

## סטודנטים יקרים

לפניכם ספר תרגילים בקורס מבוא לסטטיסטיקה ועיבוד נתונים 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  
לחויית 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!**

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

כתב ופתר - ברק קנדל ©

## תוכן

4	פרק 1 - הסקה סטטיסטית - הקדמה
7	פרק 2 - התפלגות הדגימה
7	ממוצע המדגם ומשפט הגבול המרכזי
15	התפלגות סכום תצפיות המדגם ומשפט הגבול המרכזי
18	פרק 3 - מושגים בסיסיים באמידה
25	פרק 4 - רווח סמך לתוחלת (ממוצע האוכלוסייה)
25	רווח סמך כששונות האוכלוסייה ידועה
31	קביעת גודל מדגם באמידת תוחלת עם שונות אוכלוסייה ידועה
34	רווח סמך לתוחלת (ממוצע האוכלוסייה) כששונות האוכלוסייה אינה ידועה
40	פרק 5 - רווח סמך לפרופורציה
43	קביעת גודל מדגם באמידת פרופורציה
46	פרק 6 - רווח סמך להפרש פרופורציות
49	פרק 7 - רווח סמך להפרש תוחלות ממדגמים בלתי תלויים
49	כששונות האוכלוסייה ידועות
52	כששונות האוכלוסייה אינן ידועות אך שוות והמדגמים בלתי תלויים
55	פרק 8 - רווח סמך לתוחלת ההפרש במדגם מזווג
58	פרק 9 - תרגול מסכם ברווחי סמך
61	פרק 10 - בדיקת השערות על פרמטרים
61	הקדמה
64	טעויות בבדיקת השערות
66	פרק 11 - בדיקת השערות על תוחלת (ממוצע)
66	כאשר שונות האוכלוסייה ידועה
71	סיכוי לטעויות ועוצמה כאשר שונות האוכלוסייה ידועה
78	מובהקות התוצאה (p-value) בבדיקת השערות על תוחלת עם שונות ידועה
83	בדיקת השערות על תוחלת (ממוצע) כאשר שונות האוכלוסייה אינה ידועה
88	מובהקות התוצאה (p-value) כאשר שונות האוכלוסייה לא ידועה
92	ניתוח פלטי SPSS
97	הקשר בין רווח סמך לבדיקת השערות על תוחלת

100	פרק 12 - בדיקת השערות על פרופורציה
100	התהליך
104	סיכוי לטעויות ועוצמה
109	מובהקות התוצאה
113	פרק 13 - בדיקת השערות על הפרש פרופורציות
117	פרק 14 - בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים
117	כשהשונויות של האוכלוסייה ידועות
121	כשהשונויות האוכלוסייה לא ידועות ומניחים שהן שוות
125	ניתוח פלטים
132	פרק 15 - בדיקת השערות על תוחלת ההפרשים במדגמים מזווגים (תלויים)
132	בדיקת השערות למדגמים מזווגים
138	מדגמים מזווגים - ניתוח פלטים
144	פרק 16 - הקשר בין רווח סמך לבדיקת השערות על הפרש תוחלות
148	פרק 17 - שאלות מסכמות בבדיקת השערות על פרמטרים
148	שאלות מסכמות בבדיקת השערות על פרמטרים
154	שאלות מסכמות בסגנון רב ברירה ( אמריקאיות) על בדיקת השערות
170	פרק 18 - ניתוח שונות חד כיוונית
182	פרק 19 - מבחני חי בריבוע
182	מבחן טיב התאמה
189	הקשר בין מבחן טיב התאמה לבדיקת השערות על הפרופורציה
192	מבחן חי בריבוע לאי תלות בין משתנים
192	דעה
194	שביעות רצון
197	הקשר בין מבחן לאי תלות ובדיקת השערות להפרש פרופורציות
200	פרק 20 - מדדי קשר - מדד הקשר הלינארי (פירסון)
208	פרק 21 - מדדי קשר - רגרסיה ליניארית

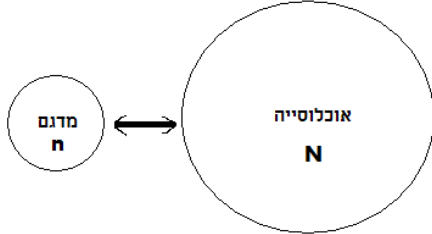
## פרק 1 - הסקה סטטיסטית - הקדמה

### 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 סיכוי 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 אנו לא יודעים אותו סטטיסטי הוא משתנה ממדגם למדגם ולכן יש לו התפלגות הנקראת התפלגות הדגימה.

**דוגמה (פתרון בהקלטה):**

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

א. מי האוכלוסייה?

ב. מה המשתנה?

ג. מה הפרמטרים?

ד. מהו גודל המדגם?

ה. מהו הסטטיסטי שמתכננים להוציא מהמדגם?

ו. האם הפרמטר או הסטטיסטי הוא משתנה מקרי?

**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. להלן התפלגות מספר מקלטי הטלוויזיה למשפחה בישוב "העוגן".

נגדיר את  $x$  להיות מספר המקלטים של משפחה אקראית.

מתכננים לדגום מאוכלוסיה זו 4 משפחות ולהתבונן בממוצע מספר מקלטי הטלוויזיה במדגם.

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

א. מיהי האוכלוסייה ומהו המשתנה הנחקר?

ב. 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.

א. מהי האוכלוסייה ?

ב. מה המשתנה באוכלוסייה?

ג. מהם הפרמטרים?

ד. מהו הסטטיסטי?

## פרק 2 - התפלגות הדגימה

### ממוצע המדגם ומשפט הגבול המרכזי

#### רקע:

$$\bar{x} = \frac{\sum x_i}{n} : \text{ בפרק זה נדון בהתפלגות של ממוצע המדגם :}$$

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

גדלים המתארים התפלגות כלשהי או אוכלוסייה כלשהי נקראים פרמטרים. להלן רשימה של פרמטרים החשובים לפרק זה:

ממוצע האוכלוסייה נסמן ב  $\mu$  (נקרא גם תוחלת).

שונות אוכלוסייה נסמן ב-  $\sigma^2$ .

סטיית תקן של אוכלוסייה:  $\sigma$ .

#### א. תכונות התפלגות

ממוצע כל ממוצעי המדגם האפשריים שווה לממוצע האוכלוסייה:

$$E(\bar{x}) = \mu_{\bar{x}} = \mu$$

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

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

יש יחס הפוך בין גודל המדגם לבין שונות ממוצעי המדגם.

אם נוציא שורש לשונות נקבל סטיית תקן של ממוצע המדגם שנקראת גם טעות תקן:

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

#### דוגמה: (פתרון בהקלטה)

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

א. מי אוכלוסיית המחקר? מהו המשתנה הנחקר?

ב. מהם הפרמטרים של האוכלוסייה?

ג. מה התוחלת ומהי סטיית התקן של ממוצע המדגם?

**ב. דגימה מהתפלגות נורמאלית**

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

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

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

**דוגמה: (פתרון בהקלטה)**

משקל תינוק ביום היוולדו מתפלג נורמאלית עם ממוצע 3400 גרם וסטיית תקן של 400 גרם. מה ההסתברות שבמדגם של 4 תינוקות אקראיים בעת הולדתם המשקל הממוצע של התינוקות יהיה מתחת ל-3.5 ק"ג?

**ג. משפט הגבול המרכזי**

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

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

**דוגמה: (פתרון בהקלטה)**

משקל חפיסת שוקולד בקו ייצור מתפלג עם ממוצע 100 גרם וסטיית תקן של 4 גרם. דגמו מקו הייצור 36 חפיסות שוקולד אקראיות. מה ההסתברות שהמשקל הממוצע של חפיסות השוקולד שנדגמו יהיה מתחת ל 102 גרם?



**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?

ה. מהו תוחלת ממוצע המדגם?

ו. מהי טעות התקן?

2. להלן התפלגות מספר מקלטי הטלויזיה למשפחה בישוב מסוים :

מספר מקלטים	מספר המשפחות
0	500
1	2500
2	3500
3	3000
4	500
	<b>סך הכול <math>N = 10000</math></b>

נגדיר את  $x$  להיות מספר המקלטים של משפחה אקראית.

א. בנו את פונקציית ההסתברות של  $x$ .

ב. חשבו את התוחלת, השונות וסטיית התקן של  $x$ .

ג. אם נדגום 4 משפחות מהישוב עם החזרה מה תהיה התוחלת, מהי השונות ומהי סטיית התקן של ממוצע המדגם?

3. אם נטיל קובייה פעמיים ונתבונן בממוצע התוצאות שיתקבלו, מה תהיה התוחלת ומה תהיה סטיית התקן של ממוצע זה?

4. 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 כי ביום 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?

ה. הסבירו ללא חישוב כיצד התשובה לסעיף הקודם הייתה משתנה אם היה מדובר על יותר מ-4 תינוקות?

5. הגובה של המתגייסים לצה"ל מתפלג נורמאלית עם תוחלת של 175 ס"מ וסטיית תקן של 10 ס"מ. ביום מסוים התגייסו 16 חיילים.

א. מה ההסתברות שהגובה הממוצע שלהם יהיה לפחות 190 ס"מ?

ב. מה ההסתברות שהגובה הממוצע שלהם יהיה בדיוק 180 ס"מ?

ג. מה ההסתברות שהגובה הממוצע שלהם יסטה מתוחלת הגבהים בפחות מ-5 ס"מ?

ד. מהו הגובה שבהסתברות של 90% הגובה הממוצע של המדגם יהיה נמוך ממנו?

6. הזמן הממוצע שלוקח לאדם להגיע לעבודתו 30 דקות עם שונות של 16 דקות רבועות. 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 גבוה ממנו?

ג. מה ההסתברות שממוצע משך הנסיעה השבועי יהיה מרוחק מ-30 דקות בלפחות 2 דקות?

ד. כיצד התשובה לסעיף הקודם הייתה משתנה אם האדם היה נוסע לעבודה 6 פעמים בשבוע?

7. 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?

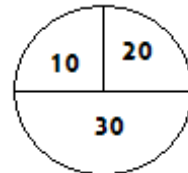
ד. בקבוקי היין שבארגז נמזגים לקערה עם קיבולת של שלושה ליטר. מה ההסתברות שהיין יגלוש מהקערה?

8. 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.

9. בקזינו ישנה רולטה. על הרולטה רשומים המס' הבאים כמוראה בשרטוט :



אדם מסובב את הרולטה וזוכה בסכום הרשום על הרולטה.

א. בנו את פונקציית ההסתברות של סכום הזכייה במשחק בודד.

ב. מה התוחלת ומה השונות של סכום הזכייה?

ג. אם האדם ישחק את המשחק 5 פעמים מה התוחלת ומה השונות של ממוצע סכום הזכייה בחמשת המשחקים?

ד. אם האדם משחק את המשחק 50 פעם מה ההסתברות שבסה"כ יזכה ב-1050 ₪ ומעלה?

10. 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 ₪?

11. מטילים קובייה 50 פעמים בכל פעם מתבוננים בתוצאה של הקובייה. מה ההסתברות שהממוצע של התוצאות יהיה לפחות 3.72 ב-50 ההטלות?

12. 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.

13. נתון משתנה מקרי בדיד בעל פונקציית ההסתברות הבאה :

2	4	6	8	X
$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	P(X)

מתוך התפלגות זו נלקח מדגם מקרי בגודל 50 . מה הסיכוי שממוצע המדגם יהיה קטן מ-5?

14. נתון ש  $X \sim N(\mu, \sigma^2)$  דגמו 5 תצפיות מאותה התפלגות והתבוננו בממוצע המדגם  $\bar{X}$  :

לכן  $P(\bar{X} > \mu)$  יהיה : ( בחר בתשובה הנכונה )

א. 0

ב. 0.5

ג. 1

ד. לא ניתן לדעת.

15. נתון ש  $X$  מתפלג כלשהו עם תוחלת  $\mu$  ושונות  $\sigma^2$  .

החליטו לבצע מדגם בגודל 200 מתוך ההפלגות הנתונה לפי משפט הגבול המרכזי מתקיים ש :  
( בחר בתשובה הנכונה )

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

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

ג.  $\bar{X} \sim N(\mu, \sigma^2)$

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

16. נתון ש  $X \sim N(\mu, \sigma^2)$  . אם נדגום  $n$  תצפיות מתוך ההתפלגות ונגדיר  $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$  אזי :

( בחר בתשובה הנכונה )

א.  $\mu$  ו-  $\bar{X}$  יהיו משתנים מקריים.

ב.  $\mu$  יהיה משתנה מקרי ו  $\bar{X}$  קבוע.

ג.  $\bar{X}$  יהיה משתנה מקרי ו  $\mu$  קבוע.

ד.  $\mu$  ו  $\bar{X}$  יהיו קבועים.

**פתרונות:****שאלה 2**

א.

4	3	2	1	0	X
0.05	0.3	0.35	0.25	0.05	P(x)

$$\mu = 2.05 \quad \sigma^2 = 0.9475 \quad \sigma = 0.973 \quad \text{ב.}$$

$$\mu_{\bar{x}} = 2.05 \quad \sigma_{\bar{x}}^2 = 0.2369 \quad \text{ג.}$$

$$\sigma(\bar{X}) = 0.486$$

**שאלה 3**

$$\mu_{\bar{x}} = 3.5$$

$$\sigma(\bar{X}) = 1.21$$

**שאלה 4**

א. 0.8413

ב. 0.0013

ג. 0

ד. 0.1974

**שאלה 6**

א. 0.0465

ב. 27.71

ג. 0.2628

**שאלה 7**

א. 0

ב. 0.1587

ג. 0.1587

ד. 0.5

**שאלה 8**

א. 0.5468

ב. 0.6826

**שאלה 9**

א.

30	20	10	
0.5	0.25	0.25	P(x)

ב. התוחלת: 22.5

השונות: 68.75

ג. התוחלת: 22.5

השונות: 13.75

ד. 0.8997

**שאלה 10**

0.0475

**שאלה 11**

0.1814

**שאלה 12**

א. 0.9772

ב. 0.0228

ג. 271

**שאלה 14**

התשובה ב

**שאלה 15**

התשובה ד

**שאלה 16**

התשובה ג

## התפלגות סכום תצפיות המדגם ומשפט הגבול המרכזי

### רקע:

$$T = \sum_{i=1}^n X_i$$

כעת נדון בסטטיסטי המבטא את סכום התצפיות במדגם

כאשר כל התצפיות נדגמו באקראי מאותה אוכלוסייה.

כלומר, היו  $X_1, \dots, X_n$  - משתנים מקריים בלתי תלויים בעלי התפלגות זהה שתוחלתה  $\mu$  ושונותה  $\sigma^2$  אזי:

### א. התוחלת והשונות של סכום התצפיות:

$$E(T) = n\mu$$

$$V(T) = n\sigma^2$$

### ב. דגימה מתוך התפלגות נורמלית:

$$T \sim N(n\mu, n\sigma^2)$$

אזי

$$Z = \frac{T - n\mu}{\sqrt{n\sigma^2}}$$

אם  $X \sim N(\mu, \sigma^2)$

### ג. משפט הגבול המרכזי:

$$E(X) = \mu$$

$$V(X) = \sigma^2$$

אם  $X$  מתפלג כלשהו וידוע

אזי עבור מדגם מספיק גדול (לפחות 30)

$$T \rightsquigarrow N(n\mu, n\sigma^2)$$

### דוגמה: (פתרון בהקלטה)

בעיר מסוימת המשכורת הממוצעת של עובד הינה 8000 ₪. עם סטיית תקן של 2000 ₪. נדגמו 100 עובדים מהעיר שמפקידים את משכורתיהם לסניף בנק.

