

## סטודנטים יקרים

לפניכם ספר תרגילים בקורס סטטיסטיקה ב'. 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).

**GOOL**  
בשביל התירגול

אנו מאחלים לכם הצלחה מלאה בבחינות

צוות האתר GooL

**גול זה בול. בשבילך!**

## תוכן

3	פרק 1 - סטטיסטיקה תיאורית - סיווג משתנים וסולמות מדידה.....
6	פרק 2 - מדדי קשר - מדד הקשר של קרמר .....
9	פרק 3 - מדדי קשר - מדד הקשר של ספירמן .....
13	פרק 4 - מדדי קשר - מדד הקשר הלינארי (פירסון).....
21	פרק 5 - מדדי קשר - רגרסיה ליניארית.....
24	פרק 6 - הסקה סטטיסטית - הקדמה.....
27	פרק 7 - מושגים בסיסיים באמידה .....
34	פרק 8 - רווח סמך להפרש תוחלות ממדגמים בלתי תלויים .....
34	כששונויות האוכלוסייה ידועות .....
37	כששונויות האוכלוסייה אינן ידועות אך שוות .....
39	פרק 9 - רווח סמך לתוחלת ההפרש במדגם מזווג .....
42	פרק 10 - בדיקת השערות על פרמטרים.....
42	הקדמה.....
45	טעויות בבדיקת השערות .....
47	פרק 11 - בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים .....
47	כשהשונויות של האוכלוסייה ידועות.....
51	כששונויות האוכלוסייה לא ידועות ומניחים שהן שוות.....
55	ניתוח פלטים.....
62	פרק 12 - בדיקת השערות על תוחלת ההפרשים במדגמים מזווגים (תלויים) .....
62	בדיקת השערות למדגמים מזווגים.....
68	מדגמים מזווגים - ניתוח פלטים .....
74	פרק 13 - הקשר בין רווח סמך לבדיקת השערות על הפרש תוחלות.....
78	פרק 14 - מבחנים אפרמטריים .....

## 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 הישיות באותה קבוצה.

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

חלוקה אחת של המשתנים הנמדדים היא לפי סולמות מדידה :

### מיון משתנים לפי סולמות המדידה :

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

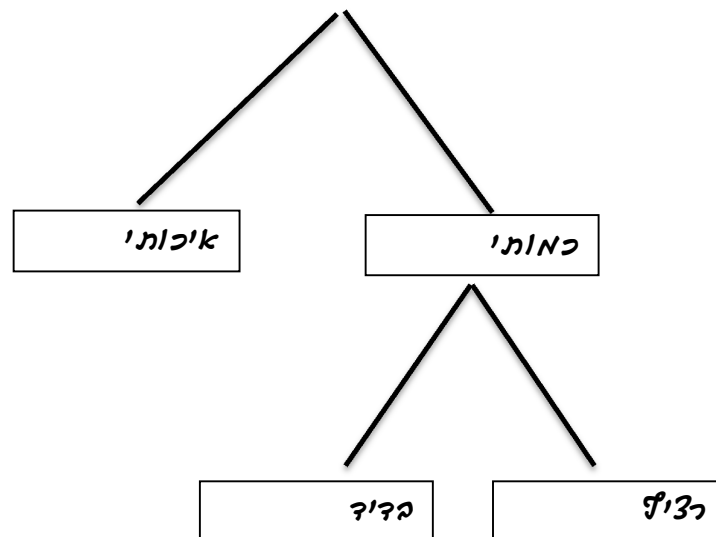
משתנה כמותי הוא משתנה שערכיו הם מספרים להם יש משמעות כמותית כמו : גובה אדם בס"מ, ציון בבחינה וכדומה.

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 - מדדי קשר - מדד הקשר של קרמר

הערה: הפרק ללא הסקה

### רקע:

מתי 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/נשא. 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

שלב ב' 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)

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

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

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 3 - מדדי קשר - מדד הקשר של ספירמן

הערה: הפרק ללא הסקה

#### 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	Rx	ציון שופט ב	Ry	$d=r_x-r_y$	$d^2$
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 4 - מדדי קשר - מדד הקשר הלינארי (פירסון)

