסמך שהתקבל הוא (99,105). שחזרו את :
  - א. ממוצע המדגם.
  - ב. שגיאת האמידה המקסימאלית.
  - ג. רמת הסמך.

6. זמן החלמה מאנגינה מתפלג עם סטיית תקן 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.
- א. בנו רווח סמך לתוחלת זמן ההחלמה תחת האנטיביוטיקה החדשה ברמת סמך של 90%.
- ב. מה היה קורה לאורך רווח הסמך אם היה תקציב להגדלת גודל המדגם פי 4? הסבירו.
- ג. מה היה קורה לאורך רווח הסמך אם היינו בונים את רווח הסמך ברמת סמך גדולה יותר? הסבירו.
7. חוקר בנה רווח סמך לממוצע וקיבל את רווח הסמך הבא:  $82 < \mu < 92$ . נתון שסטיית התקן בהתפלגות שווה ל-10 ושהמדגם מתבסס על 16 תצפיות. התפלגות המשתנה היא נורמאלית.
- א. מהו ממוצע המדגם?
- ב. מהי רמת הסמך של רווח הסמך שנבנה?
- ג. מה הסיכוי ששגיאת האמידה באמידת ממוצע האוכלוסייה תעלה על 5 ?
8. חוקר בנה רווח סמך לתוחלת כאשר השונות בהתפלגות ידועה ברמת סמך של 95%. אם החוקר כעת יבנה על סמך אותם נתונים רווח סמך ברמת סמך קטנה מ-95%, מי מהמשפטים הבאים אינו יהיה נכון.
- א. אורך רווח הסמך החדש יהיה קטן יותר.
- ב. גודל המדגם יהיה כעת קטן יותר.
- ג. המרחק בין ממוצע המדגם לקצות רווח הסמך יהיו קטנים יותר ברווח הסמך החדש.
- ד. רמת הביטחון לבנות רווח הסמך החדש תהיה קטנה יותר.
9. חוקר בנה רווח סמך ל- $\mu$  וקיבל  $48 < \mu < 54$  מה נכון בהכרח:
- א.  $\mu = 51$
- ב.  $\bar{X} = 6$
- ג.  $\bar{X} = 51$
- ד. אורך רווח הסמך הינו 3.
10. איזה מהגורמים הבאים אינו משפיע על גודלו של רווח בר סמך, כאשר שונות האוכלוסייה ידועה? (בחר בתשובה הנכונה)
- א. רמת הביטחון.
- ב. סטיית התקן באוכלוסייה.
- ג. מספר המשתתפים.
- ד. סטיית התקן במדגם.



**פתרונות :****שאלה 2**

$$4920.6 < \mu < 4979.4$$

**שאלה 3**

$$א. \quad 223.42 < \mu < 236.58$$

$$ב. \quad 222.16 < \mu < 237.84$$

**שאלה 5**

$$א. \quad 102$$

$$ב. \quad 3$$

$$ג. \quad 0.9544$$

**שאלה 6**

$$א. \quad 83.5 < \mu < 4.42$$

$$ב. \quad \text{יקטן פי } 2$$

$$ג. \quad \text{גדל}$$

**שאלה 7**

$$א. \quad 87$$

$$ב. \quad 5$$

$$ג. \quad 0.9544$$

**שאלה 8**

$$א. \quad 139$$

$$ב. \quad 21 < \mu < 25$$

**שאלה 9**

התשובה היא : ב

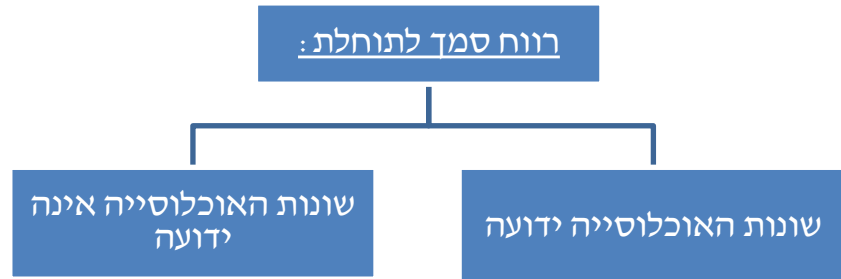
**שאלה 10**

התשובה היא : ג

## רווח סמך לתוחלת (ממוצע האוכלוסייה) כששונות האוכלוסייה אינה ידועה

רקע:

בבואנו לבנות רווח סמך לתוחלת אנו צריכים להתמקד בשני המצבים הבאים:



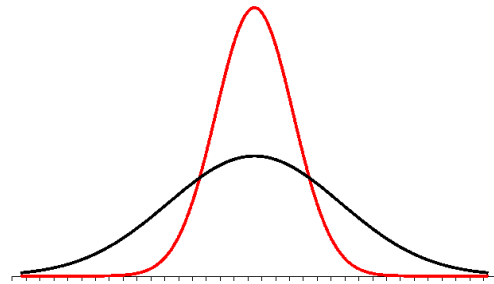
בפרק זה נעסוק במקרה ששונות האוכלוסייה אינה ידועה לנו. מקרה יותר פרקטי.

התנאי:  $X \sim N$  או שהמדגם גדול

$$\bar{X} \pm t_{\alpha/2}^{(n-1)} \cdot \frac{\hat{S}}{\sqrt{n}} \quad \text{רווח סמך:}$$

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

התפלגות T:



הינה התפלגות סימטרית פעמונית שהתוחלת שלה היא 0. ההתפלגות דומה להתפלגות Z רק שהיא יותר רחבה ולכן הערכים שלה יהיו יותר גבוהים. התפלגות T תלויה במושג שנקרא דרגות חופש. דרגות החופש הן  $df=n-1$ . ככל שדרגות החופש עולות ההתפלגות הופכת להיות יותר גבוהה וצרה. כשדרגות החופש שואפות לאינסוף התפלגות T שואפת להיות כמו התפלגות Z.

**דוגמה :** (פתרון בהקלטה)

הזמן שלוקח לפתור שאלה מסוימת בחשבון מתפלג אצל תלמידי כיתות ח' נורמאלית. במטרה לאמוד את תוחלת זמן הפתרון נדגמו 4 תלמידים בכיתה ח'. להלן התוצאות שהתקבלו בדקות : 4.7, 5.2, 4.6, 5.3.

בנו רווח סמך ברמת סמך של 95% לממוצע זמן הפתרון לשאלה בקרב תלמידי כיתה ח'.

**פתרון :**

$$4.39 < \mu < 5.51$$

**תרגילים:**

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, 88, 84, 79, 89. הערה: לצורך פתרון הנח שקצב פעימות הלב מתפלג נורמאלית בקירוב.
- א. בנו רווח סמך ברמת סמך של 95% לתוחלת הדופק של נוטלי התרופה הנ"ל.
- 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 רווח הסמך ברמת ביטחון של 99% כיצד הדבר היה משפיע על רווח הסמך?
2. במדגם שנעשה על 25 מתגייסים לצבא האמריקאי התקבל כי: גובה ממוצע של חייל הינו 178 ס"מ עם סטיית תקן  $\hat{S} = 13$  ס"מ. בנו רווח סמך ברמת סמך של 90% לתוחלת גובה המתגייסים לצבא האמריקאי. מה יש להניח לצורך פתרון?
3. אדם מעוניין לאמוד את זמן הנסיעה הממוצע שלו לעבודה. לצורך כך הוא דוגם 5 ימים שזמן הנסיעה בהם בדקות הוא: 27, 34, 32, 40, 30.
- א. ברמת ביטחון של 95% אמוד את זמן הנסיעה הממוצע. מהי ההנחה הדרושה לצורך פתרון?
- ב. איך גודל רווח הסמך היה משתנה אם היו דוגמים עוד ימים?
4. ציוני מבחן אינטליגנציה מתפלגים נורמאלית. נדגמו 25 מבחנים והתקבל ממוצע ציונים 102 וסטיית תקן מדגמית 13.
- א. בנו רווח 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 הסעיפים הנ"ל.
5. נשקלו 60 תינוקות אשר נולדו בשבוע ה-40 של ההיריון. המשקל נמדד בקילוגרמים. להלן התוצאות שהתקבלו:  $\sum_{i=1}^{60} X_i = 195$ ,  $\sum_{i=1}^{60} X_i^2 = 643.19$ . בנו רווח סמך ברמת סמך של 95% לתוחלת משקל תינוק ביום היוולדו.

6. שני סטטיסטיקאים בנו רווח בר-סמך לאותו פרמטר  $\mu$ . לכל אחד מהסטטיסטיקאים מדגם אחר, אך באותו גודל 10. שניהם קבעו אותה רמת סמך.

סטטיסטיקאי א : הניח  $\sigma = 20$

סטטיסטיקאי ב : חישב לפי המדגם וקיבל  $\hat{S} = 20$

למי משני הסטטיסטיקאים יהיה רווח סמך ארוך יותר? (בחר בתשובה הנכונה)

א. סטטיסטיקאי א

ב. סטטיסטיקאי ב

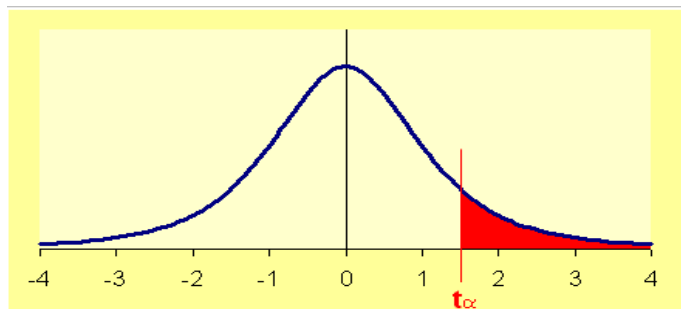
ג. אותו אורך רווח סמך לשני הסטטיסטיקאים.

ד. תלוי בתוצאות המדגם של כל סטטיסטיקאי.

7. נתון ש :  $X \sim N(\mu, \sigma^2)$  ביצעו מדגם בגודל 16 וקיבלו סטיית תקן מדגמית 10. אורך רווח

הסמך שהתקבל הוא : 8.765. מהי רמת הביטחון של רווח הסמך?

טבלת ערכים קריטיים לפי התפלגות t ראה איור מטה.							
דרגות חופש	$\alpha$						
	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
1	3.078	6.314	12.706	31.821	63.657	318.309	636.619
2	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
50	1.299	1.676	2.009	2.403	2.678	3.261	3.496
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
90	1.291	1.662	1.987	2.368	2.632	3.183	3.402
120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
$\infty$	1.282	1.645	1.960	2.326	2.576	3.090	3.291



**פתרונות:****שאלה 1**

$$79.88 < \mu < 89.72 \text{ א.}$$

**שאלה 4**

$$96.63 < \mu < 107.37 \text{ א.}$$

$$96.90 < \mu < 107.10 \text{ ב.}$$

**שאלה 5**

$$3.149 < \mu < 3.351$$

**שאלה 7**

90%

## פרק 28 - רווח סמך לפרופורציה

### רקע:

מטרה: לאמוד את P – פרופורציה באוכלוסייה.

האומד הנקודתי:  $\hat{p} = \frac{y}{n}$  (Y- מספר ההצלחות שבמדגם)

רווח הסמך ל P:  $\hat{p} \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

התנאי לבנות את רווח הסמך הינו מדגם של לפחות 30 תצפיות (לעיתים נותנים תנאי של מספר הצלחות ומספר כשלוונות לפחות 5 או לפחות 10)

האומד לטעות התקן:  $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

מתקיים ש:  $\hat{P} = \frac{A+B}{2}$   $L = 2\varepsilon$

דוגמה: (פתרון בהקלטה)

1. במטרה לאמוד את אחוז המובטלים במשק נדגמו 200 אזרחים. מתוכם התקבל ש 24 היו מובטלים.

א. בנו רווח סמך לאחוז המובטלים באוכלוסייה ברמת סמך של 95%.

ב. מהו האומד לטעות התקן?

פתרון:

א.  $7.5\% < p < 16.5\%$

ב. 2.29%



**תרגילים:**

1. נדגמו 200 דירות בעיר חיפה. 48 מתוכן נמצאו כבעלות ממ"ד.
  - א. בנו רווח סמך ברמת סמך של 95% לאחוז הדירות בחיפה עם ממ"ד.
  - ב. על סמך סעיף א' מה ניתן לומר על שגיאת האמידה המקסימאלית?
  - ג. בהנחה ובחיפה 80 אלף דירות, בנו רווח סמך ברמת סמך של 95% למספר הדירות בחיפה עם ממ"ד בפועל.
  
2. במדגם של 300 אנשי היי-טק התקבל ש-180 מהם אקדמאים.
  - א. בנו רווח סמך לפרופורציית אקדמאים ברמת סמך של 95% (בקרב אנשי היי-טק).
  - ב. כיצד רווח הסמך של סעיף א' היה משתנה אם היינו מקטינים את רמת הסמך?
  - ג. כיצד רווח הסמך היה משתנה אם הינו מגדילים את גודל המדגם?
  
3. במדגם של 400 נהגים התקבל רווח סמך לפרופורציית הנהגים החדשים:  $0.08 < p < 0.18$ 
  - א. כמה נהגים במדגם היו נהגים חדשים?
  - ב. מהי רמת הסמך של רווח הסמך שנבנה?
  
4. במסגרת מערכת הבחירות בארה"ב נשאלו 840 אנשים עבור איזה מועמד יצביעו.
  - א. 510 אנשים ענו כי יצביעו בעד ברק אובמה. בסקר פורסם שתתכן סטייה של  $\pm 3\%$  מתוצאות האמת. באיזו רמת ביטחון הסקר השתמש?
  
5. במדגם של 300 נשים בגילאי 40-35 נמצא ש-140 היו נשואות, 80 היו גרושות, 60 רווקות והיתר אלמנות.
  - א. מצאו רווח סמך ברמה של 90% לאחוז הגרושות באוכלוסייה הנחקרת.
  - ב. מצאו רווח סמך ברמה של 99% לסיכוי שבאוכלוסייה הנחקרת תמצא אישה לא נשואה?
  
6. ביצעו מדגם באוכלוסייה. שיעור ההצלחות במדגם היה 10% ורווח הסמך ניבנה ברמת סמך של 95%. אורכו הינו 0.3156.
  - א. מהו גודל המדגם שנלקח?

**פתרונות:****שאלה 3**

א. 52

ב. 0.997

**שאלה 5**א.  $22.5\% < p < 30.9\%$ ב.  $45.91\% < p < 60.72\%$ **שאלה 6**

200

## פרק 29 - רווח סמך להפרש תוחלות ממדגמים בלתי תלויים

### כששונויות האוכלוסייה אינן ידועות אך שוות

#### רקע:

מטרה: לאמוד את פער התוחלות:  $\mu_1 - \mu_2$ , כלומר ההבדלים של הממוצעים בין שתי האוכלוסיות.

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

התנאים לבניית רווח הסמך:

$$1. \sigma_1^2 = \sigma_2^2$$

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

3. מדגמים בלתי תלויים.

השוונת המשוקללת: כיוון שאנו מניחים שבין שתי האוכלוסיות השונויות שוות אנו אומדים את השוונת הזו על ידי שקלול שתי השונויות של שני המדגמים על ידי הנוסחה הבאה:

$$\hat{S}_p^2 = \frac{(n_1 - 1)\hat{S}_1^2 + (n_2 - 1)\hat{S}_2^2}{n_1 + n_2 - 2}$$

דרגות החופש:  $d.f = n_1 + n_2 - 2$

רווח סמך:

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\frac{\alpha}{2}}^{n_1+n_2-2} \cdot \sqrt{\frac{\hat{S}_p^2}{n_1} + \frac{\hat{S}_p^2}{n_2}}$$

אם הערך אפס נופל בגבולות רווח הסמך נגיד שבביטחון של  $1 - \alpha$  לא קיים הבדל בין התוחלות.

#### דוגמה: (פתרון בהקלטה)

מחקר מעוניין לבדוק האם קיים הבדל בין תל אביב לבאר שבע מבחינת ההכנסה הממוצעת של אקדמאים. להלן תוצאות המדגם שנעשה:

באר שבע	תל אביב	
10	20	מספר האקדמאים
9500	11,000	ממוצע הכנסות של אקדמאים
250	200	סטיית התקן של הכנסות אקדמאים

בנו רווח סמך ברמת ביטחון של 90% להפרש תוחלות ההכנסה בשני האזורים. הניחו שהשכר מתפלג נורמלית עם אותה שונות בכל אחד מהאזורים.

פתרון: (1357,1643)

**תרגילים:**

1. נדגמו 15 ישראלים ו-15 אמריקאים.

כל הנדגמים נגשו למבחן IQ. להלן תוצאות המדגם:

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

מצאו רווח סמך ברמת סמך של 95% לסטייה בין ממוצע הציונים בישראל לממוצע הציונים בארה"ב. רשמו את כל ההנחות הדרושות לצורך פתרון התרגיל.

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

25	21	20	22	X
12	17	25	18	Y

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

## פרק 30 - רווח סמך לתוחלת ההפרש במדגם מזווג

### רקע:

מדגם מזווג: מדגם אחד שבו יש  $n$  צמדנים.

כל תצפית במדגם תנפק זוג ערכים:  $X$  ו- $Y$ .