- מה התוחלת וסטיית התקן של סך המשכורות שיופקדו לסניף הבנק על ידי העובדים הללו?
- מה ההסתברות שלסניף יופקד פחות מ-780 אלף ₪ ע"י אותם עובדים? (0.1587)

### תרגילים:

1. המשקל באוכלוסייה מסוימת מתפלג נורמאלית עם תוחלת של 60 ק"ג וסטיית תקן של 10 ק"ג.
  - א. מה הסיכוי שאדם אקראי מהאוכלוסייה ישקול מתחת ל-65 ק"ג?
  - ב. מה הסיכוי שהמשקל הממוצע של 4 אנשים אקראיים יהיה מתחת ל-65 ק"ג?
  - ג. מה הסיכוי שהמשקל הכולל של 4 אנשים אקראיים יהיה מתחת ל-240 ק"ג?
  
2. נפח יין בבקבוק מתפלג נורמאלית עם תוחלת של 750 מ"ל וסטיית תקן של 20 מ"ל. אדם קנה מארז של 4 בקבוקי יין.
  - א. מהי התוחלת ומהי סטיית התקן של נפח היין במארז?
  - ב. את היין שבמארז האדם מזוג לכלי שקיבולתו 3.1 ליטר. מה ההסתברות שהיין יגלוש מהכלי?
  - ג. אם לא היה נתון שנפח היין מתפלג נורמאלית. האם התשובה לסעיף א הייתה משתנה? האם התשובה לסעיף ב הייתה משתנה?
  
3. בספר כלשהו 500 עמודים. קצב הקריאה הממוצע הוא עמוד אחד ב 4 דקות עם סטיית תקן של 1 דקות.
  - א. מה ההסתברות לסיים את הפרק הראשון (40 עמודים) תוך שעתיים וחצי?
  - ב. מהו האחוזון ה-95 לזמן סיום קריאת הספר?



**פתרונות:****שאלה 1**

א. 0.6915

ב. 0.8413

ג. 0.5

**שאלה 2**

א. תוחלת 3000 מ"ל וסטיית תקן 40 מ"ל

ב. 0.0062.

### פרק 3 - 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?

- $\hat{\theta}$  יהיה אומד חסר הטיה ל  $\theta$  אם התוחלת של  $\hat{\theta}$  תהיה שווה ל  $\theta$  :  $E(\hat{\theta}) = \theta$
- טעות התקן של אומד היא סטיית התקן שלו, כלומר :  $\sigma(\hat{\theta}) = S.E$

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:
- 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 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 4 - 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  $\sigma^2$  (שוונות האוכלוסייה) ידועה

הפרמטר שנרצה לאמוד :  $\mu$

האומד נקודתי :  $\bar{x}$

התנאים לבניית רווח הסמך :

1  $X \sim N$  או  $n \geq 30$

2  $\sigma^2$  (שוונות 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.
  - 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.
  
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

### קביעת גודל מדגם באמידת תוחלת עם שונות אוכלוסייה ידועה

#### רקע:

אם מעוניינים לאמוד את ממוצע האוכלוסייה כאשר סטיית התקן של האוכלוסייה ידועה:  $\sigma$   
ברמת סמך של  $1 - \alpha$  ושגיאת אמידה שלא תעלה על  $\varepsilon$  מסוים, נציב בנוסחה הבאה:

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

כדי להציב בנוסחה צריך שהמשתנה הנחקר יתפלג נורמלית או שהמדגם ייצא בגודל של לפחות 30 תצפיות.

#### דוגמה: (פתרון בהקלטה)

חברת תעופה מעוניינת לאמוד את תוחלת משקל המטען של נוסע. נניח שמשקל מטען של נוסע מתפלג נורמלית עם סטיית תקן של 2 ק"ג. כמה נוסעים יש לדגום אם מעוניינים שבביטחון של 98% הסטייה המרבית בין ממוצע המדגם לממוצע האמיתי לא יעלה על 0.5 ק"ג? (תשובה: 87)

**תרגילים:**

1. משתנה מקרי מתפלג נורמאלית עם סטיית תקן ידועה 12. מה צריך להיות גודל המדגם כדי לבנות רווח סמך ברמת סמך של 98% שאורכו לא יעלה על 2?
2. מעוניינים לאמוד את הדופק הממוצע של מתגייסים לצבא. מעוניינים שבביטחון של 95% שגיאת האמידה המרבית תהיה 0.5. נניח שהדופק מתפלג נורמאלית על סטיית תקן של 3 פעימות לדקה.  
א. כמה מתגייסים יש לדגום?  
ב. אם ניקח מדגם הגדול פי 4 מהמדגם של סעיף א ונאמוד את הממוצע באותה רמת סמך כיצד הדבר ישפיע על שגיאת האמידה?
3. יהי  $X$  משתנה מקרי עם ממוצע  $\mu$  וסטיית תקן  $\sigma$ . חוקר רוצה לבנות רווח בר סמך ל- $\mu$  ברמת ביטחון של 0.95 כך שהאורך של הרווח יהיה  $0.5\sigma$ . מהו גודל המדגם הנדרש?



**פתרונות :****שאלה 1**

780

**שאלה 2**

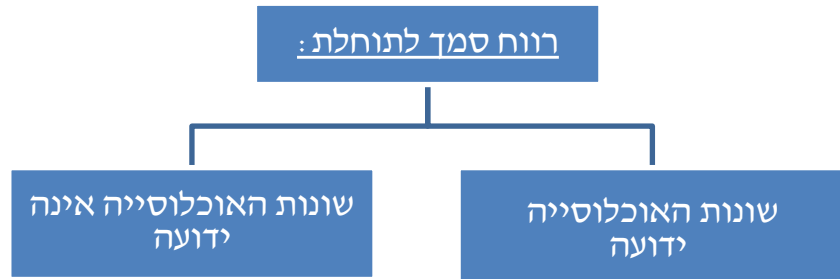
א. 139

ב. הדבר יקטין את  $\varepsilon$  פי 2.**שאלה 3** $n = 62$

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

רקע:

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



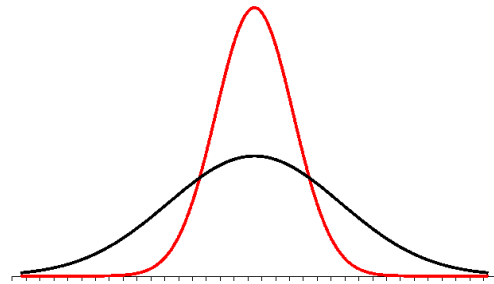
בפרק זה נעסוק במקרה ששונות האוכלוסייה ( $\sigma^2$ ) אינה ידועה לנו. מקרה יותר פרקטי.

התנאי:  $X \sim N$  או שהמדגם גדול

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

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

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 כיתה 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: 89, 79, 84, 88, 84. 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  $S=13$  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. נדגמו 120 אנשים אקראיים מעל גיל 50. עבור כל אדם נבדק מספר שנות השכלתו. להלן התוצאות שהתקבלו:  $\bar{x} = 13.8$ ,  $S = 2$ . בנו רווח סמך ברמת סמך של 96% לממוצע ההשכלה של אזרחים מעל גיל 50.

7. 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  $S = 20$

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

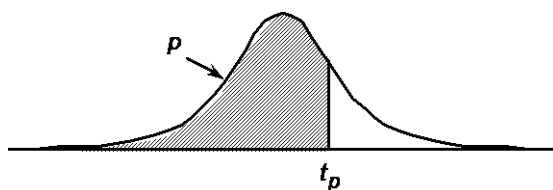
א. סטטיסטיקאי א

ב. סטטיסטיקאי ב

ג. אותו אורך רווח סמך לשני הסטטיסטיקאים.

ד. תלוי בתוצאות המדגם של כל סטטיסטיקאי.

8. 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?



P

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

**פתרונות:****שאלה 1**

$$א. \quad 79.88 < \mu < 89.72$$

**שאלה 4**

$$א. \quad 96.63 < \mu < 107.37$$

$$ב. \quad 96.90 < \mu < 107.10$$

**שאלה 5**

$$3.149 < \mu < 3.351$$

**שאלה 8**

90%

## פרק 5 - 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

### קביעת גודל מדגם באמידת פרופורציה

#### רקע:

בפרק זה נדון איך קובעים גודל מדגם שבאים לאמוד פרופורציה באוכלוסייה מסוימת:  
 החוקר קובע מראש את רמת הסמך הרצויה:  $1 - \alpha$ .  
 החוקר קובע מראש את הטעות הסטטיסטית המרבית שבה הוא מעוניין:  $\varepsilon$  (או את אורך רווח הסמך).

$$L = 2\varepsilon - \text{אורך רווח הסמך.}$$

$\varepsilon$  - טעות אמידה מרבית: המרחק המקסימאלי (הסטייה) בין הפרמטר  $(p)$  לאומד  $(\hat{p})$ .

$$\varepsilon = z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

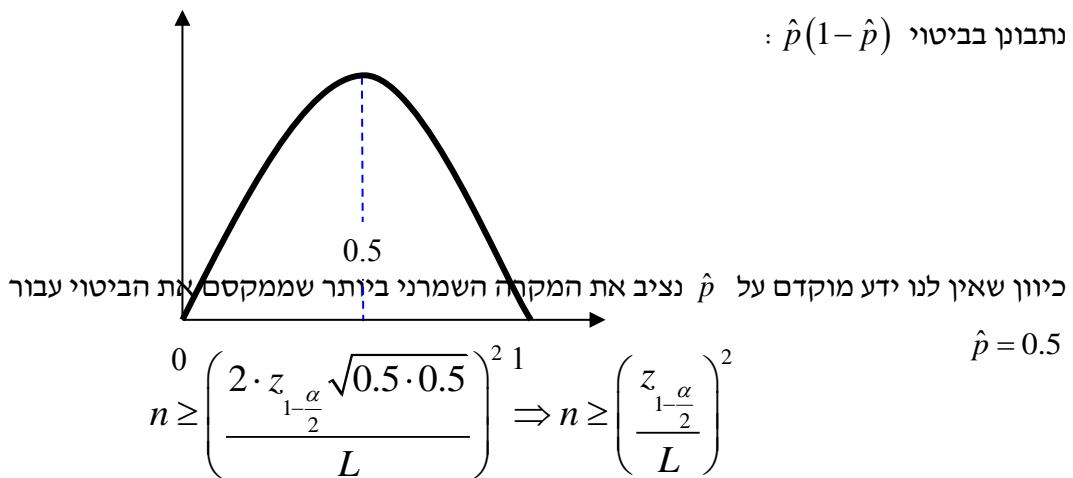
ויתעניין לדעת מהו גודל המדגם הרצוי לשם כך.

נקבל ש:

$$n \geq \left( \frac{2 \cdot z_{1-\alpha/2} \cdot \sqrt{\hat{p}(1-\hat{p})}}{L} \right)^2$$

הבעיה שאין לנו יודעים את  $\hat{p}$ .

נתבונן בביטוי  $\hat{p}(1-\hat{p})$ :



**דוגמה:** (פתרון בהקלטה)

מעוניינים לאמוד את שיעור האבטלה במשק. האמידה צריכה להתבצע ברמת סמך של 90% ועם שגיאת אמידה שלא תעלה על 4%.  
 מהו גודל המדגם המינימאלי שיש לקחת?

#### פתרון:

423

### תרגילים :

1. הממשלה אומדת מדי חודש את אחוז התמיכה בה. מהו גודל המדגם אשר יש לקחת אם דורשים שהאומדן לא יסטה מהאחוז האמתי באוכלוסייה ביותר מ-3%, וזאת בביטחון של 95%?
2. ענה על הסעיפים הבאים :
  - א. כמה אזרחים יש לדגום כדי לאמוד את אחוז התמיכה בממשלה עם אורך רווח הסמך שלא עולה על 9% ברמת סמך של 90%?
  - ב. בהנחה ובוצע מדגם שאת גודלו חישובתם בסעיף א והתקבל שאחוז התמיכה בממשלה במדגם הנו 42%. בנו רווח סמך לאחוז התמיכה בממשלה ברמת סמך של 95%.
  - ג. על סמך סעיף ב'. האם תקבל את הטענה שמיעוט האוכלוסייה תומך הממשלה?
3. משרד הבריאות מתכנן לבצע מדגם שמטרתו לבדוק את הסיכוי לחלות בשפעת עם לקיחת חיסון נגד שפעת. הוא מעוניין שבסיכוי של 98% טעות האמידה לא תעלה על 3%.
  - א. כמה מחוסנים יש לדגום ?
  - ב. משרד הבריאות ביצע את המדגם שאת גודלו חישובת בסעיף הקודם וקיבל ש 15% מבין אלה שקיבלו חיסון נגד שפעת בכל זאת חלו במשך החורף בשפעת. בנו ברמת סמך של 98% את הסיכוי לחלות בחורף בשפעת עם לקיחת חיסון נגד שפעת.

**פתרונות :**

**שאלה 1 :**

1068

## פרק 6 - רווח סמך להפרש פרופורציות

### רקע:

המטרה: לאמוד את  $p_1 - p_2$ : הפרש פרופורציות בין שתי אוכלוסיות שונות.

האומד הנקודתי:  $\hat{p}_1 - \hat{p}_2$

התנאי לבניית רווח הסמך: כל מדגם מעל 30 או לבדוק שמספר ההצלחות ומספר הכישלונות בכל מדגם לפחות 5 בכל מדגם (יש כאלה שבודקים לפחות 10).

### רווח סמך:

$$(\hat{p}_1 - \hat{p}_2) \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

רק שאפס נופל בתחומי רווח הסמך להפרש הפרופורציה נאמר שלא ניתן לקבוע שקיים הבדל מובהק בין הפרופורציות באוכלוסיות.

### דוגמה: (פתרון בהקלטה)

במטרה להשוות בין שתי תרופות נדגמו 200 איש שלקחו תרופה x. מתוכם 180 טענו שהתרופה עזרה להם. כמו כן נלקחו 300 איש שלקחו את תרופה y. מתוכם 150 טענו שהתרופה עזרה להם. בנו רווח סמך להפרש אחוזי ההצלחה של התרופות ברמת סמך של 95%. מה ניתן לומר על סמך רווח הסמך על ההבדלים בין התרופות?

### פתרון:

(47%, 33%)

### תרגילים:

1. מתוך 150 נשים שנדגמו באקראי 30% תמכו בהצעת חוק מסוימת. מתוך 200 גברים שנדגמו באקראי 25% תמכו בהצעת החוק.  
 א. בנו רווח סמך לפער בין אחוזי התמיכה של הנשים לעומת הגברים ברמת סמך של 96%.  
 ב. בנו רווח סמך ברמת סמך של 95% לאחוז התמיכה בהצעת החוק.
  
2. במחקר רפואי השתתפו 200 אנשים הסובלים מכאבים כרוניים. הם חולקו באקראי ל-2 קבוצות שוות בגודלן.  
 קבוצה 1 קיבלה את תרופה A וקבוצה שנייה קיבלה את תרופה B.  
 בקרב לוקחי תרופה A 90 טענו שמצבם השתפר. בקרב לוקחי תרופה B 70 טענו שמצבם השתפר.  
 א. בנו רווח סמך ברמת סמך של 95% להפרש בין שיעורי ההצלחה של שתי התרופות.  
 ב. האם על סמך סעיף א ניתן לקבוע שקיים הבדל בין התרופות מבחינת שיעורי ההצלחה?
  
3. נדגמו 200 משפחות מגוש דן. ל-70% מתוכן מכשיר DVD בבית.  
 נדגמו 300 משפחות מאזור הצפון ל-65% מתוכן מכשיר DVD בבית.  
 א. בנו רווח סמך ברמת סמך של 98% לפרופורציות המשפחות בגוש דן עם DVD בבית.  
 ב. בנו רווח סמך ברמת סמך של 95% להפרש בין פרופורציות המשפחות בגוש דן עם DVD לבין פרופורציות המשפחות בצפון עם DVD.

