

סטודנטים יקרים

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

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

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

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

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

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

כתב ופתר - ברק קנדל ©

תוכן

3.....	פרק 1 - רווח סמך לפרופורציה
10.....	פרק 2 - בדיקת השערות על פרופורציה
18.....	פרק 3- מבחני חי בריבוע
37.....	פרק 4 - רווח סמך לתוחלת (ממוצע)
43.....	פרק 5 - בדיקת השערות על תוחלת (ממוצע)
52.....	פרק 6 - רווח סמך להפרש תוחלות (ממוצעים) במדגמים בלתי תלויים
55.....	פרק 7 - בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים
66.....	פרק 8 - רווח סמך לתוחלת ההפרש במדגם מזווג
70.....	פרק 9 - בדיקת השערות על תוחלת ההפרשים במדגמים מזווגים (תלויים)
82.....	פרק 10 - הקשר בין רווח סמך לבדיקת השערות על הפרש תוחלות
86.....	פרק 11 - תרגול מסכם ברווחי סמך
89.....	פרק 12-שאלות מסכמות בבדיקת השערות על פרמטרים
109.....	פרק 13- SPSS ניתוח פלטים
123.....	פרק 14- ניתוח שונות חד גורמי
123.....	1. מתי משתמשים בניתוח שונות חד גורמי?
123.....	2. השערות
123.....	3. אמידת גודל האפקט
125.....	4. ביצוע המבחן הסטטיסטי
127.....	5. הנחות המודל
129.....	6. ניתוח פלטים
131.....	7. פרופורציית השונות המוסברת- eta^2
132.....	8. השוואות מרובות (POST-HOC TESTS)
135.....	פרק 15- רגרסיה ליניארית חד משתנית
136.....	מבוא- מקדם המתאם הליניארי
144.....	פרק 16 - קו הרגרסיה במדגם
152.....	פרק 17 - מובהקות הרגרסיה באוכלוסיה

פרק 1 - 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 כשלונות לפחות 5 או לפחות 10)

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

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 התקון?

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.

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.

מהו גודל המדגם שנלקח?

פתרונות:**שאלה 3**

א. 52

ב. 0.997

שאלה 5א. $22.5\% < p < 30.9\%$ 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: ε (או את אורך רווח הסמך).

$L = 2\varepsilon$ - אורך רווח הסמך.

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

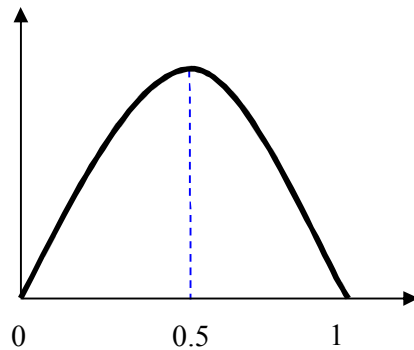
$$\varepsilon = z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

ויתעניין לדעת מהו גודל המדגם הרצוי לשם כך.

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

הבעיה שאין לנו יודעים את \hat{p} .

נתבונן בביטוי $\hat{p}(1-\hat{p})$:



כיוון שאין לנו ידע מוקדם על \hat{p} נציב את המקרה השמרני ביותר שממקסם את הביטוי עבור

$$\hat{p} = 0.5$$

$$n \geq \left(\frac{2 \cdot z_{1-\alpha/2} \cdot \sqrt{0.5 \cdot 0.5}}{L} \right)^2 \Rightarrow n \geq \left(\frac{z_{1-\alpha/2}}{L} \right)^2$$

אך אם תהיה לנו אינפורמציה מוקדמת על הפרופורציה נציב את הערך הקרוב ביותר ל-0.5 האפשרי.

דוגמה: (פתרון בהקלטה)

מעוניינים לאמוד את שיעור האבטלה במשק. האמידה צריכה להתבצע ברמת סמך של 90% ועם שגיאת אמידה שלא תעלה על 4%.
א. מהו גודל המדגם המינימאלי שיש לקחת?
ב. חזור לסעיף א' אם ידוע שהאבטלה לא אמורה לעלות על 20%.

פתרון:

א. 423

ב. 271

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. משרד התקשורת מעוניין לדעת מה שיעור בתי האב עם אינטרנט.
 - א. כמה בתי אב יש לדגום אם מעוניינים שבביטחון של 90% אורך רווח הסמך לא יעלה על 8%?
 - ב. חזרו על סעיף א. אם ידעו שלפני חמש שנים ל-80% מבתי האב היה אינטרנט וכיום יש להניח שיש ליותר אינטרנט.
3. ערוץ טלוויזיה מעוניין לאמוד את הרייטינג של הערוץ בפריים טיים. המטרה שבביטחון של 95% הסטייה המרבית בין האומד לרייטינג האמתי לא תעלה על 4%.
 - א. כמה מכשירי PEOPLE METER יש להתקין לצורך האמידה?
 - ב. לפי הערכה מוקדמת הרייטינג של הערוץ לא יכול לעלות על 20%. בהנחה ומכשיר כזה עולה 500 ₪ ליחידה מה החיסכון הכספי מאינפורמציה זאת?
4. השאלות הבאות מתייחסות לסעיף 4 :
 - 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 חישבתם 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?
5. 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 חישבת 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.
 - ג. בהמשך לסעיף הקודם. מהי טעות האמידה המרבית בביטחון של 98% ? מדוע הוא קטן מ-3% ?

פתרונות:

שאלה 1

1068

שאלה 3

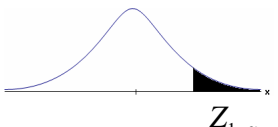
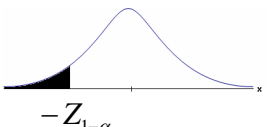
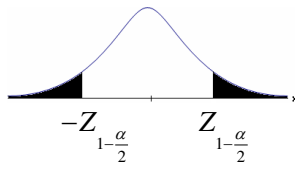
א. 601

ב. 108000 ₪.

פרק 2 - בדיקת השערות על פרופורציה

התהליך

רקע:

$H_0 : p = p_0$ $H_1 : p > p_0$	$H_0 : p = p_0$ $H_1 : p < p_0$	$H_0 : p = p_0$ $H_1 : p \neq p_0$	השערת האפס : השערה אלטרנטיבית:
$np_0 \geq 5 \ \& \ n(1-p_0) \geq 5$			תנאים:
$Z_{\hat{p}} > Z_{1-\alpha}$  דוחים את H_0	$Z_{\hat{p}} < -Z_{1-\alpha}$  דוחים את H_0	או $Z_{\hat{p}} < -Z_{1-\frac{\alpha}{2}}$ $Z_{\hat{p}} > Z_{1-\frac{\alpha}{2}}$  דוחים את H_0	כלל ההכרעה: אזור הדחייה של :

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 $\hat{p} < p_0 - Z_{1-\alpha/2} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$	כלל ההכרעה: אזור הדחייה של H_0
--	--	--	--

דוגמה: (פתרון בהקלטה)

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

תרגילים:

1. במשך שנים אחוז המועמדים שהתקבל לפקולטה מסוימת היה 25%. השנה מתוך מדגם של 120 מועמדים התקבלו 22. ברמת מובהקות של 5% האם השנה הקשו על תנאי הקבלה?
2. במדגם של 300 אזרחים 57% מתנגדים להצעת חוק מסוימת. לאור נתונים אלה האם רוב האזרחים מתנגדים להצעת החוק? בדקו ברמת מובהקות של 10%.
3. הטילו מטבע 50 פעמים וקיבלו 28 פעמים עץ. האם המטבע הוגן ברמת מובהקות של 5%?
4. קפיטריה במכללה מסוימת מעריכה כי אחוז הסטודנטים שקונים קפה בקפיטריה הינו 20%. נערך סקר אשר כלל 200 סטודנטים. התברר כי 33 מהם רוכשים קפה בקפיטריה. מטרת הסקר הייתה לבדוק את אמיתות הערכה של הקפיטריה.
 - א. רשמו את ההשערות.
 - ב. בדוק את ההשערות ברמת מובהקות של 10%.
 - ג. מה תהיה המסקנה אם נקטין את רמת המובהקות?
5. חבר כנסת רוצה להעביר חוק. לצורך כך הוא דוגם 400 אזרחים במטרה לבדוק האם רוב האזרחים תומכים בחוק. במדגם התקבל ש-276 אזרחים תומכים בחוק.
 - א. מה מסקנתכם ברמת מובהקות של 5%?
 - ב. האם ניתן לדעת מה תהיה המסקנה אם רמת המובהקות תהיה גדולה יותר? הסבירו.
6. שני חוקרים בדקו את ההשערות הבאות:

$$H_0 : p = p_0$$

$$H_1 : p > p_0$$
 חוקר א השתמש ברמת מובהקות α_1 וחוקר ב ברמת מובהקות α_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:**

נדחה H_0

שאלה 2:

נדחה H_0

שאלה 3:

נקבל H_0

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.

מובהקות התוצאה

רקע:

דרך נוספת להגיע להכרעות שלא דרך כלל הכרעה, היא דרך חישוב מובהקות התוצאה:
 באמצעות תוצאות המדגם מחשבים את מובהקות התוצאה שמסומן ב- p_v .
 את רמת המובהקות החוקר קובע מראש לעומת זאת, את מובהקות התוצאה החוקר יוכל לחשב רק אחרי שיהיו לו את התוצאות.
 המסקנה של המחקר תקבע לפי העיקרון הבא:

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

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

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

אם ההשערה היא דו צדדית:

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

מובהקות התוצאה היא גם האלפא המינימלית לדחיית השערת האפס.

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

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

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 ארוז 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 :**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 3- מבחני חי בריבוע

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: $\chi^2_{\alpha, (r-1)(c-1)}$, כלומר הערך שהשטח מימינו הוא α בהתפלגות חי בריבוע שדרגות החופש הן $(r-1)(c-1)$. אם $\chi^2 > \chi^2_{\alpha, (r-1)(c-1)}$ 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.

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

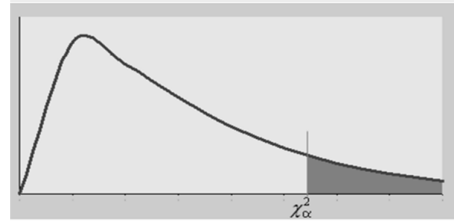
הערה: תנאי כדי לבצע את המבחן הוא $E_i \geq 5$ לכל i . במידה ותנאי זה לא מתקיים יש אפשרות לאחד קטגוריות סמוכות עד שהתנאי יתקיים. תנאי חלופי: אין E קטן מ-1 וגם אין ביותר מ-20% מהתאים E קטן מ-5.

דוגמה : (הפתרון בהקלטה)

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

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

טבלת חי בריבוע על סמך השטח מצד ימין

 α

df	0.995	.99	.975	.95	.90	.75	.50	.25	.10	.05	.025	.01	.005
1	0.0 ⁴ 393	0.0 ³ 157	0.0 ³ 982	0.0 ² 393	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 :

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

מה המסקנה ברמת מובהקות של 2.5%?

2. מפעל עובד בשלוש משמרות. להלן מספר המוצרים 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
143		253	השבוע
243	314		סה"כ
			550
			1050

- א. מהי רמת המובהקות המינמלית 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 1:1:1:3 לטובת הכחול.
- ב. בדוק ברמת מובהקות של 2.5% האם קיים הבדל בין החנייות מבחינת התפלגות הצבעים של הפריטים הנמכרים.
6. סטודנט קיבל בבדיקת השערות ערך χ^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

פלטטים על מבחן לאי תלות

רקע:

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

מבנה המבחן:

השערות:

H_0 : אין תלות בין המשתנים

H_1 : יש תלות בין המשתנים

דרגות חופש:

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

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

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: (פתרון בהקלטה)

במחקר רצו לבדוק את הקשר בין צבע שיער לבין צבע עיניים של אנשים. הפלטים שהתקבלו מצורפים.