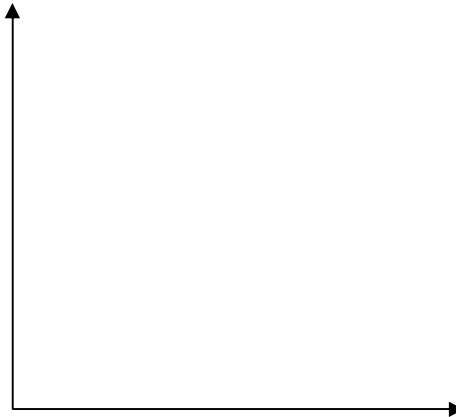
הערה: הפרק ללא הסקה

### רקע:

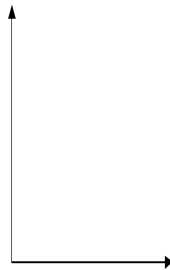
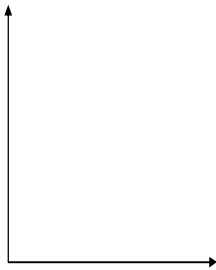
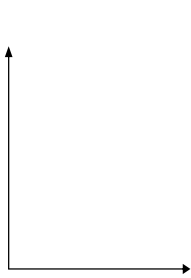
המטרה היא לבדוק האם קיים קשר (קורלציה, מתאם) של קו ישר בין שני משתנים כמותיים. מבחינת סולמות המדידה קשר בין סולמות רווחים ומנה. בדרך כלל,  $X$  הוא המשתנה המסביר (הבלתי תלוי) ו  $Y$  הוא המשתנה המוסבר (התלוי). למשל, נרצה להסביר כיצד השכלה של אדם הנמדדת בשנות לימוד  $X$  מסבירה את ההכנסה שלו  $Y$ . במקרה 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) שבדוק עד כמה קיים קשר לינארי בין שני המשתנים. המדד (ניקרא גם מדד הקשר של פירסון) מכמת את מה שניראה בשלב הראשון רק בעין.

המדד בודק את כיוון הקשר (חיובי או שלילי).

ואת עוצמת הקשר (חלש עד חזק).

מקדם מתאם זה מקבל ערכים בין -1 ל 1.

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

$$y = bx + a$$

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 שמשנתנה אחד עולה לשני יש נטייה לעלות בערכו אבל לא קיימת

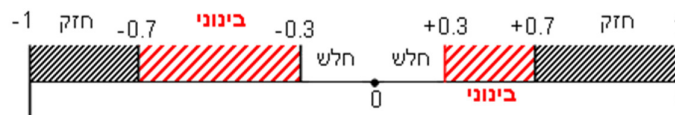
נוסחה לינארית שמקשרת את  $X$  ל-  $Y$  באופן מוחלט ואילו מתאם שלילי חלקי אומר שככל

שמשנתנה 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 :

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

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

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

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

ב. מה המשמעות של התוצאה שקיבלת בסעיף א?

4. נסמן ב- $X$  את ההכנסה של משפחה באלפי ₪. נסמן ב- $Y$  את ההוצאות של משפחה באלפי ₪.

נלקחו 20 משפחות והתקבלו התוצאות הבאות :

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

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

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

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.

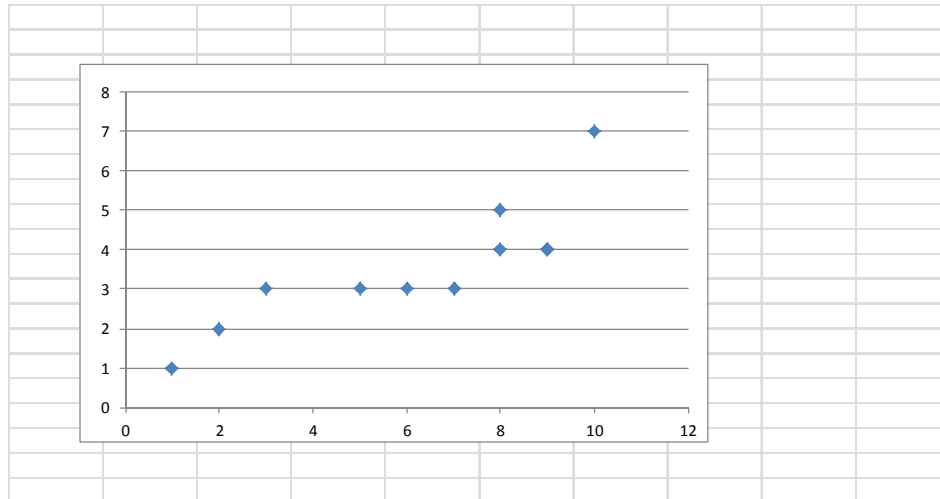
שאלות אמריקאיות:

7. נמצא שקיים מקדם מתאם שלילי בין הציון בעברית לציון בחשבון בבחינה לכן :
- הדבר מעיד שהציונים בכתה היו שליליים.
  - ככל שהציון של תלמיד יורד בחשבון יש לו נטייה לרדת בעברית.
  - ככל שהציון של תלמיד עולה בחשבון יש לו נטייה לרדת בעברית.
  - 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 ₪ ) מהו מקדם המתאם בין המחיר בדולר למחיר בש"ח?