**פתרונות:****שאלה 2**

$$.א \quad 0.093 < P_A - P_B < 0.307$$

**שאלה 3**

$$.א \quad 0.625 < p < 0.7754$$



## פרק 7 - רווח סמך להפרש תוחלות ממדגמים בלתי תלויים

### כששוניות 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

## כששונות האוכלוסייה אינן ידועות אך שוות והמדגמים בלתי תלויים

### 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 השונות שוות אנו אומדים את השונות הזו על ידי שקלול שתי השונות של שני המדגמים על ידי הנוסחה הבאה:

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

דרגות החופש:  $d.f = n_1 + n_2 - 2$

רווח סמך:

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

אם הערך אפס נופל בגבולות 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 8 - 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:

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 9 - תרגול מסכם ברווחי סמך

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

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

- א. תנו רווח סמך לממוצע 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 הני"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. להלן תוצאות מדגם שבדק עבור כל משפחה האם יש לה בבית מכשיר טאבלט.

אזור מגורים	גוש דן	שאר הארץ
גודל המדגם	200	240
מספר משפחות בעלי טאבלט	160	168

- א. בנו רווח סמך להבדל בין אחוז המשפחות עם טאבלט בגוש דן ואחוז המשפחות בעלי טאבלט בשאר חלקי הארץ. ברמת סמך של 98%.
- ב. בנו רווח סמך לפרופורצית משפחות בעלות טאבלט בכלל הארץ ברמת סמך של 95%.

5. 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%.

6. 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.  $0.5\% < p_1 - p_2 < 19.5\%$

ב.  $0.704 \leq p \leq 0.768$

**שאלה 5**

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

ב.  $6467 \leq \mu \leq 7133$

**שאלה 6**

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

## פרק 10 - 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?



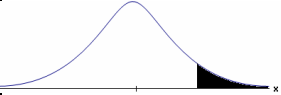


## תרגילים:

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 11 - 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 שסטית 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

## 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 נוסיף 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:**


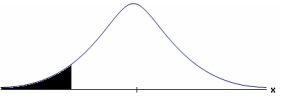

תשובה: א

**שאלה 8:**

תשובה: ג

**בדיקת השערות על תוחלת (ממוצע) כאשר שונות האוכלוסייה אינה ידועה**

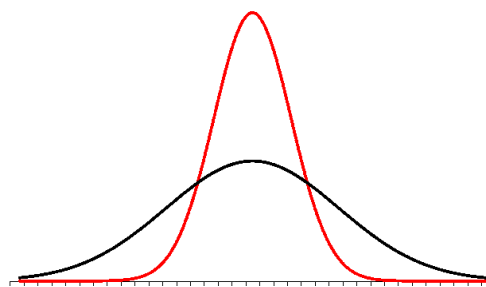
**רקע:**

$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$	<b>השערת האפס:</b> <b>השערה אלטרנטיבית:</b>
7. $\sigma$ אינה ידועה 8. $X \sim N$ או מדגם מספיק גדול			<b>תנאים:</b>
$t_{\bar{x}} > t_{1-\alpha}^{(n-1)}$  $t_{1-\alpha, n-1}$ <b>דוחים את <math>H_0</math></b>	$t_{\bar{x}} < -t_{1-\alpha}^{(n-1)}$  $-t_{1-\alpha, n-1}$ <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> <b>אזור הדחייה של <math>H_0</math>:</b>
$\bar{X} > \mu_0 + t_{1-\alpha}^{n-1} \cdot \frac{S}{\sqrt{n}}$	$\bar{X} < \mu_0 - t_{1-\alpha}^{n-1} \cdot \frac{S}{\sqrt{n}}$	$\bar{X} > \mu_0 + t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S}{\sqrt{n}}$ $\bar{X} < \mu_0 - t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S}{\sqrt{n}}$	<b>חלופה לכלל הכרעה:</b> <b>נדחה <math>H_0</math> אם מתקיים:</b>

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 ופתר - 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 ) כאשר שונות האוכלוסייה לא ידועה

### רקע:

נזכיר שהמסקנה 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.

$H_0 : \mu = \mu_0$ $H_1 : \mu > \mu_0$	$H_0 : \mu = \mu_0$ $H_1 : \mu < \mu_0$	$H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$	<b>השערת האפס :</b> <b>השערה אלטרנטיבה :</b>
9. $\sigma$ אינה ידועה 10. $X \square N$ או מדגם מספיק גדול			<b>תנאים :</b>
$P_{H_0}(\bar{X} \geq \bar{x})$	$P_{H_0}(\bar{X} \leq \bar{x})$	אם $2 \cdot P_{H_0}(\bar{X} \geq \bar{x}) \leftarrow \bar{x} > \mu_0$ אם $2 \cdot P_{H_0}(\bar{X} \leq \bar{x}) \leftarrow \bar{x} < \mu_0$	<b>p-value</b>

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

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

$$d.f = n - 1$$



**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 :

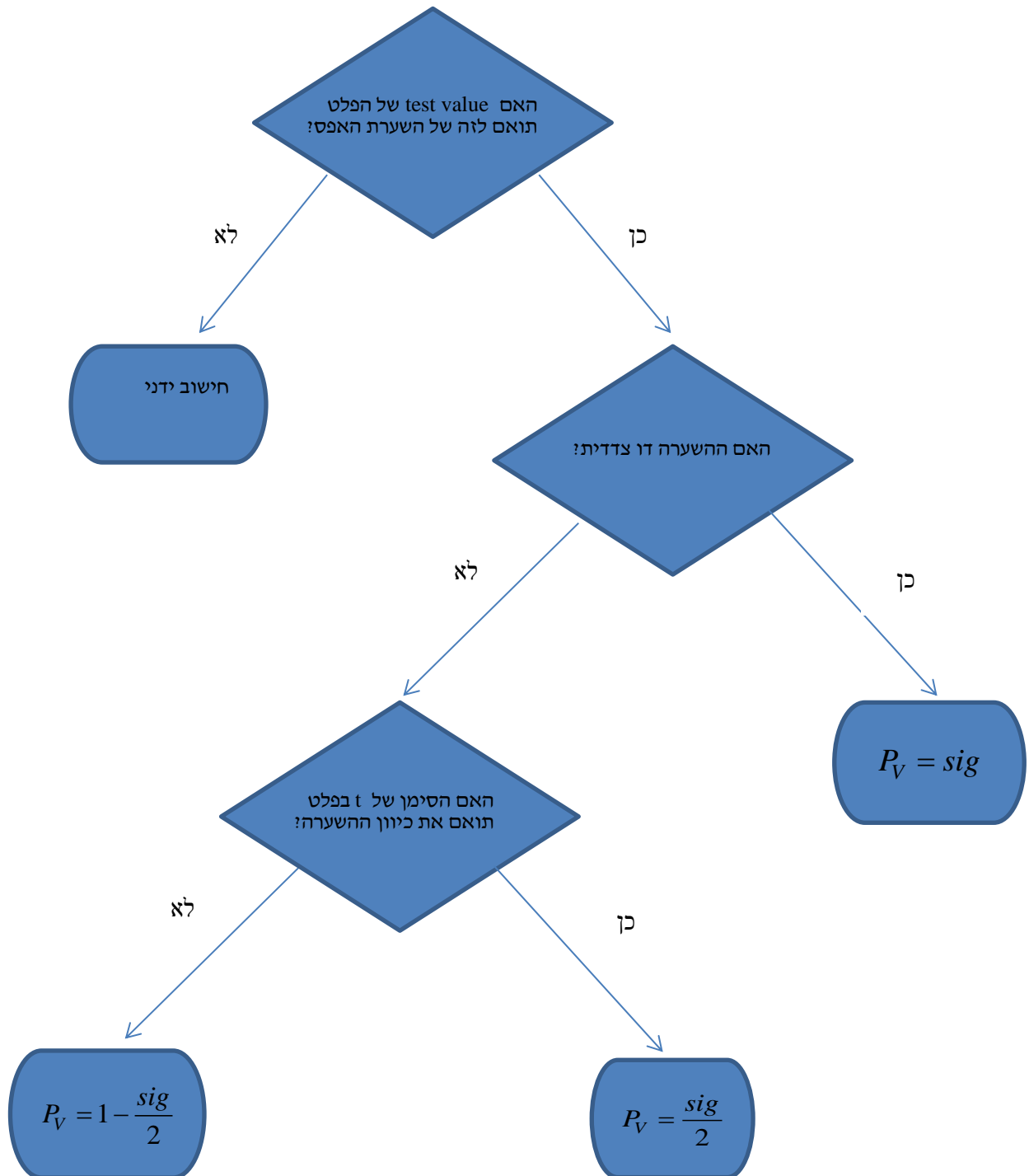
שאלה 3:

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:**

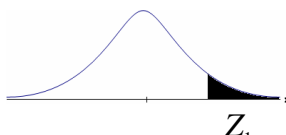
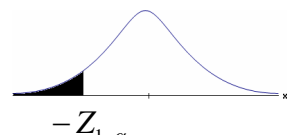
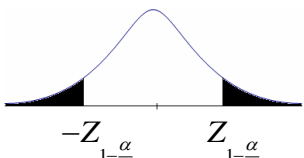
א.  $191.8 \leq \mu \leq 200.2$

ב. נכריע שיש אמת בפרסום.

## פרק 12 - 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. בדקו 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



$H_0 : p = p_0$ $H_1 : p > p_0$	$H_0 : p = p_0$ $H_1 : p < p_0$	$H_0 : p = p_0$ $H_1 : p \neq p_0$	<b>השערת האפס :</b> <b>השערה</b> <b>אלטרנטיבית :</b>
$np_0 \geq 5 \& n(1-p_0) \geq 5$			<b>תנאים :</b>
$\hat{p} > p_0 + Z_{1-\alpha} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$	$\hat{p} < p_0 - Z_{1-\alpha} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$	$\hat{p} > p_0 + Z_{1-\alpha/2} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$ או $\hat{p} < p_0 - Z_{1-\alpha/2} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$	<b>כלל ההכרעה :</b> <b>אזור הדחייה של</b> $H_0$
$P_{H_1}(\hat{p} < p_0 + Z_{1-\alpha} \cdot \sqrt{\frac{p_0(1-p_0)}{n}})$	$P_{H_1}(\hat{p} > p_0 - Z_{1-\alpha} \cdot \sqrt{\frac{p_0(1-p_0)}{n}})$	$P_{H_1}(p_0 - Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0(1-p_0)}{n}} < \hat{p} < p_0 + Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0(1-p_0)}{n}})$	<b>חישוב <math>\beta</math> :</b>

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.

### 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:

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.

$H_0 : p = p_0$ $H_1 : p > p_0$	$H_0 : p = p_0$ $H_1 : p < p_0$	$H_0 : p = p_0$ $H_1 : p \neq p_0$	<b>השערת האפס :</b> <b>השערה אלטרנטיבית:</b>
$np_0 \geq 5 \ \& \ n(1-p_0) \geq 5$			<b>תנאים:</b>
$P_{H_0}(\hat{P} \geq \hat{p})$	$P_{H_0}(\hat{P} \leq \hat{p})$	אם $2 \cdot P_{H_0}(\hat{P} \geq \hat{p}) \Leftarrow \hat{p} > p_0$ אם $2 \cdot P_{H_0}(\hat{P} \leq \hat{p}) \Leftarrow \hat{p} < p_0$	<b>p-value</b>

$$\hat{P} \sim N(p_0, \frac{p_0(1-p_0)}{n}) : \text{ כאשר בהנחת השערת האפס :}$$

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

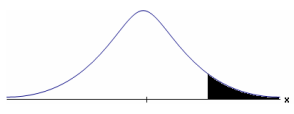
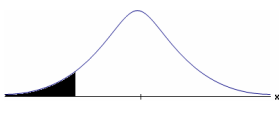

**שאלה 6:**

0.7486



**פרק 13 - בדיקת השערות על הפרש פרופורציות**

**רקע:**

$H_0 : p_1 - p_2 = 0$ $H_1 : p_1 - p_2 > 0$	$H_0 : p_1 - p_2 = 0$ $H_1 : p_1 - p_2 < 0$	$H_0 : p_1 - p_2 = 0$ $H_1 : p_1 - p_2 \neq 0$	<b>השערת האפס :</b> <b>השערה אלטרנטיבית:</b>
2. מדגמים גדולים		1. מדגמים בלתי תלויים	<b>תנאים:</b>
$Z_{\hat{p}_1 - \hat{p}_2} > Z_{1-\alpha}$  $Z_{1-\alpha}$ - דוחים את $H_0$	$Z_{\hat{p}_1 - \hat{p}_2} < -Z_{1-\alpha}$  $-Z_{1-\alpha}$ - דוחים את $H_0$	או $Z_{\hat{p}_1 - \hat{p}_2} < -Z_{1-\frac{\alpha}{2}}$ $Z_{\hat{p}_1 - \hat{p}_2} > Z_{1-\frac{\alpha}{2}}$  $-Z_{1-\frac{\alpha}{2}}$ $Z_{1-\frac{\alpha}{2}}$ - דוחים את $H_0$	<b>כלל ההכרעה:</b> אזור הדחייה של :

**סטטיסטי המבחן :**

כאשר הפרופורציה המשוקללת:  $\hat{p} = \frac{y_1 + y_2}{n_1 + n_2} = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2}$

$$Z_{\hat{p}_1 - \hat{p}_2} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}}$$

**חלופה אחרת לכלל הכרעה:**

$\hat{p}_1 - \hat{p}_2 > 0 + Z_{1-\alpha} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}$	$\hat{p}_1 - \hat{p}_2 < 0 - Z_{1-\alpha} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}$	$\hat{p}_1 - \hat{p}_2 > 0 + Z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}$ או $\hat{p}_1 - \hat{p}_2 < 0 - Z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}$	<b>כלל ההכרעה:</b> אזור הדחייה של $H_0$
---	---	--	--

$$\hat{p}_1 - \hat{p}_2 \sim N(p_1 - p_2, \frac{p_1 \cdot q_1}{n_1} + \frac{p_2 \cdot q_2}{n_2})$$

התפלגות של  $\hat{p}_1 - \hat{p}_2$  :

$$Z_{\hat{p}_1 - \hat{p}_2} = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}}$$

$$Z_{\hat{p}_1 - \hat{p}_2} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}}$$

תקנון:

**דוגמה:** (פתרון בהקלטה)

נדגמו 80 סטודנטים שנבחנו במיקרו-כלכלה. מתוכם 60 עברו את הבחינה. נדגמו 100 סטודנטים שנבחנו בסטטיסטיקה א'. מתוכם 82 עברו את הבחינה. האם שיעור העוברים את הבחינה בסטטיסטיקה גבוה מאשר מהבחינה במיקרו כלכלה? בדקו ברמת מבוהקות של 10%.