- א. מהו המבחן הסטטיסטי המתאים ?
- ב. כמה קטגוריות יש לכל משתנה?
- ג. רשמו את השערות המחקר.
- ד. מה מספר דרגות החופש ?
- ה. כמה אנשים במדגם נמצאו עם שיער חום?
- ו. כמה אנשים היית מצפה במדגם שיהיה להם שיער חום ועיניים ירוקות בהנחה ואין קשר בין צבע שיער לצבע עיניים?
- ז. מתוך הבלונדינים מה אחוז בעלי עיניים כחולות במדגם?
- ח. מתוך בעלי עיניים ירוקות מה אחוז הבלונדינים במדגם ?
- ט. מה ערכו של סטטיסטי המבחן ומהי מובהקות התוצאה?
- י. מה מסקנת המחקר? $\alpha = 5\%$

להלן הפלטים שהתקבלו :

Case Processing Summary

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

hair_color * eye_color Crosstabulation

		eye_color			Total
		brown	green	Blue	
black	Count	13	7	7	27
	Expected Count	10.7	8.3	8.0	27.0
	% within hair_color	48.1%	25.9%	25.9%	100.0%
	% within eye_color	41.9%	29.2%	30.4%	34.6%
brown	Count	12	12	6	30
	Expected Count	11.9	9.2	8.8	30.0
	% within hair_color	40.0%	40.0%	20.0%	100.0%
	% within eye_color	38.7%	50.0%	26.1%	38.5%
blond	Count	6	5	10	21
	Expected Count	8.3	6.5	6.2	21.0
	% within hair_color	28.6%	23.8%	47.6%	100.0%
	% within eye_color	19.4%	20.8%	43.5%	26.9%
Total	Count	31	24	23	78
	Expected Count	31.0	24.0	23.0	78.0
	% within hair_color	39.7%	30.8%	29.5%	100.0%
	% within eye_color	100.0%	100.0%	100.0%	100.0%

Chi-Square Tests

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

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

תרגילים:

1. בסקר שנעשה על ידי משרד ראש הממשלה נדגמו 60 אזרחים. כל אזרח נשאל על מגדרו והאם הוא בעד הקמת מדינה פלסטינית.

א. מה ההשערות הנבדקות ומהו סטטיסטי המבחן?

ב. אם סטטיסטי המבחן היה גדל כיצד הדבר היה משפיע על SIG שבפלט.

ג. האם קיים קשר בין מגדר ודעה ברמת מובהקות של 5% ?

ד. מהו האומדן לאחוז התומכים במדינה פלסטינית מתוך הגברים?

ה. איזה אחוז מהנשאלים שהיו בעד מדינה פלסטינית הם גברים?

להלן הפלטים :

Chi-Square Tests

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

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

b. minimum expected count is 7.25.

gender * opinion Crosstabulation

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

2. להלן פלט על סמך סקר שנעשה בקרב סטודנטים, בסקר נשאלו הסטודנטים על המוזיקה אותה הם מעדיפים וצורת הבילוי המועדפת עליהם.

Crosstab

Count		בילוי			Total
		קריאה	ספורט	מועדון	
מוזיקה	רוק	0	0	11	11
	פופ	1	6	8	15
	קלאסי	5	6	9	20
Total		6	12	28	46

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	11.929 ^a	?	.018
N of Valid Cases	46		

a. 5 cells (55.6%) have expected count less than 5. The minimum expected count is 1.43.

- א. בין אלו משתנים נבדק הקשר? כמה קטגוריות לכל משתנה?
 ב. האם התנאים של המודל מתקיימים?
 ג. מה מספר דרגות החופש במבחן הני"ל?
 ד. מה ההשערות של המבחן?

3. מחקר התעניין לבדוק את הקשר בין רמת הכנסה של משפחה לבין צריכת עגבניות אורגניות. הפלטים מצורפים.
- א. השלימו את שלושת המספרים החסרים בטבלה (היכן שיש סימני שאלה).
- ב. מה ערכו של חי בריבוע הסטטיסטי.
- ג. תנו הערכה למובהקות התוצאה לבדיקת הקשר בין רמת הכנסה של משפחה לבין צריכת עגבניות אורגניות.

Crosstabulation רמת_הכנסה * צרכן עגבניות

		צרכן עגבניות		Total
		אורגני	לא אורגני	
הרבה מתחת לממוצע רמת_הכנסה	Count	17	42	59
	% within רמת_הכנסה	28.8%	?	100.0%
	% within צרכן עגבניות	13.6%	33.6%	23.6%
מתחת לממוצע	Count	27	22	49
	% within רמת_הכנסה	55.1%	44.9%	100.0%
	% within צרכן עגבניות	?	17.6%	19.6%
ממוצע	Count	31	29	60
	% within רמת_הכנסה	51.7%	48.3%	100.0%
	% within צרכן עגבניות	24.8%	23.2%	24.0%
מעל הממוצע	Count	44	26	70
	% within רמת_הכנסה	62.9%	37.1%	100.0%
	% within צרכן עגבניות	35.2%	20.8%	28.0%
הרבה מעל הממוצע	Count	?	6	12
	% within רמת_הכנסה	50.0%	50.0%	100.0%
	% within צרכן עגבניות	4.8%	4.8%	4.8%
Total	Count	125	125	250

4. חוקר בדק את הקשר בין צבע השיער לבין צבע העיניים בעזרת מבחן חי בריבוע בקרב 52 נבדקים. תוצאות המבחן מוצגות בטבלה. בנוסף ידוע כי סטטיסטי המבחן שהתקבל מעיבוד הנתונים הוא 8.08.

- א. מה תהיה מסקנת המחקר ברמת מובהקות של 1%?
 ב. מה ערכו של E עבור עיניים כחולות וצבע שיער כהה.
 ג. מה יהיה בקירוב ערכו של מקדם המתאם של קרמר?
 ד. מהי פרופורציית בעלי צבע השיער הבהיר מקרב בעלי העיניים הירוקות?

להלן הפלט:

Crosstabulation **צבע עיניים * צבע שיער**

			צבע שיער		Total
			כהה	בהיר	
צבע עיניים	כחול	Count			
		% within	50.0%	50.0%	100.0%
		% within	21.6%	53.3%	30.8%
		% of Total	15.4%	15.4%	30.8%
חום		Count			
		% within	83.3%	16.7%	100.0%
		% within	27.0%	13.3%	23.1%
		% of Total	19.2%	3.8%	23.1%
ירוק		Count			
		% within	79.2%	20.8%	100.0%
		% within	51.4%	33.3%	46.2%
		% of Total	36.5%	9.6%	46.2%
Total		Count			
		% within	71.2%	28.8%	100.0%
		% within	100.0%	100.0%	100.0%
		% of Total	71.2%	28.8%	100.0%

5. במחקר מסוים רצו לבדוק האם יש קשר בין המגדר להוצאה על לבוש במשך שנה. דגמו באופן מקרי גברים ונשים ובדקו את רמת ההוצאה שלהם על לבוש בשנה האחרונה. חוקר א' בדק האם קיים הבדל בתוחלות ההוצאה בין גברים לנשים. חוקר ב' קיבץ את ההוצאה לקטגוריות ובאופן הזה בדק האם קיים הבדל בהתפלגות ההוצאה בין גברים לנשים. הקטגוריות חולקו לשלוש קבוצות הוצאה .
- א. איזה פלט מתאים לאיזה אחד מהחוקרים? נמקו.
- ב. מה מסקנתו של חוקר א'? בדקו ברמת מובהקות $\alpha = 0.05$. (רשמו השערות, נסחו הנחות, ציינו כלל החלטה ותנו מסקנה במונחי המשתנים)
- ג. איזו טעות יכולה להיות במסקנתו של חוקר א'? נסחו את הטעות במונחי השאלה.
- ד. מהי מסקנתו של חוקר ב'? בדקו ברמת מובהקות $\alpha = 0.05$. (רשמו השערות, נסחו הנחות, ציינו כלל החלטה ורשמו מסקנה במונחי המשתנים)
- ה. איזו טעות יכולה להיות במסקנתו של חוקר ב'? נסחו זאת במונחי השאלה
- ו. כיצד ניתן ליישב את מסקנות שני החוקרים?

להלן פלט ראשון :

T-Test
Group Statistics

gender	N	Mean	Std. Deviation	Std. Error Mean
female	40	2.9000	1.15025	.18187
male	40	2.6000	2.52982	.40000

dimensio
n1

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
expose	Equal variances assumed	16.805	.000	.683	78	.497	.30000	.43941	-.57479	1.17479
	Equal variances not assumed			.683	54.464	.498	.30000	.43941	-.58078	1.18078

Crosstabs
Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
gender * category	80	100.0%	0	.0%	80	100.0%

gender * category Crosstabulation

		category			Total	
		a	b	c		
gender	Female	Count	2	30	8	40
		Expected Count	11.0	21.0	8.0	40.0
		% within gender	5.0%	75.0%	20.0%	100.0%
		% within category	9.1%	71.4%	50.0%	50.0%
	Male	Count	20	12	8	40
		Expected Count	11.0	21.0	8.0	40.0
		% within gender	50.0%	30.0%	20.0%	100.0%
		% within category	90.9%	28.6%	50.0%	50.0%
Total		Count	22	42	16	80
		Expected Count	22.0	42.0	16.0	80.0
		% within gender	27.5%	52.5%	20.0%	100.0%
		% within category	100.0%	100.0%	100.0%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	22.442 ^a	2	.000
Likelihood Ratio	25.064	2	.000
N of Valid Cases	80		

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

The minimum expected count is 8.00.

פתרונות:**שאלה 1:**

- א. 1.973
 ב. קטן
 ג. לא נדחה H_0
 ד. 34.5%
 ה. 38.5%

שאלה 2:

- א. בילוי מועדף ו- מוזיקה מועדפת עם 3 קטגוריות לכל משתנה.
 ב. לא.
 ג. 4

שאלה 3:

- א. 71.2% 21.6% 6
 ב. 15.8
 ג. קטן מ- 0.005

שאלה 4:

- א. לא נדחה H_0
 ב. 11.4
 ג. 0.394
 ד. 20.8%

שאלה 5:

נקבל H_0

שאלה 6:

א

שאלה 7:

ג

שאלה 8:

- א. אותה מסקנה
 ב. לא ניתן לדעת.

שאלה 3:

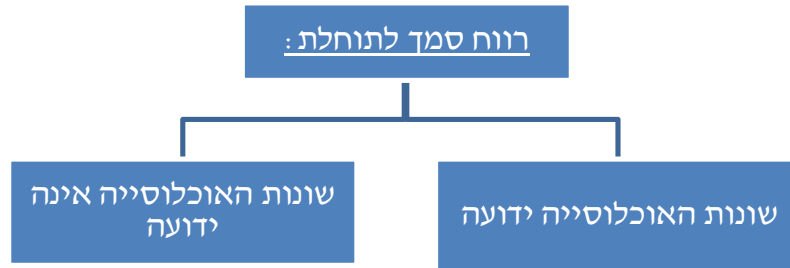
- א. יש קשר
 ב. 0.394
 ג. 0.208

פרק 4 – רווח סמך לתוחלת (ממוצע)

רווח סמך לתוחלת (ממוצע האוכלוסייה) כששונות האוכלוסייה אינה ידועה

רקע:

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



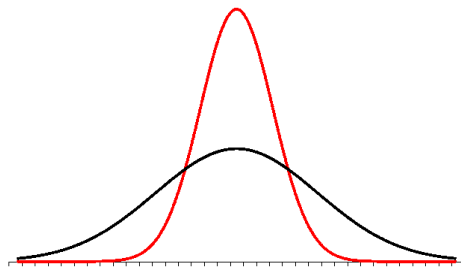
בפרק זה נעסוק במקרה ששונות האוכלוסייה אינה ידועה לנו. מקרה יותר פרקטי.

התנאי: $X \sim N$ או שהמדגם גדול

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

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

התפלגות T:



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

דוגמה : (פתרון בהקלטה)

הזמן שלוקח לפתור שאלה מסוימת בחשבון מתפלג אצל תלמידי כיתות ח' נורמאלית. במטרה לאמוד את תוחלת זמן הפתרון נדגמו 4 תלמידים בכיתה ח' . להלן התוצאות שהתקבלו בדקות : 4.7, 5.2, 4.6, 5.3 . בנו רווח סמך ברמת סמך של 95% לממוצע זמן הפתרון לשאלה בקרב תלמידי כיתה ח' .

פתרון :

$$4.39 < \mu < 5.51$$

תרגילים:

1. מחקר מעוניין לדעת כיצד תרופה מסוימת משפיעה על קצב פעימות הלב. ל-5 אנשים שנטלו את התרופה מדדו את הדופק והתקבל מספר פעימות לדקה: 84, 88, 84, 79, 89. הערה: לצורך פתרון הנח שקצב פעימות הלב מתפלג נורמאלית בקירוב.
 - א. בנו רווח סמך ברמת סמך של 95% לתוחלת הדופק של נוטלי התרופה הני"ל.
 - ב. נתון שהדופק הממוצע ללא לקיחת התרופה הינו 70. לאור זאת, האם בביטחון של 95% התרופה משפיעה על הדופק?
 - ג. בהמשך לסעיף א, אם היינו בונים את רווח הסמך ברמת ביטחון של 99% כיצד הדבר היה משפיע על רווח הסמך?

2. במדגם שנעשה על 25 מתגייסים לצבא האמריקאי התקבל כי: גובה ממוצע של חייל הינו 178 ס"מ עם סטיית תקן $\hat{\sigma} = 13$ 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 הני"ל.