ניצור משתנה חדש:

$$D = x - y$$

הפרמטר שנרצה לאמוד:  $\mu_D$

התנאים לבניית רווח הסמך:

•  $x, y \sim N$

• המדגם מזווג

נוסחת רווח הסמך:

$$\bar{D} \pm t_{1-\frac{\alpha}{2}}^{n-1} \frac{S_D}{\sqrt{n}}$$

כאשר דרגות החופש:  $d.f = n - 1$

**דוגמה:** (פתרון בהקלטה)

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

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

הניחו כי זמני הריצות מתפלגים נורמלית.  
מצאו רווח סמך של 95% להפרש תוחלת הזמן בין שתי התוכנות.

**תרגילים:**

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

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

נניח שהציונים מתפלגים נורמאלית.

- א. בנו רווח סמך ברמת סמך של 95% לתוחלת פער הציונים בין סמסטר א לבין סמסטר ב.  
 ב. האם על סמך רווח הסמך קיים הבדל בין הסמסטרים מבחינת תוחלת הציונים?  
 ג. מה צריך לשנות בנתונים כדי שהמדגמים יהיו בלתי תלויים?

2. במטרה לבדוק האם קיים הבדל בין קווי זהב לבזק ממוצע המחירים לשיחות בינ"ל. נגדמו באקראי 7 מדינות ועבור כל מדינה נבדקה עלות דקת שיחה. להלן התוצאות:

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

בהנחה והמחירים מתפלגים נורמלית עבור כל חברה בנו רווח סמך ברמת סמך של 90% לתוחלת הפרש המחירים של שתי החברות.

### פרק 31 - תרגול מסכם ברווחי סמך

1. 200 אנשים נשאלו כמה פעמים ביום הם שותים כוס קפה. להלן התפלגות התשובות:

5	4	3	2	1	0	מספר פעמים
10	20	22	28	34	86	מספר אנשים

א. תנו רווח סמך למוצע מספר כוסות הקפה שאנשים נוהגים לשתות ביום.  $\alpha = 0.05$   
 ב. אדם השותה לפחות 4 כוסות קפה ביום נקרא "מכור לקפה". בנו רווח סמך לאחוז

"המכורים לקפה"  $\alpha = 0.1$

2. חוקר בנה רווח סמך לאחוז האנשים שהתקררו לפחות פעם אחת בשנה. רווח הסמך שהתקבל הוא  $81 < p < 91$  רווח הסמך הנ"ל התבסס על מדגם של 500 איש.

א. כמה אנשים במדגם טענו שכלל לא התקררו השנה?

ב. באיזו רמת סמך נבנה רווח הסמך?

ג. בנו רווח סמך לאחוז האנשים שהתקררו לפחות פעם אחת השנה ברמת סמך של 95% על

סמך

תוצאות המדגם.

3. ציוני IQ בארה"ב מתפלגים נורמאלית עם תוחלת 100. במדגם של 20 ישראלים שנבחנו במבחן

$$\sum_{i=1}^{20} x_i = 2040$$

ה-IQ התקבלו התוצאות הבאות:

$$\sum_{i=1}^{20} x_i^2 = 210740$$

א. אמדו ברמת ביטחון של 90% את ממוצע ציוני בחינת ה-IQ בישראל – מהי ההנחה

הדרושה

לפתרון?

ב. על סמך רווח הסמך של סעיף א האם תקבלו את הטענה שבישראל ממוצע הציונים שונה

מארה"ב?

ג. מה היה קורה לרווח הסמך אם הינו מגדילים את רמת הסמך שלו?



4. בנק מתלבט האם לפתוח סניף באזור A או באזור B. לצורך פתרון נניח שסטית התקן של המשכורת באזור A היא 1200 ובאזור B 1500. הבנק דגם 50 אנשים מאזור A, המשכורת הממוצעת שהתקבלה במדגם היא 6,800 ₪. כמו כן נדגמו 40 אנשים מאזור B, המשכורת הממוצעת שהתקבלה במדגם היא 6,600 ₪.
- א. בנו רווח סמך ברמת סמך של 95% להפרש הממוצעים של המשכורות בשני האזורים. האם על סמך רווח הסמך ניתן להמליץ לבנק היכן לפתוח את הסניף. אם כן, היכן?
- ב. בנו רווח סמך לתוחלת המשכורת באזור A ברמת סמך של 95%.

5. להלן מדגם של שכר הדירה בשייח של 5 דירות שלושה חדרים בשכונת בבלי בתל אביב :

שנת 2012	8000	7500	7000	6500	7500
שנת 2013	8000	8200	7800	6800	7700

- בנו רווח סמך ברמת סמך של 95% לתוחלת עליית שכר הדירה משנת 2012 לשנת 2013 בשכונת בבלי. ניתן להניח ששכר הדירה בשכונה מתפלג נורמלית.

**פתרונות:****שאלה 1**

א.  $1.21 \leq \mu \leq 1.65$

ב.  $10.85\% \leq p \leq 19.15\%$

**שאלה 2**

א. 70

ב. 99.88%

ג.  $83\% \leq p \leq 89\%$

**שאלה 3**

א.  $97.4 \leq \mu \leq 106.6$

ב. לא

ג. יגדל

**שאלה 4**

א.  $-372 \leq \mu_A - \mu_B \leq 772$

ב.  $6467 \leq \mu \leq 7133$

**שאלה 5**

$-21 \leq \mu_{2013} - \mu_{2012} \leq 821$

## פרק 32 - בדיקת השערות כללית

### רקע:

תהליך של בדיקת השערות הוא תהליך מאד נפוץ בעולם הסטטיסטיקה. בתהליך זה ישנן שתי השערות שנבדקות :

השערת האפס המסומנות ב-  $H_0$

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

בדרך כלל השערת האפס מסמנת את אשר היה מקובל עד עכשיו , את השגרה הנורמה ואילו ההשערה האלטרנטיבית את החדשנות בעצם ההשערה האלטרנטיבית מדברת על הסיבה שהמחקר נעשה .

למשל ,

ישנה תרופה קיימת למחלה A אשר גורמת ל – 10% מהמשתמשים בה לתופעות לוואי . חברת תרופות טוענת שפיתחה תרופה שיעילה באותה מידה , אך מקטינה את הסיכוי לתופעות הלוואי. לכן יש לבצע מחקר שעל סמך תוצאותיו ננסה להכריע איזה השערה נקבל :

$H_0$  : התרופה החדשה הנה קונבנציונאלית וגורמת ל-10% תופעות לוואי.

$H_1$  : התרופה החדשה מקטינה את אחוז הסובלים מתופעות לוואי מתחת ל -10%.

בתהליך של בדיקת השערות יוצרים כלל שניקרא כלל הכרעה :

הכלל יוצר אזור שניקרא אזור דחייה ( דחייה של השערת האפס כלומר קבלה של האלטרנטיבה) ואזור קבלה ( קבלה של השערת האפס ודחייה של האלטרנטיבה). כלל ההכרעה מתבסס על איזשהו סטטיסטי .

בתהליך יש ללכת לתוצאות המדגם ולבדוק האם התוצאות נופלות באזור הדחייה או הקבלה וכך להגיע למסקנה – המסקנה היא בעירבון מוגבל כיוון שהיא תלויה בכלל ההכרעה ובתוצאות המדגם. נשנה את כלל ההכרעה אנחנו יכולים לקבל מסקנה אחרת . נבצע מדגם חדש אנחנו עלולים לקבל תוצאה אחרת.

לכן יתכנו טעויות במסקנות שלנו :

		הכרעה	
		H0	H1
מציאות	H0	אין טעות	טעות מסוג 1
	H1	טעות מסוג 2	אין טעות

הגדרת הטעויות:

טעות מסוג ראשון- להכריע לדחות את  $H_0$  למרות שבמציאות  $H_0$  נכונה.

טעות מסוג שני- להכריע לקבל את  $H_0$  למרות שבמציאות  $H_1$  נכונה.

מה הן הטעויות האפשריות במחקר של התרופות? ( בהקלטה )

נגדיר את ההסתברויות הבאות:

הסיכוי לבצע טעות מסוג 1 ( רמת מובהקות )

$$\alpha = P(H_0 \text{ לדחות} | H_0 \text{ נכונה}) = P_{H_0}(H_0 \text{ לדחות})$$

הסיכוי לבצע טעות מסוג 2:

$$\beta = P(H_0 \text{ לקבל} | H_1 \text{ נכונה}) = P_{H_1}(H_0 \text{ לקבל})$$

רמת בטחון:

$$(1 - \alpha) = P(H_0 \text{ לקבל} | H_0 \text{ נכונה}) = P_{H_0}(H_0 \text{ לקבל})$$

עוצמה:

$$\pi = (1 - \beta) = P(H_1 \text{ נכונה} | H_0 \text{ לדחות}) = P_{H_1}(H_0 \text{ לדחות})$$

דוגמה: ( פתרון בהקלטה )

בכד יש 10 כדורים. יתכן ש- 5 מהם לבנים והיתר שחורים (כד א- השערת האפס) או ש- 7 מהם לבנים והיתר שחורים (כד ב- השערה אלטרנטיבית).

כדי להחליט איזה מהכדים ברשותנו, הוחלט להוציא כדור ולהשתמש בכלל ההחלטה הבא: אם הכדור שהוצא הוא לבן שזהו כד ב' ( $H_1$ ).

א. חשבו את רמת המובהקות ואת רמת הביטחון של המבחן המוצע.

ב. חשבו את הסיכוי לטעות מסוג שני והעוצמה של המבחן המוצע.

**תרגילים:**

1. אדם חשוד בביצוע פשע. מהן הטעויות האפשריות בהכרעת הדין?
2. ילד קנה שקית סוכריות אטומה שבה ציפה ל-10 סוכריות תות ו-5 לימון. ישנה שקית אחרת אותה הוא לא רצה בה 6 סוכריות תות ו-9 לימון. הוא החליט להוציא באקראי סוכרייה אם היא תהיה לימון הוא יחזיר את השקית לחנות. מה הסיכויים לכל סוג של טעות בהכרעתו?

3. יהי  $X$  מספר שלם הנבחר באקראי מבין המספרים השלמים. הסיכוי ש- $X$  יקבל ערך

$$p(X = k) = \frac{1}{n} \quad \text{עבור } k = 1, 2, \dots, n$$

נתונות ההשערות הבאות לגבי התפלגות של  $X$ :

$$H_0 : n = 4$$

$$H_1 : n = 6$$

- כמו כן נתון כלל ההכרעה הבא: נדחה את השערת האפס אם  $X > 3$ .  
חשבו את הסיכוי לטעות מסוג ראשון וטעות מסוג שני ואת העוצמה?

4. איכות של מוצר מסווגת ל-4 רמות איכות: מצוין, טוב, בינוני וירוד. להלן התפלגות טיב המוצר בשני מפעלים:

מפעל	מצוין	טוב	בינוני	ירוד
"היוצר"	0.6	0.2	0.2	0
"שמשון"	0.1	0.2	0.3	0.4

בוחרים ממשלוח מוצר באקראי, אך לא יודעים מאיזה מפעל המשלוח הגיע. על סמך בדיקת האיכות מנסים להכריע האם מדובר במפעל "היוצר" (השערת האפס) או במפעל "שמשון" (השערה אלטרנטיבית).

- א. להלן כלל החלטה: אם מדובר במוצר שטיבו "טוב" נכריע שהמוצר בא ממפעל "שמשון", מהן ההסתברויות לסוגי הטעויות השונים?
- ב. להלן כלל החלטה: אם מדובר במוצר שטיבו "בינוני" או גרוע מכך נכריע שהמוצר בא ממפעל "שמשון", מה מהן ההסתברויות לסוגי הטעויות השונים?
- ג. איזה כלל החלטה עדיף? נמק!

5. במטרה לבדוק האם מטבע תקין הטילו אותו 8 פעמים. הוחלט שאם מספר העצים יהיה בין 1 ל 7 כולל יוחלט שהמטבע תקין, אחרת נחליט שהמטבע מזויף.
- א. רשמו את השערות המחקר.
- ב. מה ההסתברות לטעות מסוג ראשון?
- ג. מהי עצמת המבחן אם במציאות אכן המטבע אינו תקין כי הסיכוי לעץ בו הוא 20%.
6. להלן השערות:
- $$H_0 : X \sim t(5) \quad (\text{התפלגות } t \text{ עם } 5 \text{ דרגות חופש})$$
- $$H_1 : X \sim Z \quad (\text{התפלגות נורמאלית סטנדרטית})$$
- כלל החלטה: נדחה את השערת האפס אם  $X$  גדול מ-2.015.
- א. מהי רמת המובהקות של כלל החלטה?
- ב. מהי העוצמה של כלל החלטה?
7. במפעל מסוים נפליטים לאוויר חומרים רעילים. במצב שיגרה העוצמה הממוצעת של החומר הרעיל אמורה להיות 6,000 יחידות עם סטיית תקן 900. במצב חירום העוצמה הממוצעת היא 7,000 עם סטיית תקן 900. במפעל מערכת התראה נתמכת על ידי 9 חיישנים. אם ממוצע העוצמה של החומר הרעיל לפי תשעת החיישנים עולה על 6600 יחידות מופעלת מערכת ההתראה. נתון שעוצמת הזיהום מתפלגת נורמאלית.
- א. מה הסיכוי להתראת שווא? (באיזה סוג טעות מדובר)?
- ב. מה הסיכוי שבמצב חירום מערכת ההתראה לא תפעל? (באיזה סוג טעות מדובר)?
- ג. מה ההסתברות שאם המצב הוא מצב חירום מערכת ההתראה תפעל? (איך קוראים להסתברות זו)?
- ד. בסעיפים הבאים נשנה בכל סעיף נתון מסוים. כל סעיף עומד בפני עצמו, כיצד השינוי ישנה את הסיכוי לטעות מסוג ראשון ושני?
1. המפעל יקנה עוד 4 חיישנים.
  2. מצב חרום מוגדר כעת בתוחלת של 7500 יחידות.
  3. מערכת ההתראה תופעל אם ממוצע של תשעת החיישנים יהיה מעל 6700.
8. במטרה לבדוק האם במקום עבודה מסוים פרופורציית הבנים נמוכה מפרופורציית הבנות נדגמו באקראי 10 עובדים. הוחלט שאם מספר הבנים במדגם יהיה לכל היותר 2 תתקבל הטענה שפרופורציית הבנים נמוכה מפרופורציית הבנות.
- א. מה רמת המובהקות של כלל ההכרעה הנ"ל?
- ב. מהי העוצמה בהנחה ובחברה 30% בנים?

9. זמן ההשפעה של משכך הכאבים "אופטלנוס" מתפלג נורמאלית עם תוחלת של 40 דקות וסטיית תקן של 12 דקות.
- חברת התרופות המייצרת את התרופה מנסה לשפר את התרופה כך שתוחלת הזמן עד להשפעה תתקצר. לצורך כך, דגמו 25 מטופלים שיקבלו את התרופה "אופטלנוס פורטה", ממוצע זמן התגובה של המטופלים היה 34.5 דקות. חברת התרופות החליטה מראש שאם ממוצע הזמן עד להשפעה יהיה נמוך מ-35 דקות, היא תמשיך בתהליך שיווק "אופטלנוס פורטה".
- א. מהי רמת המובהקות של המבחן המוצע?
- ב. על סמך תוצאות המדגם. מהי המסקנה ומהי הטעות האפשרית במסקנה?
- ג. מהי עצמת המבחן המוצע אם במציאות התרופה "אופטלנוס פורטה" מפחיתה את התוחלת לכדי 32 דקות?
- ד. כיצד תשתנה התשובה לסעיף ג' אם החברה הייתה מחליטה שהיא תמשיך בתהליך שיווק התרופה החדשה כאשר ממוצע המדגם יהיה נמוך מ-36 דקות?
10. ציוני פסיכומטרי מתפלגים נורמלית עם סטיית תקן 120. מכון טוען שלימודים אצלו מעלים את ממוצע הציונים ביותר מ-30 נקודות. נלקחו 20 שלמדו במכון ו-20 שניגשו לבחינה בלמידה עצמית. הוחלט במשרד פרסום לקבל את טענת המכון רק אם במדגם ממוצע הציונים של אלה שלמדו במכון יהיה גבוהה בלפחות 50 נקודות מאלה שלא היו.
- א. מהי רמת המובהקות של המחקר?
- ב. מה הסיכוי לעשות טעות מסוג שני II בהנחה שהמכון מעלה את ממוצע הציונים ב-60 נקודות?
- ג. כיצד התשובות לסעיף א ו ב' היו משתנות אם מסתבר שסטיית התקן בציוני הפסיכומטרי הינה 100. הסבירו ללא חישוב.
11. קו ייצור נחשב תקין אם יש בו לכל היותר 4% פגומים, ונחשב שאינו תקין אחרת. מנהל האיכות דוגם בכל יום מקו הייצור 500 מוצרים. אם במדגם יהיה לפחות 30 מוצרים פגומים יפסיקו באותו היום את קו הייצור.
- א. מה ההסתברות להפסיק את קו הייצור כשהוא תקין. איך קוראים להסתברות זאת?
- ב. מה ההסתברות להמשיך ביום מסוים את קו הייצור למרות שאינו תקין כי היו 8% פגומים בקו הייצור. איך קוראים להסתברות זאת?
12. מעוניינים לבדוק האם בפקולטה מסוימת ישנה העדפה לגברים. הוחלט לדגום 200 מתקבלים ועל סמך מספר הבנים לקבוע אם טענת המחקר מתקבלת.
- חוקר א' קבע רמת מובהקות של 5% וחוקר ב' החליט לקבל את טענת המחקר אם במדגם יהיו לפחות 120 בנים. למי מבין החוקרים רמת מובהקות גדולה יותר?