**תרגילים:**

1. במדגם של 200 גברים. 8% מהם היו מובטלים. המדגם של 180 נשים 10% מהן היו מובטלות. האם קיים הבדל מובהק בין פרופורציית המובטלים לפרופורציית המובטלות. בדוק ברמת מובהקות של 5%.
2. אחוז בעלי רישיון נהיגה בקרב האוכלוסייה הבוגרת הינו 60%. במדגם של 300 בוגרים מתל אביב 204 היו בעלי רישיון נהיגה. במדגם של 220 בוגרים מירושלים 100 היו בעלי רישיון נהיגה.
  - א. ברמת מובהקות של 5% האם תקבלו את הטענה שאחוז בעלי הרישיון בתל אביב גבוה מהאחוז הארצי?
  - ב. ברמת מובהקות של 10% האם תקבלו את הטענה שאחוז בעלי הרישיון נהיגה בתל אביב גבוה מאחוז בעלי רישיון הנהיגה בירושלים?
3. נדגמו 500 בוגרים מתוכם 200 גברים והיתר נשים. במדגם התקבל: מתוך הגברים ל-48% תעודת בגרות. מתוך הנשים ל-58% תעודת בגרות. מטרת המחקר היא לבדוק האם שיעור הזכאיות לבגרות גבוה משיעור הזכאים.
  - א. מהי מובהקות התוצאה?
  - ב. מה תהיה המסקנה ברמת מובהקות של 8%?
4. במדגם שנערך על 100 פרות מחוות בדרום הארץ התקבל כי 20 פרות נושאות וירוס מסוים. במדגם שנערך על 200 פרות מחוות בצפון הארץ התקבל כי 10 מתוכן נושאות וירוס גם כן.
  - א. בנו מבחן ברמת מובהקות של 5% לבדיקת הטענה כי הווירוס תקף את פרות הדרום באופן משמעותי יותר מאשר את הפרות בצפון הארץ.
  - ב. מהי המסקנה לבדיקת הטענה של סעיף א ומהי הטעות האפשרית במסקנה?
  - ג. מהי עוצמת המבחן אם שיעור הפרות בדרום עם הווירוס גבוה ב-10% משיעור הפרות בצפון עם הווירוס?
  - ד. כיצד העוצמה תשתנה אם נגדיל את רמת המובהקות?

**פתרונות:**

**שאלה 1:**

לא נדחה את  $H_0$

**שאלה 2:**

א. נדחה  $H_0$

ב. נדחה  $H_0$

**שאלה 3:**

א. 0.0139

ב. נדחה  $H_0$

**שאלה 4:**

ב. נדחה  $H_0$



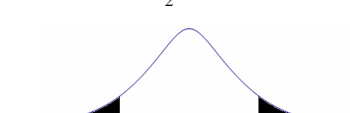
ג. 0.8238

ד. תגדל

## פרק 14 - בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

### כשהשונויות של האוכלוסייה ידועות

#### רקע:

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

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



**בששונות האוכלוסייה 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>
<p>4. מדגמים בלתי תלויים</p> <p>5. <math>\sigma_1, \sigma_2</math> לא ידועות אך שוות</p> <p>6. המשתנים בכל אוכלוסייה מתפלגים נורמלית</p>			<b>תנאים:</b>
$t_{\bar{x}_1 - \bar{x}_2} > t_{1-\alpha}^{(n_1+n_2-2)}$  <p align="center"><math>t_{1-\alpha, n_1+n_2-2}</math></p> <p align="center">- דוחים את <math>H_0</math></p>	$t_{\bar{x}_1 - \bar{x}_2} < -t_{1-\alpha}^{(n_1+n_2-2)}$  <p align="center"><math>-t_{1-\alpha, n_1+n_2-2}</math></p> <p align="center">- דוחים את <math>H_0</math></p>	<p>או <math>t_{\bar{x}_1 - \bar{x}_2} &lt; -t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)}</math></p> <p><math>t_{\bar{x}_1 - \bar{x}_2} &gt; t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)}</math></p>  <p align="center"><math>-t_{1-\frac{\alpha}{2}, n_1+n_2-2}</math>      <math>t_{1-\frac{\alpha}{2}, n_1+n_2-2}</math></p> <p align="center">- דוחים את <math>H_0</math></p>	<b>אזור הדחייה של <math>H_0</math> :</b>

**סטטיסטי המבחן :**

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

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

**חלופה אחרת לכלל הכרעה:**

$\bar{x}_1 - \bar{x}_2 > c + t_{1-\alpha}^{(n_1+n_2-2)} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$	$\bar{x}_1 - \bar{x}_2 < c - t_{1-\alpha}^{(n_1+n_2-2)} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$	$\bar{x}_1 - \bar{x}_2 > c + t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$ <p align="center">או</p> $\bar{x}_1 - \bar{x}_2 < c - t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$	<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:**

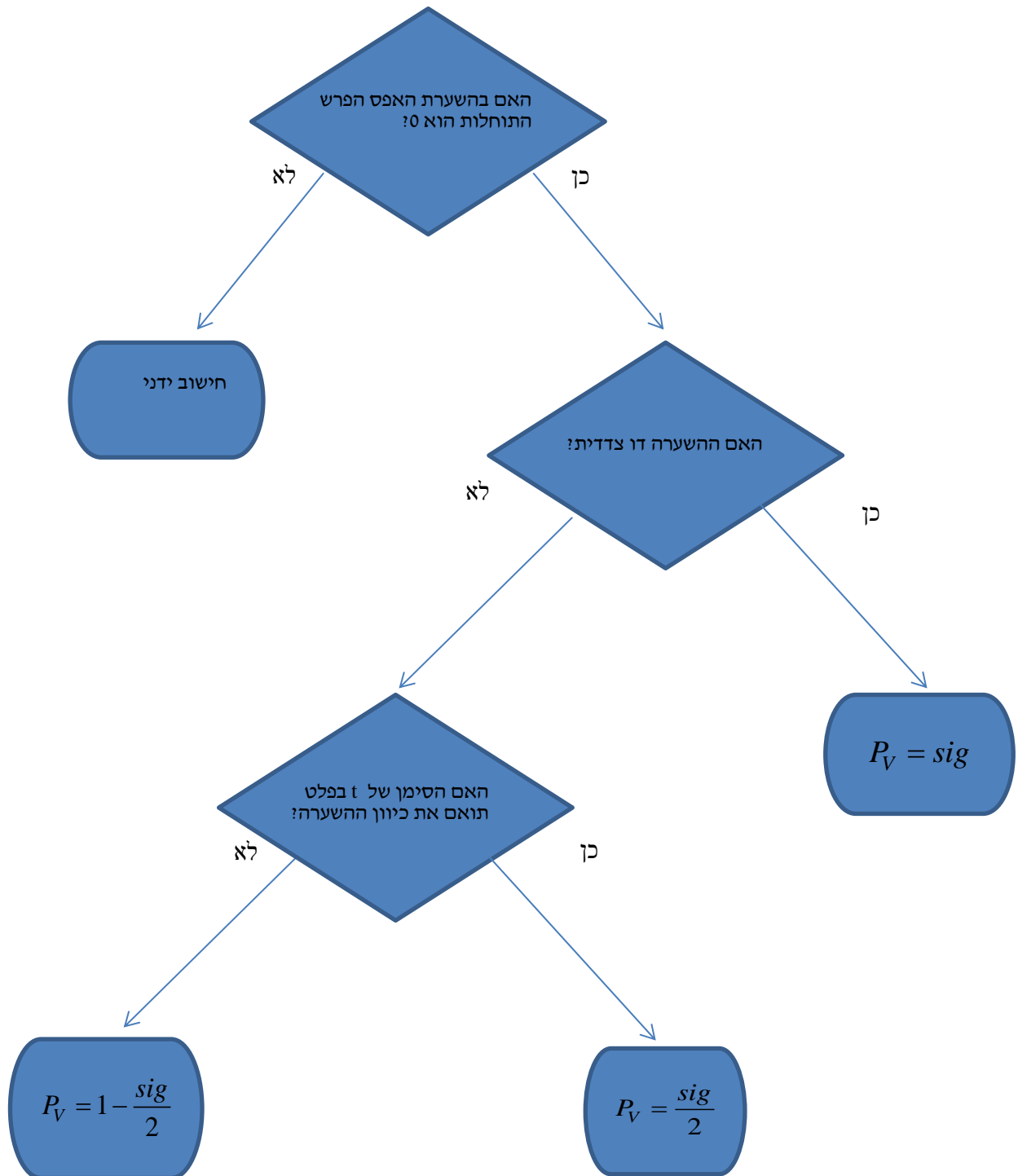
א. נדחה  $H_0$

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.

- א. רשמו את השערות המחקר והסבירו מהו המבחן המתאים כאן.
  - ב. כיצד הייתה משתנה התשובה לסעיף הקודם אם היה מדובר על אותם אנשים שציונם נבדק פעם אחרי המבחן הראשון שעשו ופעם אחרי המבחן השני?
  - ג. 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  $\alpha=0.05$ .
  - ד. 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 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.



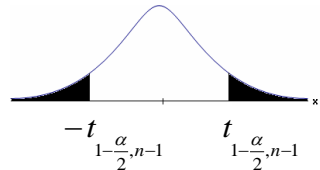
**שאלה 4:**

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

**פרק 15 - 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 $D \square N$ 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 הכרעה:</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 :

שופרסל	מגה בעיר	המוצר
18	17	שמפו
57	48	גיל כביסה
35	35	עוגת גבינה
10	12	לחם
47	49	קפה נמס
142	113	בקבוק יין
26	20	גבינה בולגרית

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:**

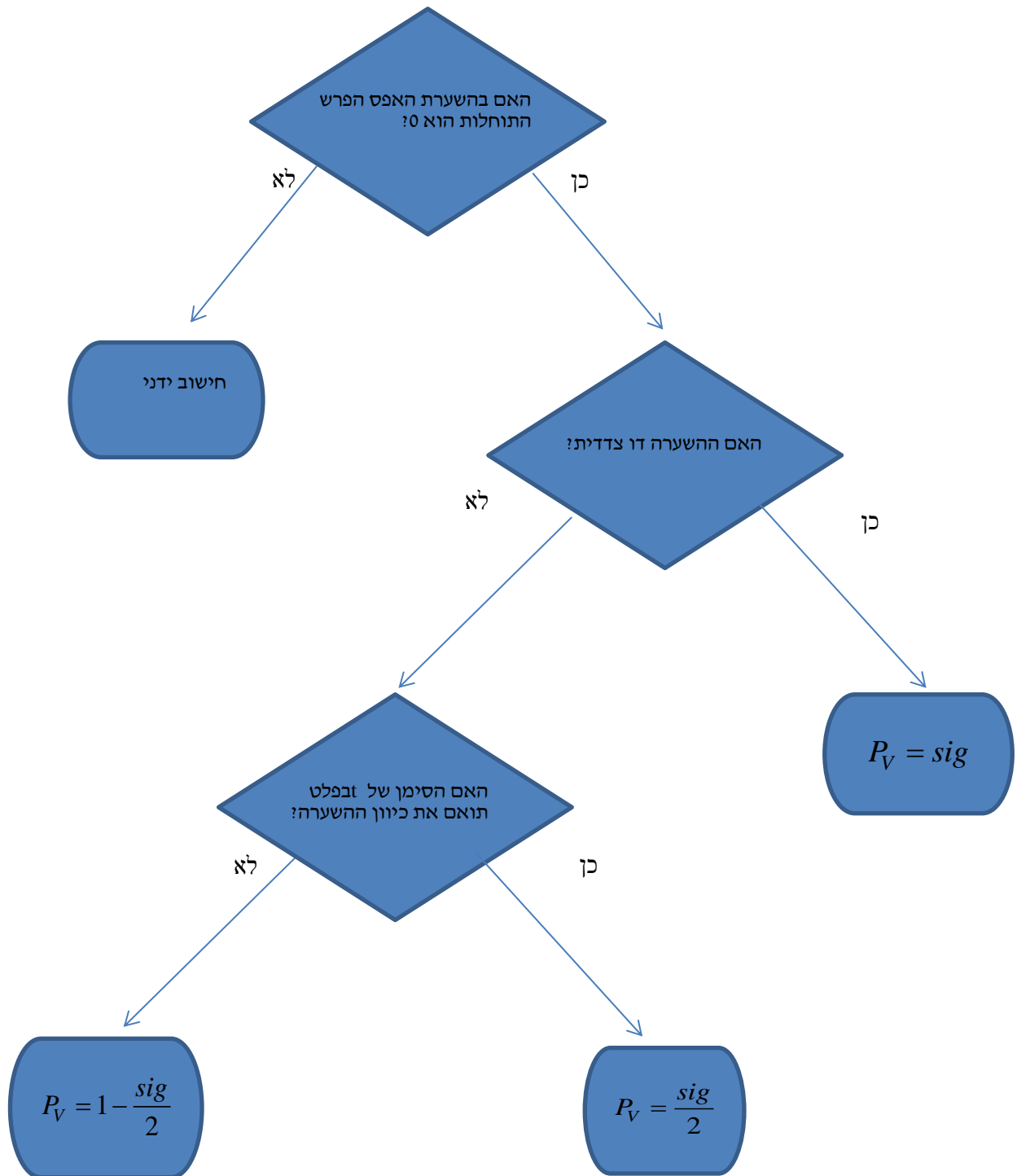
התשובה היא ד.

**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 16 - 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. נדגמו 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 17- שאלות מסכמות בבדיקת השערות על פרמטרים

### שאלות מסכמות בבדיקת השערות על פרמטרים

1. שני חוקרים נתבקשו לבדוק את ההשערות הבאות:

$$H_0 : \mu = 520$$

$$H_1 : \mu > 520$$

כל חוקר בדק מדגם של 225 נחקרים. ידוע ש-  $\sigma = 20$ .

חוקר א' קבע את כלל ההכרעה לפי  $\alpha = 0.05$ .

חוקר ב' מחליט לדחות  $H_0$  אם  $\bar{X} > 522$ .

א. למי מהחוקרים הסתברות לטעות מסוג ראשון קטנה יותר?

ב. מהי ההסתברות לטעות מסוג שני של חוקר ב' עבור  $\mu_1 = 525$ .

ג. הסבר ללא חישוב נוסף, האם ההסתברות לטעות מסוג שני עבור  $\mu_1 = 525$ , של חוקר א' שווה/קטנה/גדולה לזו של חוקר ב'.

ד. חוקר א' קיבל במדגם שלו  $\bar{X} = 523$ . מהי מסקנתו?

2. ידוע כי תוחלת מספר ה"לייקים" היומי של דנה היא 12 עם סטיית תקן 5.

דני טוען שהוא יותר פופולארי מדנה בכך שהוא מקבל יותר "לייקים" מדנה ביום. על-מנת לבדוק זאת ספר דני כמה "לייקים" הוא קיבל בכל יום במהלך 7 שבועות (כלומר, ב- 49 ימים) וקיבל סך-הכול 637 "לייקים". נניח כי סטיית התקן של מספר ה"לייקים" שדני מקבל ביום זהה לסטיית התקן של דנה.

א. מהי רמת המובהקות שכדאי לדני לדרוש, כדי שדנה תשתכנע בצדקת טענתו (שדני פופולארי יותר בכך שהוא מקבל יותר "לייקים" מדנה ביום).

ב. אם דני משער שתוחלת מספר ה"לייקים" שהוא מקבל ביום היא 14 וקובע רמת מובהקות 2.5%, מהי עוצמת המבחן של דני?

3. ברצוננו להשוות בין רשתות A לבין B. לשם כך בחרנו 4 מוצרים, ובדקנו את מחיריהם בשתי הרשתות. להלן התוצאות:

מוצר	A	B
1	5	5
2	4	5
3	5	3
4	7	4

- הניחו כי המחירים מתפלגים נורמלית.  
 אם יש הנחות נוספות כדי לבצע את המבחן הפרמטרי רשמו אותן.  
 א. בדקו האם קיים הבדל בין הרשתות מבחינת תוחלת המחירים. רמת מובהקות של 5%.  
 ב. חזרו על הסעיף הקודם בהנחה ונבחרו בכל רשת מוצרים באקראי ולא בהכרח אותם מוצרים.