5. נשקלו 60 תינוקות אשר נולדו בשבוע ה-40 של ההיריון. המשקל נמדד בקילוגרמים. להלן התוצאות שהתקבלו: $\sum_{i=1}^{60} X_i = 195$, $\sum_{i=1}^{60} X_i^2 = 643.19$. בנו רווח סמך ברמת סמך של 95% לתוחלת משקל תינוק ביום היוולדו.

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

סטטיסטיקאי א : הניח $\sigma = 20$

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

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

א. סטטיסטיקאי א

ב. סטטיסטיקאי ב

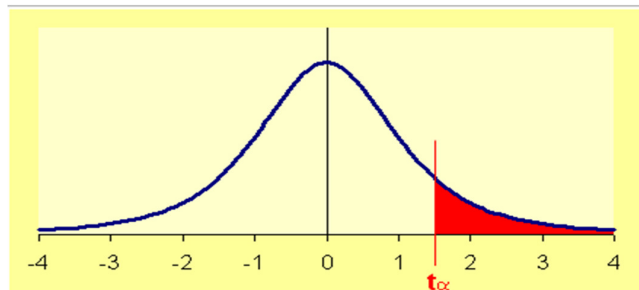
ג. אותו אורך רווח סמך לשני הסטטיסטיקאים.

ד. תלוי בתוצאות המדגם של כל סטטיסטיקאי.

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

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

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



פתרונות:**שאלה 1**

$$79.88 < \mu < 89.72 \text{ א.}$$

שאלה 4

$$96.63 < \mu < 107.37 \text{ א.}$$

$$96.90 < \mu < 107.10 \text{ ב.}$$

שאלה 5

$$3.149 < \mu < 3.351$$

שאלה 7

90%

פרק 5 - בדיקת השערות על תוחלת (ממוצע)

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 :
1. σ ידועה 2. $X \sim N$ או מדגם מספיק גדול			תנאים :
$P_{H_0}(\bar{X} \geq \bar{x})$	$P_{H_0}(\bar{X} \leq \bar{x})$	אם $2 \cdot P_{H_0}(\bar{X} \geq \bar{x}) \leftarrow \bar{x} > \mu_0$ אם $2 \cdot P_{H_0}(\bar{X} \leq \bar{x}) \leftarrow \bar{x} < \mu_0$	p-value

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:

**מובהקות התוצאה (p-value) בבדיקת השערות על תוחלת עם שונות אוכלוסייה
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.

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

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

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

$d.f = n - 1$

דוגמה : (פתרון בהקלטה)

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

א. רשום את השערות המחקר.

ב. מצא חסמים למובהקות התוצאה.

ג. מה המסקנה ברמת מובהקות של 5% ?

תרגילים :

1. קו ייצור אריזות סוכר נארזות כך שהמשקל הממוצע של אריזות הסוכר צריך להיות אחד קילוגרם. בכל יום דוגמים מקו הייצור 5 אריזות במטרה לבדוק האם קו הייצור תקין. בבדיקה דגמו 5 אריזות סוכר ולהלן משקלן בגרמים:
 $1008, 1024, 996, 1005, 997$
 א. רשמו את השערות המחקר.
 ב. מהי מובהקות התוצאה? הצג חסמים.
 ג. מה המסקנה ברמת מובהקות של 5%?
2. חוקר בדק את הטענה כי פועלים העובדים במשמרת לילה איטיים יותר מפועלים העובדים ביום. ידוע כי משך הזמן הממוצע הדרוש לייצר מוצר מסוים ביום הוא 6 שעות. במדגם מיקרי של 25 פועלים שעבדו במשמרת לילה נמצא כי הזמן הממוצע לייצר אותו מוצר הוא 7 שעות עם סטית תקן של 3 שעות.
 מהי ה- α 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

פרק 6 - רווח סמך להפרש תוחלות (ממוצעים) במדגמים בלתי תלויים

כששונויות האוכלוסייה אינן ידועות אך שוות

רקע:

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

האומד נקודתי: $\bar{x}_1 - \bar{x}_2$

התנאים לבניית רווח הסמך:

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

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

3. מדגמים בלתי תלויים.

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

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

דרגות החופש: $d.f = n_1 + n_2 - 2$

רווח סמך:

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

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

דוגמה: (פתרון בהקלטה)

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

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

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

פתרון: (1357,1643)

תרגילים:

1. נדגמו 15 ישראלים ו-15 אמריקאים.

כל הנדגמים נגשו למבחן IQ. להלן תוצאות המדגם:

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

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

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


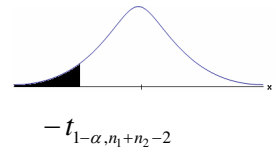

X	22	20	21	25
Y	18	25	17	12

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

פרק 7 - בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

כששונות האוכלוסייה לא ידועות ומניחים שהן שוות

רקע:

$H_0 \quad \mu_1 - \mu_2 = c$ $H_1 \quad \mu_1 - \mu_2 > c$	$H_0 \quad \mu_1 - \mu_2 = c$ $H_1 \quad \mu_1 - \mu_2 < c$	$H_0 \quad \mu_1 - \mu_2 = c$ $H_1 \quad \mu_1 - \mu_2 \neq c$	השערת האפס : השערה אלטרנטיבית:
תנאים: 3. מדגמים בלתי תלויים 4. σ_1, σ_2 לא ידועות אך שוות 5. המשתנים בכל אוכלוסייה מתפלגים נורמלית			
$t_{\bar{x}_1 - \bar{x}_2} > t_{1-\alpha}^{(n_1+n_2-2)}$  $t_{1-\alpha, n_1+n_2-2}$ דוחים את H_0	$t_{\bar{x}_1 - \bar{x}_2} < -t_{1-\alpha}^{(n_1+n_2-2)}$  $-t_{1-\alpha, n_1+n_2-2}$ דוחים את H_0	או $t_{\bar{x}_1 - \bar{x}_2} < -t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)}$ $t_{\bar{x}_1 - \bar{x}_2} > t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)}$  $-t_{1-\frac{\alpha}{2}, n_1+n_2-2} \quad t_{1-\frac{\alpha}{2}, n_1+n_2-2}$ דוחים את H_0	אזור הדחייה של H_0:

סטטיסטי המבחן :

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

חלופה אחרת לכלל הכרעה:

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

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	2012
	69	74	105	91	82	100	2013

בדקו שבשנת 2013 הייתה ירידה משמעותית בשטחי הדירות לעומת שנת 2012 עבור רמת מובהקות של 5%. הניחו ששטחי הדירות בכל שנה מתפלגים נורמלית עם אותה שונות.

2. נדגמו 15 ישראלים ו-15 אמריקאים. כל הנתונים 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 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

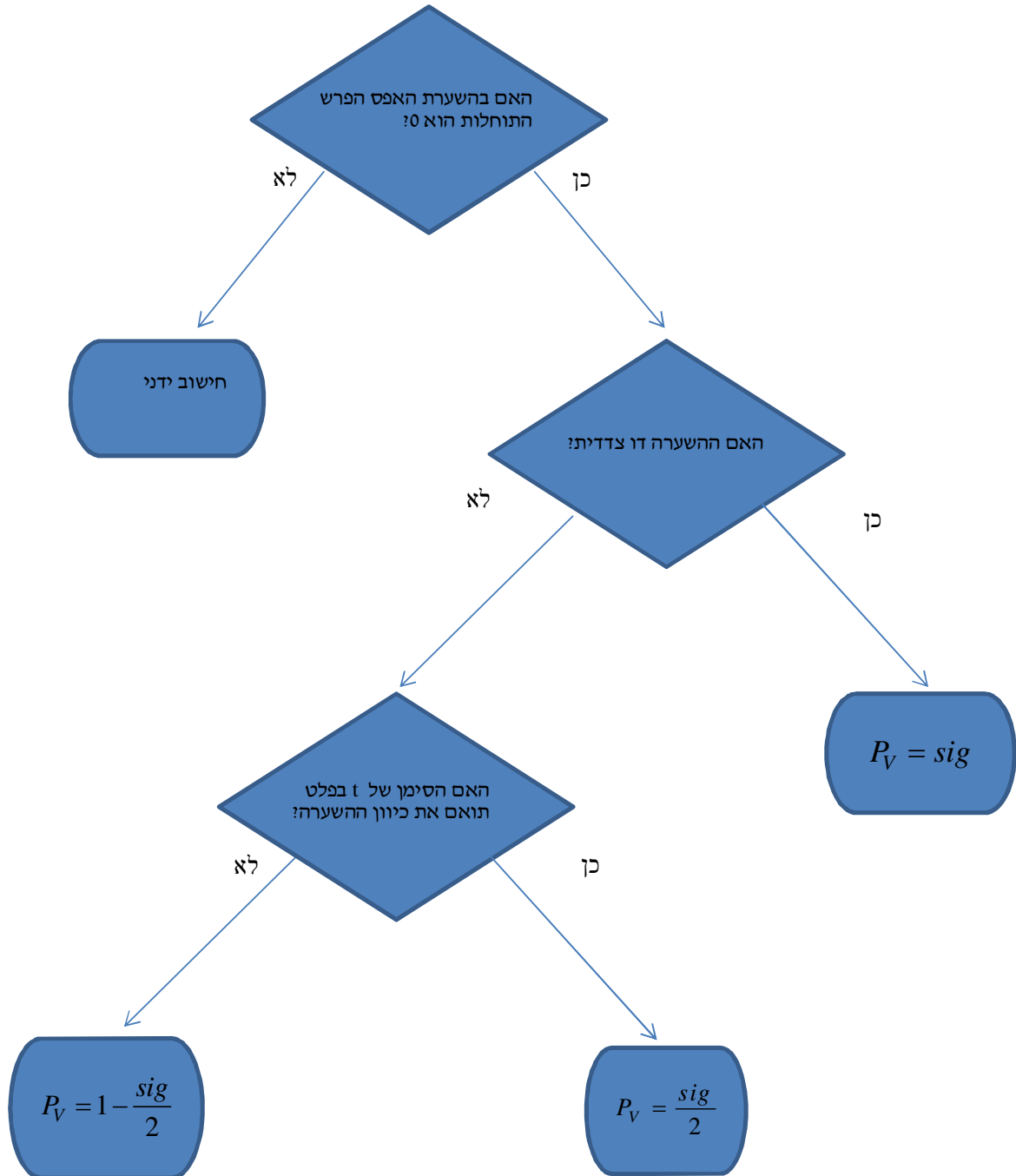
ב. רמות מובהקות של לפחות 5%

ג. לא נדחה H_0

ניתוח פלטים

רקע:

מובהקות התוצאה על סמך הפלט:



דוגמה: (פתרון בהקלטה)

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

Group Statistics

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

Independent Samples Test

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

א. מהו המבחן הסטטיסטי שנעשה כאן?

ב. בדוק ברמת מובהקות של 5% האם קיים שוויון שונויות בין שני האזורים?

ג. בדוק האם קיים הבדל בין "South East" ל-"North East" ברמת מובהקות של 5% מבחינת מספר האחים והאחיות הממוצע.

ד. מהי מובהקות התוצאה לבדיקת הטענה שהפרש הממוצע בין-"South East"

לבין-"North East" חיובי ?

תרגילים :

1. להלן פלט מתוכנת SPSS מתוך מחקר שבחן את רמת האופטימיות של גברים ונשים.
רמת האופטימיות נמדדה בסקאלת ציונים של 1 עד 5.

Group Statistics

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

Independent Samples Test

	Levene's Test for Equality of Variances	t-test for Equality of Means								
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
optimizm	Equal variances assumed	.383	.536	1.935	1199	.053	.05500	.02842	-.00076	?
	Equal variances not assumed			1.938	190.977	.053	.05500	.02838	-.00068	.11067

- א. האם ניתן להניח ששוונות האופטימיות של נשים וגברים שווה ברמת מובהקות של 5%?
 ב. ברמת מובהקות של 5% האם קיים הבדל בין הנשים לגברים ברמת האופטימיות הממוצעת שלהם?
 ג. מצא את הגבול העליון של רווח הסמך המסומן בסימן שאלה בפלט. דייק עד 5 ספרות אחרי הנקודה.
 ד. בנה רווח סמך לתוחלת רמת האופטימיות של הגברים ברמת סמך של 95%.

2. פסיכולוגים טוענים שאנשים שניגשים למבחן אינטליגנציה יותר מפעם אחת נוטים לקבל ציונים גבוהים יותר.
להלן הפלט שהתקבל:

T-Test

Group Statistic:

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

Independent Samples Test

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

מקרא:

A = נגשו פעם אחת.

B = נגשו יותר מפעם אחת.