13. מספר המכונניות הנכנסות לחניון "עזרים" מתפלג פואסונית. בשנה שעברה המכונניות נכנסו לחניון בקצב של 2 מכונניות לדקה. בעקבות תלונות על עומס יתר בכניסה לחניון מעוניין מנהל החניון לבדוק האם קצב כניסת המכונניות לחניון גדל השנה. מנהל החניון החליט לספור את מספר המכונניות שיכנסו לחניון בדקה אקראית. אם מספר המכונניות שישפרו יהיה לפחות 4 יפתח מנהל החניון שער נוסף לחניון.

א. רשום את השערות מנהל החניון ואת כלל ההחלטה שלו. האם כלל ההכרעה הגיוני?  
 ב. מהי רמת המובהקות של כלל ההכרעה?  
 ג. מהי העוצמה של כלל ההחלטה, אם כיום קצב כניסת המכונניות לחניון גדל ל-4 מכונניות בדקה?

14. עודד עובד במפעל שבו מתחילים לעבוד בשעה 8:00. עודד בדרך כלל מאחר לעבודה והמנהל החליט לרשום את שעת בואו לעבודה. המנהל טוען שמשך האיחור של עודד (דקות),  $X$ , היא משתנה אחיד  $U(0, 60)$ . עודד טוען שהוא לא מגיע באיחור כה גדול, אלא שהתפלגות  $X$  היא בעלת התפלגות מעריכית עם תוחלת איחור של 20 דקות. לבדיקת טענת המנהל ( $H_0$ ) כנגד טענת עודד ( $H_1$ ), המבוסס על משך האיחור של חגי ביום אחד.

מוצאים שני ככלי הכרעה:

כלל 1: דחה את השערת האפס אם משך האיחור יהיה לפחות 40 דקות.

כלל 2: דחה את השערת האפס אם משך האיחור יהיה לכל היותר 20 דקות.

חשב את הסיכוי לטעות מסוג ראשון ושני לכל אחת מכללי ההכרעה. מי עדיף?



פתרונות:שאלה 2

$$\beta = \frac{2}{5} \quad \alpha = \frac{1}{3}$$

שאלה 3

$$\beta = 0.5 \quad \alpha = 0.25$$

שאלה 4

$$\beta = 0.8 \quad \alpha = 0.2 \quad \text{א.}$$

$$\beta = 0.3 \quad \alpha = 0.2 \quad \text{ב.}$$

ג. כלל ב'

שאלה 5

ב. 0.00781

ג. 0.1678

שאלה 6

א. 0.05

ב. 0.022

שאלה 7

א. 0.0228

ב. 0.0918

ג. 0.9082

שאלה 8

א. 0.055

ב. 0.383

שאלה 10

א. 0.2981

ב. 0.3974

ג. קטן

שאלה 11

א. 0.0113

ב. 0.0495

**שאלה 12**

חוקר א

**שאלה 13**

ב. 0.1428

ג. 0.566

## פרק 33 - בדיקת השערות על פרמטרים

### הקדמה

#### רקע:

תהליך של בדיקת השערות הוא תהליך מאוד נפוץ בעולם הסטטיסטיקה. בבדיקת השערות על פרמטרים נעבוד לפי השלבים הבאים:

**שלב א:** נזהה את הפרמטר הנחקר.

**שלב ב:** נרשום את השערות המחקר.

**השערת האפס** המסומנות ב-  $H_0$

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

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

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

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

**שלב ד:** נרשום את כלל ההכרעה.

בתהליך של בדיקת השערות יוצרים כלל שניקרא **כלל הכרעה**:

הכלל יוצר אזור שניקרא **אזור דחייה** (דחייה של השערת האפס כלומר קבלה של האלטרנטיבה) ו**אזור קבלה** (קבלה של השערת האפס ודחייה של האלטרנטיבה). כלל ההכרעה מתבסס על איזשהו סטטיסטי.

אזור הדחייה מוכתב על ידי סיכון שלוקח החוקר מראש שניקרא רמת מובהקות ומסומן ב-  $\alpha$ .

#### שלב ה:

בתהליך יש ללכת לתוצאות המדגם ולחשב את הסטטיסטי המתאים ולבדוק האם התוצאות נופלות באזור הדחייה או הקבלה.

#### שלב ו:

להסיק מסקנה בהתאם לתוצאות המדגם.

**דוגמה:** ( פתרון בהקלטה)

משרד הבריאות פרסם שמשקל ממוצע של תינוקות ביום היוולדם בישראל 3300 גר'. משרד הבריאות רוצה לחקור את הטענה שנשים מעשנות בזמן ההיריון יולדות תינוקות במשקל נמוך מהממוצע. במחקר השתתפו 20 נשים מעשנות בהריון. להלן תוצאות המדגם שבדק את המשקל של התינוקות בעת הלידה:

$$n = 20$$

$$\bar{X} = 3120$$

$$S = 280$$

א. מהי אוכלוסיית המחקר?

ב. מה המשתנה הנחקר?

ג. מה הפרמטר הנחקר?

ד. מהן השערות המחקר?

### תרגילים:

1. ממוצע הציונים בבחינת הברורות 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 אקראיים. ממוצע הציונים של אותם תלמידים לאחר שלמדו בשיטתו היה 75.5.

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

2. לפי הצהרת היצרן של חברת משקאות מסוימת נפח הנוזל בבקבוק מתפלג נורמלית עם תוחלת 500 סמ"ק וסטיית תקן 20 סמ"ק. אגודת הצרכנים מתלוננת על הפחתת נפח המשקה בבקבוק מהכמות המוצהרת. במדגם שעשתה אגודת הצרכנים התקבל נפח ממוצע של 492 סמ"ק במדגם בגודל 25.

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

3. במשך שנים אחוז המועמדים שהתקבל לפקולטה למשפטים היה 25%. השנה מתוך מדגם של 120 מועמדים התקבלו 22. מחקר מעוניין לבדוק האם השנה מקשים על הקבלה לפקולטה למשפטים.

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

4. בחודש ינואר השנה פורסם שאחוז האבטלה במשק הוא 8% במדגם עכשווי התקבל שמתוך 200 אנשים 6.5% מובטלים. רוצים לבדוק ברמת מובהקות של 5% האם כיום אחוז האבטלה הוא כמו בתחילת השנה.

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

## טעויות בבדיקת השערות

### רקע:

בתהליך של בדיקת השערות יוצרים כלל שניקרא כלל הכרעה :

הכלל יוצר אזור שניקרא אזור דחייה ( דחייה של השערת האפס כלומר קבלה של האלטרנטיבה) ואזור קבלה ( קבלה של השערת האפס ודחייה של האלטרנטיבה). כלל ההכרעה מתבסס על איזשהו סטטיסטי .

בתהליך יש ללכת לתוצאות המדגם ולבדוק האם התוצאות נופלות באזור הדחייה או הקבלה וכך להגיע למסקנה – המסקנה היא בעירבון מוגבל כיוון שהיא תלויה בכלל ההכרעה ובתוצאות המדגם. נשנה את כלל ההכרעה אנחנו יכולים לקבל מסקנה אחרת . נבצע מדגם חדש אנחנו עלולים לקבל תוצאה אחרת.

לכן יתכנו טעויות במסקנות שלנו :

		הכרעה	
		H0	H1
מציאות	H0	אין טעות	טעות מסוג 1
	H1	טעות מסוג 2	אין טעות

### הגדרת הטעויות:

טעות מסוג ראשון- להכריע לדחות את  $H_0$  למרות שבמציאות  $H_0$  נכונה.

טעות מסוג שני- להכריע לקבל את  $H_0$  למרות שבמציאות  $H_1$  נכונה.

דוגמה: (פתרון בהקלטה)

אדם חשוד בביצוע עבירה ונתבע בבית המשפט. אילו סוגי טעויות אפשריות בהכרעת הדין?

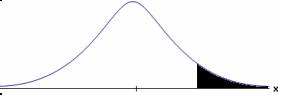


## תרגילים:

1. לפי הצהרת היצרן של חברת משקאות מסוימת נפח הנוזל בבקבוק מתפלג נורמלית עם תוחלת 500 סמ"ק וסטיית תקן 20 סמ"ק. אגודת הצרכנים מתלוננת על הפחתת נפח המשקה בבקבוק מהכמות המוצהרת. במדגם שעשתה אגודת הצרכנים התקבל נפח ממוצע של 492 סמ"ק במדגם בגודל 25. בסופו של דבר הוחלט להכריע לטובת חברת המשקאות.
  - א. רשמו את השערות המחקר.
  - ב. מה מסקנת המחקר?
  - ג. איזו סוג טעות יתכן וביצעו במחקר?
  
2. במחקר על פרמטר מסוים הוחלט בסופו של דבר לדחות את השערת האפס.
  - א. האם ניתן לדעת אם בוצע טעות במחקר?
  - ב. 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 שכיום ממוצע מספר הילדים במשפחה קטן יותר. לצורך כך הוחלט לדגום 121 משפחות. במדגם התקבל ממוצע 2.17 ילדים למשפחה. על סמך תוצאות המדגם נקבע שלא ניתן לקבוע שבאופן מובהק תוחלת מספר הילדים למשפחה קטנה כיום.
  - א. מהי אוכלוסיית המחקר?
  - ב. מה המשתנה הנחקר?
  - ג. מה הפרמטר הנחקר?
  - ד. מה השערות המחקר?
  - ה. מה מסקנת המחקר?
  - ו. מהי סוג הטעות האפשרית במחקר?

## פרק 34 - בדיקת השערות על תוחלת (ממוצע)

### כאשר שונות האוכלוסיה ידועה

#### רקע:

$H_0 : \mu = \mu_0$ $H_1 : \mu > \mu_0$	$H_0 : \mu = \mu_0$ $H_1 : \mu < \mu_0$	$H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$	<b>השערות האפס :</b> <b>השערה אלטרנטיבה :</b>
1. $\sigma$ ידועה 2. $X \sim N$ או מדגם מספיק גדול			<b>תנאים :</b>
$Z_{\bar{x}} > Z_{1-\alpha}$  $Z_{1-\alpha}$ - דוחים את $H_0$	$Z_{\bar{x}} < -Z_{1-\alpha}$  $-Z_{1-\alpha}$ - דוחים את $H_0$	$Z_{\bar{x}} < -Z_{1-\frac{\alpha}{2}}$ או $Z_{\bar{x}} > Z_{1-\frac{\alpha}{2}}$  $-Z_{1-\frac{\alpha}{2}}$ $Z_{1-\frac{\alpha}{2}}$ - דוחים את $H_0$	<b>כלל ההכרעה :</b> <b>אזור הדחייה של <math>H_0</math> :</b>

#### סטטיסטי המבחן :

$$Z_{\bar{x}} = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

חלופה אחרת לכלל הכרעה :

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



**דוגמה** : (פתרון בהקלטה)

יבול העגבניות מתפלג נורמלית עם תוחלת של 10 טון לדונם וסטיית תקן של 2.5 טון לדונם בעונה. משערים ששיטת זיבול חדשה תעלה את תוחלת היבול לעונה מבלי לשנות את סטיית התקן. נדגמו 4 חלקות שזובלו בשיטה החדשה. היבול הממוצע שהתקבל היה 12.5 טון לדונם. בדוק את ההשערה ברמת מובהקות של 1%.

### תרגילים:

1. ממוצע הציונים בבחינת 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 אקראיים. ממוצע הציונים של אותם תלמידים לאחר שלמדו בשיטתו היה 75.5. בהנחה שגם בשיטתו סטיית התקן תהייה 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. אגודת הצרכנים מתלוננת על הפחתת נפח המשקה בבקבוק מהכמות המוצהרת. במדגם שעשתה אגודת הצרכנים התקבל נפח ממוצע של 492 סמ"ק במדגם בגודל 25.
  - א. מה מסקנתכם ברמת מובהקות של 2.5%?
  - ב. האם ניתן לדעת מה תהיה המסקנה עבור רמת מובהקות הגבוהה מ-5%?
3. מהנדס האיכות מעוניין לבדוק אם מכונה מכוילת (מאופסת). המכונה כוונה לחתוך מוטות באורך 50 ס"מ. לפי נתוני היצרן סטיית התקן בחיתוך המוטות היא 0.5 ס"מ. במדגם של 50 מוטות התקבל ממוצע אורך המוט 50.93 ס"מ. מה מסקנתכם ברמת מובהקות של 5%?
4. המשקל הממוצע 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 שצריכה להביא להורדת המשקל. לשם בדיקת יעילות הדיאטה נלקח מדגם מקרי של 50 ספורטאים ובתום שנה של שימוש בדיאטה התברר שהמשקל הממוצע במדגם 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. במדגם של 25 ברגים העובי הממוצע 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? בחר בתשובה הנכונה.
- הגדלת רמת המובהקות לא תשתנה את מסקנת המחקר.
  - הגדלת רמת המובהקות תשנה את מסקנת המחקר.
  - הקטנת רמת המובהקות לא תשנה את מסקנת המחקר.
  - הקטנת רמת המובהקות תשנה את מסקנת המחקר.
7. חוקר ערך מבחן דו צדדי ברמת מובהקות של  $\alpha$  והחליט לדחות את השערת האפס.
- אם החוקר היה עורך מבחן צדדי ברמת מובהקות של  $\frac{\alpha}{2}$  אזי בהכרח: (בחר בתשובה הנכונה)
- השערת האפס הייתה נדחית.
  - השערת האפס הייתה לא נדחית.
  - לא ניתן לדעת מה תהיה מסקנתו במקרה זה.
8. שני סטטיסטיקאים בדקו השערות  $H_0: \mu = \mu_0$  כנגד  $H_1: \mu > \mu_0$  עבור שונות ידועה ובאותה רמת מובהקות. שני החוקרים קבלו אותו ממוצע במדגם אך לחוקר א' היה מדגם בגודל 100 ולחוקר ב' מדגם בגודל 200.
- אם חוקר א' החליט לדחות את  $H_0$ , מה יחליט חוקר ב'? נמקו.
  - אם חוקר א' יחליט לא לדחות את  $H_0$ , מה יחליט חוקר ב'? נמקו.

**פתרונות :****שאלה 1:** $H_0$  נקבל**שאלה 2:** $H_0$  נדחה**שאלה 3:** $H_0$  נדחה**שאלה 4:** $H_0$  נדחה**שאלה 5:** $H_0$  נקבל**שאלה 6:**

א

**שאלה 7:**

ג

**שאלה 8:**

א. אותה מסקנה

ב. לא ניתן לדעת.

**סיכוי לטעויות ועוצמה כאשר שונות האוכלוסייה ידועה**

**רקע:**

	הכרעה		
		H0	H1
מציאות	H0	אין טעות	טעות מסוג 1
	H1	טעות מסוג 2	אין טעות

נגדיר את ההסתברויות הבאות:

הסיכוי לבצע טעות מסוג 1 (רמת מובהקות)

$$\alpha = P(H_0 \text{ לדחות את } H_0 | \text{לדחות את } H_0) = P_{H_0}(H_0)$$

הסיכוי לבצע טעות מסוג 2:

$$\beta = P(H_0 \text{ לקבל את } H_1 | \text{לקבל את } H_0) = P_{H_1}(H_0)$$

רמת בטחון:

$$(1 - \alpha) = P(H_0 \text{ לקבל את } H_0 | \text{לקבל את } H_0) = P_{H_0}(H_0)$$

עוצמה:

$$\pi = (1 - \beta) = P(H_1 \text{ נכונה} | \text{לדחות את } H_0) = P_{H_1}(H_0)$$

התהליך לחישוב סיכוי לטעות מסוג שני:

$H_0 : \mu = \mu_0$ $H_1 : \mu > \mu_0$	$H_0 : \mu = \mu_0$ $H_1 : \mu < \mu_0$	$H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$	השערת האפס : השערה אלטרנטיבה :
3. $\sigma$ ידועה 4. $X \sim N$ או מדגם מספיק גדול			תנאים :
$\bar{X} > \mu_0 + Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$	$\bar{X} < \mu_0 - Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$	$\bar{X} > \mu_0 + Z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$ או $\bar{X} < \mu_0 - Z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$	כלל ההכרעה : אזור הדחייה של $H_0$ :
$P_{H_1}(\bar{X} < \mu_0 + Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}})$	$P_{H_1}(\bar{X} > \mu_0 - Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}})$	$P_{H_1}(\mu_0 - Z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \bar{X} < \mu_0 + Z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}})$	חישוב $\beta$ :

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

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

**דוגמה :** (פתרון בהקלטה)

בתחילת השנה חשבון הטלפון הסלולארי הממוצע לאדם היה 200 ₪ עם סטיית תקן של 80 ₪ לחודש. בעקבות כניסתן של חברות טלפון סלולארית חדשות מעוניינים לבדוק האם כיום ממוצע חשבון הטלפון הסלולארי פחת. לצורך בדיקה דגמו באקראי 36 אנשים וחשבון הטלפון הסלולארי שלהם היה 150 ₪ בממוצע לחודש.