- 1
- 0
- 4.2
- לא ניתן לדעת.

9. להלן דיאגרמת פיזור :



מה יהיה מקדם המתאם בין שני המשתנים?

- 1
- 0.85
- 0.15
- 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 5 - 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 הליניארי 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. משפחות כהן הכניסה 15,000 ₪, מה ההוצאה הצפויה שלה?

2. 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

הרגסיה 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:**

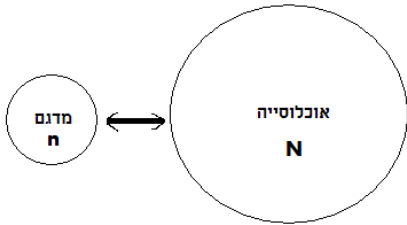
-0.2

## פרק 6 - הסקה סטטיסטית - הקדמה

### רקע:

אוכלוסייה – קבוצה שאליה מפנים שאלה מחקרית.

למשל, חברת תרופות שמעוניינת לפתח תרופה למחלת הסוכרת מתעניינת באוכלוסיית חולי הסוכרת בעולם.



מדגם – חלק מתוך האוכלוסייה.

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

במקרים רבים אין אפשרות לחקור את כל האוכלוסייה כיוון שאין גישה לכולה, היא גדולה מידי, או 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 הם שונים

למשל:

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

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



**דוגמה (פתרון בהקלטה):**

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

א. מי האוכלוסייה?

ב. מה המשתנה?

ג. מה הפרמטרים?

ד. מהו גודל המדגם?

ה. מהו הסטטיסטי שמתכננים להוציא מהמדגם?

ו. האם הפרמטר או הסטטיסטי הוא משתנה מקרי?

**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.
- מי האוכלוסייה?
  - מה המשתנה?
  - מהם הפרמטרים?
  - מהו גודל המדגם?
2. להלן התפלגות מספר מקלטי הטלויזיה למשפחה בישוב "העוגן".  
נגדיר את  $x$  להיות מספר המקלטים של משפחה אקראית.  
מתכננים לדגום מאוכלוסייה זו 4 משפחות ולהתבונן בממוצע מספר מקלטי הטלויזיה במדגם.

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

- מיהי האוכלוסייה ומהו המשתנה הנחקר?
  - מהו הסטטיסטי שיילקח מהמדגם ומה סימונו?
3. נתון כי 20% מהשכירים במדינה הם אקדמאיים. נבחרו באקראי 10 שכירים באותה אוכלוסייה ומתכננים לפרסם את מספר האקדמאיים שנדגמו.
- מהי האוכלוסייה?
  - מה המשתנה באוכלוסייה?
  - מהם הפרמטרים?
  - מהו הסטטיסטי?

## פרק 7 - מושגים בסיסיים באמידה

### רקע:

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

כמו ממוצע הגבהים בקרב מתגייסים 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- $\hat{\theta}$ . 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  $\theta$
- $\sigma(\hat{\theta}) = S.E$  : טעות התקן של אומד היא סטיית התקן שלו, כלומר

להלן פרמטרים מרכזיים והאומדים שלהם:

ממוצע האוכלוסייה:  $\mu$

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

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

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

פרופורציה באוכלוסייה:  $p$

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

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

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

שונות האוכלוסייה:  $\sigma^2$

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

$E(S^2) = \sigma^2$  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, 184, 192, 171, 180, 177, 187, 168, 177 ו-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 וחישבו 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 תהיה:

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

$$\sigma^2 = 64. \text{ טעות התקן של האומדן ל- } \mu \text{ היא:}$$

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 8 - רווח סמך להפרש תוחלות ממדגמים בלתי תלויים

### כששונות האוכלוסייה ידועות

#### 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.  $\sigma_1^2, \sigma_2^2$  ידועות.

2.  $X_1, X_2 \sim N$  או  $n_1, n_2 > 30$

3. שני מדגמים בלתי תלויים.

#### רווח סמך:

$$(\bar{x}_1 - \bar{x}_2) \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

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

#### דוגמה: (פתרון בהקלטה)

נדגמו 100 תושבים מאזור a והמשכורת הממוצעת הייתה שם 9200 ₪.

כמו כן נדגמו 120 תושבים מאזור b וממוצע המשכורות שהתקבל שם 8700 ₪.

לצורך פתרון נניח שסטיית התקן של המשכורות באוכלוסיית שני האזורים היא 1800 ₪.

אמדו ברמת סמך של 90% את הפרש השכר הממוצע בין אזור a לאזור b.