- רשמו את השערות המחקר והסבירו מהו המבחן המתאים כאן.
- כיצד הייתה משתנה התשובה לסעיף הקודם אם היה מדובר על אותם אנשים שציונם נבדק פעם אחרי המבחן הראשון שעשו ופעם אחרי המבחן השני?
- האם ניתן לומר כי מידת הפיזור של ציוני האנשים הנבחרים בפעם הראשונה שונה ממידת הפיזור של ציוני האנשים אשר נבחרים בפעם השנייה. בדוק ברמת מובהקות של $\alpha=0.05$.
- האם נכונה טענת הפסיכולוגים ברמת מובהקות של $\alpha=0.01$.

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

T-Test

Group Statistics

מקום המגורים	N	Mean	Std. Deviation	Std. Error Mean
גיל הנישואין עיר א	50	24.8072	1.38978	.19654
עיר ב	50	23.0131	1.62070	.22920

Independent Samples Test

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

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

4. להלן פלט של תוכנת SPSS:

T-Test

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

Independent Samples Test

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

א. השלם את סימני השאלה בטבלה.

ב. מהי מובהקות התוצאה לבדיקת הטענה שקיים הבדל בין השונות של X לזה של Y?

ג. מהי מובהקות התוצאה לבדיקת הטענה שהתוחלת של X גדולה מהתוחלת של Y?

ד. מהי מובהקות התוצאה לבדיקת הטענה שהתוחלת של X קטנה מהתוחלת של Y?

פתרונות:**שאלה 1:**

- א. נכריע שוויון שוניות.
 ב. נקבע שלא קיים הבדל בין נשים לגברים מבחינת האופטימיות הממוצעת.
 ג. 0.11076
 ד. $2.5665 \leq \mu \leq 2.6441$

שאלה 2:

- א. מבחן t למדגמים בלתי תלויים.
 ב. מבחן t למדגמים מזווגים.
 ג. נקבע שקיים שוויון שוניות.
 ד. לא נקבל את טענת הפסיכולוגים.

שאלה 3:

- א. מבחן t להשוואת תוחלת במדגמים בלתי תלויים.
 ב. $1.19497 \leq \mu_1 - \mu_2 \leq 2.39332$
 ג. כן.

שאלה 4:

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

פרק 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: μ_D

התנאים לבניית רווח הסמך:

- $x, y \sim N$

- המדגם מזווג

נוסחת רווח הסמך:

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

כאשר דרגות החופש: $d.f = n - 1$

דוגמה: (פתרון בהקלטה)

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

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

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

תרגילים:

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

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

ניח שהציונים מתפלגים נורמאלית.

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

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



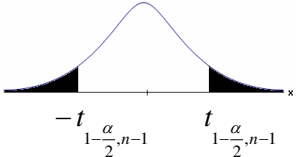
המדינה	בזק-X	קווי זהב-Y
ארה"ב	1.5	1.4
קנדה	2.1	2
הולנד	2.2	1.9
פולין	3	3.1
מצרים	3.5	3.3
סין	3.2	3.2
יפן	4.2	4.2

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

**פרק 9 - בדיקת השערות על תוחלת הפרשים במדגמים מזווגים
(תלויים)**

בדיקת השערות למדגמים מזווגים

רקע:

$H_0: \mu_D = C$ $H_1: \mu_D > C$	$H_0: \mu_D = C$ $H_1: \mu_D < C$	$H_0: \mu_D = C$ $H_1: \mu_D \neq C$	השערת האפס : השערה אלטרנטיבית:
<p align="center">.6 σ_D אינה ידועה</p> <p align="center">.7 $D \sim N$ או מדגם מספיק גדול</p>			תנאים:
$t_{\bar{D}} > t_{1-\alpha}^{(n-1)}$  $t_{1-\alpha, n-1}$ 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 -	כלל ההכרעה: אזור הדחייה של H_0 :
$\bar{D} > C + t_{1-\alpha}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$	$\bar{D} < C - t_{1-\alpha}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$	$\bar{D} > C + t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$ או $\bar{D} < C - t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$	חלופה לכלל הכרעה : 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 בממוצע?

2. מכון המכין לפסיכומטרי טוען שהוא מעלה את ממוצע הציונים ביותר מ-30 נקודות. 8 נבחנים נבדקו לפני ואחרי שהם למדו במכון. להלן התוצאות שהתקבלו:

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

- מה מסקנתכם ברמת מובהקות 5%? הניחו שציוני פסיכומטרי מתפלג נורמלית.

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

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

- פורסם שתלמידים שמסיימים את סמסטר ב משפרים במוצע את הציונים ב-5 נקודות לעומת סמסטר א'. הנח שהציונים מתפלגים נורמלית.
- א. מהי מובהקות התוצאה לבדיקת הטענה שהשיפור הוא יותר מ 5 נקודות?
- ב. על סמך הסעיף הקודם, מהי רמת המובהקות המינימלית להכרעה שהשיפור הוא יותר מ- 5 נקודות?
- ג. לאור זאת, מה המסקנה ברמת מובהקות של 10% ?

4. לצורך בדיקת השפעת היפנוזה על לימוד אנגלית, נבחרו 10 זוגות תאומים זהים. אחד התאומים למד אנגלית בהשפעת היפנוזה, והשני ללא היפנוזה. לאחר מכן נערך לשניהם מבחן באנגלית. נניח שציוני המבחן מתפלגים נורמאלית ללא ידיעת השונות האמתית.
- המבחן שיש לבצע כאן הוא:
- א. מבחן Z למדגם יחיד.
- ב. מבחן T למדגם יחיד.
- ג. מבחן T למדגמים בלתי תלויים.
- ד. מבחן T למדגמים מזווגים.

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

משקל במכשיר 1	5.4	6.9	7.0	5.2
משקל במכשיר 2	5.3	6.9	7.1	5.0

נניח שהמשקלים מתפלגים נורמלית.

המבחן שיש לבצע כאן הוא :

א. מבחן Z למדגם יחיד.

ב. מבחן T למדגם יחיד.

ג. מבחן T למדגמים בלתי תלויים.

ד. מבחן T למדגמים מזווגים.

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

המבחן שיש לבצע כאן הוא :

א. מבחן Z למדגם יחיד.

ב. מבחן T למדגם יחיד.

ג. מבחן T למדגמים בלתי תלויים.

ד. מבחן T למדגמים מזווגים.

פתרונות:**שאלה 1:**לא נדחה H_0 **שאלה 2:**לא נדחה H_0 **שאלה 3:**א. $0.25 \leq p \leq 0.5$

ב. 0.5

ג. לא נדחה H_0 **שאלה 4:**

התשובה היא ד.

שאלה 5:

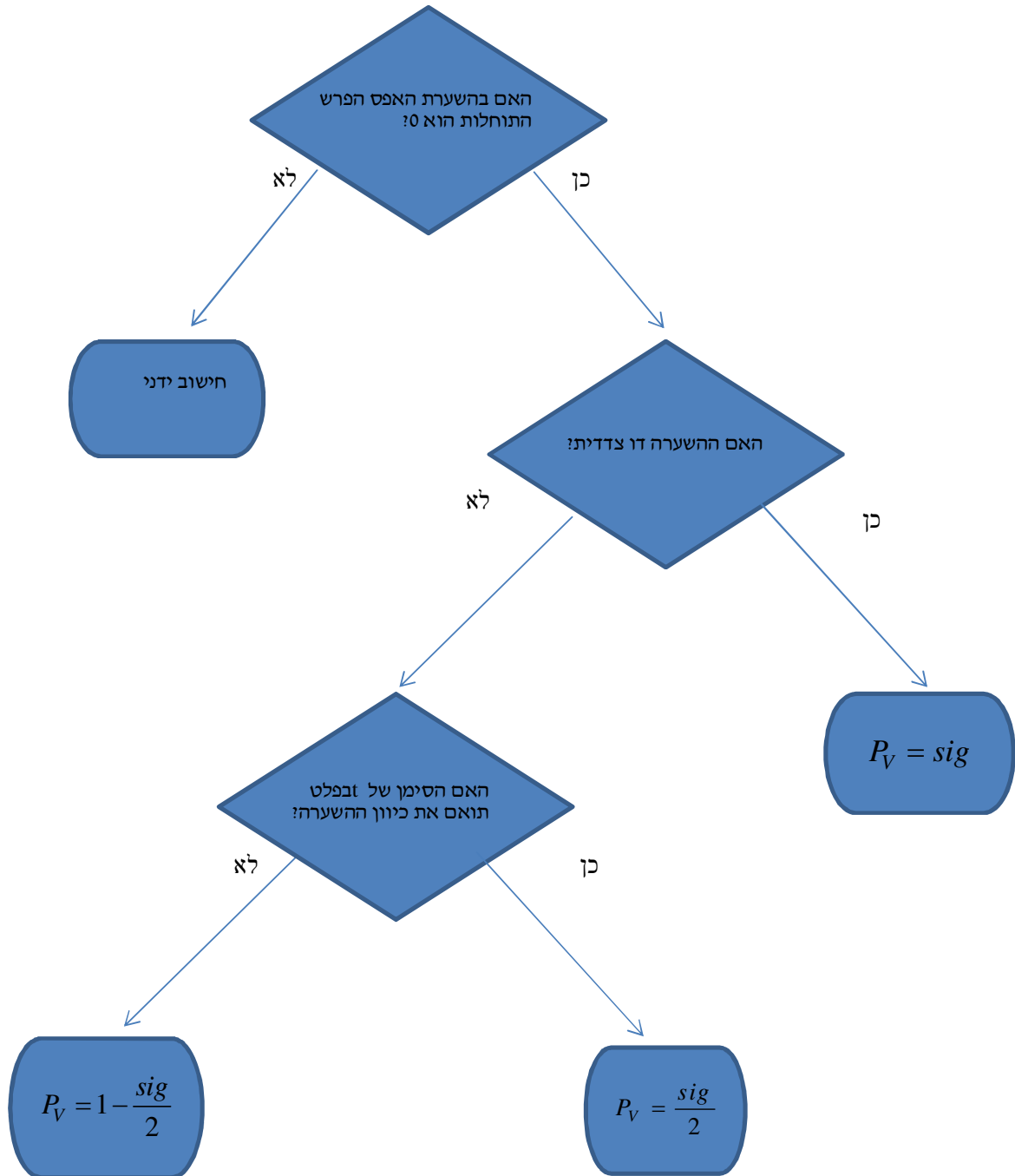
התשובה היא ד.

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:

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

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

- א. השלימו את החלקים החסרים בפלט. (מסומנים בסימני שאלה).
 ב. בדוק את הטענה שהשופט הדרום קוריאני נותן בממוצע 0.2 נקודות יותר מאשר השופט האיטלקי ברמת מובהקות של 5%.
 ג. מהו רווח הסמך ברמת סמך של 95% לתוחלת פער הציונים בין השופטים.
 ד. בנה את הרווח כעת ברמת סמך של 98% לתוחלת פער בציונים בין השופטים.

3. בדקו את ציוניהם של 44 נבדקים אקראיים במבחן הפסיכומטרי. פעם אחת לפני הכנה (Before) ופעם אחת אחרי הכנה (After).

Paired Samples Test

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

- א. רשום מהו המבחן הסטטיסטי ונסח את ההשערות אליהם מתייחס הפלט.
- ב. בדוק את ההשערה שממוצע ציונים משתפרים לאחר ההכנה ברמת מובהקות של 5%.
- ג. בדוק את ההשערה שממוצע ציונים משתפרים לאחר ההכנה ביותר מ-5 נקודות ברמת מובהקות של 5%.
- ד. מצא רווח סמך לתוחלת שיפור ממוצע הציונים לאחר ההכנה ברמת ביטחון של 95%.

T-Test

Paired Samples Statistics

	Mean	N	Std. Deviation	Std. Error Mean
Pair 1 x	54.0000	6	5.86515	2.39444
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 10 - הקשר בין רווח סמך לבדיקת השערות על הפרש תוחלות

רקע:

ניתן לבצע בדיקת השערות דו צדדית ברמת מובהקות α 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	מרצה X
64	81	68	מרצה Y

א. חשבו רווח סמך ברמת סמך 90% להפרש בין התוחלות של הציונים אצל שני המרצים.
 ב. האם ברמת מובהקות של 10% נכריע שיש הבדל בין תוחלות הציונים אצל שני המרצים?

שאלות אמריקאיות:

3. סטטיסטיקאי נתבקש לאמוד את הפרש הממוצעים של שני טיפולים לפי שני מדגמים מקריים בלתי תלויים.
 הוא חישב רווח סמך להפרש ברמת סמך 0.98, וקיבל את הרווח $-2 < \mu_1 - \mu_2 < 4.5$.
 אילו יתבקש החוקר לבדוק לפי אותם נתונים את ההשערות:
 $H_0: \mu_1 - \mu_2 = 0$; $H_1: \mu_1 - \mu_2 \neq 0$, ברמת מובהקות 0.05 מסקנתו תהיה:
 א. לדחות את השערת האפס.
 ב. לא לדחות את השערת האפס.
 ג. שלא ניתן לדעת את המסקנה עבור רמת מובהקות 0.05.
 ד. שלא נתונות בשאלה סטיות התקן של האוכלוסיות, ולכן לא ניתן להסיק דבר.