4. במטרה לבדוק האם סטודנטים הלומדים במכללות משקיעים יותר זמן ללימודים מאשר סטודנטים באוניברסיטאות נדגמו 12 סטודנטים ובדקו לכל סטודנט את הזמן שהוא משקיע ביום ללימודים. הזמנים נמדדו בדקות:

180	140	171	189	156	176	סטודנטים באוניברסיטאות
150	204	186	191	190	180	סטודנטים במכללות

- א. נסח את ההשערות ובדוק אותן ברמת מובהקות של 5%. רשום את כלל ההכרעה ואת ההנחות הדרושות לביצוע המבחן הפרמטרי.  
 ב. חשב את P-value.  
 ג. ישנה טענה שממוצע זמן ההשקעה בלימודים במכללות הוא 3.5 שעות ביום. בדוק את הטענה כאשר רמת המובהקות הינה 5%.

5. במטרה להשוות בין אחוזי הצפייה של גברים ונשים בתוכנית טלוויזיה מסוימת בוצע סקר ובו התקבלו תוצאות הבאות:

לא צופים	צופים	
42	320	נשים
120	72	גברים

- א. האם יש הבדל בין אחוזי הצפייה של גברים ונשים ברמת מובהקות של 1%?
- ב. עבור רמת מובהקות של 5% בדוק טענה שמבין הצופים בתוכנית הטלוויזיה אחוז הנשים גדול פי 2 מאחוז הגברים.
6. בשנת 2000 ל-60% היה מדיח כלים בבית. מחקר רוצה לבדוק האם כיום פרופורציית המשפחות עם מדיח כלים עלה. הוחלט לבצע מדגם אקראי של 150 משפחות.
- א. רשמו את השערות המחקר.
- ב. מה היא מסקנת המחקר ברמת מובהקות של 5% אם במדגם ל-102 משפחות היה מדיח כלים.
- ג. מהי הטעות האפשרית במסקנה מהסעיף הקודם. האם ניתן לדעת את הסתברותה?
7. נערך מחקר על הקשר בין עישון ויתר לחץ דם. נבדק מדגם מקרי של 200 מעשנים ונמצא כי 30 סבלו מיתר לחץ דם. ידוע שבאוכלוסייה 18% סובלים מיתר לחץ דם.
- א. בדוק ברמת מובהקות 0.1 את השערה כי אחוז הסובלים מיתר לחץ דם בקרב המעשנים גדול מאשר כלל האוכלוסייה.
- ב. מהי רמת המובהקות המינימלית לקבלת הטענה שאחוז הסובלים מיתר לחץ דם בקרב המעשנים גדול מאשר כלל האוכלוסייה.
- ג. מהי עצמת המבחן, אם אחוז הסובלים מיתר לחץ דם בקרב אוכלוסיית המעשנים היא בפועל 25%.

8. להלן התפלגות מספר הנסיעות לחופשה השנתית במדגם של משפחות ישראליות :

4	3	2	1	0	מספר הנסיעות
12	20	26	102	84	מספר המשפחות

בדוק ברמת מובהקות של 5% :

א. באיטליה משפחות נוסעות בממוצע פעמיים בשנה לחופשה. האם בישראל משפחות נוסעות פחות מאשר באיטליה?

ב. בהולנד 80% מהמשפחות נוסעות לפחות פעם אחת בשנה לחופשה, האם בישראל אחוז המשפחות שנוסעות לפחות פעם אחת בשנה לחופשה נמוך מאשר בהולנד?

9. נתון כי :

$$X \sim N(\mu, \sigma^2 = 10^2)$$

מעוניינים לבדוק את ההשערות :

$$H_0 : \mu = 40$$

$$H_1 : \mu > 40$$

דגמו 25 תצפיות מהאוכלוסייה והתקבל  $\bar{X} = 45$ .

א. חשבו את P-value (מובהקות התוצאה).

ב. חזור על סעיף א אם ההשערה האלטרנטיבית הייתה :

$$H_1 : \mu < 40$$

ג. חזור על סעיף א אם ההשערה האלטרנטיבית הייתה :

$$H_1 : \mu \neq 40$$

**פתרונות:****שאלה 1:**

- א. חוקר א
- ב. 0.0122
- ג. גדלה
- ד. נדחה  $H_0$

**שאלה 2:**

- א. לפחות 0.0808
- ב. 0.7995

**שאלה 3:**

- א. לא נדחה  $H_0$
- ב. לא נדחה  $H_0$

**שאלה 4:**

- א. לא נדחה  $H_0$
- ב. בין 5% ל- 10%
- ג. נדחה  $H_0$

**שאלה 5:**

- א. נדחה  $H_0$
- ב. נדחה  $H_0$

**שאלה 6:**

- ב. נדחה  $H_0$
- ג. טעות מסוג ראשון בסיכוי של 0.05

**שאלה 7:**

- א. לא נדחה  $H_0$
- ב. 0.8643
- ג. 0.8749



שאלה 8:א. נדחה  $H_0$ ב. נדחה  $H_0$ שאלה 9:

א. 0.0062

ב. 0.9938

ג. 0.0124

### שאלות מסכמות בסגנון רב ברירה ( אמריקאיות) על בדיקת השערות

1. בבדיקת השערה חד-צדדית ימנית ברמת מובהקות  $\alpha=0.01$ , נדחתה השערת האפס. מה הייתה המסקנה לו נבדקה אותה ההשערה באמצעות אותם נתונים ברמת מובהקות  $\alpha=0.05$ ?
- השערת האפס הייתה נדחית.
  - השערת האפס לא הייתה נדחית.
  - ההשערה המחקרית הייתה נדחית.
  - בהעדר נתונים נוספים, לא ניתן לדעת.
2. על מנת לבדוק האם ההסתברות ללידת בן הנה חצי, נבחר מדגם מקרי של 200 ילדים, ונמצא שישנם 120 בנים.
- מהן ההשערות האלטרנטיביות להשערת האפס?
- $H_1 : p = 0.5$
  - $H_1 : p = 0.6$
  - $H_1 : p > 0.5$
  - $H_1 : p \neq 0.5$
3. 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 האמיתית.
- המבחן שיש לבצע כאן הוא :
- מבחן Z למדגם יחיד.
  - מבחן T למדגם יחיד.
  - מבחן T למדגמים בלתי תלויים.
  - מבחן T למדגמים מזווגים.

4. כדי לבדוק את הטענה שגברים רווקים שוקלים פחות מגברים נשואים לקח חוקר מדגם מקרי של 4 גברים ומדד את משקלם לפני נישואיהם ולאחר נישואיהם. הנה התוצאות:

68	82	93	69	לפני הנישואין-X
71	84	88	80	לאחר הנישואין-Y

מהן ההשערות הנבדקות? (ההפרש חושב X-Y)

א.  $H_1 : \mu_d < 0, H_0 : \mu_d = 0$

ב.  $H_1 : \mu_x - \mu_y < 0, H_0 : \mu_x - \mu_y = 0$

ג.  $H_1 : \mu_x - \mu_y < 0, H_0 : \mu_x - \mu_y = 0$

ד.  $H_1 : \mu_d > 0, H_0 : \mu_d = 0$

5. 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.

6. ידוע כי ילד בגיל שנתיים ישן בממוצע 9 שעות בלילה. במדגם של 20 תינוקות בני שנתיים המתגוררים בצפון נמצא, כי ממוצע שעות השינה בלילה הינו 10 עם סטיית תקן של 1.1 במדגם של 10 תינוקות בדרום נמצא, כי ממוצע שעות השינה בלילה הינו 7.9 עם סטיית תקן של 1.1. על מנת להשוות בין ממוצע שעות השינה של ילדים מהצפון לבין זה של כלל הילדים יש לערוך \_\_\_\_\_, ועל מנת להשוות בין ממוצע שעות השינה של ילדים מהדרום לזה של ילדים המתגוררים בצפון יש לערוך \_\_\_\_\_. יש להניח שההנחות הדרושות מתקיימות.

א. מבחן Z למדגם יחיד; מבחן T למדגם יחיד.

ב. מבחן T למדגם יחיד; מבחן T למדגמים תלויים.

ג. מבחן T למדגם יחיד; מבחן T למדגמים בלתי תלויים.

ד. מבחן T למדגמים בלתי תלויים; מבחן T לממוצע יחיד.

7. מובהקות התוצאה (P-value) היא גם:

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. כדי לבדוק את הטענה שגברים רווקים שוקלים פחות מגברים נשואים לקח חוקר מדגם מקרי של 4 גברים ומדד את משקלם לפני נישואיהם ולאחר נישואיהם. הנה התוצאות:

68	82	93	69	לפני הנישואין
71	84	88	80	לאחר הנישואין

באיזה התפלגות משתמשים לבדיקת ההשערות, ובכמה דרגות חופש:

א. ההתפלגות  $Z$  ללא דרגות חופש.

ב. ההתפלגות  $T$  ו-3 דרגות חופש.

ג. ההתפלגות  $T$  ו-6 דרגות חופש.

ד. ההתפלגות  $\chi^2$  ו-3 דרגות חופש.

9. שני סטטיסטיקאים בודקים השערות ברמת מובהקות  $\alpha = 0.05$ , על סמך אותו מדגם.

סטטיסטיקאי א' בודק את ההשערה:  $H_0 : \mu = 20$  כנגד האלטרנטיבה

$H_1 : \mu \neq 20$  ומחליט לא לדחות את השערת האפס.

סטטיסטיקאי ב' בודק את ההשערה  $H_0 : \mu \leq 20$  כנגד האלטרנטיבה  $H_1 : \mu > 20$

מה יחליט סטטיסטיקאי ב'?

א. לדחות את השערת האפס.

ב. לא לדחות את השערת האפס.

ג. ללא נתונים נוספים אי אפשר לדעת מה יחליט.

10. חוקר בדק השערה מסוימת והחליט לדחות את השערת האפס ברמת מובהקות 5%. מה נכון לומר?  
 א. הוא בוודאות ידחה את השערת האפס ברמת מובהקות 9% ואילו ברמת מובהקות 2% יש לבדוק מחדש.  
 ב. הוא בוודאות לא ידחה את השערת האפס ברמת מובהקות 9% ואילו ברמת מובהקות 2% יש לבדוק מחדש.  
 ג. הוא בוודאות ידחה את השערת האפס ברמת מובהקות 9% וברמת מובהקות 2%.  
 ד. הוא בוודאות לא ידחה את השערת האפס ברמת מובהקות 9% ואילו ברמת מובהקות 2% יש לבדוק מחדש.

11. רמת הכולסטרול בדמם של אנשים מתפלג נורמאלית עם תוחלת של 180 מ"ג (ל 100 סמ"ק דם). וסטיית

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

כתב ופתר - ברק קנדל ©

תקן של 10 מ"ג. מעוניינים לבדוק את הטענה שצמחונים הם בעלי רמת כולסטרול נמוכה יותר. נניח שסטיית התקן אצל צמחונים זהה לסטיית התקן של כלל האנשים. במדגם של 20 צמחונים התקבל ממוצע רמת כולסטרול 174.5 מ"ג.

אם הוחלט לקבל את הטענה שצמחונים הם בעלי רמת כולסטרול נמוכה יותר איזה סוג טעות אפשרית במסקנה?  
 א. טעות מסוג ראשון.  
 ב. טעות מסוג שני.  
 ג. טעות מסוג שלישי.  
 ד. לא ניתן לדעת כיוון שאנו לא יודעים מה התוחלת האמיתית אצל הצמחוניים.

12. בסקר שנערך התקבל ש 60% מתוך 220 נשאלים מבקרים אצל השיננית לפחות פעם אחת בשנה. עבור אילו רמות מובהקות ניתן יהיה לקבוע שרוב האוכלוסייה מבקרת אצל השיננית לפחות פעם בשנה?  
 א. רמת מובהקות הגדולה מ-5%.  
 ב. רמת מובהקות הקטנה מ-5%.  
 ג. רמת מובהקות הגדלה מ-0.0015.  
 ד. רמת מובהקות הקטנה מ-0.0015.

13. שני חוקרים העוסקים בתחום מחקרי משותף החליטו להסתמך על נתונים של מדגם שפורסם על ידי הלשכה המרכזית לסטטיסטיקה. חוקר א' ניסח השערה דו צדדית ואילו חוקר ב' ניסח השערה חד צדדית. מסקנתו של איזה מבין המשפטים הבאים הוא הנכון בנוגע למסקנות החוקרים?  
 א. אם חוקר א' ידחה את השערת האפס לא ניתן לדעת מה יחליט חוקר ב' באותה רמת מובהקות.  
 ב. אם חוקר א' יקבל את השערת האפס גם חוקר ב' יקבל את השערת האפס באותה רמת מובהקות.  
 ג. אם חוקר ב' ידחה את השערת האפס גם חוקר א' ידחה את השערת האפס באותה רמת מובהקות.  
 ד. אם חוקר א' ידחה את השערת האפס גם חוקר ב' ידחה את השערת האפס בתנאי שרמת המובהקות כפולה בגודלה.

14. ידוע מנתוני העבר כי תוחלת הציונים בבחינה בפסיכולוגיה היא 79. הועלתה השערה כי תוחלת הציונים

בקרב העולים החדשים נמוכה יותר. לצורך בדיקת הטענה נלקח מדגם מקרי של 47 סטודנטים עולים ונמצא ממוצע של 75. מה משמעות הפרמטר בניסוח ההשערות?

- א. תוחלת ציוני העולים באוכלוסייה.
- ב. ממוצע ציוני העולים במדגם.
- ג. תוחלת ציוני האוכלוסייה מנתוני העבר.
- ד. ממוצע ציוני שאר האוכלוסייה במדגם.

15. חוקר ביצע מחקר וידוע כי עשה טעות מסוג 1. מה מהבאים נכון?

- א. החוקר דחה את השערת  $H_0$  כאשר היא הייתה נכונה.
- ב. החוקר דחה את השערת  $H_1$  כאשר היא הייתה נכונה.
- ג. החוקר לא דחה את השערת  $H_0$  כאשר היא הייתה לא נכונה.
- ד. המדגם של החוקר שייך בפועל להתפלגות הדגימה של  $H_1$ .

16. חוקר ביקש לבחון האם תאומים זהים אשר הופרדו בילדותם שונים מתאומים זהים אשר גדלו יחדיו מבחינת מידת הפער בין התאומים בלחץ הדם. הוא דגם 20 זוגות תאומים מכל אוכלוסייה ומדד את הפרש בין לחץ הדם בכל זוג תאומים. מהו המבחן הסטטיסטי המתאים?

- א. מבחן  $t$  למדגמים בלתי תלויים עם 38 דרגות חופש.
- ב. מבחן  $t$  למדגמים מזווגים, עם 39 דרגות חופש.
- ג. מבחן  $t$  למדגמים בלתי תלויים עם 39 דרגות חופש.
- ד. מבחן  $t$  למדגמים מזווגים עם 38 דרגות חופש.

17. בינואר השנה פורסם שהשכר הממוצע במשק הוא 8,900 ₪. במדגם שנעשה בחודש יוני על 60 עובדים נרשם עבור כל עובד במדגם האם השכר שלו נמוך או לא נמוך מהשכר הממוצע שפורסם בחודש ינואר. מהו המבחן המתאים כדי לבדוק שרוב העובדים בחודש יוני קיבלו שכר הנמוך מהשכר הממוצע שפורסם בחודש ינואר?

- א. מבחן  $Z$  על פרופורציה.
- ב. מבחן  $t$  על תוחלת אחת.
- ג. מבחן  $t$  על שתי תוחלות במדגמים בלתי תלויים.
- ד. מבחן  $t$  על שתי תוחלות במדגמים תלויים.