## תרגילים:

1. מעוניינים לבדוק האם קיים הבדל בין ממוצע ציוני הפסיכומטרי של חיילים לממוצע ציוני הפסיכומטרי של תלמידי תיכון. ידוע שציוני הפסיכומטרי מתפלגים נורמאלית עם סטיית תקן 100. במדגם של 16 נבחנים חיילים התקבל ממוצע 543. במדגם של 20 תלמידי תיכון התקבל ממוצע 508. בנו רווח סמך לפער תוחלות הציונים בין חיילים לתלמידי תיכון ברמת סמך של 90%. מה ניתן להסיק מרווח סמך זה?
  
2. ציוני I.Q. מתוכננים כך שיתפלגו נורמאלית עם סטיית תקן של 15. במדגם של 20 נבחנים ישראלים התקבל ממוצע ציונים 104. במדגם של 23 נבחנים אמריקאיים התקבל ממוצע ציונים 99.
  - א. בנו רווח סמך ברמת סמך של 95% לפער בין ישראל לארה"ב בממוצע הציונים במבחן ה-IQ.
  - ב. האם קיים הבדל בין ישראלים לאמריקאים מבחינת ממוצע הציונים?
  
3. חברה להנדסת בניין מעוניינת להשוות ברמת הקשיות של שני סוגי ברגים. ידוע שרמת הקשיות של ברגים מתפלגת נורמלית עם סטיית תקן של 4 יחידות. במדגם של 15 ברגים מסוג א' התקבל רמת קשיות ממוצעת של 28 יחידות ובמדגם של 12 ברגים מסוג ב' התקבל רמת קשיות ממוצעת של 25. עבור אילו רמות בטחון יקבע שאין הבדל בין שני סוגי הברגים מבחינת ממוצע רמת הקשיות שלהם?

**פתרונות :****שאלה 1**

(-20,90)

**שאלה 3:**

רמות בטחון הגבוהות מ: 0.9476

### 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.

השונות המשוקללת: כיוון שאנו מניחים שבין שתי האוכלוסיות השונויות שוות אנו אומדים את השונות 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:

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

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  $N(\mu_y, \sigma^2)$ .

X	22	20	21	25
Y	18	25	17	12

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

## פרק 9 - 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

ניח 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 מבחינת ממוצע המחירים לשיחות בינ"ל. נגדמו באקראי 7 מדינות ועבור כל מדינה נבדקה עלות דקת שיחה. להלן התוצאות:

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

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

## פרק 10 - בדיקת השערות על פרמטרים

### הקדמה

#### רקע:

תהליך של בדיקת השערות הוא תהליך מאד 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 )  
**ואזור 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 )

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

כתב ופתר - ברק קנדל ©

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 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.

- 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 איזושהו 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)



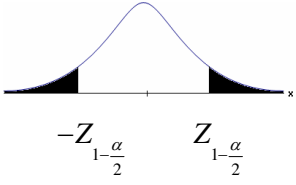
אדם חשוד בביצוע עבירה ונתבע בבית המשפט. אילו סוגי טעויות אפשריות בהכרעת הדין?

## 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 11 - בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

### כשהשונויות של האוכלוסייה ידועות

<u>רקע:</u>			
$H_0 \mu_1 - \mu_2 = c$ $H_1 \mu_1 - \mu_2 > c$	$H_0 \mu_1 - \mu_2 = c$ $H_1 \mu_1 - \mu_2 < c$	$H_0 \mu_1 - \mu_2 = c$ $H_1 \mu_1 - \mu_2 \neq c$	<b>השערת האפס:</b> <b>השערה אלטרנטיבית:</b>
1. מדגמים בלתי תלויים 2. $\sigma_1, \sigma_2$ ידועות 3. $X_1, X_2 \sim N$ או מדגמים מספיק גדולים			<b>תנאים:</b>
$Z_{\bar{x}_1 - \bar{x}_2} > Z_{1-\alpha}$  $Z_{1-\alpha}$ - דוחים את $H_0$	$Z_{\bar{x}_1 - \bar{x}_2} < -Z_{1-\alpha}$  $-Z_{1-\alpha}$ - דוחים את $H_0$	$Z_{\bar{x}_1 - \bar{x}_2} > Z_{1-\frac{\alpha}{2}}$ או $Z_{\bar{x}_1 - \bar{x}_2} < -Z_{1-\frac{\alpha}{2}}$  $-Z_{1-\frac{\alpha}{2}}$ $Z_{1-\frac{\alpha}{2}}$ - דוחים את $H_0$	<b>כלל ההכרעה:</b> אזור הדחייה של $H_0$ :

#### סטטיסטי המבחן:

$$Z_{\bar{x}_1 - \bar{x}_2} = \frac{\bar{x}_1 - \bar{x}_2 - c}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