4. במטרה לבדוק האם קיים הבדל בין קווי זהב לבזק מבחינת ממוצע המחירים לשיחות בינ"ל. נגדמו באקראי 7 מדינות ועבור כל מדינה נבדקה עלות דקת שיחה. בהנחה והמחירים מתפלים נורמלית בנו רווח סמך לממוצע ההפרשים וקיבלו :
- $$-0.0293 < \mu_D < 0.2145$$
- רווח הסמך הוא ברמת סמך של 95%.
- לכן מסקנת המחקר היא :
- ברמת מובהקות של 5% לא נוכל לקבוע שקיים הבדל בין החברות.
 - ברמת מובהקות של 5% נקבע שקיים הבדל מובהק בין החברות.
 - לא ניתן לדעת מה המסקנה ברמת מובהקות של 5% כיוון שלא נאמר מה ההגדרה של D.

פתרונות:**שאלה 1:**

א. $-3.8 \leq \mu_D \leq 19$

ב. נכריע שיש אמת בפרסום.

שאלה 2:

א. $-8.5 \leq \mu_X - \mu_Y \leq 26.5$

ב. נכריע שאין הבדל.

שאלה 3:

התשובה היא ג.

שאלה 4:

התשובה היא א.

פרק 11 - תרגול מסכם ברווחי סמך

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

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

א. תנו רווח סמך 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 מדגם של שכר הדירה בשי"ח של 5 דירות שלושה חדרים בשכונת בבלי בתל אביב:

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

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

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

פרק 12- שאלות מסכמות בבדיקת השערות על פרמטרים

שאלות מסכמות בבדיקת השערות על פרמטרים

1. 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.

2. 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%.

3. 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?

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

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

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?

5. 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. לא נדחה H_0 ב. לא נדחה H_0 **שאלה 2:**א. לא נדחה H_0

ב. בין 5% ל- 10%

ג. נדחה H_0 **שאלה 3:**ב. נדחה H_0

ג. טעות מסוג ראשון בסיכוי של 0.05

שאלה 4:א. נדחה H_0 ב. נדחה H_0 **שאלה 5:**

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 $\alpha=0.05$?

- א. השערת האפס הייתה נדחית.
- ב. השערת האפס לא הייתה נדחית.
- ג. ההשערה המחקרית הייתה נדחית.
- ד. 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?

א. $H_1 : p = 0.5$

ב. $H_1 : p = 0.6$

ג. $H_1 : p > 0.5$

ד. $H_1 : p \neq 0.5$

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. ניח שציוני המבחן מתפלגים נורמאלית ללא ידיעת השונות האמיתית.

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.

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	X- לפני הנישואין
71	84	88	80	Y- לאחר הנישואין

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 מההנחות 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 (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.

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 השערות, ובכמה דרגות חופש:

- ההתפלגות Z ללא דרגות חופש.
 - ההתפלגות T ו-3 דרגות חופש.
 - ההתפלגות T ו-6 דרגות חופש.
 - ההתפלגות χ^2 ו-3 דרגות חופש.
9. שני סטטיסטיקאים בודקים השערות ברמת מובהקות $\alpha = 0.05$, על סמך אותו מדגם.

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

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

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

מה יחליט סטטיסטיקאי ב'?

- לדחות את השערת האפס.
- לא לדחות את השערת האפס.
- ללא נתונים נוספים אי אפשר לדעת מה יחליט.

10. חוקר בדק השערה מסוימת והחליט לדחות את השערת האפס ברמת מובהקות 5%. מה נכון לומר?
- הוא בוודאות ידחה את השערת האפס ברמת מובהקות 9% ואילו ברמת מובהקות 2% יש לבדוק מחדש.
 - הוא בוודאות לא ידחה את השערת האפס ברמת מובהקות 9% ואילו ברמת מובהקות 2% יש לבדוק מחדש.
 - הוא בוודאות ידחה את השערת האפס ברמת מובהקות 9% וברמת מובהקות 2%.
 - 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 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 זו צדדית ואילו חוקר ב' ניסח השערה חד צדדית. מסקנתו של איזה מבין המשפטים הבאים הוא הנכון בנוגע למסקנות החוקרים?

א. אם חוקר א' ידחה את השערת האפס לא ניתן לדעת מה יחליט חוקר ב' באותה רמת מובהקות.

ב. אם חוקר א' יקבל את השערת האפס גם חוקר ב' יקבל את השערת האפס באותה רמת מובהקות.

ג. אם חוקר ב' ידחה את השערת האפס גם חוקר א' ידחה את השערת האפס באותה רמת מובהקות.

ד. אם חוקר א' ידחה את השערת האפס גם חוקר ב' ידחה את השערת האפס בתנאי שרמת המובהקות כפולה בגודלה.

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?
- תוחלת ציוני העולים באוכלוסייה.
 - ממוצע ציוני העולים במדגם.
 - תוחלת ציוני האוכלוסייה מנתוני העבר.
 - ממוצע ציוני שאר האוכלוסייה במדגם.
15. חוקר ביצע מחקר וידוע כי עשה טעות מסוג 1. מה מהבאים נכון?
- החוקר דחה את השערת H_0 כאשר היא הייתה נכונה.
 - החוקר דחה את השערת H_1 כאשר היא הייתה נכונה.
 - החוקר לא דחה את השערת H_0 כאשר היא הייתה לא נכונה.
 - 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?
- מבחן t למדגמים בלתי תלויים עם 38 דרגות חופש.
 - מבחן t למדגמים מזווגים, עם 39 דרגות חופש.
 - מבחן t למדגמים בלתי תלויים עם 39 דרגות חופש.
 - 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?
- מבחן Z על פרופורציה.
 - מבחן t על תוחלת אחת.
 - מבחן t על שתי תוחלות במדגמים בלתי תלויים.
 - 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:
- שלושתם במבחן t למדגמים בלתי תלויים.
 - עידו במבחן t למדגמים מזווגים, ורון ויואב במבחן t למדגמים בלתי תלויים.
 - עידו במבחן t למדגמים מזווגים, רון במבחן t למדגמים בלתי תלויים ויואב במבחן t למדגם יחיד.
 - עידו במבחן t למדגמים מזווגים, רון ויואב במבחן t למדגם יחיד.

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

- 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.

20. חוקר ערך מבחן דו צדדי ברמת מובהקות של α והחליט לדחות את השערת האפס.

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

- השערת האפס הייתה נדחית.
- השערת האפס הייתה לא נדחית.
- 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 השתמשה בטבלה של התפלגות t .

מה נוכל לומר בנוגע להחלטת המחקר שלהם?

- אם ליאור ידחה את השערת האפס אז גם בהכרח רוני.
- 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.
 22. 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. σ^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 :

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

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 α , והן $(1 - \beta)$, יקטנו.

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

ג. α יגדל ואילו $(1 - \beta)$ יקטן.

ד. הן α 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. σ^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
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:

$$H_0: \mu_1 - \mu_2 = 0 ; H_1: \mu_1 - \mu_2 \neq 0, \text{ ברמת מובהקות } 0.05 \text{ מסקנתו תהיה:}$$

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:

$$-0.0293 < \mu_D < 0.2145 \text{ רווח הסמך הוא ברמת סמך של } 95\%.$$

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 פורסם 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 (β).
- 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 α . במידה וחוקר זה היה בודק השערה זו

ברמת מובהקות של 2α 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 σ^2 . רמת המובהקות היא 5%. נמצא שהעוצמה היא 0.9. להלן 3 טענות:

(1) עבור מדגם בגודל n וברמת מובהקות 5% לבדיקת ההשערות: $H_0 : \mu = 55$
 $H_1 : \mu = 60$ העוצמה

תהיה גדולה מ- 0.9.

(2) עבור מדגם בגודל $2n$ ורמת מובהקות 5% לבדיקת ההשערות: $H_0 : \mu = 55$
 $H_1 : \mu = 65$ העוצמה

תהיה גדולה מ- 0.9.

(3) עבור מדגם בגודל n ורמת מובהקות 10% לבדיקת ההשערות: $H_0 : \mu = 55$
 $H_1 : \mu = 65$ העוצמה

תהיה קטנה מ- 0.9.

א. שלושת הטענות אינן נכונות.

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

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

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

פתרונות :

א	26	א	1
ב	27	ד	2
א	28	ד	3
א	29	א	4
ג	30	ג	5
ד	31	ג	6
א	32	א	7
א	33	ב	8
ד	34	ג	9
ג	35	א	10
א	36	א	11
ג	37	ג	12
א	38	א	13
ג	39	א	14
א	40	א	15
א	41	א	16
א	42	א	17
ה	43	ג	18
ד	44	א	19
ד	45	ג	20
א	46	ב	21
ד	47	ג	22
ג	48	ד	23
ב	49	ג	24
ד	50	ג	25

פרק 13 - SPSS ניתוח פלטים

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

ד. תן אומדן להפרש הממוצעים.

ה. תן אומדן לטעות התקן של הפרש הממוצעים.

ו. האם קיים הבדל מובהק בין השכלת האבות להשכלת האמהות 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	???	???

- ה. השלימו את החלקים החסרים בפלט. (מסומנים בסימני שאלה).
- ו. בדוק את הטענה שהשופט הדרום קוריאני נותן בממוצע 0.2 נקודות יותר מאשר השופט האיטלקי ברמת מובהקות של 5%.
- ז. מהו רווח הסמך ברמת סמך של 95% לתוחלת פער הציונים בין השופטים.
- ח. 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

- ה. רשום מהו המבחן הסטטיסטי ונסח את ההשערות אליהם מתייחס הפלט.
- ו. בדוק את ההשערה שממוצע ציונים משתפרים לאחר ההכנה ברמת מובהקות של 5%.
- ז. בדוק את ההשערה שממוצע ציונים משתפרים לאחר ההכנה ביותר מ-5 נקודות ברמת מובהקות של 5%.
- ח. מצא רווח סמך לתוחלת שיפור ממוצע הציונים לאחר ההכנה ברמת ביטחון של 95%.

4. 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

- ד. מהו המבחן הסטטיסטי שנעשה כאן?
- ה. בדוק ברמת מובהקות של 5% האם קיים שוויון שונויות בין שני האזורים?
- ו. בדוק האם קיים הבדל בין "South East" ל-"North East" ברמת מובהקות של 5% מבחינת מספר האחים והאחיות הממוצע.
- ז. בדוק את הטענה שהפרש הממוצע בין-"South East" לבין-"North East" הינו 0.1 מבחינת מספר האחים והאחיות. ברמת מובהקות של 5%.

5. להלן פלט מתוכנת SPSS מתוך מחקר שבחן את רמת האופטימיות של גברים ונשים.
רמת האופטימיות נמדדה בסקאלת ציונים של 1 עד 5.

Group Statistics

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

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
optimizm	Equal variances assumed	.383	.536	1.935	1199	.053	.05500	.02842	-.00076	?
	Equal variances not assumed			1.938	190.977	.053	.05500	.02838	-.00068	.11067

- ד. האם ניתן להניח ששוונות האופטימיות של נשים וגברים שווה ברמת מובהקות של 5%?
- ה. ברמת מובהקות של 5% האם קיים הבדל בין הנשים לגברים ברמת האופטימיות הממוצעת שלהם?
- ו. מצא את הגבול העליון של רווח הסמך המסומן בסימן שאלה בפלט. דייק עד 5 ספרות אחרי הנקודה.
- ז. בנה רווח סמך לתוחלת רמת האופטימיות של הגברים ברמת סמך של 95%.

בסקר שנעשה על ידי משרד ראש הממשלה נדגמו 60 אזרחים. כל אזרח נשאל על מינו והאם הוא בעד הקמת מדינה פלשתינאית. מצורף הפלט .

א. הסבירו מהי שאלת המחקר בפלט זה.

ב. הגדירו את המשתנה הבלתי תלוי ואת המשתנה התלוי.

ג. רשמו במונחי השאלה את ההנחות והתנאים הדרושים לביצוע המבחן הסטטיסטי הבודק את ההשפעה של המשתנה הבלתי תלוי על התלוי באוכלוסייה.

ד. האם ניתן להסיק כי באוכלוסייה יש השפעה של המשתנה הבי"ת על התלוי ? בדקו ברמת מובהקות $\alpha = 0.05$. (נסחו השערות, ציינו כלל החלטה ונסחו מסקנה במונחי המשתנים)

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
gender * opinion	60	100.0%	0	.0%	60	100.0%

gender * opinion Crosstabulation

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

Chi-Square Tests

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

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 7.25.