18. שלושה חוקרים רצו לבדוק את השפעתו של שידור פרסומות נגד תאונות דרכים על מהירות הנהיגה של

- נהגים בישראל (השוונות של מהירות הנהיגה בישראל אינה ידועה).  
 עידו השווה את מהירות הנהיגה של קבוצת נהגים אחת, חודש לפני שידור הפרסומות וחודש לאחר שידור הפרסומות.  
 רון השווה את מהירות הנהיגה של קבוצת נהגים, שראו את הפרסומות, למהירות הנהיגה של קבוצת נהגים, שלא ראו את הפרסומות.  
 יואב השווה את מהירות הנהיגה של קבוצת נהגים בחודש בו שודרו הפרסומות, למהירות הנהיגה הממוצעת בישראל על פי נתוני משרד התחבורה.  
 המבחנים בהם צריכים החוקרים להשתמש הם:  
 א. שלושתם במבחן  $t$  למדגמים בלתי תלויים.  
 ב. עידו במבחן  $t$  למדגמים מזווגים, ורון ויואב במבחן  $t$  למדגמים בלתי תלויים.  
 ג. עידו במבחן  $t$  למדגמים מזווגים, רון במבחן  $t$  למדגמים בלתי תלויים ויואב במבחן  $t$  למדגם יחיד.  
 ד. עידו במבחן  $t$  למדגמים מזווגים, רון ויואב במבחן  $t$  למדגם יחיד.

19. במחקר נמצא שתוצאה היא מובהקת ברמת מובהקות של 5%. מה תמיד נכון?

- א. הגדלת רמת המובהקות לא תשתנה את מסקנת המחקר.  
 ב. הגדלת רמת המובהקות תשנה את מסקנת המחקר.  
 ג. הקטנת רמת המובהקות לא תשנה את מסקנת המחקר.  
 ד. הקטנת רמת המובהקות תשנה את מסקנת המחקר.

20. חוקר ערך מבחן דו צדדי ברמת מובהקות של  $\alpha$  והחליט לדחות את השערת האפס.

אם החוקר היה עורך מבחן חד צדדי ברמת מובהקות של  $\frac{\alpha}{2}$  אזי בהכרח:

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

21. 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 השתמש בטבלה של התפלגות  $t$ .

מה נוכל לומר בנוגע להחלטת המחקר שלהם?

- א. אם ליאור ידחה את השערת האפס אז גם בהכרח רוני.  
 ב. אם רוני ידחה 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.

22. נתון ש  $X \sim N(\mu, \sigma^2)$  כמו כן נתונות ההשערות הבאות:

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

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu < \mu_0$$

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.

א. כעת בברור הוא ידחה את השערת האפס.

ב. כעת הוא דווקא יקבל את השערת האפס.

ג. כעת לא ניתן לדעת מה תהיה מסקנתו.

23. 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.

24. 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.

25. מה המצב הרצוי לחוקר המבצע בדיקת השערה :

$1 - \beta$	$\alpha$
גדולה	ה. גדולה
קטנה	ו. גדולה
גדולה	ז. קטנה
קטנה	ח. קטנה

26. 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.

27. 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.

28. בבדיקת השערות התקבל שה-  $p\text{-value}=0.02$ .

מה תהיה מסקנת חוקר המשתמש ברמת מובהקות 1%? בחר בתשובה הנכונה.

ה. יקבל את השערת האפס בכל מקרה.

ו. ידחה את השערת האפס מקרה.

ז. ידחה את השערת האפס רק אם המבחן הנו דו צדדי.

ח. לא ניתן לדעת כי אין מספיק נתונים.

29. מובהקות התוצאה ( $PV$ ) היא גם :

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.

30. בבדיקת השערות מסוימת התקבל  $p\text{ value}=0.0254$  לכן :

א. ברמת מובהקות של 0.01 אך לא של 0.05 נדחה את  $H_0$ .

ב. ברמת מובהקות של 0.01 ושל 0.05 לא נדחה את  $H_0$ .

ג. ברמת מובהקות של 0.05 אך לא של 0.01 נדחה את  $H_0$ .

ד. ברמת מובהקות של 0.01 ושל 0.05 נדחה את  $H_0$ .

31. רמת המובהקות במחקר הייתה 2% לכן.

א. בסיכוי של 2% נדחה את השערת האפס.

ב. בסיכוי של 2% לא נדחה את השערת האפס.

- ג. בסיכוי של 2% השערת האפס לא נכונה.  
ד. אף תשובה לא נכונה.

32. 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%. אם הוא היה מגדיל את רמת המובהקות ל-10% אזי:
- א. כעת בברור הוא ידחה את השערת האפס.  
ב. כעת הוא דווקא יקבל את השערת האפס.  
ג. כעת לא ניתן לדעת מה תהיה מסקנתו.

33. 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. מספר דרגות החופש במבחן הוא :

א. 9

ב. 19

ג. 18

ד. 8

34. 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
---------------	-----	-----	-----	-----

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

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.

35. 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.

36. סטטיסטיקאי ערך מבחן סטטיסטי. הוא חישב את עוצמת המבחן וקיבל 0. המשמעות של

תוצאה זו היא:

א. לעולם לא לדחות את השערת האפס כאשר היא לא נכונה.

ב. תמיד לדחות את השערת האפס כאשר היא נכונה.

ג. לעולם לא לדחות את השערת האפס כאשר היא נכונה.

ד. תמיד לדחות את השערת האפס כאשר היא לא נכונה.

37. 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 :

- ה. לדחות את השערת האפס.
- ו. לא לדחות את השערת האפס.
- ז. שלא ניתן לדעת את המסקנה עבור רמת מובהקות 0.05 .
- ח. שלא נתונות בשאלה סטיות התקן של האוכלוסיות, ולכן לא ניתן להסיק דבר.

38. 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 רווח סמך לממוצע ההפרשים וקיבלו :

$$-0.0293 < \mu_D < 0.2145$$

רווח הסמך הוא ברמת סמך של 95%.

- לכן מסקנת המחקר היא :
- ד. ברמת מובהקות של 5% לא נוכל לקבוע שקיים הבדל בין החברות.
- ה. ברמת מובהקות של 5% נקבע שקיים הבדל מובהק בין החברות.
- ו. לא ניתן לדעת מה המסקנה ברמת מובהקות של 5% כיוון שלא נאמר מה ההגדרה של D.

39. אם רמת מובהקות של מבחן סטטיסטי הינה 0, הכוונה היא :

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

40. חוקר ביצע ניסוי. הוא ניסח את ההשערות הבאות:  $H_0: \mu = 10$  לצורך בדיקה הוא לקח  $H_1: \mu \neq 10$

- מדגם מקרי בגודל 5 מתוך אוכלוסייה המתפלגת נורמאלית עם שונות לא ידועה. על סמך תוצאות המדגם הוא חישב וקיבל:  $t_{\bar{x}} = -2.63$ . לכן המסקנה היא:
- הוא ידחה  $H_0$  ברמת מובהקות 0.1 אך לא כן ברמת מובהקות 0.05.
  - הוא ידחה  $H_0$  ברמת מובהקות 0.05 אך לא כן ברמת מובהקות 0.025.
  - הוא ידחה  $H_0$  ברמת מובהקות 0.025 אך לא כן ברמת מובהקות 0.01.
  - הוא לא ידחה  $H_0$  ברמת מובהקות 0.1.

41. האיגוד האמריקני לרפואת ילדים מפרסם הנחיות חדשות הקובעות כי יש ליטול תוספת יוד במהלך תקופת ההיריון וההנקה. מחסור במינרל זה עלול לגרום לפגיעה מוחית אצל העובר והתינוק. החלטה זו נקבעה על סמך מחקר בו השתתפו 1050 נשים שנטלו יוד במהלך תקופת ההיריון וההנקה. מתוך הנשים שהשתתפו במחקר רק ל-21 נמצאו ילדים בעלי פגיעה מוחית לעומת 3% באוכלוסייה הכללית. בנוסף פרסם שהאיגוד האמריקאי מגיע למסקנותיו על סמך רמת מובהקות של 0.5%.

מה הסיכוי לבצע טעות מסוג ראשון במחקר?

- 0.005
- 0.03
- 0.0287
- 0.05

42. חוקרת שיערה, כי משקלן של נשים כשנה לאחר החתונה גבוה ממשקלן בעת החתונה. החוקרת דגמה 15 נשים, ובדקה את משקלן בשתי נקודות הזמן (בעת החתונה, ושנה לאחריה), אך לא מצאה הבדל מובהק ברמת מובהקות 0.01 בהנחה, כי **במציאות** השערתה של החוקרת נכונה, **סביר** כי אם היא תגדיל את גודל המדגם, אזי:

- יקטן הסיכוי לטעות מסוג שני ( $\beta$ ).
- תגדל רמת הביטחון ( $1-\alpha$ ).
- אף תשובה לא נכונה.
- כל התשובות נכונות.

43. איזה מהמשפטים הבאים נכון תמיד?

א.  $POWER + \beta + \alpha = 1$

ב.  $\beta - POWER = 0.05$

ג.  $POWER + \alpha = 1$

ד.  $\beta + \alpha = 1$

ה. הכול לא נכון

44. מה נכון לומר לגבי הנחת שיוויון השונויות במבחן T למדגמים בלתי תלויים?

א. היא אומרת שהשונויות המדגמיות שוות.

ב. בלעדיה אין שום דרך לבדוק השערה על הפרש בין תוחלות.

ג. היא חשובה הן עבור מדגמים מזווגים והן עבור מדגמים בלתי תלויים.

ד. אף תשובה אינה נכונה.

45. חוקר החליט לא לדחות השערה ברמת מובהקות של  $\alpha$ . במידה וחוקר זה היה בודק השערה זו

ברמת מובהקות של  $2\alpha$  על סמך אותם נתונים האם ההשערה תדחה?

א. ההשערה תדחה.

ב. ההשערה לא תדחה.

ג. התשובה תלויה בעוצמת המבחן.

ד. לא ניתן לדעת בוודאות אם ההשערה תדחה או לא.

46. חוקרת שיערה, כי בגילאי הגן בנות יותר תקשורתיות מבנים. אם החוקרת תדגום אקראית 30 בנים ו-

30 בנות, ובמדגם יתקבל אותו ממוצע של ציון תקשורת. סטטיסטי המבחן יהיה:

א. אפס

ב. חיובי

ג. שלילי

ד. לא ניתן לדעת

47. עוצמה שווה ל-1 פרושה:

א. לעולם לא לדחות את השערת האפס כאשר היא נכונה.

ב. תמיד לדחות את השערת האפס כאשר היא נכונה.

ג. לעולם לא לדחות את השערת האפס כאשר היא לא נכונה.

48. מה מהבאים נכון לגבי מבחן t מדגמים מזווגים?

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

49. לבדיקת ההשערה החד צדדית על התוחלת של התפלגות נורמלית נלקח מדגם

$$H_0 : \mu \geq 10$$

$$H_1 : \mu < 10$$

והתקבלה רמת מובהקות מינימאלית לדחיית השערת האפס 0.058. לו רצינו לבדוק את ההשערה

הדו צדדית  $H_0 : \mu = 10$ , אז על סמך תוצאת אותו המדגם ברמת מובהקות 0.05:

$$H_1 : \mu \neq 10$$

- א. ניתן להכריע בין ההשערות רק אם שונות האוכלוסייה נתונה.
- ב. מקבלים את השערת האפס.
- ג. דוחים את השערת האפס.
- ד. לא ניתן להכריע בין ההשערות שכן חסרים נתונים.

50. לבדיקת ההשערה החד צדדית ימנית נלקח מדגם מקרי בגודל  $n$   $H_0 : \mu = 55$   
 $H_1 : \mu = 65$

מאוכלוסייה בעלת התפלגות נורמלית ושונות  $\sigma^2$ . רמת המובהקות היא 5%. נמצא שהעוצמה היא 0.9. להלן 3 טענות:

(1) עבור מדגם בגודל  $n$  וברמת מובהקות 5% לבדיקת ההשערות:  $H_0 : \mu = 55$   
 $H_1 : \mu = 60$  תהיה גדולה מ- 0.9.

(2) עבור מדגם בגודל  $2n$  ורמת מובהקות 5% לבדיקת ההשערות:  $H_0 : \mu = 55$   
 $H_1 : \mu = 65$  תהיה גדולה מ- 0.9.

(3) עבור מדגם בגודל  $n$  ורמת מובהקות 10% לבדיקת ההשערות:  $H_0 : \mu = 55$   
 $H_1 : \mu = 65$  תהיה קטנה מ- 0.9.

א. שלושת הטענות אינן נכונות.

ב. טענות 2 ו-3 אינן נכונות וטענה 1 נכונה.

ג. טענות 1 ו-2 נכונות וטענה 3 אינה נכונה.

ד. טענות 1 ו-3 אינן נכונות וטענה 2 נכונה.



**פתרונות :**

א	2	א	1
ב	2	ב	2
א	2	ב	3
א	2	א	4
ג	3	ג	5
ב	3	ג	6
א	3	א	7
א	3	ב	8
ב	3	ג	9
ג	3	א	1
א	3	א	1
ג	3	ג	1
א	3	א	1
ג	3	א	1
א	4	א	1
א	4	א	1
א	4	א	1
ה	4	ג	1
ב	4	א	1
ב	4	ג	2
א	4	ב	2
ב	4	ג	2
ג	4	ב	2
ב	4	ג	2
ב	5	ג	2

## פרק 18 - ניתוח שונות חד כיוונית

### רקע תיאורטי:

ניתוח שונות (חד כיוונית) הוא מבחן להשוואת תוחלות ( $\mu_1, \dots, \mu_k$ ) של k אוכלוסיות שונות.

ולכן בניתוח שונות השערות המחקר הן:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k \quad (\text{התוחלות של כל האוכלוסיות שוות})$$

$$H_1 : \quad \text{אחרת} \quad (\text{לפחות שתיים מהתוחלות שונות})$$

ההנחות 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.

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

$$SS_B = \sum_{i=1}^k n_i [\bar{X}_i - \bar{X}]^2 \quad \text{סכום ריבועים בין הקבוצות}$$

$$SS_W = \sum_{i=1}^k [n_i - 1] \cdot \hat{S}_i^2 \quad \text{סכום ריבועים בתוך הקבוצות}$$

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

$$SST = SSB + SSW$$

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

- א. מהו המבחן הסטטיסטי המתאים? רשמו את ההשערות ואת ההנחות של המבחן.
- ב. מהו גודל המדגם? מהו המשתנה הבלתי תלוי (FACTOR) כמה רמות יש לו?
- ג. חשבו את הממוצע ואת סטיית התקן של הציונים בכל אחד מהמדגמים.
- ד. מלאו את טבלת ANOVA.
- ה. רשמו את כלל ההכרעה למבחן שהוצע בסעיף א ברמת מובהקות של 5%.
- ו. האם קיים הבדל בין בתי הספר בעיר מבחינת רמת הצלחת התלמידים במקצוע המתמטיקה? ענה על סמך הסעיפים הקודמים.

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

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

- א. רשמו את השערות המחקר וההנחות הדרושות כדי לבצע את המבחן המתאים.  
 ב. מה מסקנת המחקר ברמת מובהקות של 5%?  
 ג. האם יש צורך לבצע השוואות מרובות?  
 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			

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

### ANOVA

pulse

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

### Post Hoc Tests

#### Multiple Comparisons

pulse  
Tukey HSD

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

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



## Homogeneous Subsets

## pulse

Tukey HSD<sup>a</sup>

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

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 5.000.

- א. בדוק ברמת מובהקות של 5% האם קיים הבדל בין המינונים השונים מבחינת תוחלת הדופק של האנשים? רשמו את ההשערות וההנחות הדרושות לצורך פתרון.
- ב. הסבירו ללא חישוב כיצד הייתה משתנה התשובה לסעיף הקודם אם הינו מעלים את הדופק של כל התצפיות במחקר ב- 2.
- ג. האם יש צורך במחקר בהשוואת מרובות. נמק!
- ד. לטבלת ה ANOVA צורפו טבלאות של השוואות מרובות בשיטה הנקראת "טוקיי". ברמת בטחון של 95% מה הם הממצאים לפי שיטה זו?