#### חלופה אחרת לכלל הכרעה:

$\bar{x}_1 - \bar{x}_2 > c + Z_{1-\alpha} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	$\bar{x}_1 - \bar{x}_2 < c - Z_{1-\alpha} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	$\bar{x}_1 - \bar{x}_2 > c + Z_{1-\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ <b>או</b> $\bar{x}_1 - \bar{x}_2 < c - Z_{1-\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	<b>נדחה <math>H_0</math> אם מתקיים:</b>
---	---	---	---

#### התפלגות הפרש הממוצעים:

$$\bar{x}_1 - \bar{x}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

#### התקנון:

$$Z_{\bar{x}_1 - \bar{x}_2} = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

**דוגמה : (פתרון בהקלטה)**

בשנת 2004 הפער בין השכר הממוצע של הגברים לנשים היה 3000 ₪ לטובת הגברים.  
מעוניינים לבדוק האם כיום הצטמצם הפער בין הגברים לנשים מבחינת השכר הממוצע.  
נדגמו 100 עובדים גברים. שכרם הממוצע היה 9,072 ₪. נדגמו 80 עובדות, שכרן הממוצע היה  
7809 ₪. לצורך פתרון נניח שסטיות התקן של השכר ידועות ושוות ל-2000 ₪ באוכלוסיית הנשים  
ו-3000 ₪ באוכלוסיית הגברים. מה המסקנה ברמת מבוהקות של 5%?




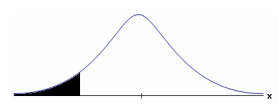

### תרגילים :

1. מחקר טוען שאנשים החיים במרכז הארץ צופים בממוצע בטלוויזיה יותר מאנשים שלא חיים במרכז. נדגמו 100 אנשים מהמרכז ו-107 אנשים לא מהמרכז. אנשים אלו נשאלו כמה שעות ביום הם נוהגים לצפות בטלוויזיה.  
 במדגם של מרכז הארץ התקבל ממוצע 2.7 שעות.  
 במדגם של מחוץ למרכז הארץ התקבל ממוצע 1.8 שעות.  
 לצורך פתרון הניחו שבכל אזור, סטיית התקן היא שעה 1 ביום. בדקו את טענת המחקר ברמת מובהקות של 1%.
  
2. ציוני פסיכומטרי מתפלגים נורמלית עם סטיית תקן 100. מכון ללימוד פסיכומטרי טוען שהוא יכול לשפר את ממוצע הציונים ביותר מ-30 נקודות. במדגם של 20 נבחנים שניגשו למבחן ללא הכנה במכון התקבל ממוצע 508. במדגם של 25 נבחנים שעברו הכנה במכון התקבל ממוצע ציונים 561. מה מסקנתכם ברמת מובהקות של 5%.

**פתרונות:****שאלה 1:**נדחה  $H_0$ **שאלה 2:**לא נדחה את  $H_0$

### 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> 4. מדגמים בלתי תלויים 5. $\sigma_1, \sigma_2$ לא ידועות אך שוות 6. 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 :

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2} \quad \text{השונויות המשוקללות} \quad t_{\bar{x}_1 - \bar{x}_2} = \frac{(\bar{x}_1 - \bar{x}_2) - c}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}}$$

#### 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. 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

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.

W	f	h

א. בדקו ברמת מובהקות של 5% האם נורות מסוג W60 דולקות בממוצע יותר מאשר נורות מסוג W100. רשמו את כל ההנחות הדרושות לפתרון.

ב. עבור איזו רמת מובהקות ניתן לקבוע שנורות מסוג W60 דולקות בממוצע 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  $H_0$