7. במחקר מסוים רצו לבדוק האם יש קשר בין המגדר להוצאה על לבוש במשך שנה. דגמו באופן מקרי גברים ונשים ובדקו את רמת ההוצאה שלהם על לבוש בשנה האחרונה. חוקר א' בדק האם קיים הבדל בתוחלות ההוצאה בין גברים לנשים. חוקר ב' קיבץ את ההוצאה לקטגוריות ובאופן הזה בדק האם קיים הבדל בהתפלגות ההוצאה בין גברים לנשים. הקטגוריות חולקו לשלוש קבוצות הוצאה .
- א. איזה פלט מתאים לאיזה אחד מהחוקרים? נמקו.
- ב. מה מסקנתו של חוקר א'? בדקו ברמת מובהקות $\alpha = 0.05$. (רשמו השערות, נסחו הנחות, ציינו כלל החלטה ותנו מסקנה במונחי המשתנים)
- ג. איזו טעות יכולה להיות במסקנתו של חוקר א'? נסחו את הטעות במונחי השאלה.
- ד. מהי מסקנתו של חוקר ב'? בדקו ברמת מובהקות $\alpha = 0.05$. (רשמו השערות, נסחו הנחות, ציינו כלל החלטה ורשמו מסקנה במונחי המשתנים)
- ה. איזו טעות יכולה להיות במסקנתו של חוקר ב'? נסחו זאת במונחי השאלה
- ו. כיצד ניתן ליישב את מסקנות שני החוקרים?

פלט ראשון :

T-Test**Group Statistics**

gender	N	Mean	Std. Deviation	Std. Error Mean
female	40	2.9000	1.15025	.18187
male	40	2.6000	2.52982	.40000

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
expose	Equal variances assumed	16.805	.000	.683	78	.497	.30000	.43941	-.57479	1.17479
	Equal variances not assumed			.683	54.464	.498	.30000	.43941	-.58078	1.18078

פלט שני:

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
gender * category	80	100.0%	0	.0%	80	100.0%

gender * category Crosstabulation

			category			Total
			a	b	c	
gender	female	Count	2	30	8	40
		Expected Count	11.0	21.0	8.0	40.0
		% within gender	5.0%	75.0%	20.0%	100.0%
		% within category	9.1%	71.4%	50.0%	50.0%
male	male	Count	20	12	8	40
		Expected Count	11.0	21.0	8.0	40.0
		% within gender	50.0%	30.0%	20.0%	100.0%
		% within category	90.9%	28.6%	50.0%	50.0%
Total	Total	Count	22	42	16	80
		Expected Count	22.0	42.0	16.0	80.0
		% within gender	27.5%	52.5%	20.0%	100.0%
		% within category	100.0%	100.0%	100.0%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	22.442 ^a	2	.000
Likelihood Ratio	25.064	2	.000
N of Valid Cases	80		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 8.00.

13. במחקר בדקו את זמני ההמתנה למוקד טלפוני. מטרת המחקר הייתה לבדוק האם תוחלת זמני ההמתנה השתפרו כתוצאה משיפור תהליך העבודה בחברה. להלן פלט שנשעה על הנתונים:

Normality Kolmogorov-Smirnov Test

	Minutes to Respond
N	1963
Kolmogorov-Smirnov Z	5.776
Asymp. Sig. (2-tailed)	.000

א. איך קוראים למבחן הסטטיסטי שנעשה פה?

ב. מה מבחן זה בודק?

ג. מה המסקנה?

14. במחקר השתתפו אנשים שנשאלו האם הם מעשנים והאם הם חלו בחורף האחרון בשפעת. מטרת המחקר היא לבדוק האם קיים קשר בין עישון לבין קיומה של מחלת השפעת.
- א. מה המשתנה התלוי ומהו המשתנה הבלתי תלוי?
- ב. רשמו את השערות המחקר ובדקו האם הפלט מקיים את האילוצים הנדרשים.
- ג. השלימו את כל סימני השאלה שבפלט (בסך הכול 5)

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
smoking * flu	30	100.0%	0	.0%	30	100.0%

smoking * flu Crosstabulation

			flu		Total
			yes	no	
smoking	smoking	Count	11	9	20
		Expected Count	?	?	20.0
not smoking	not smoking	Count	3	7	10
		Expected Count	?	?	10.0
Total		Count	14	16	30
		Expected Count	14.0	16.0	30.0

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	?	1	.196
Continuity Correction ^b	.820	1	.365
Likelihood Ratio	1.713	1	.191
Fisher's Exact Test			
Linear-by-Linear Association	1.618	1	.203
N of Valid Cases	30		

a. 1 cells (25.0%) have expected count less than 5. The minimum expected count is 4.67.

b. Computed only for a 2x2 table

תשובות סופיות ל-SPSS

<p><u>שאלה 2</u></p> <p>א. 25</p> <p>ב. (4.3056, 4.5104)</p> <p>ד. נדחה H_0</p>	<p><u>שאלה 1</u></p> <p>ב. 23</p> <p>ג. נקבל H_0</p> <p>ד. (22.104, 24.771)</p>
<p><u>שאלה 4</u></p> <p>ב. נקבל H_0</p> <p>ג. נקבל H_0</p> <p>ד. נדחה H_0</p>	<p><u>שאלה 3</u></p> <p>א. $n = 44$</p> <p>$\bar{X} = 67.66$</p> <p>$t = -2.48$</p> <p>ב. 0.017</p> <p>ג. 0.0085</p> <p>ד. 0.9915</p>
<p><u>שאלה 6</u></p> <p>א. (-0.007)</p> <p>ב. (0.1)</p> <p>ג. אין הבדל מובהק</p>	<p><u>שאלה 5</u></p> <p>ב. 2.959</p> <p>ג. 73</p> <p>ד. 2.017</p>
<p><u>שאלה 8</u></p> <p>ב. נדחה H_0</p> <p>ג. נקבל H_0</p> <p>ד. (1.592, 13.317)</p>	<p><u>שאלה 7</u></p> <p>א. $d.f = 299$</p> <p>$n = 300$</p> <p>$\bar{X} = 8.496$</p> <p>Sig=0</p> <p>ב. נדחה H_0</p> <p>ג. (-0.46341, -0.38126)</p> <p>ד. (-0.4708, -0.3738)</p>
<p><u>שאלה 10</u></p> <p>א. נקבל H_0</p> <p>ב. נקבל H_0</p> <p>ג. (0.11076)</p> <p>ד. 2.6053 ± 0.0395</p>	<p><u>שאלה 9</u></p> <p>ב. נקבל H_0</p> <p>ג. נקבל H_0</p> <p>ד. נקבל H_0</p>
<p><u>שאלה 12</u></p> <p>ב. נקבל H_0</p> <p>ד. נדחה H_0</p>	<p><u>שאלה 11</u></p> <p>ד. לא נקבל את הטענה שקיים קשר בין המין לדעה.</p>

שאלה 14

$$\chi^2 = 1.74.g$$

שאלה 13

H_0 נדחה

פרק 14- ניתוח שונות חד גורמי

1. מתי משתמשים בניתוח שונות חד גורמי?

בניתוח שונות חד גורמי משתמשים כאשר רוצים לבדוק האם קיים הבדל בין מספר קבוצות במשתנה מסוים.

דוגמא מס' 1: חוקר רצה לבחון השפעתן של שלוש שיטות טיפול פיזיותרפיה שונות על רמת ההתפתחות המוטורית של ילדים קטנים (ציוני ההתפתחות המוטורית נעים בין 1 "נמוכה מאד" ל- 10 "גבוהה מאד").

1.1 זיהוי המודל המחקרי

זיהוי המודל המחקרי יתבצע על ידי הגדרת המשתנים: התלוי והבלתי-תלוי

שאלה מס' 1 (מבוססת על דוגמא מס' 1)

הגדר את המשתנים וזהה את המודל המחקרי

תשובה:

2. השערות

המשך דוגמא מס' 1: לצורך הבדיקה נבדקו 14 ילדים בעלי רמת התפתחות זהה והם חולקו באקראי ל- 3 קבוצות. לאחר 8 חודשי טיפול עבר כל אחד מהילדים הערכה של רמת ההתפתחות המוטורית שלו. להלן התוצאות שהתקבלו:

	שיטה ג'	שיטה ב'	שיטה א'	
	2	5	8	
	4	4	5	
	3	7	4	
	2	7	7	
	4		6	
$n = 14$	$n_3 = 5$	$n_2 = 4$	$n_1 = 5$	n_j
$\bar{x} = 4.86$	$\bar{x}_3 = 3$	$\bar{x}_2 = 5.75$	$\bar{x}_1 = 6$	\bar{x}_j
$\hat{s}_T = 1.92$	$\hat{s}_3 = 1$	$\hat{s}_2 = 1.5$	$\hat{s}_1 = 1.58$	\hat{s}_j

השערות:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_K$$

$H_1: otherwise$

אפקט הטיפול = Δ = הבדל בין התוחלות באוכלוסיה

3. אמידת גודל האפקט

כיצד נסיק על פי ההבדל בין הממוצעים במדגם (\bar{x}_j) שקיים הבדל בין התוחלות באוכלוסיה (μ_j)?

לפתרון מלא בסרטון וידאו היכנסו ל- www.Gool.co.il

כתב ופתר - ברק קנדל ©

התשובה: באמצעות ניתוח השונות (סטטיסטי המבחן F):

$$F = \frac{MSB}{MSW}$$

$$E(MSB) = \Delta + \sigma_e^2$$

$$E(MSW) = \sigma_e^2$$

בכדי להבין את הרציונאל עליו נשען סטטיסטי המבחן F נעזר בדוגמא ההיפותטית הבאה:

נתונות שתי אוכלוסיות שבשתייהן $\Delta = 0$

אוכ' מס' 1: $\sigma_j^2 = 0$ אוכ' מס' 2: $\sigma_j^2 \neq 0$

שיטה	שיטה	שיטה	שיטה	שיטה	שיטה
א'	ב'	ג'	א'	ב'	ג'
1	2	2	6	6	6
3	5	4	6	6	6
6	6	5	6	6	6
6	6	7	6	6	6
10	7	8	6	6	6
8	10	10	6	6	6
$\mu_1 = 6$	$\mu_2 = 6$	$\mu_3 = 6$	$\mu_3 = 6$	$\mu_2 = 6$	$\mu_1 = 6$

$$\bar{x}_j \neq \mu_j$$

$$\bar{x}_j = \mu_j$$

שאלה מס' 2

חוה/י דעתך על הטענות הבאות:

א. MSB הוא אומד חסר הטיה ל:

1. שונות הטעויות.

תשובה:

2. לגודל האפקט

תשובה:

ב. במבחן F הבודק הבדלים בין תוחלות, ה-MSB וה-MSW הינם משתנים מקריים בעלי תוחלת E(MSB) ו-E(MSW) בהתאמה. במידה ולא קיים הבדל בין התוחלות של המשתנה התלוי, ציין מה מבין שלהלן מתקיים ונמק תשובתך:

a. $E(MSW) = E(MSB)$

b. MSB בהכרח שווה או קטן מ-MSW

c. MSB בהכרח שווה ל-MSW

d. $E(MSB) < E(MSW)$

e. אף אחת מהתשובות הנ"ל

תשובה:

4. ביצוע המבחן הסטטיסטי

4.1 חישוב סטטיסטי המבחן F

את סטטיסטי המבחן $F = \frac{MSB}{MSW}$ נחשב באמצעות טבלת מקור שונות (ANOVA):

מקור השונות	df (degrees of freedom) דרגות חופש	SS (sum of squares) סכומי ריבועים	MS (Mean of squares) שונויות	F
B	C-1	$SSB = \sum_{j=1}^c nj(\bar{x}_j - \bar{x})^2$	MSB=SSB/C-1	$F = \frac{MSB}{MSW}$
W	n-C	$SSW = \sum_{j=1}^c \hat{S}_j^2(n_j - 1)$	MSW=SSW/n-C	
T	n-1	$SST = (N - 1)\hat{S}_T^2$ $= SSB + SSW$	MST=SST/n-1 $= \hat{S}_T^2$ <u>הערה חשובה:</u> MST \neq MSB+MSW	

4.2 כלל הכרעה

דוחים את H0 אם: $F > F_{\alpha}(c-1, n-c)$

4.3 מסקנה

יש/אין עדות להבדל במשתנה התלוי בין קבוצות המשתנה הב"ת באוכלוסייה, ברמת מובהקות של אלפא

שאלה מס' 3 (מבוססת על דוגמא מס' 1)

האם שיטת הטיפול משפיעה על רמת ההתפתחות המוטורית של הילדים? בדוק ברמת מובהקות של 0.05.