8. בעיר מסוימת רצו לבדוק האם קיים הבדל ברמה של התלמידים בין בתי הספר השונים בעיר. ביצעו מדגם מכל בית ספר ונתנו מבחן זהה לכל הנדגמים. לאחר מכן ריכזו את הנתונים בתוכנה סטטיסטית והפעילו ניתוח שונות. מצורפים הפלטים שהתקבלו.  
ענו על הסעיפים הבאים:
- כמה בתי ספר יש בעיר?
  - כמה תלמידים השתתפו בסך הכול במחקר?
  - האם קיים הבדל בין בתי הספר בעיר מבחינה רמת הציונים? בדקו ברמת מובהקות של 1%
  - בביטחון של 95% אילו בתי ספר שונים זה מזה ברמת התלמידים? נמקו והסבירו.

### Oneway

#### ANOVA

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

### Post Hoc Tests

## Multiple Comparisons

grade

Scheffe

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

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

## Homogeneous Subsets

grade

Scheffe<sup>a</sup>

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

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 5.000.

**פתרונות סופיים חלקיים - ניתוח שונות חד כיוונית**

2. אם חישוב 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

7. א. נדחה את השערת האפס. ב. לא משתנה. ג. כן

8. א. 5 ב. 25 ג. כן

## פרק 19 - מבחני חי בריבוע

### מבחן טיב התאמה

#### רקע:

מבחן זה הוא מבחן הבא לבדוק האם אוכלוסייה מסוימת מתפלגת לפי התפלגות נתונה. המשתנה הנחקר מחולק למספר קטגוריות ויש לבדוק האם תוצאות המדגם תואמות להתפלגות הנתונה.

#### מבנה המבחן:

#### השערות:

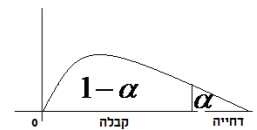
$H_0$ : המשתנה מתפלג לפי התפלגות מסוימת

$H_1$ : אחרת

#### כלל הכרעה:

הערך הקריטי נקבע על סמך התפלגות חי בריבוע. התפלגות זו היא אסימטרית חיובית ותלויה בדרגות החופש.

כאשר  $K$  - מספר הקטגוריות,  $d.f = K - 1$ .



הערך הקריטי הוא:  $\chi^2_{1-\alpha, K-1}$ , כלומר האחוזון ה- $1 - \alpha$  בהתפלגות חי בריבוע שדרגות החופש הן  $K - 1$ .

אם  $\chi^2 > \chi^2_{1-\alpha, K-1}$ , דוחים את השערת האפס.

#### סטטיסטי המבחן:

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

$O_i$  - השכיחות שנצפתה במדגם בקטגוריה  $i$ .

$p_i$  - הסתברות לקטגוריה  $i$  לפי השערת האפס.

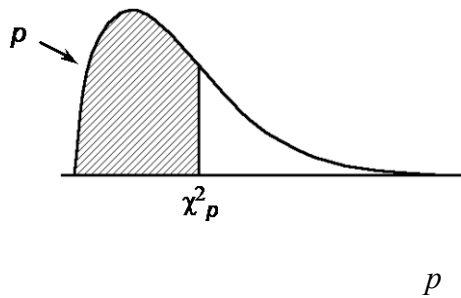
$E_i = np_i$  שכיחות צפויה במדגם לקטגוריה  $i$  בהנחת השערת האפס.

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

קטגוריות סמוכות עד שהתנאי יתקיים.

**זוגמה:** (פתרון בהקלטה)

במדינה מסוימת שלוש מפלגות. בפרלמנט הנוכחי התפלגות מספר המושבים היא 30% למפלגה A, 60% למפלגה B ו-10% למפלגה C. לקראת הבחירות המתוכננות בשבוע הבא נעשה סקר שכלל 300 אזרחים. בסקר התקבל ש-40% יצביעו למפלגה A, 50% למפלגה B ו-10% למפלגה C. האם תוצאות הסקר תואמות להתפלגות המושבים בפרלמנט הנוכחי? בדקו ברמת מובהקות של 5%.

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

df	.005	.01	.025	.05	.10	.25	.50	.75	.90	.95	.975	.99	.995
1	0.0 <sup>4</sup> 393	0.0 <sup>3</sup> 157	0.0 <sup>3</sup> 982	0.0 <sup>2</sup> 393	0.0158	0.102	0.455	1.32	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.103	0.211	0.575	1.39	2.77	4.61	5.99	7.38	9.21	10.6
3	0.0717	0.115	0.216	0.352	0.584	1.21	2.37	4.11	6.25	7.81	9.35	11.3	12.8
4	0.207	0.297	0.484	0.711	1.06	1.92	3.36	5.39	7.78	9.49	11.1	13.3	14.9
5	0.412	0.554	0.831	1.15	1.61	2.67	4.35	6.63	9.24	11.1	12.8	15.1	16.7
6	0.676	0.872	1.24	1.64	2.20	3.45	5.35	7.84	10.6	12.6	14.4	16.8	18.5
7	0.989	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.0	14.1	16.0	18.5	20.3
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.2	13.4	15.5	17.5	20.1	22.0
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.4	14.7	16.9	19.0	21.7	23.6
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.5	16.0	18.3	20.5	23.2	25.2
11	2.60	3.05	3.82	4.57	5.58	7.58	10.3	13.7	17.3	19.7	21.9	24.7	26.8
12	3.07	3.57	4.40	5.23	6.30	8.44	11.3	14.8	18.5	21.0	23.3	26.2	28.3
13	3.57	4.11	5.01	5.89	7.04	9.30	12.3	16.0	19.8	22.4	24.7	27.7	29.8
14	4.07	4.66	5.63	6.57	7.79	10.2	13.3	17.1	21.1	23.7	26.1	29.1	31.3
15	4.60	5.23	6.26	7.26	8.55	11.0	14.3	18.2	22.3	25.0	27.5	30.6	32.8
16	5.14	5.81	6.91	7.96	9.31	11.9	15.3	19.4	23.5	26.3	28.8	32.0	34.3
17	5.70	6.41	7.56	8.67	10.1	12.8	16.3	20.5	24.8	27.6	30.2	33.4	35.7
18	6.26	7.01	8.23	9.39	10.9	13.7	17.3	21.6	26.0	28.9	31.5	34.8	37.2
19	6.84	7.63	8.91	10.1	11.7	14.6	18.3	22.7	27.2	30.1	32.9	36.2	38.6
20	7.43	8.26	9.59	10.9	12.4	15.5	19.3	23.8	28.4	31.4	34.2	37.6	40.0
21	8.03	8.90	10.3	11.6	13.2	16.3	20.3	24.9	29.6	32.7	35.5	38.9	41.4
22	8.64	9.54	11.0	12.3	14.0	17.2	21.3	26.0	30.8	33.9	36.8	40.3	42.8
23	9.26	10.2	11.7	13.1	14.8	18.1	22.3	27.1	32.0	35.2	38.1	41.6	44.2
24	9.89	10.9	12.4	13.8	15.7	19.0	23.3	28.2	33.2	36.4	39.4	43.0	45.6
25	10.5	11.5	13.1	14.6	16.5	19.9	24.3	29.3	34.4	37.7	40.6	44.3	46.9
26	11.2	12.2	13.8	15.4	17.3	20.8	25.3	30.4	35.6	38.9	41.9	45.6	48.3
27	11.8	12.9	14.6	16.2	18.1	21.7	26.3	31.5	36.7	40.1	43.2	47.0	49.6
28	12.5	13.6	15.3	16.9	18.9	22.7	27.3	32.6	37.9	41.3	44.5	48.3	51.0
29	13.1	14.3	16.0	17.7	19.8	23.6	28.3	33.7	39.1	42.6	45.7	49.6	52.3
30	13.8	15.0	16.8	18.5	20.6	24.5	29.3	34.8	40.3	43.8	47.0	50.9	53.7



### תרגילים:

1. במטרה לבדוק האם קובייה הוגנת, מטילים אותה 120 פעמים. התקבל 17 פעמים 1, 23 פעמים 2, 20 פעמים 3, 25 פעמים 4, 18 פעמים 5 ו- 17 פעמים 6. מה מסקנתכם ברמת מובהקות של 5%?
2. מפעל מייצר סוכריות בצבעים כחול, אדום, ירוק וכתום. מעוניינים לבדוק שפרופורציית הסוכריות הכחולות גדולה פי 2 מכל צבע אחר. לצורך כך נדגמו באקראי 200 סוכריות והתקבל: 70 כחולות, 50 אדומות, 40 ירוקות והיתר כתומות. מה מסקנתכם ברמת מובהקות של 5%?
3. משרד החינוך טוען שבקרב השכירים במשק היחס בין השכירים בעלי השכלה נמוכה, תיכונית ואקדמאית הוא 1:2:1 בהתאמה. במדגם של 200 שכירים התקבלו 56 אנשים בעלי השכלה נמוכה, 105 בעלי השכלה תיכונית והיתר בעלי השכלה גבוהה.
  - א. על סמך תוצאות המדגם האם התפלגות ההשכלה היא כמו שמשדרד החינוך מפרסם? בדוק ברמת מובהקות של 5%.
  - ב. בנו רווח סמך ברמת סמך של 95% לפרופורציית השכירים במשק בעלי השכלה אקדמאית.
4. 200 איש נתבקשו לבחור ספרה באקראי והנה התוצאות שהתקבלו :
 

18 איש בחרו בספרה 0
24 איש בחרו בספרה 1
17 איש בחרו בספרה 2
19 איש בחרו בספרה 3
20 איש בחרו בספרה 4
18 איש בחרו בספרה 5
22 איש בחרו בספרה 6
היתר בחרו בספרות 7-9

  - א. על סמך התוצאות הללו האם בחירת הספרות אקראית? רמת מובהקות של 5%.
  - ב. תנו הערכה למובהקות התוצאה.
  - ג. אם נגדיל את גודל המדגם פי 2 ונשמור על אותם יחסים של כמות האנשים במדגם שבחרו בספרות, כיצד הדבר ישפיע על ערכו של הסטטיסטי  $\chi^2$ ? מה תהיה המסקנה במקרה זה?

5. מעוניינים לבדוק האם קובייה היא הוגנת. הטילו את הקובייה פעמיים והתבוננו בסכום הוצאות. חזרו על התהליך 72 פעמים. להלן התוצאות שהתקבלו במדגם:

מספר הטלות	סכום התוצאות
20	2-5
17	6-8
20	9-10
15	11-12

מה המסקנה ברמת מובהקות של 5%.

6. בפנס יש 4 סוללות. בבדיקה שנערכה ב-400 פנסים נמצאו סוללות פגומות לפי השכיחויות הבאות:

מספר הסוללות הפגומות	0	1	2	3 ומעלה
שכיחות	276	104	12	8

מעוניינים לבדוק על סמך תוצאות מדגם אלה האם הסיכוי לסוללה פגומה הוא 20%. מהי רמת המובהקות המינימלית עבורה נכריע שהסיכוי לסוללה פגומה אינו 20%?

7. מטילים מטבע עד שלאשונה מתקבל "ראש". חוזרים על התהליך 120 פעמים. נסמן ב- $X$  את מספר ההטלות עד קבלת הראש. להלן התוצאות שהתקבלו:

$x$	1	2	3	4	5	6
מספר החזרות על התהליך	54	20	16	22	6	2

א. בהנחה והמטבע הוגן, מהי ההתפלגות של  $X$ ?

ב. בודק האם המטבע הוגן, על סמך תוצאות המדגם ברמת מובהקות של 5%.

$$H_0 : X \sim N(40, 2^2)$$

$$H_1 : \text{else}$$

תוצאות המדגם הן :

מעל 44	40-44	36-40	מתחת 36	X
2A	45A	50A	3A	מספר הדגימות

מהו ערכו המקסימלי של A עבורו נקבל את  $H_0$  ברמת מובהקות של 5% :

פתרונות:שאלה 1 :לא נדחה  $H_0$ שאלה 2 :לא נדחה  $H_0$ שאלה 3 :א. לא נדחה  $H_0$  ב.  $(0.14, 0.25)$ שאלה 4 :א. לא נדחה  $H_0$  ב. בין 0.95 ל- 0.975 ג. יגדל פי 2. מסקנה לא תשתנה.שאלה 5 :

נכריע שהקובייה אינה הוגנת.

שאלה 6 :

0.005

שאלה 7 :א.  $X \sim G(0.5)$  ב. נסיק שהמטבע לא הוגן.שאלה 8 :

14

### הקשר בין מבחן טיב התאמה לבדיקת השערות על הפרופורציה

#### רקע:

אם אנו רוצים לבצע מבחן טיב התאמה על משתנה שיש לו שתי קטגוריות בלבד (משתנה דיכוטומי), הדבר זהה לתהליך של בדיקת השערות דו צדדית על פרופורציה בודדת.

#### דוגמה: (פתרון בהקלטה)

הטילו מטבע 80 פעמים וקיבלו 48 פעמים את התוצאה "ראש". בדקו האם המטבע הוא הוגן ברמת מובהקות של 5%.

א. באמצעות מבחן טיב התאמה.

ב. באמצעות מבחן  $Z$  לפרופורציה בודדת.

**תרגילים:**

1. בסקר שנעשה על 320 נשאלים, 43.75% טענו שהחיה המועדפת עליהם היא כלב. עד היום היה נהוג לחשוב ש40% מהאנשים מעדיפים כלבים.

בדקו ברמת מובהקות של 5% האם הסקר יישנה את הסברה שהייתה נהוגה עד היום לגבי העדפת כלב.

- א. באמצעות מבחן טיב התאמה.  
ב. באמצעות מבחן על פרופורציה.

2. לסוכנות מכוניות שלושה סניפים ברחבי הארץ. המכוניות נמכרות בסניפים השונים. מתוך 100 מכוניות נמצא ש- 65 נמכרו בסניף תל-אביב, 23 בסניף ירושלים והיתר בסניף חיפה.

א. בדקו ברמת מובהקות של 5% האם שיעור המכוניות שנמכרות בסניף ת"א גדול פי 2 מכל סניף אחר.

ב. בדקו באמצעות מבחן טיב התאמה האם 60% מהמכוניות נהוגות להימכר בסניף תל אביב. האם יש דרך אחרת לבדוק את ההשערה?

3. בתחרות ריצה בית ספרית שלושה מסלולי ריצה. ב-50 תחרויות בדקו באיזה מסלול היה הניצחון. התוצאות שהתקבלו מסוכמות בטבלה הבאה:

המסלול	1	2	3
מספר הניצחונות	20	15	15

א. בדקו ברמת מובהקות של 5% האם יש מסלול מועדף לניצחון.

ב. בדקו ברמת מובהקות של 5% האם הסיכוי לנצח במסלול מספר 1 גבוה מ- $\frac{1}{3}$ .

**פתרונות:****שאלה 1:**לא נדחה  $H_0$ **שאלה 2:**א. נדחה  $H_0$     ב. לא נדחה  $H_0$ **שאלה 3:**א. לא נדחה  $H_0$     ב. לא נדחה  $H_0$

## מבחן חי בריבוע לאי תלות בין משתנים

### רקע:

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

### מבנה המבחן:

### השערות:

$H_0$ : אין תלות בין המשתנים

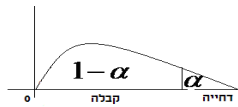
$H_1$ : יש תלות בין המשתנים

### כלל הכרעה:

הערך הקריטי נקבע על סמך התפלגות חי בריבוע. התפלגות זו היא אסימטרית חיובית ותלויה בדרגות החופש.

$$d.f = (r-1)(c-1)$$

כאשר  $r$  - מספר הקטגוריות של המשתנה שבשורות.  
 $c$  - מספר הקטגוריות של המשתנה שבעמודות.