א. רשמו את השערות המחקר ובנו כלל הכרעה במונחי חשבון ממוצע מדגמי ברמת מובהקות של 5%.

ב. מה מסקנתכם? איזה סוג טעות אפשרית במסקנה?

ג. נניח שבמציאות כיום החשבון הממוצע הוא 160 ₪. מה הסיכוי לבצע טעות מסוג שני?

ד. אם נקטין את רמת המובהקות מסעיף א', כיצד הדבר ישפיע על התשובה מסעיף ג'?

**תרגילים:**

לפתרון מלא בסרטון וידאו היכנסו ל- [www.GooL.co.il](http://www.GooL.co.il)

כתב ופתר - ברק קנדל ©

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 כלל הכרעה המתבסס על הסמך תצפית בודדת כך שרמת המובהקות תהיה 5%.

א. עבור אילו ערכים של  $X$  שידגם נדחית 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 סטיית תקן

0.4. מעוניינים לבדוק אם כיום 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 קריטי ברמת מובהקות של 5%.

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

ג. אם באמת ממוצע מספר הילדים במשפחה פחת לכדי 2.1 מהי העצמה של הכלל

מסעיף א?

3. להלן נתונים על תהליך של בדיקת השערות על תוחלת :

$$H_0 : \mu = 200$$

$$H_1 : \mu \neq 200$$

$$\sigma = 30$$

$$n = 225$$

א. רשום כלל הכרעה במונחי ממוצע מדגם קריטי וברמת מובהקות של 10%.

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

ג. הסבר ללא חישוב איך העצמה תשתנה אם רמת המובהקות תהייה 5%?

4. מפעל לייצור צינורות מייצר צינור שקוטרו מתפלג נורמלית עם תוחלת של 50 מ"מ וסטית תקן של 6 מ"מ. במחלקת ביקורת האיכות דוגמים בכל יום 81 צינורות ומודדים את קוטרים,



בכדי לבדוק, בעזרת מבחן סטטיסטי, האם מכונת הייצור מכוילת כנדרש או שקוטר הצינורות קטן מהדרוש.

- א. רשום את ההשערות ואת כלל ההכרעה ברמת מובהקות של 5%.
- ב. אם ביום כלשהו מכונת הייצור התקלקלה והיא מייצרת את הצינורות בקוטר שתוחלתו 48 מ"מ בלבד (סטית התקן לא השתנתה), מה ההסתברות שהתקלה לא תתגלה בביקורת האיכות? כיצד נקראת הסתברות זו?
- ג. הסבר ללא חישוב כיצד התשובה לסעיף ב תשתנה אם רמת המובהקות תגדל.
- ד. הסבר ללא חישוב כיצד התשובה לסעיף ב תשתנה 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 תצפיות. ידוע שסטיית 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) ?
1. סטיית התקן הייתה יותר גדולה.
  2. הסיכוי לטעות מסוג שני גדול יותר.

השאלות שלהלן הן שאלות רב בררתיות. בחר בכל שאלה את התשובה הנכונה ביותר :

6. אם חוקר החליט להגדיל את רמת המובהקות במחקר שלו אזי :

א. הסיכוי לטעות מסוג ראשון גדל.

ב. העוצמה של המבחן גדלה.

ג. הסיכוי לטעות מסוג שני גדל.

ד. תשובות א ו-ב נכונות.

7. חוקר ביצע מחקר ובו עשה טעות מסוג שני לכן :

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 בדיקת השערה :

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

9. נערך שינוי בכלל ההחלטה של בדיקת השערה מסוימת ובעקבותיו אזור דחיית

$H_0$  קטן. כל שאר הגורמים נשארו ללא שינוי. כתוצאה מכך :

א. הן  $\alpha$ , והן  $(1 - \beta)$ , יקטנו.

ב.  $\alpha$  יישאר ללא שינוי ואילו  $(1 - \beta)$  יגדל.

ג.  $\alpha$  יגדל ואילו  $(1 - \beta)$  יקטן.

ד. הן  $\alpha$  והן  $(1 - \beta)$  יגדלו.

10. ידוע כי לחץ דם תקין באוכלוסייה הוא 120. רופא מניח שלחץ הדם בקרב

עיתונאים גבוה יותר מהממוצע באוכלוסייה. הוא לקח מדגם של 60 עיתונאים

וקיבל ממוצע 137.

על סמך המדגם, הוא בודק טענתו ברמת מובהקות 0.02 ומסיק שלחץ הדם בקרב

העיתונאים אינו גבוה יותר. מה הטעות האפשרית שהרופא עושה ?

א. טעות מסוג ראשון.

ב. טעות מסוג שני.

ג. טעות מסוג שלישי.

ד. אין טעות במסקנתו.

**פתרונות :****שאלה 1:**

א. מעל 6.645

ב. 0.3632

**שאלה 2:**א. נדחה  $H_0$  אם  $\bar{X} < 2.24$ ב. נדחה  $H_0$ 

ג. 1

**שאלה 3:**א. נדחה  $H_0$  אם  $\bar{X} > 203.29$  או  $\bar{X} < 196.71$ 

ב. 0.8051

ג. תקטן.

**שאלה 4:**א. נדחה  $H_0$  אם  $\bar{X} < 48.9$ 

ב. 0.0885

ג. תקטן.

ד. תקטן.

**שאלה 6:**

ד

**שאלה 7:**

ג

**שאלה 8:**

ג

**שאלה 9:**

א

**שאלה 10:**

ב

### קביעת גודל מדגם כששונות האוכלוסייה ידועה

#### רקע:

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu = \mu_1$$

השערות המחקר הן :

סטיית התקן של האוכלוסייה ידועה  $\sigma$  ומעוניינים לבצע מחקר שרמת המובהקות לא תעלה על  $\alpha$  והסיכוי לטעות מסוג שני לא יעלה על  $\beta$ .

הנוסחה הבאה נותנת את גודל המדגם הרצוי :

$$n \geq \left( \frac{(Z_{1-\alpha} + Z_{1-\beta}) \times \sigma}{\mu_0 - \mu_1} \right)^2$$

דוגמה: (פתרון בהקלטה)

משרד החינוך מפעיל בגן חובה שיטת חינוך שפותחה בשנת 1995. לפי שיטת חינוך זו תוחלת הציון במבחן אוצר מילים לגיל הרך הוא 70. אנשי חינוך החליטו לבדוק שיטת חינוך שפותחה בהולנד הנותנת שם תוחלת ציון אוצר מילים של 80.

נניח שציוני מבחן זה מתפלגים נורמאלית עם  $\sigma = 17$ .

כדי לבדוק האם גם בישראל הפעלת שיטת החינוך ההולנדית תעבוד בגנים, רוצים לבנות מחקר ברמת מובהקות של 5%. כמו כן, מעוניינים שאם בהפעלת השיטה ההולנדית תוחלת הציונים תעלה לכדי 80, המחקר יגלה זאת בסיכוי של 90%. כמה ילדי גן חובה דרושים למחקר?

**תרגילים:**

1. במבחן אינטליגנציה הציונים מתפלגים נורמאלית עם סטיית תקן 8 וממוצע 100. פסיכולוג מעוניין לבדוק את הטענה שבאוכלוסיות במצב סוציו אקונומי נמוך תוחלת הציונים היא 95. אם מעוניינים לגלות את הטענה בהסתברות של לפחות 99% כשרמת המובהקות היא 5% מהו גודל המדגם הדרוש?

2. משרד התקשורת טוענים שאדם מדבר בממוצע 180 דקות בחודש בטלפון הסלולרי. חברות הטלפון הסלולרי טוענות שאינפורמציה זו אינה נכונה ואדם מדבר בממוצע פחות : כ-160 דקות. לצורך פתרון נניח שסטיית התקן של זמן השיחה החודשי ידוע ושווה ל-60 דקות. כמה אנשים יש לדגום כך שאם טענת משרד התקשורת נכונה נדחה אותה בסיכוי של 5% (איך קוראים להסתברות זאת?) כמו כן אם טענת חברות הטלפון הסלולרית נכונה המחקר יגלה זאת בסיכוי של 90% (איך קוראים להסתברות זאת?)

$$3. \begin{aligned} H_0 : \mu = \mu_0 \\ H_1 : \mu = \mu_1 \end{aligned} \quad \text{השערות המחקר הן :}$$

כמו כן נתון שהמשתנה מתפלג נורמלית עם סטיית התקן ידועה  $\sigma$  מעוניינים לבצע מחקר שרמת המובהקות לא תעלה על  $\alpha$  והסיכוי לטעות מסוג שני לא יעלה על  $\beta$ .

הוכח שגודל המדגם הרצוי לכך יהיה :

$$n \geq \left( \frac{(Z_{1-\alpha} + Z_{1-\beta}) \times \sigma}{\mu_0 - \mu_1} \right)^2$$

**פתרונות :**

**שאלה 1:**

41

**שאלה 2 :**

78

**שאלה 3:**

הוכחה

## מובהקות התוצאה ( p-value ) בבדיקת השערות על תוחלת עם שונות ידועה

### רקע:

דרך נוספת להגיע להכרעות שלא דרך כלל הכרעה, היא דרך חישוב מובהקות התוצאה :  
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 שיהיו לו את התוצאות.  
המסקנה של המחקר תקבע לפי העיקרון הבא :

$$\text{אם } p_v \leq \alpha \text{ דוחים את } H_0$$

מובהקות התוצאה זה הסיכוי לקבלת תוצאות המדגם וקיצוני מתוצאות אלה בהנחת השערת האפס.

$$p_v = P_{H_0} \text{ (לקבל את תוצאות המדגם וקיצוני)}$$

אם ההשערה היא דו צדדית :

$$p_v = 2 P_{H_0} \text{ (לקבל את תוצאות המדגם וקיצוני)}$$

מובהקות התוצאה היא גם האלפא המינימלית לדחיית השערת האפס.

$H_0 : \mu = \mu_0$ $H_1 : \mu > \mu_0$	$H_0 : \mu = \mu_0$ $H_1 : \mu < \mu_0$	$H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$	השערת האפס : השערה אלטרנטיבה :
.5 $\sigma$ ידועה			תנאים :
.6 $X \square N$ או מדגם מספיק גדול			
$P_{H_0}(\bar{X} \geq \bar{x})$	$P_{H_0}(\bar{X} \leq \bar{x})$	אם $2 \cdot P_{H_0}(\bar{X} \geq \bar{x}) \leftarrow \bar{x} > \mu_0$ אם $2 \cdot P_{H_0}(\bar{X} \leq \bar{x}) \leftarrow \bar{x} < \mu_0$	p-value

כאשר בהנחת השערת האפס :  $\bar{X} \sim N(\mu_0, \frac{\sigma^2}{n})$

$$Z_{\bar{x}} = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

**דוגמה:** (פתרון בהקלטה)

המשקל הממוצע של מתגייסים לצבא לפני 20 שנה היה 65 ק"ג. מחקר מעוניין לבדוק האם כיום המשקל הממוצע של מתגייסים גבוה יותר. נניח שמשקל המתגייסים מתפלג נורמאלית עם סטיית תקן של 12 ק"ג. במדגם של 16 מתגייסים התקבל משקל ממוצע של 71 ק"ג.

א. מהי מובהקות התוצאה?

ב. מה המסקנה אם רמת המובהקות היא 5% ואם רמת המובהקות היא 1%?



**תרגילים:**

1. לפניך השערות של מחקר :

$$H_0 : \mu = 70$$

$$H_1 : \mu > 70$$

המשתנה הנחקר מתפלג נורמלית עם סטיית תקן 20. במדגם מאותה אוכלוסייה התקבלו התוצאות הבאות :

$$n = 100$$

$$\bar{x} = 74$$

מהי מובהקות התוצאה?

2. השכר הממוצע במשק בשנת 2012 היה 8800 ₪ עם סטיית תקן 2000. במדגם שנעשה אתמול על 100 עובדים התקבל שכר ממוצע 9500 ₪. מטרת המחקר היא לבדוק האם כיום חלה עליה בשכר. עבור אילו רמות מובהקות שיבחר החוקר יוחלט שחלה עליה בשכר הממוצע במשק?

3. אדם חושד שחברת ממתקים לא עומדת בהתחייבויותיה, ומשקלו של חטיף מסוים אותו הוא קונה מדי בוקר נמוך מ – 100 גרם. חברת הממתקים טוענת מצידה שהיא אכן עומדת בהתחייבויותיה. ידוע כי סטית התקן של משקל החטיף היא 12 גרם. האדם מתכוון לשקול 100 חפיסות חטיפים ולאחר מכן 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 נקבל את השערת האפס?

ד. מה המסקנה ברמת מובהקות של 5%?

4. מכונה לחיתוך מוטות במפעל חותכת מוטות באורך שמתפלג נורמאלית עם תוחלת אליה כוונה המכונה וסטיית תקן 2 ס"מ. ביום מסוים כוונה המכונה לחתוך מוטות באורך 80 ס"מ. אחראי האיכות מעוניין לבדוק האם המכונה מכוילת. לצורך כך נדגמו מקו הייצור 16 מוטות שנחתכו אורכן הממוצע היה 81.7 ס"מ.

א. מהי רמת המובהקות המינימלית עבורה נכריע שהמכונה לא מכוילת?

ב. אם נוסף עוד תצפית שערכה יהיה 82 ס"מ, כיצד הדבר ישפיע על התשובה של הסעיף הקודם?

ג. הכרע ברמת מובהקות של 5% האם המכונה מכוילת.

5. אם מקבלים בחישובים אלפא מינימלית (P value) קטנה מאוד, סביר להניח כי החוקר ידחה את השערת האפס בקלות. נכון? לא נכון? נמק.
6. בבדיקת השערות התקבל שה-  $p\text{-value}=0.02$ .  
 מה תהיה מסקנת חוקר המשתמש ברמת מובהקות 1%? בחר בתשובה הנכונה.  
 א. יקבל את השערת האפס בכל מקרה.  
 ב. ידחה את השערת האפס מקרה.  
 ג. ידחה את השערת האפס רק אם המבחן הנו דו צדדי.  
 ד. לא ניתן לדעת כי אין מספיק נתונים.
7. מובהקות התוצאה (PV) היא גם : ( בחר בתשובה הנכונה )  
 א. רמת המובהקות המינימאלית לדחות השערת האפס.  
 ב. רמת המובהקות המקסימאלית לדחיית השערת האפס.  
 ג. רמת המובהקות שנקבעת מראש על ידי החוקר טרם קיבל את תוצאות המחקר.  
 ד. רמת המובהקות המינימאלית לאי דחיית השערת האפס.
8. בבדיקת השערות מסוימת התקבל  $p\text{ value}=0.0254$  לכן (בחר בתשובה הנכונה):  
 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  $H_0$ .

**פתרונות :****שאלה 1:**

0.0228

**שאלה 2:**

עבור כל רמת מובהקות סבירה.

**שאלה 3:**

ב. 0.1056

ג. 0.1056

ד. נכריע שיש עמידה בהתחייבות של החברה.

**שאלה 4:**

א. 0.0006

ב. יקטן.

ג. נכריע שאין כיוול.

**שאלה 5:**

נכון

**שאלה 6:**

תשובה: א

**שאלה 7:**


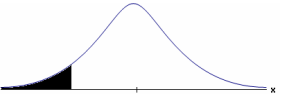
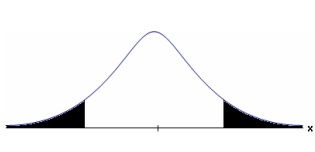
תשובה: א

**שאלה 8:**

תשובה: ג

**בדיקת השערות על תוחלת (ממוצע) כאשר שונות האוכלוסייה אינה ידועה**

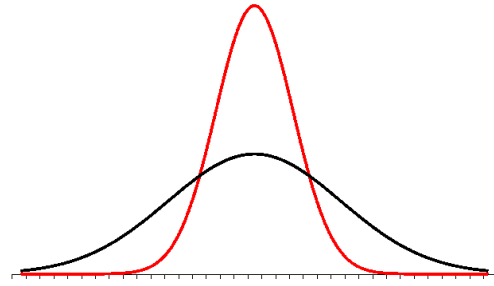
**רקע:**

$H_0 : \mu = \mu_0$ $H_1 : \mu > \mu_0$	$H_0 : \mu = \mu_0$ $H_1 : \mu < \mu_0$	$H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$	<b>השערת האפס :</b> <b>השערה אלטרנטיבית:</b>
<b>תנאים:</b> 7. $\sigma$ אינה ידועה 8. $X \square N$ או מדגם מספיק גדול			
$t_{\bar{x}} > t_{1-\alpha}^{(n-1)}$  $t_{1-\alpha, n-1}$ <b>דוחים את <math>H_0</math> -</b>	$t_{\bar{x}} < -t_{1-\alpha}^{(n-1)}$  $-t_{1-\alpha, n-1}$ <b>דוחים את <math>H_0</math> -</b>	$t_{\bar{x}} < -t_{1-\frac{\alpha}{2}}^{(n-1)}$ או $t_{\bar{x}} > t_{1-\frac{\alpha}{2}}^{(n-1)}$  $-t_{1-\frac{\alpha}{2}, n-1}$ $t_{1-\frac{\alpha}{2}, n-1}$ <b>דוחים את <math>H_0</math> -</b>	<b>כלל ההכרעה:</b> אזור הדחייה של $H_0$ :
$\bar{X} > \mu_0 + t_{1-\alpha}^{n-1} \cdot \frac{S}{\sqrt{n}}$	$\bar{X} < \mu_0 - t_{1-\alpha}^{n-1} \cdot \frac{S}{\sqrt{n}}$	$\bar{X} > \mu_0 + t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S}{\sqrt{n}}$ $\bar{X} < \mu_0 - t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S}{\sqrt{n}}$	<b>חלופה לכלל הכרעה :</b> נדחה $H_0$ אם מתקיים:

סטטיסטי המבחן :

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

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

**התפלגות T:**

הינה התפלגות סימטרית פעמונית שהתוחלת שלה היא 0. ההתפלגות דומה להתפלגות Z רק שהיא יותר רחבה ולכן הערכים שלה יהיו יותר גבוהים. התפלגות T תלויה במושג שנקרא דרגות חופש. דרגות החופש הן  $df=n-1$ . ככל שדרגות החופש עלות ההתפלגות הופכת להיות יותר גבוהה וצרה. כשדרגות החופש שואפות 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 נדגמו 10 משטחים ונמצא שהעובי הממוצע הוא 0.104 עם אומדן לסטיית תקן 0.002 ס"מ.  
 א. מהן השערות המחקר?  
 ב. מה ההנחה הדרושה לצורך פתרון?  
 ג. בדוק ברמת מובהקות של 5%.