תשובה:

5. הנחות המודל

5.1 רישום ההנחות במונחי השאלה

דגימה מקרית של האובייקט שנדגם.
 מדגמים בלתי תלויים.
 התפלגות נורמאלית של המשתנה התלוי בתוך כל אוכלוסיית המשתנה ה"ב"ת.
 שונות המשתנה התלוי שווה בכל אוכלוסיות המשתנה ה"ב"ת.
 מודל קבוע.

5.2 הנחת שוויון שונות

הנחת שוויון שונות גורסת כי כל השונות של האוכלוסיות שוות אחת לשנייה ולשונות

$$\text{הטעויות } (\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \dots = \sigma_k^2 = \sigma_e^2).$$

תחת הנחת שוויון שונות :

(1) כל אחת מן האומדים לשונות מתוך נתוני המדגם $(\hat{\sigma}_j^2)$ וה-MSW (המהווה ממוצע

$$\text{משוקלל של האומדים לשונות של כל קבוצה } (MSW = \frac{\sum \hat{\sigma}_j^2 \cdot (n_j - 1)}{n - c}) \text{ אומדים את כל}$$

אחת משונות האוכלוסייה (σ_j^2) וגם את השונות הטעויות (σ_e^2) .

(2) בכדי לאמוד כל אחת מן השונות באוכלוסייה (σ_j^2) כמו גם את שונות הטעויות (σ_e^2) ,

נעדיף להשתמש באומד לשונות MSW במקום באומדים לשונות $\hat{\sigma}_j^2$.

הסיבה להעדפה זו היא שחישוב ה-MSW מבוסס על כל תצפיות המדגם ואילו חישוב

ה- $\hat{\sigma}_j^2$ מבוסס רק על מס' התצפיות של הקבוצה הספציפית. ככל שהאומד לשונות

מבוסס על יותר תצפיות כך הוא יהיה יעיל ומדויק יותר כאומדן לפרמטר.

(3) חישוב ה-MSW (האומד לשונות הטעויות (σ_e^2) נשען על הנחת שוויון שונות.

שאלה מס' 4

חוה/י דעתך על הטענות הבאות:

א. תחת הנחות המודל כל אחת משונויות המדגמים היא אומדן חסר הטיה ל- σ_e^2 .

תשובה:

ב. הנחת שוויון שונויות משמשת להערכת גודל האפקט.

תשובה:

ג. תחת הנחות המודל, בכדי לאמוד את השונות של אחת האוכלוסיות נעדיף להשתמש ב-

MSW במקום באומד לשונות הספציפי של אותה אוכלוסייה (\hat{s}_j^2).

תשובה:

6. ניתוח פלטים

6.1 ביצוע המבחן הסטטיסטי על סמך פלט ה-Descriptives

דוגמא מס' 2:

במחקר מסוים רצו לבדוק האם קיימים הבדלים בהישגים בבחינה בסטטיסטיקה 3 של הסטודנטים במכללה כפונקציה של שיטות למידה שונות. לשם כך נדגמו באופן מקרי מהמכללה מס' תלמידים מכל אחת משיטות הלימוד ולהלן הנתונים שהתקבלו:

GRADES = ציונים בבחינה

TRAINING METHOD = שיטת הלימוד

שאלה מס' 5

בדוק את שאלת המחקר ברמת מובהקות של 0.05, לפי השלבים הבאים:
א. הצב השערות

ב. רשום את ההנחות הדרושות למודל במונחי השאלה המחקרית

ג. הצג את הטבלה המקובלת לניתוח הנתונים וסכם בה את הממצאים

ד. נסח מסקנה מילולית

Descriptives

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
					1 alone	20.00		
2 class	20.00	73.50	10.51	2.35	68.58	78.42	48.00	90.00
3 small groups	20.00	78.45	6.35	1.42	75.48	81.42	65.00	90.00
Total	60.00	71.88	12.06	1.56	68.77	75.00	33.00	90.00

תשובה:

6.2 בדיקת ההנחות על סמך הפלטים

שתי ההנחות הניתנות לבדיקה באמצעות הפלטים הן הנחת שוויון שונויות והנחת הנורמאליות. נדגים את בדיקת ההנחות ברמת מובהקות של 0.05 על נתוני דוגמא מס' 2.

6.2.1 בדיקת הנחת שוויון שונויות

הפלט המתאים לבדיקת הנחת שוויון שונויות הוא הפלט הבא:

Test of Homogeneity of Variances

GRADES Score on training exam

Levene Statistic	df1	df2	Sig.
2.604	2	57	.083

השערות:

$$H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2$$

$H_1: \text{otherwise}$

כלל הכרעה (על סמך ה- sig שבפלט):

$sig = 0.083 > \alpha = 0.05$ לכן אין סיבה מספקת לדחות את H_0 ברמת מובהקות של 0.05

מסקנה:

ניתן להניח שוויון שונות של הציונים בבחינה בכל אוכלוסיית שיטות הלימוד ברמת מובהקות של 0.05.

6.2.2 בדיקת הנחת הנורמאליות

הפלט המתאים לבדיקת הנחת הנורמאליות הוא הפלט הבא:

Tests of Normality

GROUP Sales training method	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
GRADES Score 1 alone	.147	20	.200*	.956	20	.470
on training exam 2 class	.131	20	.200*	.954	20	.433
3 small groups	.156	20	.200*	.944	20	.281

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

****יש להתייחס רק למבחן Kolmogorov-Smirnov**

השערות:

עבור כל אחת משלושת אוכלוסיות שיטות הלימוד יש להניח:

H_0 : התפלגות נורמאלית

H_1 : התפלגות אחרת

כלל הכרעה (על סמך ה- sig שבפלט):

עבור כל אחת משלושת אוכלוסיות שיטות הלימוד:

$sig = 0.2 > \alpha = 0.05$ ולכן אין סיבה מספקת לדחות את H_0 ברמת מובהקות של 0.05

מסקנה:

ניתן להניח התפלגות נורמאלית של הציונים בבחינה עבור כל אחת מאוכלוסיות שיטות הלימוד ברמת מובהקות של 0.05.

****הערה:** מספיק שנדחה את H_0 עבור אחת האוכלוסיות בכדי שהנחת הנורמאליות תופר.

שאלה נוספת (מבוססת על דוגמא מס' 2)

סטודנט שעדיין בפלטים הציע 2 אומדנים חסרי הטיה לשונות ציוני הבחינה בקבוצת הלומדים לבד. הערך האחד הוא: 13.5^2 והערך השני הוא: 111.02. רשום מתי היית מעדיף את אומדן א', מתי את אומדן ב' ומדוע?

תשובה:

7. פרופורציית השונות המוסברת - η^2

מדד זה מהווה פרופורציה של השונות המוסברת מתוך כלל השונות:

$$0 \leq \eta^2 = \frac{SSB}{SST} \leq 1$$

פרופורציית השונות הלא מוסברת:

$$0 \leq 1 - \eta^2 = \frac{SSW}{SST} \leq 1$$

שאלה מס' 6 (מתייחסת לדוגמא מס' 2)

א. כמה אחוזים מן השונות של ציוני הבחינה מוסברים ע"י שיטות הלימוד?

תשובה:

ב. כמה אחוזים מן השונות של ציוני הבחינה אינם מוסברים ע"י שיטות הלימוד?

תשובה:

ג. מהי עוצמת הקשר בין הציונים בבחינה לבין שיטות הלימוד?

תשובה:

8. השוואות מרובות (POST-HOC TESTS)

תוצאה מובהקת בניתוח השוונות משמעה כאמור - לפחות אחד מן הממוצעים באוכלוסייה שונה מהאחרים.

בכדי לדעת איזה מן הממוצעים שונה מהאחרים יש לבצע מבחני המשך (post-hoc tests).

אנו נשווה כל זוג ממוצעים זה לזה.

מספר ההשוואות הזוגיות האפשריות עבור C קבוצות: $\frac{c \cdot (c-1)}{2}$

בדוגמא מס' 2:

$$\frac{3 \cdot 2}{2} = 3 \text{ מס' ההשוואות המרובות:}$$

8.1 השערות

**מתייחסות לדוגמא מס' 2

$$\begin{array}{lll} H_0: \mu_2 - \mu_3 = 0 & H_0: \mu_1 - \mu_3 = 0 & H_0: \mu_1 - \mu_2 = 0 \\ H_1: \mu_2 - \mu_3 \neq 0 & H_1: \mu_1 - \mu_3 \neq 0 & H_1: \mu_1 - \mu_2 \neq 0 \end{array}$$

8.2 השיטות להשוואות מרובות: TUKY ו-L.S.D

נלמדו שתי שיטות להשוואות מרובות:

(1) TUKY

(2) L.S.D

8.2.1 ההבדל בין השיטות

TUKY מחמירה יותר בדחיית H_0 מ-L.S.D.

למה?

בשיטת L.S.D קיימת בעיית תיסוף ה- α : ככל שנעשה יותר השוואות כך הטעות מסוג ראשון (α) של כל מערכת ההשוואות תתנפח.

$$\alpha_{PE} = 1 - (1 - \alpha_{PC})^N$$

שיטת TUKY מתגברת על בעיית תיסוף ה- α על ידי כך שקובעת את ה- α הכללית או הסופית של כל מערכת ההשוואות מראש. כתוצאה מכך ה- α עבור כל השוואה קטנה וקשה יותר לדחות את H_0 לעומת שיטת L.S.D.

משמעות ההבדל בין השיטות:

אם דחינו H_0 ב-TUKY \Leftarrow נדחה את H_0 גם ב-L.S.D.

אם קיבלנו את H_0 ב-L.S.D \Leftarrow נקבל את H_0 גם ב-TUKY.

אם קיבלנו את H_0 ב-TUKY \Leftarrow לא ניתן לדעת האם נקבל או נדחה את H_0 ב-L.S.D.

אם דחינו את H_0 ב-L.S.D \Leftarrow לא ניתן לדעת האם נקבל או נדחה את H_0 ב-TUKY.

לפתרון מלא בסרטון וידאו היכנסו ל- www.GooL.co.il

כתב ופתר - ברק קנדל ©

8.3 ניתוח פלטים

ישנם שני פלטים העוסקים בהשוואות מרובות:

(1) פלט ה- Multiple Comparisons

(2) פלט ה- Homogeneous Subsets

המשך דוגמא מס' 2...

1) Post Hoc Tests

Multiple Comparisons

Dependent Variable: GRADES Score on training exam

	(I) GROUP Sales training method	(J) GROUP Sales training method	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	1 alone	2 class	-9.80*	3.33	.01	-17.82	-1.78
		3 small groups	-14.75*	3.33	.00	-22.77	-6.73
	2 class	1 alone	9.80*	3.33	.01	1.78	17.82
		3 small groups	-4.95	3.33	.31	-12.97	3.07
	3 small groups	1 alone	14.75*	3.33	.00	6.73	22.77
		2 class	4.95	3.33	.31	-3.07	12.97
LSD	1 alone	2 class	-9.80*	3.33	.00	-16.47	-3.13
		3 small groups	-14.75*	3.33	.00	-21.42	-8.08
	2 class	1 alone	9.80*	3.33	.00	3.13	16.47
		3 small groups	-4.95	3.33	.14	-11.62	1.72
	3 small groups	1 alone	14.75*	3.33	.00	8.08	21.42
		2 class	4.95	3.33	.14	-1.72	11.62

*. The mean difference is significant at the .05 level.

שאלה מס' 7

א. בהסתמך על הפלט Multiple Comparisons - מהן שיטות הלימוד אשר קיימים

ביניהן הבדלים מובהקים בהישגים בבחינה?

תשובה:

ב. מדוע קיים הבדל ב- sig בהשוואה בין הקבוצה של הלומדים בקבוצה קטנה לבין

קבוצת הלומדים בכיתה בין השיטות TUKY ו-L.S.D?

תשובה:

ג. באיזה שיטה-TUKY או L.S.D הרב"ס לאותו הפרש בין התוחלות יהיה צר

יותר? נמק.

תשובה:

2) Homogeneous Subsets

GRADES Score on training exam

GROUP Sales training method	N	Subset for alpha = .05	
		1.00	2.00
Tukey HSD ^a 1 alone	20.00	63.70	
2 class	20.00		73.50
3 small groups	20.00		78.45
Sig.		1.00	.31

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 20.000.

ד. בהסתמך על הפלט Homogeneous Subsets - מהן שיטות הלימוד אשר

קיימים ביניהן הבדלים מובהקים על פי שיטת TUKY?

תשובה:

ה. מה צפויות להיות תוצאות הבדיקה לפי שיטת L.S.D? לגבי כל השוואה ציין מה

תהיה מסקנה. במידה ואינך יכול לתת תשובה חד משמעית נמק מדוע.