הערך הקריטי הוא:  $\chi^2_{1-\alpha, (r-1)(c-1)}$ , כלומר האחוזון ה-  $1 - \alpha$  בהתפלגות חי בריבוע שדרגות

החופש הן  $(r-1)(c-1)$ . אם  $\chi^2 > \chi^2_{1-\alpha, (r-1)(c-1)}$  אז דוחים את השערת האפס.

### סטטיסטי המבחן:

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

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

$E_i$  - שכיחות צפויה במדגם בתא  $i$  בהנחת השערת האפס.

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

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

לאחד קטגוריות סמוכות עד שהתנאי יתקיים. תנאי חלופי: אין  $E$  קטן מ-1 וגם אין ביותר מ-20% מהתאים  $E$  קטן מ-5.

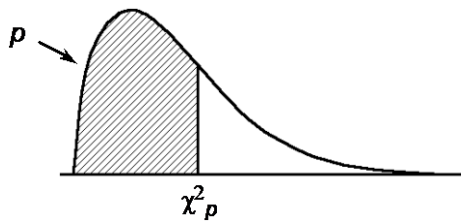
### דוגמה: (הפתרון בהקלטה)

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

דעה / המגדר	בעד	נגד	נמנע	סה"כ
גברים	50	40	10	
נשים	20	60	20	
סה"כ				



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



p

df	.005	.01	.025	.05	.10	.25	.50	.75	.90	.95	.975	.99	.995
1	0.0 <sup>4</sup> 393	0.0 <sup>3</sup> 157	0.0 <sup>3</sup> 982	0.0 <sup>2</sup> 393	0.0158	0.102	0.455	1.32	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.103	0.211	0.575	1.39	2.77	4.61	5.99	7.38	9.21	10.6
3	0.0717	0.115	0.216	0.352	0.584	1.21	2.37	4.11	6.25	7.81	9.35	11.3	12.8
4	0.207	0.297	0.484	0.711	1.06	1.92	3.36	5.39	7.78	9.49	11.1	13.3	14.9
5	0.412	0.554	0.831	1.15	1.61	2.67	4.35	6.63	9.24	11.1	12.8	15.1	16.7
6	0.676	0.872	1.24	1.64	2.20	3.45	5.35	7.84	10.6	12.6	14.4	16.8	18.5
7	0.989	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.0	14.1	16.0	18.5	20.3
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.2	13.4	15.5	17.5	20.1	22.0
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.4	14.7	16.9	19.0	21.7	23.6
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.5	16.0	18.3	20.5	23.2	25.2
11	2.60	3.05	3.82	4.57	5.58	7.58	10.3	13.7	17.3	19.7	21.9	24.7	26.8
12	3.07	3.57	4.40	5.23	6.30	8.44	11.3	14.8	18.5	21.0	23.3	26.2	28.3
13	3.57	4.11	5.01	5.89	7.04	9.30	12.3	16.0	19.8	22.4	24.7	27.7	29.8
14	4.07	4.66	5.63	6.57	7.79	10.2	13.3	17.1	21.1	23.7	26.1	29.1	31.3
15	4.60	5.23	6.26	7.26	8.55	11.0	14.3	18.2	22.3	25.0	27.5	30.6	32.8
16	5.14	5.81	6.91	7.96	9.31	11.9	15.3	19.4	23.5	26.3	28.8	32.0	34.3
17	5.70	6.41	7.56	8.67	10.1	12.8	16.3	20.5	24.8	27.6	30.2	33.4	35.7
18	6.26	7.01	8.23	9.39	10.9	13.7	17.3	21.6	26.0	28.9	31.5	34.8	37.2
19	6.84	7.63	8.91	10.1	11.7	14.6	18.3	22.7	27.2	30.1	32.9	36.2	38.6
20	7.43	8.26	9.59	10.9	12.4	15.5	19.3	23.8	28.4	31.4	34.2	37.6	40.0
21	8.03	8.90	10.3	11.6	13.2	16.3	20.3	24.9	29.6	32.7	35.5	38.9	41.4
22	8.64	9.54	11.0	12.3	14.0	17.2	21.3	26.0	30.8	33.9	36.8	40.3	42.8
23	9.26	10.2	11.7	13.1	14.8	18.1	22.3	27.1	32.0	35.2	38.1	41.6	44.2
24	9.89	10.9	12.4	13.8	15.7	19.0	23.3	28.2	33.2	36.4	39.4	43.0	45.6
25	10.5	11.5	13.1	14.6	16.5	19.9	24.3	29.3	34.4	37.7	40.6	44.3	46.9
26	11.2	12.2	13.8	15.4	17.3	20.8	25.3	30.4	35.6	38.9	41.9	45.6	48.3
27	11.8	12.9	14.6	16.2	18.1	21.7	26.3	31.5	36.7	40.1	43.2	47.0	49.6
28	12.5	13.6	15.3	16.9	18.9	22.7	27.3	32.6	37.9	41.3	44.5	48.3	51.0
29	13.1	14.3	16.0	17.7	19.8	23.6	28.3	33.7	39.1	42.6	45.7	49.6	52.3
30	13.8	15.0	16.8	18.5	20.6	24.5	29.3	34.8	40.3	43.8	47.0	50.9	53.7

**תרגילים :**

1. נבדק התלות בין גודל הארגון לבין שביעות הרצון של העובדים. להלן התוצאות :

שביעות רצון גודל המפעל	נמוכה	בינונית	גבוהה	סה"כ
גדול	182	203	215	600
קטן	154	110	136	400
סה"כ	336	313	351	1000

מה המסקנה ברמת מובהקות של 2.5%?

2. מפעל 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 יש הבדל בין שיעורי הפגומים

במשמרות השונות? הסיקו עבור רמת מובהקות  $\alpha = 0.05$ .

3. נדגמו 50 מוצרים ממפעל מסוים מתוך 30 מוצרים שיוצרו ביום 17 נבחרו לייצוא מתוך המוצרים שיוצרו בלילה 10 נבחרו לייצוא. האם יש קשר בין היות מוצר לייצוא למועד שבו הוא יוצר ? בדוק ברמת בטחון של 95%.

4. במטרה לבדוק האם השתנו דפוסי ההצבעה למפלגות השונות בין שבוע שעבר לשבוע נלקחו שני סקרים אחד מהשבוע שעבר והאחר מהשבוע. להלן דפוסי ההצבעה שהתקבלו בסקרים אלה :

	מפלגה א	מפלגה ב	מפלגות אחרות	סה"כ
שבוע שעבר				
השבוע	143		253	550
סה"כ	243	314		1050

א. מהי רמת המובהקות המינימלית עבורה ניתן להחליט שהשתנו דפוסי ההצבעה משבוע שעבר לשבוע באופן מובהק?

ב. כיצד הייתה התשובה לסעיף א משתנה אם כל השכיחויות בטבלה של תוצאות המדגם היו מוכפלות פי 2?

ג. בנו רווח סמך לשיעור המצביעים למפלגה א השבוע ברמת סמך של 95%.

5. 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 5% האם התפלגות הצבעים בחנות A היא ביחס של 3:1:1:1 לטובת הכחול.

ב. בדוק ברמת מובהקות של 2.5% האם קיים הבדל בין החנויות מבחינת התפלגות הצבעים של הפריטים הנמכרים.

6. סטודנט קיבל בבדיקת השערות ערך  $\chi^2$  (chi-square) השווה לאפס. הסטודנט הסיק כי לא קיימת תלות, בין שני המשתנים שבדק, בכל רמות מובהקות. **נכון ?**

**לא נכון? נמק/י**

7. להלן טבלת O של שני משתנים שהתקבל במדגם כלשהו :

$f(x)$	$Y_4$	$Y_3$	$Y_2$	$Y_1$	
200					$X_1$
200					$X_2$
	160	120	60	60	$f(y)$

מה צריכות להיות השכיחויות בתוך הטבלה כדי שמובהקות התוצאה (PV) תהיה 100% ?

**פתרונות:****שאלה 1:**

נסיק שיש קשר בין גודל הארגון לשביעות הרצון של העובדים.

**שאלה 2:**

נסיק שאין הבדל מובהק בין שיעור הפגומים במשמרות השונות.

**שאלה 3:**

נסיק שאין קשר בין היות מוצר לייצוא למועד שבו הוא יוצר.

**שאלה 4:**

א. 10%

ב. קטן

ג. (0.223,0.297)

**שאלה 5:**

א. נסיק שהתפלגות הצבעים בחנות היא כמו שמצוין.

ב. נסיק שיש הבדל בין החנויות מבחינת התפלגות הצבעים.

**שאלה 6:**

נכון

## הקשר בין מבחן לאי תלות ובדיקת השערות להפרש פרופורציות

### רקע:

מבחן לאי תלות שבו לכל משתנה יש שתי קטגוריות שקול לבדיקת השערות דו צדדית על הפרש פרופורציות כאשר השערת האפס היא שהפרופורציות שוות. כל זאת, כמובן, אם התנאים למבחנים מתקיימים.

### דוגמה: (פתרון בהקלטה)

בקרוב 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?

האם אפשר לבדוק זאת גם על ידי מבחן לאי תלות וגם על ידי בדיקת השערות לשתי פרופורציות?

**תרגילים:**

1. 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%.

א. על ידי מבחן לאי תלות.

ב. על ידי בדיקת השערות לשתי פרופורציות.

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 ניתן להגיד שהסיכוי להימצא במצב בריאותי תקין גבוה יותר כאשר עוסקים בפעילות גופנית סדירה לעומת המצב שלא עוסקים בפעילות גופנית סדירה? בדוק ברמת בטחון של 90%

**פתרונות:**

**שאלה 1:**

נדחה את השערת האפס.

**שאלה 2:**

ב. נדחה את השערת האפס.

## פרק 20 - 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:

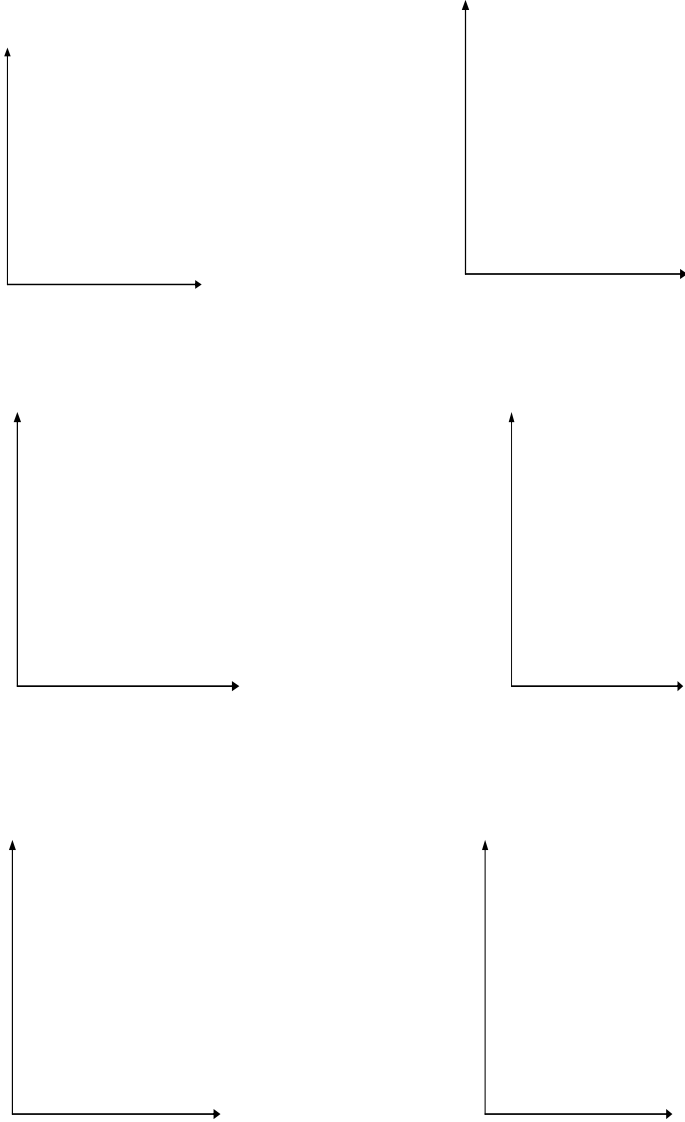
מס' דירה	X	Y
1	3	2
2	2	2
3	4	3
4	3	3
5	5	4

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

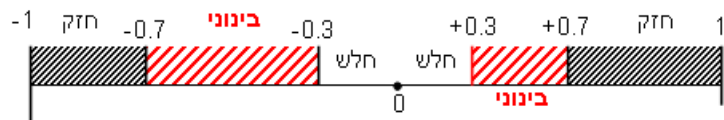
נוסחה לינארית שמקשרת את  $X$  ל- $Y$  באופן מוחלט ואילו מתאם שלילי חלקי אומר שככל

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.

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

כתב ופתר - ברק קנדל ©

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

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

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

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

**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 :

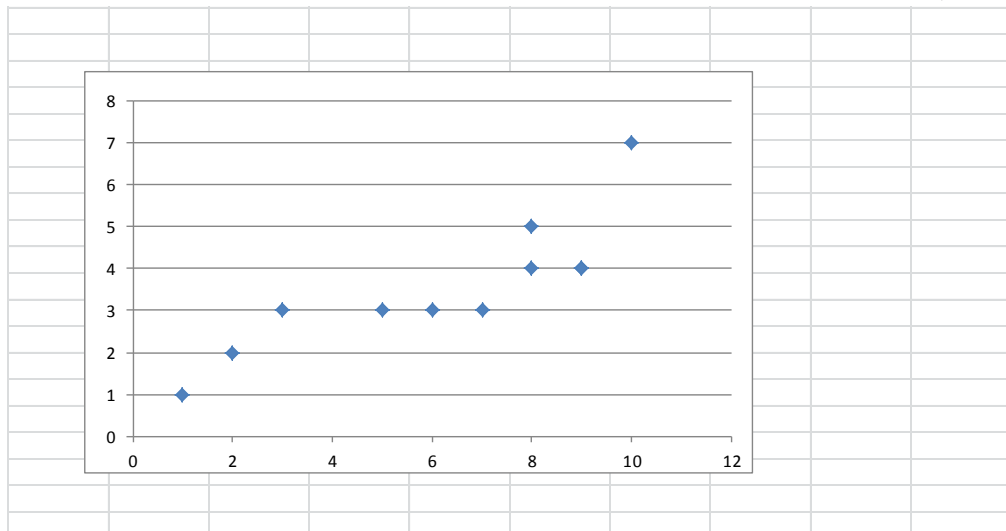
- א. הדבר מעיד שהציונים בכתה היו שליליים.
- ב. ככל שהציון של תלמיד יורד בחשבון יש לו נטייה לרדת בעברית.
- ג. ככל שהציון של תלמיד עולה בחשבון יש לו נטייה לרדת בעברית.
- ד. אף אחת מהתשובות לא נכונה.

8. נלקחו 20 מוצרים וניבדק ביום מסוים המחיר שלהם בדולרים והמחיר שלהם בש"ח ( באותו

היום ערך הדולר היה - 4.2 ₪ ) 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 21 - מדדי קשר - רגרסיה ליניארית

### רקע:

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

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

a - 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.



**תרילים:**

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?

2. נסמן ב- $X$  את ההשכלה של אדם בשנות לימוד. 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, וכך 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