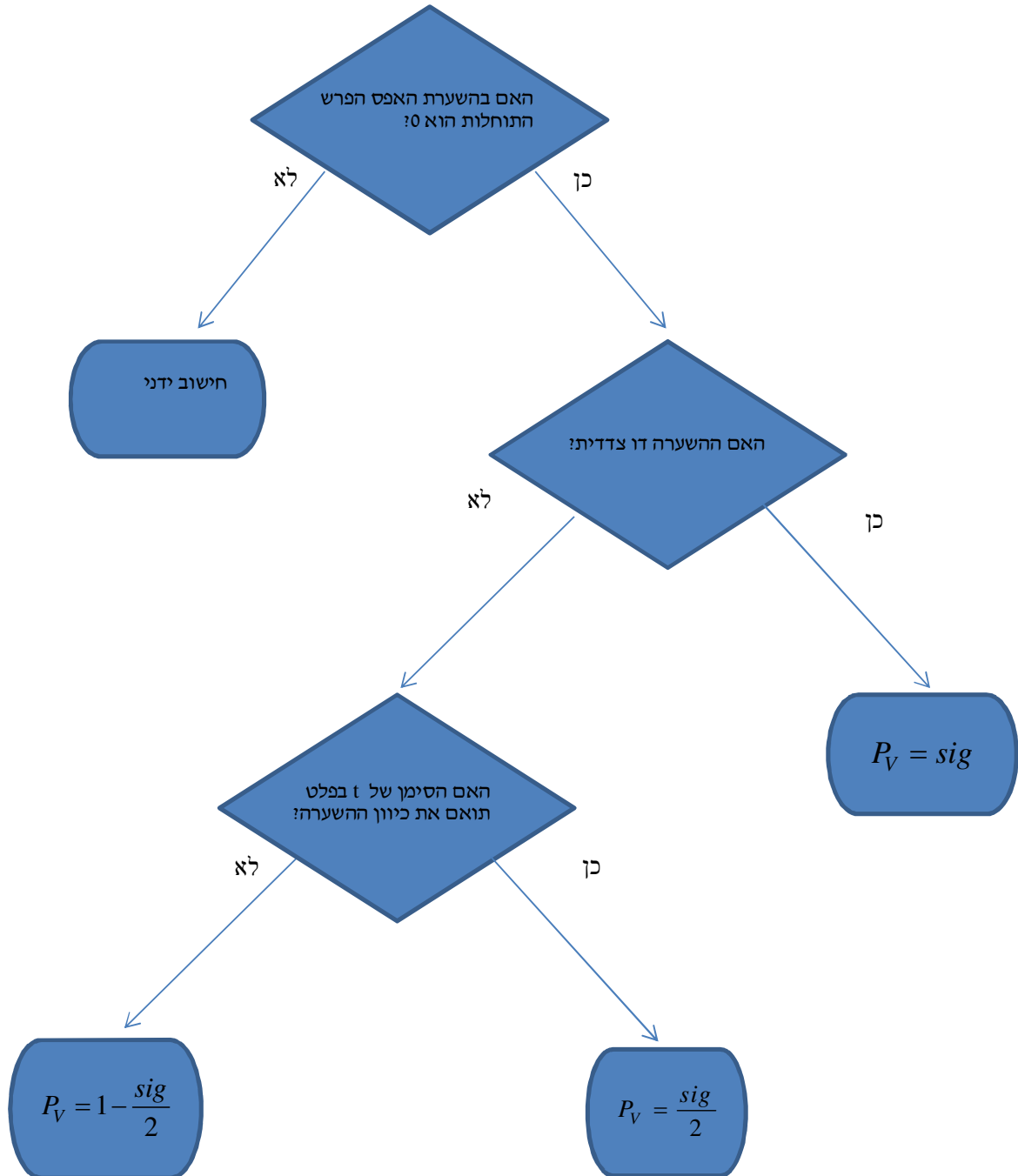
ב. רמות מובהקות של לפחות 5%

ג. לא נדחה  $H_0$

## ניתוח פלטים

### רקע:

מובהקות התוצאה על סמך הפלט:



**דוגמה:** (פתרון בהקלטה)

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	190.977	.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	.206	.656	-2.428	18	.026	-11.56566	4.76333	-21.57304	-1.55828
Equal variances assumed									
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-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variances assumed	4.446	.040	-2.164	???	.044	-10.15064	???	-20.03781	-.26347
Equal variances not assumed			-2.038	38.267	.048	???	5.00462	-20.27964	-.02164

א. 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.

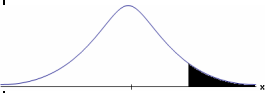

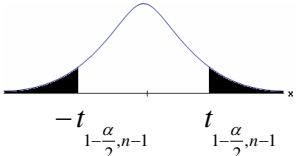
**שאלה 4:**

- א. -10.15, 4.69, 48  
 ב. 0.04  
 ג. 0.978  
 ד. 0.022

**פרק 12 - 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>
7. $\sigma_D$ אינה ידועה			<b>תנאים:</b>
8. $D \sim N$ או מדגם מספיק גדול			
$t_{\bar{D}} > t_{1-\alpha}^{(n-1)}$  $t_{1-\alpha, n-1}$ - דוחים את $H_0$	$t_{\bar{D}} < -t_{1-\alpha}^{(n-1)}$  $-t_{1-\alpha, n-1}$ - דוחים את $H_0$	$t_{\bar{D}} < -t_{1-\frac{\alpha}{2}}^{(n-1)}$ או $t_{\bar{D}} > t_{1-\frac{\alpha}{2}}^{(n-1)}$  $-t_{1-\frac{\alpha}{2}, n-1}$ $t_{1-\frac{\alpha}{2}, n-1}$ - דוחים את $H_0$	<b>כלל ההכרעה:</b> אזור הדחייה של $H_0$
$\bar{D} > C + t_{1-\alpha}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$	$\bar{D} < C - t_{1-\alpha}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$	$\bar{D} > C + t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$ או $\bar{D} < C - t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$	<b>חלופה לכלל הכרעה:</b> נדחה $H_0$ אם מתקיים:

**סטטיסטי המבחן:**

$$t_{\bar{D}} = \frac{\bar{D} - \mu_D}{\frac{S_D}{\sqrt{n}}}$$

$$S_D^2 = \frac{\sum_{i=1}^n (D_i - \bar{D})^2}{n-1} = \frac{\sum_{i=1}^n D_i^2 - n\bar{D}^2}{n-1}$$

**דוגמה:** (פתרון בהקלטה)

חברה שיווקית מעוניינת לבדוק את טענת רשת השיווק "מגה בעיר" הטוענת שמחירה 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 בממוצע?

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

- 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 במוצע 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 :

א. מבחן Z למדגם יחיד.

ב. מבחן T למדגם יחיד.

ג. מבחן T למדגמים בלתי תלויים.

ד. מבחן T למדגמים מזווגים.

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

המבחן שיש לבצע כאן הוא :

א. מבחן Z למדגם יחיד.

ב. מבחן T למדגם יחיד.

ג. מבחן T למדגמים בלתי תלויים.

ד. מבחן T למדגמים מזווגים.

**פתרונות:****שאלה 1:**לא נדחה  $H_0$ **שאלה 2:**לא נדחה  $H_0$ **שאלה 3:**א.  $0.1 \leq p \leq 0.25$ 

ב. 0.25

ג. לא נדחה  $H_0$ **שאלה 4:**

התשובה היא ד.

**שאלה 5:**

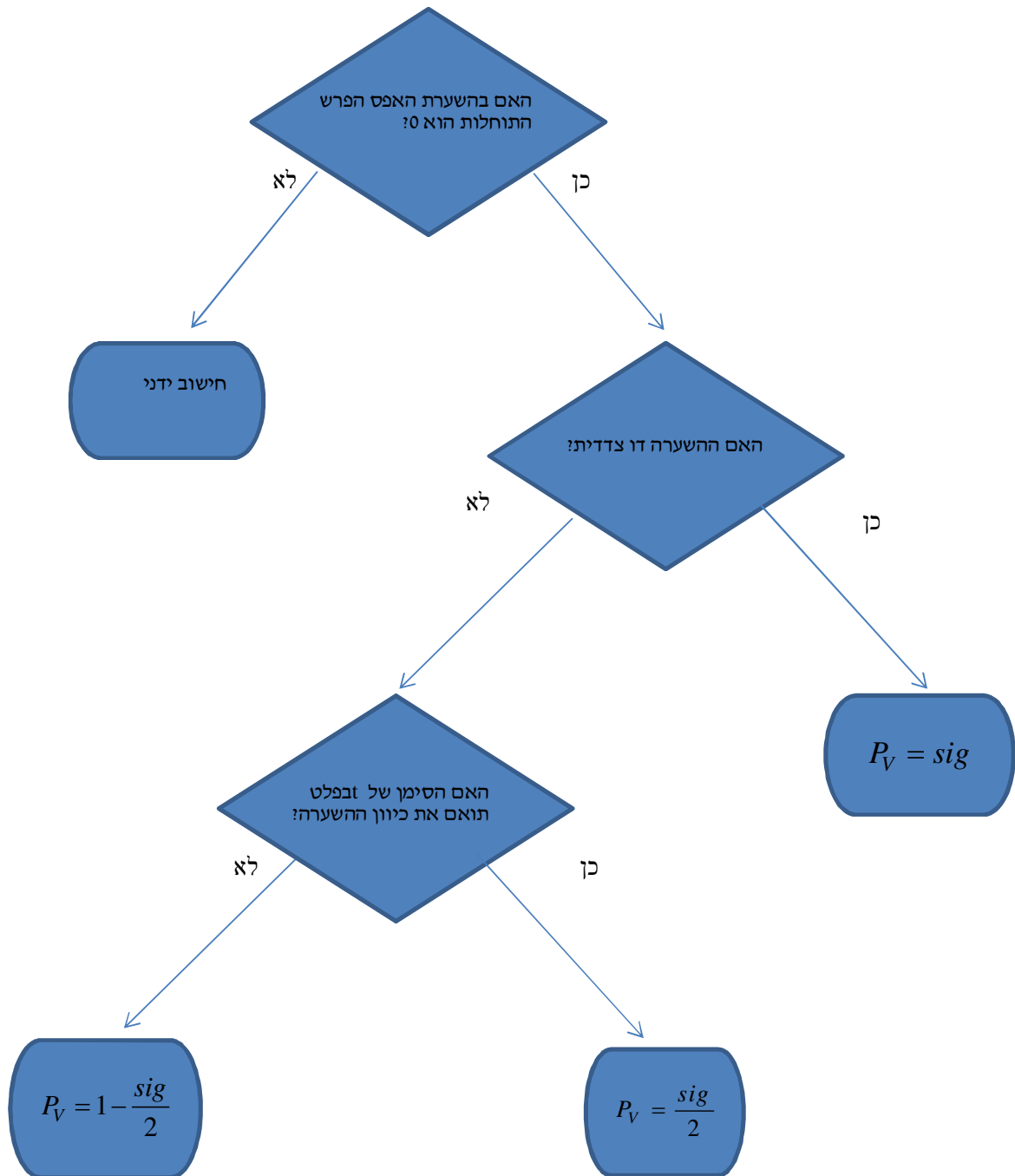
התשובה היא ד.