**תרגילים:**

1. משך זמן ההחלמה בלקיחת אנטיביוטיקה מסוימת 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 האחרת התקבלו זמני ההחלמה הבאים: 90,95,100,80,125 שעות. מה מסקנתכם ברמת מובהקות של 5%. מהי ההנחה הדרושה לצורך הפתרון?

2. משרד הבריאות פרסם שמשקל ממוצע של תינוקות ביום היוולדם בישראל 3300 גר'. משרד הבריאות רוצה לחקור את הטענה שנשים מעשנות בזמן ההיריון יולדות תינוקות במשקל נמוך מהממוצע. 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 אינטליגנציה מתפלגים נורמלית. בארה"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 מבארה"ב? הסיקו ברמת מובהקות של 5%.

4. באוכלוסייה מסוימת נדגמו 10 תצפיות 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 התבססו 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.

א. אם ליאור ידחה את השערת האפס אז גם בהכרח רוני.

ב. אם רוני תדחה את השערת האפס אז גם בהכרח ליאור.

ג. שני החוקרים בהכרח יגיעו לאותה מסקנה.

ד. לא ניתן לדעת על היחס בין דחיית השערת האפס של שני החוקרים.

6. נתון ש  $X \sim N(\mu, \sigma^2)$  כמו כן נתונות ההשערות הבאות :

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu < \mu_0$$

חוקר בדק את ההשערות הללו על סמך מדגם שכלל 10 תצפיות.  $\sigma^2$  לא הייתה ידועה לחוקר.

החוקר החליט לדחות את השערת האפס ברמת מובהקות של 5% לאחר מכן כדי לחזק את

קביעתו הוא דגם עוד 5 תצפיות ושקלל את תוצאות אלה גם למדגם כך שכלל עכשיו 15

תצפיות.

בחר בתשובה הנכונה :

א. כעת בברור הוא ידחה את השערת האפס.

ב. כעת הוא דווקא יקבל את השערת האפס.

ג. כעת לא ניתן לדעת מה תהיה מסקנתו.

### פתרונות:

לפתרון מלא בסרטון וידאו היכנסו ל- [www.GooL.co.il](http://www.GooL.co.il)

כתב ופטר - ברק קנדל ©

**שאלה 1:** $H_0$  נדחה**שאלה 2:** $H_0$  נדחה**שאלה 3:** $H_0$  נקבל**שאלה 4:** $H_0$  נקבל**שאלה 5:**

התשובה היא : ב

**שאלה 6:**

התשובה היא : ג



## מובהקות התוצאה ( p-value ) בבדיקת השערות על תוחלת עם שונות אוכלוסייה לא ידועה

### רקע:

נזכיר שהמסקנה של המחקר תקבע לפי העיקרון הבא :

אם  $p_v \leq \alpha$  דוחים את  $H_0$

מובהקות התוצאה זה הסיכוי לקבלת תוצאות המדגם וקיצוני מתוצאות אלה בהנחת השערת האפס.

$$p_v = P_{H_0} \text{ (לקבל את תוצאות המדגם וקיצוני)}$$

אם ההשערה היא דו צדדית :

$$p_v = 2 P_{H_0} \text{ (לקבל את תוצאות המדגם וקיצוני)}$$

מובהקות התוצאה היא גם האלפא המינימלית לדחיית השערת האפס.

$H_0 : \mu = \mu_0$ $H_1 : \mu > \mu_0$	$H_0 : \mu = \mu_0$ $H_1 : \mu < \mu_0$	$H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$	השערת האפס : השערה אלטרנטיבה :
9. $\sigma$ אינה ידועה 10. $X \sim N$ או מדגם מספיק גדול			תנאים :
$P_{H_0}(\bar{X} \geq \bar{x})$	$P_{H_0}(\bar{X} \leq \bar{x})$	אם $2 \cdot P_{H_0}(\bar{X} \geq \bar{x}) \leftarrow \bar{x} > \mu_0$ אם $2 \cdot P_{H_0}(\bar{X} \leq \bar{x}) \leftarrow \bar{x} < \mu_0$	p-value

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

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

$d.f = n - 1$

דוגמה : (פתרון בהקלטה)

ממוצע זמן הנסיעה של אדם לעבודה 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 בדקות הם : 27,34,32,40,30 . הנח שזמן הנסיעה מתפלג נורמלית.

א. רשום את השערות המחקר.

ב. מצא חסמים למובהקות התוצאה.

ג. מה המסקנה ברמת מובהקות של 5% ?

**תרגילים :**

1. קו 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 :
- $$1008, 1024, 996, 1005, 997$$
- א. רשמו את השערות המחקר.  
 ב. מהי מובהקות התוצאה? הצג חסמים.  
 ג. מה המסקנה ברמת מובהקות של 5%?
2. חוקר בדק את הטענה כי פועלים העובדים במשמרת לילה איטיים יותר מפועלים העובדים ביום. ידוע כי משך הזמן הממוצע הדרוש לייצר מוצר מסוים ביום הוא 6 שעות. במדגם מיקרי של 25 פועלים שעבדו במשמרת לילה נמצא כי הזמן הממוצע לייצר אותו מוצר הוא 7 שעות עם סטית תקן של 3 שעות.  
 מהי ה- $\alpha$  המינימלית שלפיה ניתן להחליט שאכן העובדים במשמרת לילה איטיים יותר ?
3. הגובה של מתגייסים לצה"ל מתפלג נורמלית . במדגם של 25 מתגייסים מדדו את הגבהים שלהם בס"מ והתקבלו התוצאות הבאות :
- $$\bar{x} = 176.2$$
- $$\sum (x_i - \bar{x})^2 = 2832$$
- מטרת המחקר היא לבדוק האם תוחלת הגבהים של המתגייסים גבוה מ-174 ס"מ באופן מובהק. מהי בקרוב מובהקות התוצאה ועל פיה מה תהיה המסקנה ברמת מובהקות של 6% ?

**פתרונות :**

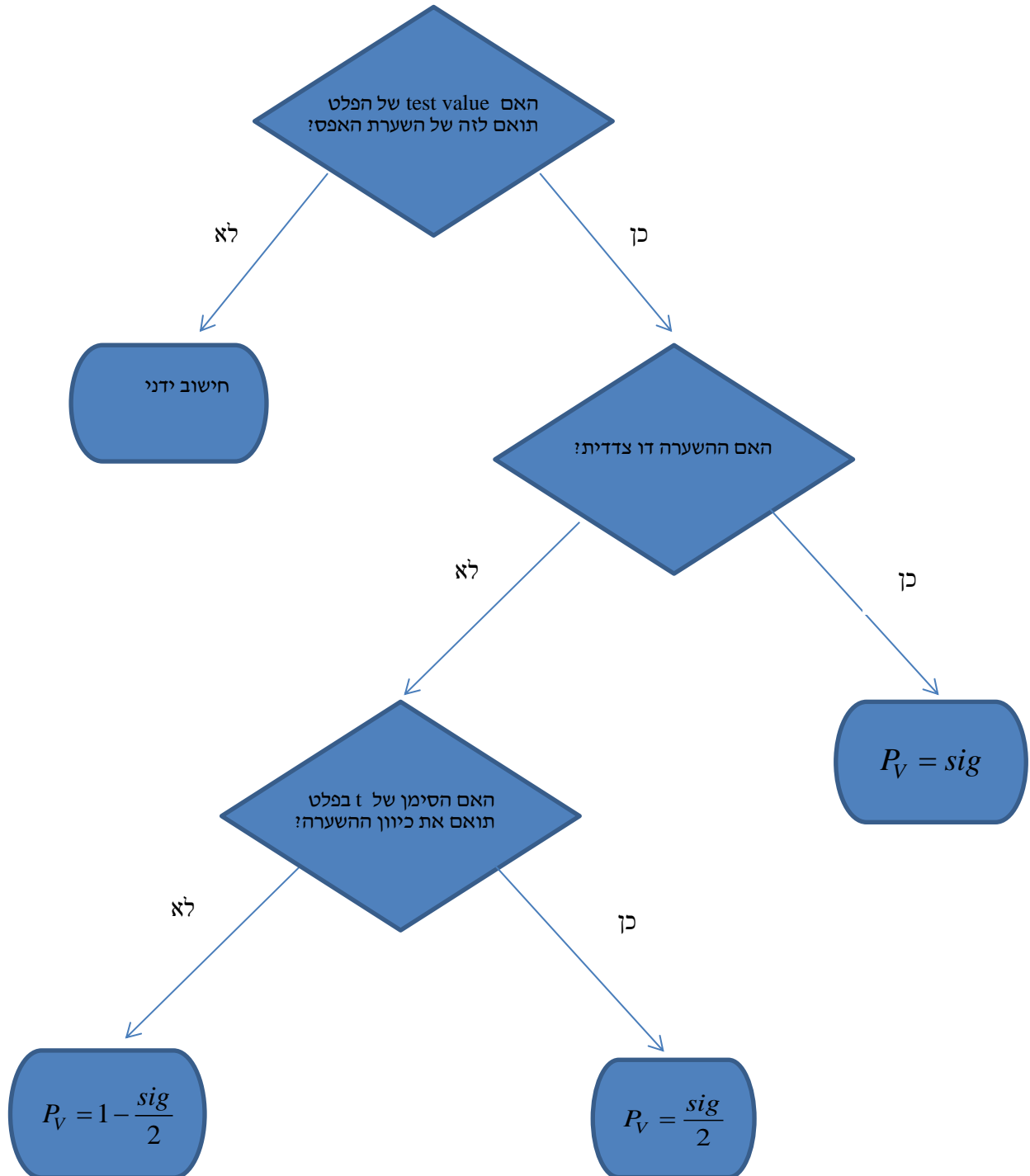
**שאלה 3:**

נקבל  $H_0$

## ניתוח פלטי SPSS

רקע:

חישוב מובהקות התוצאה באמצעות פלט תוכנת SPSS



**דוגמה:** (פתרון בהקלטה)

**One-Sample Statistics**

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

**One-Sample Test**

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

ממוצע הציונים במבחן המיצב בחשבון הוא 60. הוחלט לדגום כיתה אקראית של 25 תלמידים וללמד אותם בשיטת לימוד חדשה.

א. מהו רווח הסמך לממוצע הציונים בחשבון אם יוחלט ליישם את שיטת הלימוד החדשה?

ב. מהו  $P_V$  לבדיקת יעילותה של שיטת הלימוד החדשה?

ג. מה יוכרע ברמת מובהקות של 5% לגבי יעילותה של שיטת הלימוד החדשה?

**תרגילים :**

1. באוניברסיטה גדולה גיל הסטודנטים לתואר ראשון מתפלג נורמאלי. בעבר פורסם שהגיל הממוצע של הסטודנטים הינו 23. להלן פלט תוכנת SPSS על מדגם של 16 סטודנטים אקראיים 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 רווח סמך ברמת סמך של 95% לתוחלת גיל הסטודנטים באוניברסיטה לתואר ראשון.  
 ד. בדוק ברמת מובהקות של 5% האם הגיל הממוצע כיום שונה מבעבר?

2. קבוצת ילדים בגיל 6 קיבלה משימה לביצוע. עבור כל ילד בדקו כמה זמן לקח לו לסיים את המשימה בדקות. להלן תוצאות הניתוח 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 רווח סמך ברמת סמך של 95% לתוחלת זמן ביצוע המשימה עבור ילדים בני 6.  
 ג. מה יש להניח כדי שרווח הסמך מסעיף א' יהיה תקף?  
 ד. בדוק ברמת מובהקות של 5% שזמן ביצוע המשימה הממוצע נמוך מ-4.5 דקות.

3. לפניך פלט מחשב עבור ניתוח סטטיסטי שנעשה בתוכנת SPSS. הניתוח הוא עבור מדגם אקראי של קבוצת נבחנים בבגרות 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. השלם את הגדלים החסרים המסומנים בסמני שאלה בפלט.

ב. מהי מובהקות התוצאה לבדיקת ההשערה שהתוחלת של הציונים שונה מ-75?

ג. מהי מובהקות התוצאה לבדיקת ההשערה שהתוחלת של הציונים קטנה מ-75?

ד. מהי מובהקות התוצאה לבדיקת ההשערה שהתוחלת של הציונים גדולה מ-75?

4. יצרן סיגריות מפרסם כי תוחלת הניקוטין בסיגריות שהוא מיצר קטנה מ- 27 מ"ג. בבדיקה

מקרית של 5 סיגריות מתוצרתו נמצאו כמויות הניקוטין הבאות:

21, 21, 20, 24, 22 מ"ג. הנח כי כמות הניקוטין בסיגריות מפולג נורמאלי.

#### 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 הינו מוסיפים עוד תצפית שערכה 20. כיצד הדבר היה משפיע על הערך Sig ועל המסקנה?

ג. בדוק האם ניתן להגיד שתוחלת רמת הניקוטין שונה מ- 26 ברמת מובהקות של 5%.



**פתרונות :****שאלה 1:**

ב. 23

ג. (22.104, 24.771)

ד. נקבל  $H_0$ **שאלה 2:**

א. 25

ב. (4.3056, 4.5104)

ד. נדחה  $H_0$ **שאלה 3**א.  $n = 44$  $\bar{X} = 67.66$  $t = -2.48$  $upper = -1.3735$ 

ב. 0.017

ג. 0.0085

ד. 0.9915

**שאלה 4:**

א. נכריע שיש אמת בפרסום.

ב. מסקנה לא תשתנה.

ג. נכריע שהתוחלת שונה מ-26.

## הקשר בין רווח סמך לבדיקת השערות על תוחלת

### רקע:

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

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

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

אם  $\mu_0$  נופל ברווח  $\leftarrow$  נקבל את  $H_0$

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

### דוגמה: (פתרון בהקלטה)

חוקר ביצע בדיקת השערות לתוחלת. להלן השערותיו:

$$H_0 : \mu = 80$$

$$H_1 : \mu \neq 80$$

$$\alpha = 5\%$$

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

האם אפשר לדעת מה מסקנתו, ואם כן מהי?

**תרגילים :**

1. חוקר רצה לבדוק את ההשערות הבאות :

$$H_0 : \mu = 90$$

$$H_1 : \mu \neq 90$$

החוקר בנה רווח סמך לתוחלת ברמת סמך של 95% וקיבל את רווח הסמך הבא : (87,97).  
אם החוקר מעוניין לבצע בדיקת השערות ברמת מובהקות של 1% האם ניתן להגיע למסקנה ע"ס רווח הסמך? נמקו.

2. חוקר מעוניין לבדוק השפעת דיאטה חדשה על רמת הסוכר בדם. ידוע כי מספר מיליגרם 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 המיליגרם סוכר היה 115.5 מ"ג לסמ"ק.

א. בנה רווח סמך ברמת סמך 95% לתוחלת 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? הסבר.

3. יצרן אנטיביוטיקה רושם על גבי התרופות שכמות הפנצלין היא 200 מ"ג לקפסולה.  
משרד הבריאות ביצע מדגם של 8 קפסולות אקראיות מקו הייצור ומצא שבממוצע יש 196 מ"ג פנצלין לקפסולה עם סטיית תקן מדגמית של 5 מ"ג. בהנחה וכמות הפנצלין בקפסולה מתפלגת נורמלית.  
א. בנה רווח סמך ברמת סמך של 95% לממוצע כמות הפנצלין לקפסולה המיוצרת על ידי יצרן האנטיביוטיקה.  
ב. בדוק ברמת מובהקות של 5% האם יש אמת באינפורמציה המסופקת על ידי היצרן.

**פתרונות :****שאלה 1:**

1. נקבל השערת  $H_0$

**שאלה 2:**

א.  $112.87 \leq \mu \leq 118.13$

ב. נכריע שהדיאטה משפיעה על תוחלת רמת הסוכר בדם.

**שאלה 3:**

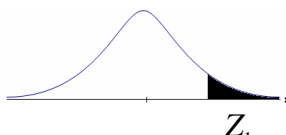
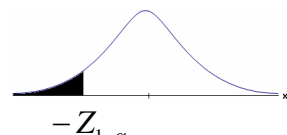
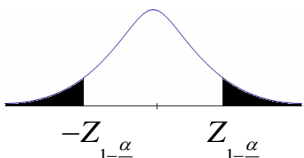
א.  $191.8 \leq \mu \leq 200.2$

ב. נכריע שיש אמת בפרסום.

## פרק 35 - בדיקת השערות על פרופורציה

### התהליך

**רקע:**

$H_0 : p = p_0$ $H_1 : p > p_0$	$H_0 : p = p_0$ $H_1 : p < p_0$	$H_0 : p = p_0$ $H_1 : p \neq p_0$	<b>השערות האפס :</b> <b>השערה אלטרנטיבית:</b>
$np_0 \geq 5 \ \& \ n(1-p_0) \geq 5$			<b>תנאים:</b>
$Z_{\hat{p}} > Z_{1-\alpha}$  $Z_{1-\alpha}$ <b>- דוחים את <math>H_0</math></b>	$Z_{\hat{p}} < -Z_{1-\alpha}$  $-Z_{1-\alpha}$ <b>- דוחים את <math>H_0</math></b>	או $Z_{\hat{p}} < -Z_{1-\frac{\alpha}{2}}$ $Z_{\hat{p}} > Z_{1-\frac{\alpha}{2}}$  $-Z_{1-\frac{\alpha}{2}}$ $Z_{1-\frac{\alpha}{2}}$ <b>- דוחים את <math>H_0</math></b>	<b>כלל ההכרעה:</b> <b>אזור הדחייה של :</b>

**סטטיסטי המבחן :**

$$Z_{\hat{p}} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

**חלופה אחרת לכלל הכרעה:**

$\hat{p} > p_0 + Z_{1-\alpha} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$	$\hat{p} < p_0 - Z_{1-\alpha} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$	$\hat{p} > p_0 + Z_{1-\alpha/2} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$ או $\hat{p} < p_0 - Z_{1-\alpha/2} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$	<b>כלל</b> <b>ההכרעה:</b> <b>אזור הדחייה</b> <b>של <math>H_0</math></b>
--	--	--	--

**דוגמה:** (פתרון בהקלטה)

בחודש ינואר השנה פורסם שאחוז האבטלה במשק הוא 8% במדגם עכשווי התקבל שמתוך 200 אנשים 6.5% מובטלים. בדקו ברמת מובהקות של 5% האם כיום אחוז האבטלה הוא כמו בתחילת השנה.

**תרגילים:**

1. במשך שנים אחוז המועמדים שהתקבל לפקולטה מסוימת היה 25%. השנה מתוך מדגם של 120 מועמדים התקבלו 22. ברמת מובהקות של 5% האם השנה הקשו על תנאי הקבלה?
2. במדגם של 300 אזרחים 57% מתנגדים להצעת חוק מסוימת. לאור נתונים אלה האם רוב האזרחים מתנגדים להצעת החוק? בדקו ברמת מובהקות של 10%.
3. הטילו מטבע 50 פעמים וקיבלו 28 פעמים עץ. האם המטבע הוגן ברמת מובהקות של 5%?
4. קפיטריה במכללה מסוימת מעריכה כי אחוז הסטודנטים שקונים קפה בקפיטריה הינו 20%. נערך סקר אשר כלל 200 סטודנטים. התברר כי 33 מהם רוכשים קפה בקפיטריה. מטרת הסקר הייתה לבדוק את אמיתות הערכה של הקפיטריה.
  - א. רשמו את ההשערות.
  - ב. בדוק את ההשערות ברמת מובהקות של 10%.
  - ג. מה תהיה המסקנה אם נקטין את רמת המובהקות?
5. חבר כנסת רוצה להעביר חוק. לצורך כך הוא דוגם 400 אזרחים במטרה לבדוק האם רוב האזרחים תומכים בחוק. במדגם התקבל ש-276 אזרחים תומכים בחוק.
  - א. מה מסקנתכם ברמת מובהקות של 5%?
  - ב. האם ניתן לדעת מה תהיה המסקנה אם רמת המובהקות תהיה גדולה יותר? הסבירו.
6. שני חוקרים בדקו את ההשערות הבאות:
 
$$H_0 : p = p_0$$

$$H_1 : p > p_0$$
 חוקר א השתמש ברמת מובהקות  $\alpha_1$  וחוקר ב ברמת מובהקות  $\alpha_2$  החוקר הראשון דחה את  $H_0$  ואילו החוקר השני קיבל את  $H_0$ . שניהם התבססו על אותם תוצאות של מדגם.
 

בחר בתשובה הנכונה:

  - א.  $\alpha_1 = \alpha_2$
  - ב.  $\alpha_1 > \alpha_2$
  - ג.  $\alpha_1 < \alpha_2$
  - ד. המצב המתואר לא אפשרי.

**פתרונות :****שאלה 1:** $H_0$  נדחה**שאלה 2:** $H_0$  נדחה**שאלה 3:** $H_0$  נקבל**שאלה 4:**ב. נקבל  $H_0$ 

ג. המסקנה לא תשתנה.

**שאלה 5:**א. נדחה  $H_0$ 

ב. המסקנה לא תשתנה.

**שאלה 6:**

התשובה היא : ג.



**סיכוי לטעויות ועוצמה****רקע:**

	הכרעה		
		H0	H1
מציאות	H0	אין טעות	טעות מסוג 1
	H1	טעות מסוג 2	אין טעות

**נגדיר את ההסתברויות הבאות:****הסיכוי לבצע טעות מסוג 1 ( רמת מובהקות ):**

$$\alpha = P(H_0 \text{ לדחות את } H_0) = P_{H_0}(H_0 \text{ לדחות})$$

**הסיכוי לבצע טעות מסוג 2:**

$$\beta = P(H_0 \text{ לקבל את } H_1) = P_{H_1}(H_0 \text{ לקבל})$$

**רמת בטחון:**

$$(1-\alpha) = P(H_0 \text{ לקבל את } H_0) = P_{H_0}(H_0 \text{ לקבל})$$

**עוצמה:**

$$\pi = (1-\beta) = P(H_1 \text{ לדחות את } H_0) = P_{H_1}(H_0 \text{ לדחות})$$

**התהליך לחישוב סיכוי לטעות מסוג שני:**לפתרון מלא בסרטון וידאו היכנסו ל- [www.GooL.co.il](http://www.GooL.co.il)

© כתב ופתר - ברק קנדל

$H_0 : p = p_0$ $H_1 : p > p_0$	$H_0 : p = p_0$ $H_1 : p < p_0$	$H_0 : p = p_0$ $H_1 : p \neq p_0$	השערת האפס : השערה אלטרנטיבית :
$np_0 \geq 5 \& n(1-p_0) \geq 5$			תנאים :
$\hat{p} > p_0 + Z_{1-\alpha} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$	$\hat{p} < p_0 - Z_{1-\alpha} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$	$\hat{p} > p_0 + Z_{1-\alpha/2} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$ או $\hat{p} < p_0 - Z_{1-\alpha/2} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$	כלל ההכרעה : אזור הדחייה של $H_0$
$P_{H_1}(\hat{p} < p_0 + Z_{1-\alpha} \cdot \sqrt{\frac{p_0(1-p_0)}{n}})$	$P_{H_1}(\hat{p} > p_0 - Z_{1-\alpha} \cdot \sqrt{\frac{p_0(1-p_0)}{n}})$	$P_{H_1}(p_0 - Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0(1-p_0)}{n}} < \hat{p} < p_0 + Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0(1-p_0)}{n}})$	חישוב $\beta$ :

כאשר :  $\hat{P} \sim N(p, \frac{p(1-p)}{n})$

והתקנון :  $Z_{\hat{p}} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$

**דוגמה:** (פתרון בהקלטה)

רופאי שיניים טוענים שיותר ממחצית האוכלוסייה הבוגרת בארץ אינם מבקרים אצל רופא שיניים באופן קבוע, כנדרש. כדי לבדוק טענה זו, נערך סקר בקרב 150 אנשים בוגרים.

א. רשמו את ההשערות וכלל הכרעה ברמת מובהקות של 10%.

ב. מהי עוצמת המבחן אם מסתבר ש 60% מהאוכלוסייה אינם מבקרים אצל רופא שיניים באופן קבוע.

### תרגילים:

1. משרד הבריאות פרסם ש 10% מתושבי המדינה סובלים ממחלת האסטמה. מחקר דורש לבדוק האם בחיפה, בגלל זיהום האוויר, שיעור הסובלים מאסטמה גבוה יותר. לצורך המחקר נבדקו 260 מתושבי חיפה.
  - א. רשמו את השערות המחקר, וצרו מבחן ברמת מובהקות של 5% לבדיקתן.
  - ב. מהי עצמת המבחן של סעיף א' בהנחה ובחיפה 16% מהתושבים סובלים מאסטמה?
  - ג. כיצד תשנה התשובה לסעיף ב' אם מסתבר שבחיפה 18% סובלים מאסטמה?
  - ד. בהמשך לסעיף א' האם נכון לומר שבהסתברות של 5% ההשערה שבחיפה 10% מהתושבים סובלים מאסטמה אינה נכונה?
  
2. אחוז הסובלים מתופעות הלוואי מתרופה מסוימת הוא 15%. חברת תרופות טוענת שפיתחה תרופה שאמורה לצמצם את אחוז הסובלים מתופעות לוואי. לצורך בדיקת הטענה הוחלט לבצע מחקר שיכלול 120 חולים שיקבלו את התרופה הנבדקת.
  - א. נניח שהתרופה נבדקת אכן מורידה את פרופורציות הסובלים מתופעות הלוואי ל-10% מהי עצמת המבחן עבור רמת מובהקות של 5%?
  
3. בעיר מסוימת היו 20% אקדמאים. בעקבות פתיחת מכללה בעיר לפני כמה שנים מעוניינים לבדוק האם אחוז האקדמאים גדל. מעוניינים שהמחקר יכלול 200 אנשים והוא יהיה ברמת מובהקות של 5%.
  - א. חשבו את הסיכוי לבצע טעות מסוג שני בהנחה והיום יש 28% אקדמאים.
  - ב. כיצד התשובה לסעיף הקודם תשתנה אם נגדיל את רמת המובהקות?
  
4. מעוניינים לבדוק האם בפקולטה מסוימת ישנה העדפה לגברים. הוחלט לדגום 200 מתקבלים ועל סמך מספר הבנים לקבוע אם טענת המחקר מתקבלת.
 

חוקר א' קבע רמת מובהקות של 5% וחוקר ב' החליט לקבל את טענת המחקר אם במדגם יהיו לפחות 120 בנים. למי מבין החוקרים רמת מובהקות גדולה יותר?
  
5. חוקר ביצע מחקר ובו עשה טעות מסוג שני לכן ( בחר בתשובה הנכונה )
  - א. השערת האפס נכונה.
  - ב. השערת האפס נדחתה.
  - ג. השערת האפס לא נדחתה.
  - ד. אף אחת מהתשובות לא נכונה בהכרח.
  
6. קבע אם הטענה הבאה נכונה:
 

"בבדיקת השערות לא ניתן לבצע בו זמנית טעות מסוג ראשון וטעות מסוג שני"

### פתרונות:

**שאלה 1:**

ב. 0.9015

ג. תגדל

ד. טענה לא נכונה.

**שאלה 2:**

0.4404

**שאלה 3:**

א. 0.1446

ב. תקטן.

**שאלה 4:**

חוקר א.

**שאלה 5:**

התשובה הנכונה היא ג.

**שאלה 6:**

נכונה.

**קביעת גודל מדגם****רקע:**

$$H_0 : p = p_0$$

$$H_1 : p = p_1$$

השערות המחקר הן :

מעוניינים לבצע מחקר שרמת המובהקות לא תעלה על  $\alpha$  והסיכוי לטעות מסוג שני לא יעלה על  $\beta$ .

הנוסחה הבאה נותנת את גודל המדגם הרצוי :

$$n \geq \left( \frac{Z_{1-\alpha} \sqrt{p_0 q_0} + Z_{1-\beta} \sqrt{p_1 q_1}}{p_0 - p_1} \right)^2$$

**דוגמה:** (פתרון בהקלטה)

רוצים לבדוק האם אחוז האנשים השוהים בשמש ללא הגנה ירד בעקבות הפרסומת על נזקי השמש. בעבר 60% מהאוכלוסייה שהתה בשמש ללא הגנה. מה גודל המדגם המינימלי שיש לקחת כדי לבדוק שהאחוז הנ"ל ירד ל-48% אם מעוניינים שהסיכוי לטעות מסוג ראשון יהיה 5% והסיכוי לטעות מסוג שני יהיה 1%?

**תרגילים:**

1. משרד התמי"ת פרסם שאחוז האבטלה במשק היום עומד על 8%. לעומתו, משרד הפנים טוען שחלה עלייה בשיעור האבטלה עד לכדי 11%.
- כדי לבדוק מי מבניהם צודק, מה צריך להיות גודל המדגם שיענה על שני התנאים הבאים:
- אם משרד התמי"ת צודק, נדחה את טענתו בסיכוי של 10%.
  - אם משרד הפנים צודק, נדחה את טענתו בסיכוי של 4%.
2. מפעיל קזינו מפרסם שהסיכוי לזכות במכונת מזל הינו 0.42.
- אדם טוען שהסיכויים לזכות במשחק נמוכים יותר. כמה פעמים יש לשחק את המשחק כדי שאם טענת מפעיל הקזינו נכונה נקבל את טענת האדם בסיכוי של 1% ואם במציאות הסיכוי לזכות במכונה הוא 0.3 נקבל את מפעיל הקזינו בסיכוי של 8%.

**פתרונות:**

**שאלה 1:**

891

**שאלה 2:**

224



### מובהקות התוצאה

#### רקע:

דרך נוספת להגיע להכרעות שלא דרך כלל הכרעה, היא דרך חישוב מובהקות התוצאה: 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 שיהיו לו את התוצאות. המסקנה של המחקר תקבע לפי העיקרון הבא:

$$\text{אם } p_v \leq \alpha \text{ דוחים את } H_0$$

מובהקות התוצאה זה הסיכוי לקבלת תוצאות המדגם וקיצוני מתוצאות אלה בהנחת השערת האפס.

$$p_v = P_{H_0} \text{ (לקבל את תוצאות המדגם וקיצוני)}$$

אם ההשערה היא דו צדדית :

$$p_v = 2 P_{H_0} \text{ (לקבל את תוצאות המדגם וקיצוני)}$$

מובהקות התוצאה היא גם האלפא המינימלית לדחיית השערת האפס.

$H_0 : p = p_0$ $H_1 : p > p_0$	$H_0 : p = p_0$ $H_1 : p < p_0$	$H_0 : p = p_0$ $H_1 : p \neq p_0$	השערת האפס : השערה אלטרנטיבית:
$np_0 \geq 5 \ \& \ n(1-p_0) \geq 5$			תנאים:
$P_{H_0}(\hat{P} \geq \hat{p})$	$P_{H_0}(\hat{P} \leq \hat{p})$	אם $2 \cdot P_{H_0}(\hat{P} \geq \hat{p}) \Leftarrow \hat{p} > p_0$ אם $2 \cdot P_{H_0}(\hat{P} \leq \hat{p}) \Leftarrow \hat{p} < p_0$	p-value

$$\hat{P} \sim N(p_0, \frac{p_0(1-p_0)}{n}) : \text{ כאשר בהנחת השערת האפס :}$$

והתקנון:

$$Z_{\hat{p}} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

**דוגמה:** (פתרון בהקלטה)

ישנה טענה שיש הבדל בין אחוז הבנים ואחוז הבנות הפונים ללמוד להנדסאי מחשבים. לשם כך נלקח מדגם מקרי של 200 תלמידים הלומדים מחשבים והתברר כי 112 מהם בנים.

א. מהי מובהקות התוצאה?

ב. מה המסקנה ברמת מובהקות של 5%?

**תרגילים:**

1. במשך שנים אחוז המועמדים שהתקבל לפקולטה מסוימת היה 25%. השנה מתוך מדגם של 120 מועמדים התקבלו 22. רוצים לבדוק האם השנה הקשו על תנאי הקבלה.
  - א. מהי מובהקות התוצאה?
  - ב. מה תהיה המסקנה ברמת מובהקות של 1% וברמת מובהקות של 5%?
  
2. נהוג לחשוב ש 60% מהילדים בגיל שלוש קמים מהמיטה במהלך הלילה לפחות פעם אחת. ישנה טענה שללא שנת צהריים פחות מ-60% מהילדים בגיל זה יקומו לפחות פעם אחת במהלך הלילה. נדגמו 80 ילדים בגיל 3 אשר אינם ישנים בצהריים מתוכם התקבל ש 41 קמו במהלך הלילה.
  - א. מהי רמת המובהקות המינימלית עבורה תתקבל הטענה במחקר?
  - ב. מהי רמת המובהקות המקסימלית עבורה לא תתקבל טענת המחקר?
  - ג. עבור אילו רמות מובהקות נקבל את טענת המחקר?
  - ד. מה תהיה מסקנת המחקר ברמת מובהקות של 6%?
  