**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:

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

		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
Pair 1 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 13 - 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 והמחירים מתפלים נורמלית בנו רוח סמך לממוצע ההפרשים וקיבלו :
- $$-0.0293 < \mu_d < 0.2145$$
- רווח הסמך הוא ברמת סמך של 95%.
- לכן מסקנת המחקר היא :
- ברמת מובהקות של 5% לא נוכל לקבוע שקיים הבדל בין החברות.
  - ברמת מובהקות של 5% נקבע שקיים הבדל מובהק בין החברות.
  - 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 14 - מבחנים אפרמטריים

### איך מענה למאן וויטני

#### א. מבחן אי תלות

1. במפעל עובד בשלוש משמרות. להלן מספר המוצרים הפגומים והתקינים בכל אחת מן המשמרות לפי מדגם שנעשה:

	יום	ערב	לילה
פגומים	50	60	70
תקינים	600	700	800

האם קיים קשר בין טיב המוצר למשמרת שלו? הסיקו עבור רמת מובהקות  $\alpha = 0.05$ .

2. בקרב מדגם של 200 נשים 120 טענו שהן תצבענה למועמד R לראשות העיר. בקרב מדגם של 200 גברים 80 טענו שהם יצביעו למועמד R האם קיים הבדל בין דפוס ההצבעה של הנשים ושל הגברים? בדוק ברמת מובהקות של 5%.

3. בחנות בגדים A בדקו את התפלגות הצבעים של הבגדים הנמכרים ביום מסוים:

צבע	שחור	לבן	אדום	כחול
מספר הפריטים	15	20	15	50

כמו כן בדקו את התפלגות הצבעים בחנות שכנה B:

צבע	שחור	לבן	אדום	כחול
מספר הכדורים	60	20	10	20

- א. בדוק ברמת מובהקות של 2.5% האם קיים הבדל בין החניות מבחינת התפלגות הצבעים של הפריטים הנמכרים.

**ב. מבחן ווילקוקסון למדגמים מזווגים:**

1. שני קונדיטורים מתחרים על מקום עבודה. נתנו לשניהם להכין 8 מאפים שונים כאשר כל אחד מהמאפים נאפה ע"י שניהם. בסופו של דבר בעל הקונדיטוריה נתן ציון לכל אחד מהאופים בעבור כל אחד מהמאפים. להלן הציונים שהתקבלו:

אופה א	אופה ב
10	9
9	8
7	7
8	9
9	6
10	6
7	5
8	4

ברמת מובהקות של 5% האם אפשר לקבוע שאופה א' טוב יותר מאופה ב'?

2. סטודנטים נתבקשו לתת חוות דעתם על רמת הקושי של הקורס (סקאלה של 1-5 כאשר 5=קשה ביותר) ועל רמת הקושי של הבחינות באותה סקאלה. מעוניינים לבדוק האם קיים הבדל בין רמת הקושי הנתפסת בעיני הסטודנט על הקורס ועל המבחנים.

להלן תוצאות המדגם:

4	5	1	2	3	4	2	3	4	1- קושי קורס
2	3	2	2	2	3	4	4	4	2 - קושי בחינה

בדקו ברמת מובהקות של 5%.

תשובות סופיות - מבחנים אפרמטריים

פרק א' - מבחן לאי תלות

<u>שאלה 1</u>	<u>שאלה 2</u>
נקבל $H_0$	נדחה $H_0$
<u>שאלה 3</u>	
נדחה $H_0$	

פרק ב' - מבחן ווילקוקסון למדגמים מזווגים

<u>שאלה 1</u>	<u>שאלה 2</u>
נדחה $H_0$	נקבל $H_0$