3. במטרה לבדוק האם מטבע הוא הוגן מטילים אותו 80 פעמים. התקבל ש 60 מההטלות הראו עץ. רשמו את השערות המחקר, חשבו את מובהקות התוצאה והסיקו מסקנה ברמת מובהקות של 5%.
  
4. בבדיקת השערות על פרופורציה התקבל שה-  $p\text{-value}=0.02$ .
  - א. יקבל את השערת האפס
  - ב. ידחה את השערת האפס.
  - ג. לא ניתן לדעת כי אין מספיק נתונים.
  
5. קבע אם הטענה הבאה נכונה:
 

"במבחן לבדיקת השערות חד-צדדי התקבל ערך  $p\text{-value}$  של 3% לכן אם היינו מבצעים מבחן דו-צדדי (כאשר יתר הנתונים ללא שינוי) היינו מקבלים ערך  $p\text{-value}$  של 6%"
  
6. במפעל 10% מהעובדים נפגעים לפחות פעם אחת בשנה מתאונות עבודה. לאור זאת, המפעל החליט לצאת בתוכנית לצמצום שיעור הנפגעים. תכנית זו נוסתה על 100 עובדים. מתוכם 12 נפגעו בתאונות עבודה במשך השנה. מהי רמת המובהקות הקטנה ביותר עבורה יוחלט שהתכנית יעילה?

**פתרונות:**

לפתרון מלא בסרטון וידאו היכנסו ל- [www.GooL.co.il](http://www.GooL.co.il)

© כתב ופתר - ברק קנדל

**שאלה 1:**

א. 0.0455

**שאלה 2:**

א. 0.0548

ב. 0.0548

ג. מעל 0.0548

ד. נכריע לטובת טענת המחקר.

**שאלה 3:**

$$p_v = 0$$

**שאלה 4:**

התשובה הנכונה : ב

**שאלה 5:**

הטענה נכונה


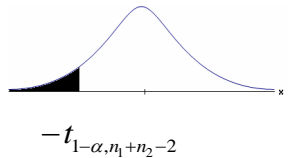

**שאלה 6:**

0.7486

## פרק 36 - בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

### כששונויות האוכלוסיה לא ידועות ומניחים שהן שוות

**רקע:**

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

**סטטיסטי המבחן :**

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

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

**חלופה אחרת לכלל הכרעה:**

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

**דוגמה :** (פתרון בהקלטה)

חברה המייצרת מוצרי בנייה טוענת שפיתחה סגסוגת (תערובת מתכות) שטמפרטורת ההתכה שלה גבוהה משמעותית מטמפרטורת ההתכה של הסגסוגת לבנייה שמשמשים בה כיום לבניית בניינים.

לצורך בדיקת טענת המחקר נדגמו 10 יחידות של מתכות מהסוג הישן ו-12 יחידות של מתכות מהסוג החדש.

**להלן תוצאות המדגם:**

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

טמפרטורת ההתכה הממוצעת במתכת החדשה 1317 מעלות עם אומד חסר הטיה לשונות  $S^2 = 260$ .

נניח לצורך פתרון שטמפרטורת ההתכה מתפלגת נורמאלית עם אותה שונות במתכות השונות. בדקו ברמת מובהקות של 5%.

**תרגילים:**

1. להלן נתונים של שטחי דירות מתוך דירות שנבנו בשנת 2012 ובשנת 2013 (מטרים רבועים):

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

בדקו שבשנת 2013 הייתה ירידה משמעותית בשטחי הדירות לעומת שנת 2012 עבור רמת מובהקות של 5%. הניחו ששטחי הדירות בכל שנה מתפלגים נורמלית עם אותה שונות.

2. נדגמו 15 ישראלים ו-15 אמריקאים.

הנדגמים נגשו למבחן IQ. להלן תוצאות המדגם:

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 במבחן ה-IQ לטובת ישראל. רשמו את כל ההנחות הדרושות לצורך פתרון התרגיל.

3. להלן תוצאות מדגם הבדק אורך חיים של נורות מסוג W60 ומסוג W100.

אורך החיים נמדד בשעות.

1-100W	2-60W	הקבוצה
956	1007	$\bar{x}$
72	80	S
15	13	n

א. בדקו ברמת מובהקות של 5% האם נורות מסוג W60 דולקות במוצע יותר מאשר נורות מסוג W100. רשמו את כל ההנחות הדרושות לפתרון.

ב. עבור איזו רמת מובהקות ניתן לקבוע שנורות מסוג W60 דולקות במוצע יותר מאשר נורות מסוג W100?

ג. בדקו ברמת מובהקות של 5% האם נורות מסוג W60 דולקות יותר מ-1000 שעות. רשמו את כל ההנחות הדרושות.

**פתרונות :****שאלה 1:**

לא נדחה  $H_0$

**שאלה 2:****שאלה 3:**

א. נדחה  $H_0$

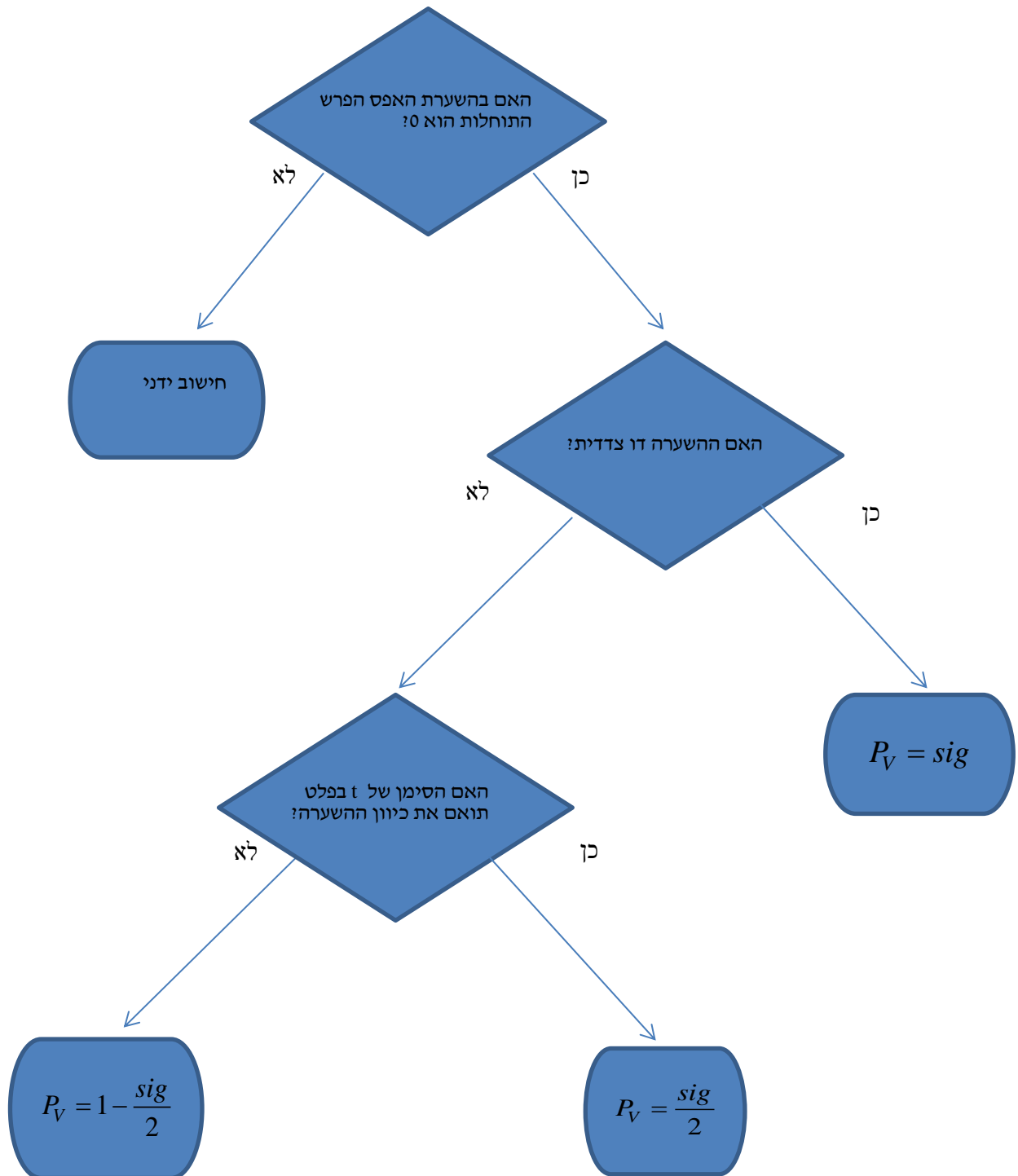
ב. רמות מובהקות של לפחות 5%

ג. לא נדחה  $H_0$



ניתוח פלטים**רקע:**

מובהקות התוצאה על סמך הפלט:

**דוגמה:** (פתרון בהקלטה)לפתרון מלא בסרטון וידאו היכנסו ל- [www.Gool.co.il](http://www.Gool.co.il)

© כתב ופתר - ברק קנדל

בסקר שנערך בארה"ב בשנת 1993 נשאלו נסקרים משני אזורים 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						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Number of Brothers and Sisters	Equal variances assumed	.173	.677	-1.583	1084	.114	-.293	.185	-.657	.070
	Equal variances not assumed			-1.576	850.945	.115	-.293	.186	-.658	.072

- א. מהו המבחן הסטטיסטי שנעשה כאן?
- ב. בדוק ברמת מובהקות של 5% האם קיים שוויון שונויות בין שני האזורים?
- ג. בדוק האם קיים הבדל בין "South East" ל-"North East" ברמת מובהקות של 5% מבחינת מספר האחים והאחיות הממוצע.
- ד. מהי מובהקות התוצאה לבדיקת הטענה שהפרש הממוצע בין-"South East" לבין-"North East" חיובי ?

**תרגילים :**

1. להלן פלט מתוכנת SPSS מתוך מחקר שבחן את רמת האופטימיות של גברים ונשים.  
רמת האופטימיות נמדדה בסקאלת 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 ששוונות 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 רווח הסמך המסומן בסימן שאלה בפלט. דייק עד 5 ספרות אחרי הנקודה.  
 ד. בנה רווח סמך לתוחלת רמת האופטימיות של הגברים ברמת סמך של 95%.

2. פסיכולוגים טוענים שאנשים שניגשים למבחן אינטליגנציה יותר מפעם אחת נוטים לקבל ציונים גבוהים יותר.  
להלן הפלט שהתקבל:

### T-Test

#### Group Statistic:

	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	Equal variances assumed	.206	.656	-2.428	18	.026	-11.56566	4.76333	-21.57304	-1.55828
	Equal variances not assumed			-2.479	17.997	.023	-11.56566	4.66583	-21.36832	-1.76299

#### מקרא:

A = נגשו פעם אחת.

B = נגשו יותר מפעם אחת.

- רשמו את השערות המחקר והסבירו מהו המבחן המתאים כאן.
  - כיצד הייתה משתנה התשובה לסעיף הקודם אם היה מדובר על אותם אנשים שציונם נבדק פעם אחרי המבחן הראשון שעשו ופעם אחרי המבחן השני?
  - האם ניתן לומר כי מידת הפיזור של ציוני אנשים הנבחנים בפעם הראשונה שונה ממידת הפיזור של ציוני האנשים אשר נבחנים בפעם השנייה. בדוק ברמת מובהקות של  $\alpha=0.05$ .
  - האם נכונה טענת הפסיכולוגים ברמת מובהקות של  $\alpha=0.01$ .
3. כחלק ממחקר בנושא הנישואין בישראל, אחד החוקרים העלה השערה שיש הבדל במוצג גיל הנישואין (הראשונים), בין נשים הגרות בערים מרכזיות לבין נשים הגרות בערים מרוחקות מהמרכז.
- לשם כך נדגמו 50 כלות מכל אחת משתי ערים עיר א'-מרכזית ועיר ב'-מרוחקת ונרשם גילן. תוצאות עיבוד הנתונים מופיעות בטבלאות שלהלן:

## T-Test

Group Statistics

מקום המגורים		N	Mean	Std. Deviation	Std. Error Mean
גיל הנישואין	עיר א	50	24.8072	1.38978	.19654
	עיר ב	50	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
	Equal variances not assumed			5.942	95.772	.000	1.79415	.30193	1.19480	2.39350

א. מהו המבחן הסטטיסטי שנעשה כאן?

ב. מצא רווח 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 בערים מרכזיות מתחתנות בגיל מאוחר יותר מאשר נשים הגרות בערים 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- ailed)	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. השלם את סימני השאלה בטבלה.

ב. מהי מובהקות התוצאה לבדיקת הטענה שקיים הבדל בין השונות של X לזה של Y?

ג. מהי מובהקות התוצאה לבדיקת הטענה שהתוחלת של X גדולה מהתוחלת של Y?

ד. מהי מובהקות התוצאה לבדיקת הטענה שהתוחלת של X קטנה מהתוחלת של Y?

**פתרונות:****שאלה 1:**

- א. נכריע שוויון שוניות.  
 ב. נקבע שלא קיים הבדל בין נשים לגברים מבחינת האופטימיות הממוצעת.  
 ג. 0.11076  
 ד.  $2.5665 \leq \mu \leq 2.6441$

**שאלה 2:**

- א. מבחן t למדגמים בלתי תלויים.  
 ב. מבחן t למדגמים מזווגים.  
 ג. נקבע שקיים שוויון שוניות.  
 ד. לא נקבל את טענת הפסיכולוגים.

**שאלה 3:**

- א. מבחן t להשוואת תוחלת במדגמים בלתי תלויים.  
 ב.  $1.19497 \leq \mu_1 - \mu_2 \leq 2.39332$   
 ג. כן.



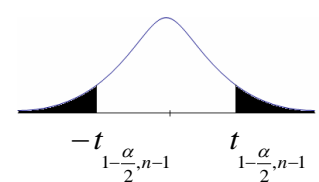
**שאלה 4:**

- א. -10.15, 4.69, 48  
 ב. 0.04  
 ג. 0.978  
 ד. 0.022

**פרק 37 - בדיקת השערות על תוחלת הפרשים במדגמים מזווגים  
(תלויים)**

**בדיקת השערות למדגמים מזווגים**

**רקע:**

$H_0: \mu_D = C$ $H_1: \mu_D > C$	$H_0: \mu_D = C$ $H_1: \mu_D < C$	$H_0: \mu_D = C$ $H_1: \mu_D \neq C$	<b>השערת האפס:</b> <b>השערה אלטרנטיבית:</b>
.4 $\sigma_D$ אינה ידועה .5 $D \square N$ או מדגם מספיק גדול			<b>תנאים:</b>
$t_{\bar{D}} > t_{1-\alpha}^{(n-1)}$  $t_{1-\alpha, n-1}$ <b>- דוחים את <math>H_0</math></b>	$t_{\bar{D}} < -t_{1-\alpha}^{(n-1)}$  $-t_{1-\alpha, n-1}$ <b>- דוחים את <math>H_0</math></b>	$t_{\bar{D}} < -t_{1-\frac{\alpha}{2}}^{(n-1)}$ או $t_{\bar{D}} > t_{1-\frac{\alpha}{2}}^{(n-1)}$  $-t_{1-\frac{\alpha}{2}, n-1}$ $t_{1-\frac{\alpha}{2}, n-1}$ <b>- דוחים את <math>H_0</math></b>	<b>כלל ההכרעה:</b> אזור הדחייה של $H_0$
$\bar{D} > C + t_{1-\alpha}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$	$\bar{D} < C - t_{1-\alpha}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$	$\bar{D} > C + t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$ $\bar{D} < C - t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$	<b>חלופה לכלל הכרעה:</b> נדחה $H_0$ אם מתקיים:

**סטטיסטי המבחן:**

$$t_{\bar{D}} = \frac{\bar{D} - \mu_D}{S_D / \sqrt{n}}$$

$$S_D^2 = \frac{\sum_{i=1}^n (D_i - \bar{D})^2}{n-1} = \frac{\sum_{i=1}^n D_i^2 - n\bar{D}^2}{n-1}$$



**דוגמה :** (פתרון בהקלטה)

חברה שיווקית מעוניינת לבדוק את טענת רשת השיווק "מגה בעיר" הטוענת שמחיריה נמוכים מהמחירים מרשת השיווק "שופרסל".

לצורך בדיקה נבחרו באקראי 4 מוצרים שונים. המחירים נבדקו בשתי הרשתות. להלן המחירים :

המוצר	מגה בעיר	שופרסל
שמפו	17	18
גיל כביסה	48	57
עוגת גבינה	35	35
לחם	12	10
קפה נמס	49	47
בקבוק יין	113	142
גבינה בולגרית	20	26

בהנחה והמחירים מתפלגים נורמאלית בדקו ברמת מובהקות של 5% את טענת רשת "מגה בעיר".

**תרגילים:**

1. במטרה לבדוק האם קיים הבדל בין חברת X לחברת Y מבחינת המחירים לשיחות בינ"ל. נגדמו באקראי 7 מדינות ועבור כל מדינה נבדקה עלות דקת שיחה. להלן התוצאות:

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

- בהנחה והמחירים מתפלגים נורמלית בכל חברה, בדקו ברמת מובהקות של 5% האם קיים הבדל בין החברות מבחינת המחירים במוצע:

2. מכון המכין לפסיכומטרי טוען שהוא מעלה את ממוצע הציונים ביותר מ-30 נקודות. 8 נבחנים נבדקו לפני ואחרי שהם למדו במכון. להלן התוצאות שהתקבלו:

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

- מה מסקנתכם ברמת מובהקות 5%? הניחו שציוני פסיכומטרי מתפלג נורמלית.

3. נדגמו 5 סטודנטים שסיימו את הקורס סטטיסטיקה ב'. להלן הציונים שלהם בסמסטר א' ו- ב':

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

פורסם שתלמידים שמסיימים את סמסטר ב משפרים בממוצע את הציונים ב-5 נקודות לעומת סמסטר א'. הנח 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 ללא היפנוזה. לאחר מכן נערך לשניהם מבחן באנגלית. נניח שציוני המבחן מתפלגים נורמאלית 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 מזווגים.

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

משקל במכשיר 1	5.4	6.9	7.0	5.2
משקל במכשיר 2	5.3	6.9	7.1	5.0

- נניח שהמשקלים מתפלגים נורמלית.  
 המבחן שיש לבצע כאן הוא:
- מבחן Z למדגם יחיד.
  - מבחן T למדגם יחיד.
  - מבחן T למדגמים בלתי תלויים.
  - מבחן T למדגמים מזווגים.
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:
- מבחן Z למדגם יחיד.
  - מבחן T למדגם יחיד.
  - מבחן T למדגמים בלתי תלויים.
  - מבחן T למדגמים מזווגים.

**פתרונות:****שאלה 1:**לא נדחה  $H_0$ **שאלה 2:**לא נדחה  $H_0$ **שאלה 3:**א.  $0.25 \leq p \leq 0.5$ 

ב. 0.5

ג. לא נדחה  $H_0$ **שאלה 4:**

התשובה היא ד.

**שאלה 5:**

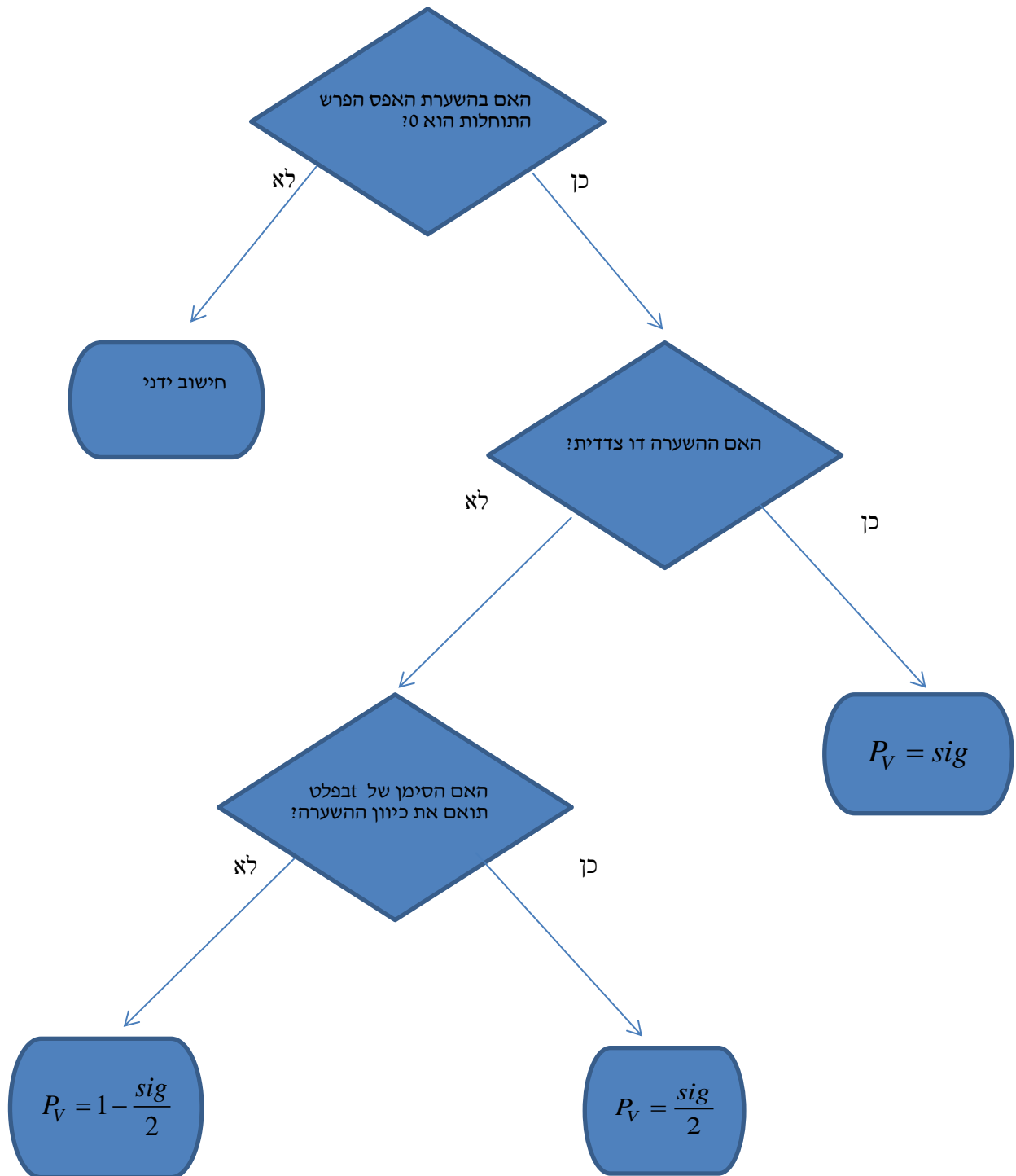
התשובה היא ד.

**שאלה 6:**

התשובה היא ג.

מדגמים מזווגים - ניתוח פלטים

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 היומית לפני הקורס וחודשיים אחריו. הניחו שצריכת הסיגריות מתפלגת נורמלית. להלן התוצאות :

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

**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

בדוק ברמת מובהקות של 5% האם הקורס יעיל .

**תרגילים :**

1. בסקר שנערך בארה"ב בשנת 1993 נשאלו נסקרים על השכלת הוריהם, להלן הפלט שהתקבל :

		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. תן אומדן להפרש הממוצעים.  
 ב. תן אומדן לטעות התקן של הפרש הממוצעים.  
 ג. האם קיים הבדל מובהק בין השכלת האבות להשכלת האימהות ברמת מובהקות של 5%?

2. בתחרות קפיצה למים שופטים באופן קבוע שופט איטלקי ושופט דרום קוריאני. להלן פלט המנתח את הציונים ששופטים אלה נתנו בתחרויות השונות :

**Paired Samples Statistics**

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Italy	???	300	.86742	.05008
	South Korea	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. בדוק את הטענה שהשופט הדרום קוריאני נותן בממוצע 0.2 נקודות יותר מאשר השופט האיטלקי ברמת מובהקות של 5%.  
 ג. מהו רווח הסמך ברמת סמך של 95% לתוחלת פער הציונים בין השופטים.  
 ד. בנה את הרווח כעת ברמת סמך של 98% לתוחלת פער בציונים בין השופטים.



3. בדקו את ציוניהם של 44 נבדקים אקראיים במבחן 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. רשום מהו המבחן הסטטיסטי ונסח את ההשערות אליהם מתייחס הפלט.
- ב. בדוק את ההשערה שממוצע ציונים משתפרים לאחר ההכנה ברמת מובהקות של 5%.
- ג. בדוק את ההשערה שממוצע ציונים משתפרים לאחר ההכנה ביותר מ-5 נקודות ברמת מובהקות של 5%.
- ד. מצא רווח סמך לתוחלת שיפור ממוצע הציונים לאחר ההכנה ברמת ביטחון של 95%.

## T-Test

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	x	54.0000	6	5.86515	2.39444
	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 x - y	7.50000	??	4.72405	-4.64356	19.64356	??	5	.173

- 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 גדלה או קטנה , אם הינו מוסיפים 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 רווח סמך לתוחלת של X ברמת סמך של 90% .

**פתרונות:**

**שאלה 1:**א.  $-0.007$ ב.  $0.1$ 

ג. אין הבדל מובהק

**שאלה 2:**א.  $d.f = 299$  $n = 300$  $\bar{X} = 8.496$ 

Sig=0

ב. נדחה  $H_0$ ג.  $(0.46341, 0.38126)$ ד.  $(0.4708, 0.3738)$ **שאלה 3:**ב. נדחה  $H_0$ ג. לא נדחה את  $H_0$ ד.  $(1.592, 13.317)$ **שאלה 4:**א.  $1.5876, 11.5715$ ב.  $0.173$ 

ג. יגדל

ד.  $0.0865$ ה.  $0.9135$ ו.  $49.18 < \mu < 58.82$

## פרק 38 - הקשר בין רווח סמך לבדיקת השערות על הפרש תוחלות

**רקע:**

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

$$H_0: \mu_1 - \mu_2 = C$$

$$H_1: \mu_1 - \mu_2 \neq C$$

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

אם C נופל ברווח  $\leftarrow$  נקבל את  $H_0$

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

**דוגמה:** (פתרון בהקלטה)

חוקר ביצע בדיקת השערות לתוחלת ההפרש במדגם מזווג. להלן השערותיו :

$$H_0: \mu_D = 80$$

$$H_1: \mu_D \neq 80$$

$$\alpha = 5\%$$

החוקר בנה רווח סמך ברמה של 90%

$$78 < \mu_D < 83$$

האם אפשר לדעת מה מסקנתו, ואם כן מהי?

**תרגילים:**

1. נדגמו 5 סטודנטים שסיימו את הקורס סטטיסטיקה ב'. להלן ציוניהם בסמסטר א' ו- ב' :

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

א. בנו רווח סמך ברמת סמך של 95% לתוחלת פער הציונים בין סמסטר א לבין סמסטר ב.  
 ב. פורסם שתלמידים שמסיימים את סמסטר ב משפרים בממוצע את הציונים ב-5 נק' לעומת סמסטר א' 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 :

68	90	82	<b>מרצה X</b>
64	81	68	<b>מרצה Y</b>

א. חשבו רווח סמך ברמת סמך 90% להפרש בין התוחלות של הציונים אצל שני המרצים.  
 ב. האם ברמת מובהקות של 10% נכריע שיש הבדל בין תוחלות הציונים אצל שני המרצים?

**שאלות אמריקאיות:**

3. סטטיסטיקאי נתבקש לאמוד את הפרש הממוצעים של שני טיפולים לפי שני מדגמים מקריים בלתי תלויים.

הוא חישב רווח סמך להפרש ברמת סמך 0.98, וקיבל את הרווח  $-2 < \mu_1 - \mu_2 < 4.5$ .  
 אילו יתבקש החוקר לבדוק לפי אותם נתונים את ההשערות :

$H_0 : \mu_1 - \mu_2 = 0$  ;  $H_1 : \mu_1 - \mu_2 \neq 0$  , ברמת מובהקות 0.05 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 נתונות בשאלה סטיות התקן של האוכלוסיות, ולכן לא ניתן להסיק דבר.

4. במטרה לבדוק האם קיים הבדל בין קווי זהב לבזק מבחינת ממוצע המחירים לשיחות בינ"ל. נגדמו באקראי 7 מדינות ועבור כל מדינה נבדקה עלות דקת שיחה. בהנחה

לפתרון מלא בסרטון וידאו היכנסו ל- [www.GooL.co.il](http://www.GooL.co.il)

והמחירים מתפלים נורמלית בנו רווח סמך לממוצע ההפרשים וקיבלו :

$$-0.0293 < \mu_b < 0.2145 \text{ רווח הסמך הוא ברמת סמך של } 95\%.$$

לכן מסקנת המחקר היא :

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

**פתרונות:****שאלה 1:**

א.  $-3.8 \leq \mu_D \leq 19$

ב. נכריע שיש אמת בפרסום.

**שאלה 2:**

א.  $-8.5 \leq \mu_X - \mu_Y \leq 26.5$

ב. נכריע שאין הבדל.

**שאלה 3:**

התשובה היא ג.

**שאלה 4:**

התשובה היא א.

## פרק 39 - ניתוח שונות חד כיוונית

### רקע תיאורטי:

ניתוח שונות (חד כיווני) הוא מבחן להשוואת תוחלות ( $\mu_1, \dots, \mu_k$ ) של k אוכלוסיות שונות. ולכן בניתוח שונות השערות המחקר הן:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k \quad (\text{התוחלות של כל האוכלוסיות שוות})$$

$$H_1 : \quad \text{אחרת} \quad (\text{לפחות שתיים מהתוחלות שונות})$$

התנחות הדרושות לביצוע התהליך הן:

1. בכל אוכלוסייה מתוך k האוכלוסיות ההתפלגות נורמלית.

2. כל האוכלוסיות הן עם אותה שונות  $\sigma^2$ .

3. המדגמים בלתי תלויים זה בזה.

ישנו משתנה המבדיל בין הקבוצות השונות, הוא המשתנה הבלתי תלוי הנקרא גורם (factor)

משתנה זה הוא קטגוריאלי עם k רמות (levels).

כדי לבצע את התהליך יש לבצע מדגם מכל אוכלוסייה:

נסמן ב-  $n_i$  את גודל המדגם בקבוצה i.

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

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

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

$$SS_B = \sum_{i=1}^k n_i [\bar{X}_i - \bar{X}]^2 \quad \text{סכום ריבועים בין הקבוצות}$$

$$SS_W = \sum_{i=1}^k [n_i - 1] \cdot \hat{S}_i^2 \quad \text{סכום ריבועים בתוך הקבוצות}$$

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

$$SST = SSB + SSW$$

[www.GooL.co.il](http://www.GooL.co.il) לפתרון מלא בסרטון וידאו היכנסו ל-

© כתב ופתר - ברק קנדל



יש למלא את טבלת ניתוח השונות הבאה :

טבלת ניתוח שונות

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

$$F = \frac{SS_B / (k - 1)}{SS_W / (n - k)} \sim F(k - 1, n - k)$$

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

**תרגילים:**

1. מחקר מעוניין להשוות בין שלוש תרופות לשיכוך כאבים במטרה לבדוק האם קיים הבדל בין התרופות מבחינת הזמן בדקות שלוקח עד שהתרופה משפיעה. לצורך הבדיקה נלקחו 15 אנשים שסובלים מכאבי ראש. אנשים אלה חולקו באקראי לשלוש : קבוצה 1 קיבלה "אקמול" קבוצה 2 קיבלה "אופטלגין" קבוצה 3 קיבלה "נורופן".
- כל אדם במחקר מסר את מספר הדקות עד שהתרופה השפיעה עליו.
- א. מהו המשתנה התלוי ומהו המשתנה הבלתי תלוי במחקר? מהו ה"גורם" וכמה רמות יש לו?
- ב. מהו המבחן הסטטיסטי המתאים כאן? רשמו את ההשערות.
- ג. מה הן ההנחות הדרושות כדי לבצע את המבחן הסטטיסטי שהצעת בסעיף הקודם?

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

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

- א. מהו המבחן הסטטיסטי המתאים? רשמו את ההשערות ואת ההנחות של המבחן.
- ב. מהו גודל המדגם? מהו המשתנה הבלתי תלוי (FACTOR) כמה רמות יש לו?
- ג. חשבו את הממוצע ואת סטיית התקן של הציונים בכל אחד מהמדגמים.
- ד. מלאו את טבלת ANOVA.
- ה. רשמו את כלל ההכרעה למבחן שהוצע בסעיף א ברמת מובהקות של 5%.
- ו. האם קיים הבדל בין בתי הספר בעיר מבחינת רמת הצלחת התלמידים במקצוע המתמטיקה? ענה על סמך הסעיפים הקודמים.

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

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

- א. רשמו את השערות המחקר וההנחות הדרושות כדי לבצע את המבחן המתאים.
- ב. מה מסקנת המחקר ברמת מובהקות של 5%?
- ג. האם יש צורך לבצע השוואות מרובות?
4. שלושה אופים נתבקשו להכין עוגת שוקולד. לכל אופה בדקו את משך זמן ההכנה בדקות.
- כל אופה נדרש לאפות בכל יום 4 עוגות.

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

האם קיים הבדל בין האופים מבחינת תוחלת זמני ההכנה של העוגות? בדקו ברמת מובהקות של 5%.

5. להלן טבלת ניתוח שונות חד כיוונית. במחקר בחנו 4 סוגי סוללות. רצו לבדוק האם לסוג הסוללה השפעה על תוחלת אורך החיים שלה. הפעילו את כל הסוללות על אותו מכשיר ובדקו את אורך החיים של כל סוללה בשעות.

### ANOVA

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

מה המסקנה ברמת מובהקות של 10%? רשמו את ההשערות וההנחות הדרושות.

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

## ANOVA

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

**פתרונות סופיים חלקיים - ניתוח שונות חד כיוונית**

2. אם חישבת נכון ה F הסטטיסטי יוצא: 0.58

3. נדחה את השערת האפס.

4. להלן טבלת הניתוח השונות המתקבלת:

	Sum of Squares	df	Mean Square	F
Between Groups	126.000	2	63.000	.756
Within Groups	750.000	9	83.333	
Total	876.000	11		

5. נקבל את השערת האפס.

6. א. 165.5 ב. 2 ג. 178.725 ד. 9.375 ה. 18.36