תשובה:

ו. מכון מחקר בדק את השאלה האם המקצוע הנלמד משפיע על המשכורת

ההתחלתית (באלפי שקלים). להלן הנתונים:

Descriptives

SALARY starting salary

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1.00 Agriculture	55	6.5911	1.75608	.23679	6.1164	7.0659	1.90	10.15
2.00 Business	61	7.5319	1.91268	.24489	7.0420	8.0217	3.48	12.59
3.00 Engineering	53	9.7779	2.33067	.32014	9.1355	10.4203	4.60	14.34
4.00 Communication	81	8.1221	2.55649	.28405	7.5568	8.6873	.74	15.02
Total	250	7.9923	2.44013	.15433	7.6883	8.2962	.74	15.02

בוצעו השוואות מרובות והתקבלו התוצאות הבאות:

Multiple Comparisons

Dependent Variable: SALARY starting salary

	(I) COLLEGE	(J) COLLEGE	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	1.00 Agriculture	2.00 Business	-.9408	.40920	.101	-1.9992	.1177
		3.00 Engineering	-3.1868*	.42359	.000	-4.2825	-2.0911
		4.00 Communication	-1.5309*	.38450	.001	-2.5255	-.5364
	2.00 Business	1.00 Agriculture	.9408	.40920	.101	-.1177	1.9992
		3.00 Engineering	-2.2460*	.41324	.000	-3.3150	-1.1771
		4.00 Communication	-.5902	.37307	.391	-1.5552	.3748
	3.00 Engineering	1.00 Agriculture	3.1868*	.42359	.000	2.0911	4.2825
		2.00 Business	2.2460*	.41324	.000	1.1771	3.3150
		4.00 Communication	1.6559*	.38880	.000	.6502	2.6616
	4.00 Communication	1.00 Agriculture	1.5309*	.38450	.001	.5364	2.5255
		2.00 Business	.5902	.37307	.391	-.3748	1.5552
		3.00 Engineering	-1.6559*	.38880	.000	-2.6616	-.6502
LSD	1.00 Agriculture	2.00 Business	-.9408*	.40920	.022	-1.7467	-.1348
		3.00 Engineering	-3.1868*	.42359	.000	-4.0211	-2.3525
		4.00 Communication	-1.5309*	.38450	.000	-2.2883	-.7736
	2.00 Business	1.00 Agriculture	.9408*	.40920	.022	.1348	1.7467
		3.00 Engineering	-2.2460*	.41324	.000	-3.0600	-1.4321
		4.00 Communication	-.5902	.37307	.115	-1.3250	.1446
	3.00 Engineering	1.00 Agriculture	3.1868*	.42359	.000	2.3525	4.0211
		2.00 Business	2.2460*	.41324	.000	1.4321	3.0600
		4.00 Communication	1.6559*	.38880	.000	.8901	2.4217
	4.00 Communication	1.00 Agriculture	1.5309*	.38450	.000	.7736	2.2883
		2.00 Business	.5902	.37307	.115	-.1446	1.3250
		3.00 Engineering	-1.6559*	.38880	.000	-2.4217	-.8901

*. The mean difference is significant at the .05 level.

שחזר את פלט ה- Homogeneous Subsets לפי שיטת TUKY ברמת מובהקות של

.0.05

תשובה:

ז. בהיסתמך על פלט ה- Homogeneous Subsets שבנית, אילו היית עורך

השוואות מרובות עבור רמת מובהקות של 0.1, בין אילו קבוצות היית מוצא הבדלים

מובהקים?

תשובה:

פרק 15- רגרסיה ליניארית חד משתנית

מטרת הרגרסיה הליניארית היא ניבוי משתנה מסוים (משתנה תלוי, המכונה "משתנה מנובא") באמצעות משתנים אחרים (משתנים ב"ת, המכונים "משתנים מנובאים"). כאשר יש לי משתנה מנובא/ב"ת אחד בלבד מדובר ברגרסיה חד משתנית ("פשוטה").

ניתוח מודל הרגרסיה מתחלק ל-3 שלבים:

(1) זיהוי צורת הקשר בין המשתנים כקשר ליניארי.

(2) בניית קו הרגרסיה במדגם.

לפתרון מלא בסרטון וידאו היכנסו ל- www.GooL.co.il

כתב ופתר - ברק קנדל ©

(3) בדיקת מובהקות קו הרגרסיה ומקדמיו באוכלוסייה.

מבוא - מקדם המתאם הליניארי

1.1 הגדרה

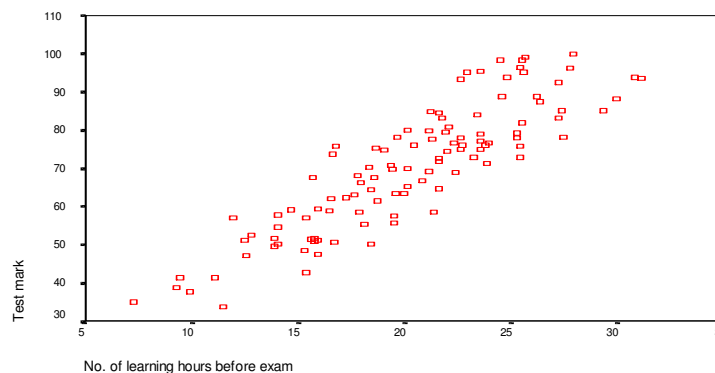
קשר בין שני משתנים כמותיים אשר ניתן להגדירו על פי נוסחת הקו הישר $(Y=a+bX)$.

1.2 דיאגרמת פיזור

דוגמא מס' 1: קשר בין שעות למידה בבחינה והציונים בבחינה:

ציונים	שעות למידה	
76	19	1
60	17	2
85	30	3
...
70	26	109

בכדי לאמוד את טיב הקשר בין המשתנים יש להציגו בדיאגרמת פיזור:



מדיאגרמת הפיזור הנ"ל נוכל להתרשם כי הקשר הוא ליניארי, חיובי וחזק.

1.3 תכונות מקדם המתאם הליניארי

$$-1 \leq r_p \leq 1$$

מקדם המתאם של פירסון מוגדר על ידי שני פרמטרים בלתי תלויים אחד בשני-כיוון ועוצמה: (1) כיוון הקשר

א. חיובי/עולה/ישר- שני המשתנים נעים באותו הכיוון. אם האחד עולה או יורד אז גם השני. מקדם המתאם (r) יקבל סימן +.

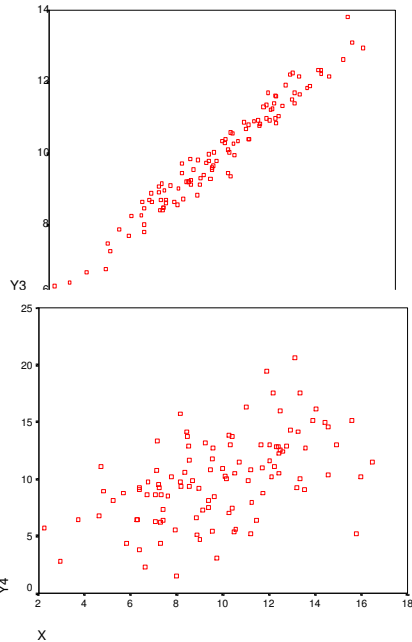
ב. שלילי/יורד/הפוך-שני המשתנים נעים בכיוונים הפוכים. אם האחד עולה השני יורד ולהפך. מקדם המתאם יקבל סימן (-).

2) עוצמת הקשר

ככל שעוצמת הקשר גבוהה יותר כך יהיה ערך מקדם המתאם קרוב יותר ל-1 בערך מוחלט. ככל שהעוצמה נמוכה יותר כל יהיה קרוב יותר ל-0 בערך מוחלט.
א. דוגמאות לקשר ליניארי חלקי:

ככל שהנקודות קרובות יותר לקו הישר כך עוצמת הקשר חזקה יותר ולהיפך:

Graph



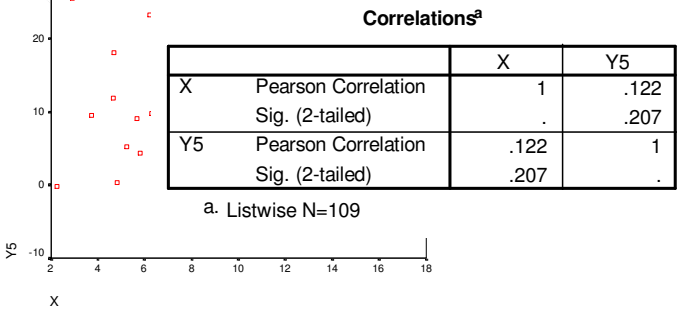
Correlations^a

	X	Y3
X	1	.979*
		.000
Y3	.979**	1
	.000	.

** Correlation is significant at the 0.01 level
a. Listwise N=109

Correlations^a

	X	Y4
Y4	1	.518**
		.000



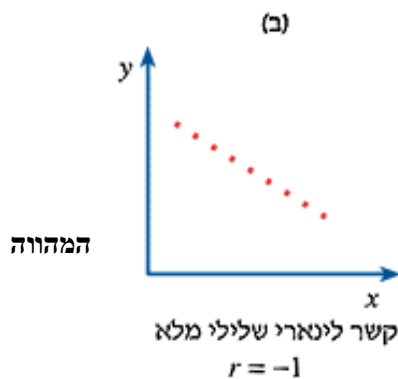
הערכת עוצמת הקשר:

$0 < |r| < 0.3$ מתאם נמוך

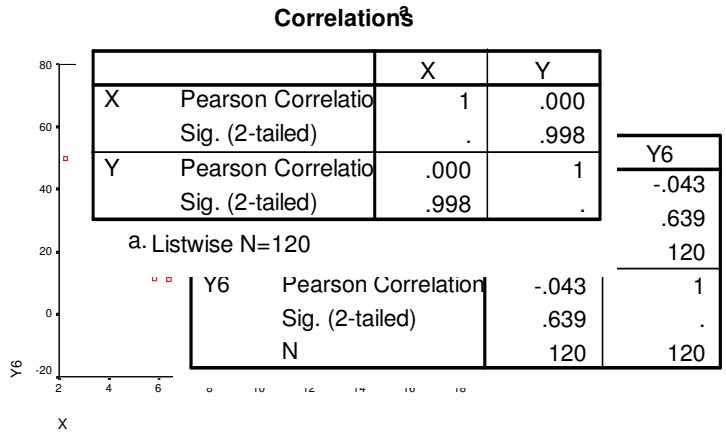
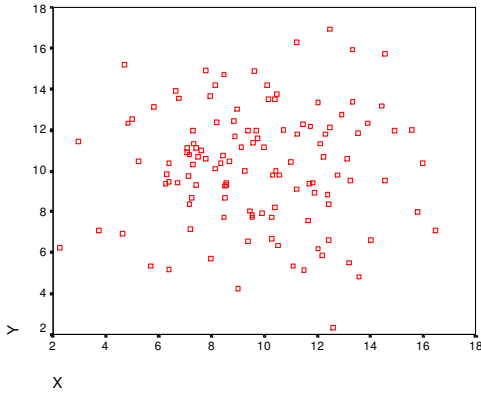
$0.3 < |r| < 0.7$ מתאם בינוני

$0.7 < |r| < 1$ מתאם חזק

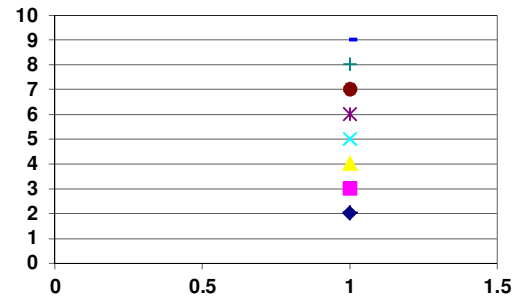
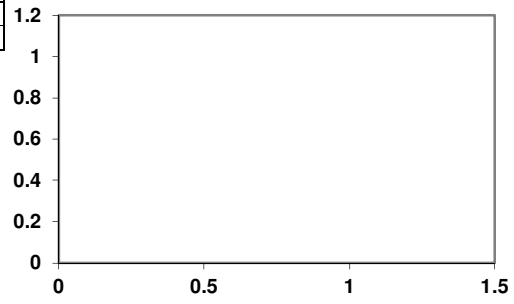
ב. דוגמאות לקשר ליניארי מושלם/מלא:



****תזכורת:** טרנספורמציה ליניארית של משתנה-הוספה/החסרה של קבוע ו/או הכפלה/חילוק בקבוע של כל ערכי המשתנה לדוגמא: כאשר $Y=X+2$, $r_{xy} = 1$ (כי Y הוא טרנספורמציה ליניארית של X).
ג. דוגמאות לחוסר קשר (ליניארי):



1	4	1
1	7	2
1	1	3
...
1	8	10



Y	X	
6	1	1
2	1	2
9	1	3
...
7	1	8

$$\hat{S}_Y^2 = 0$$

$$\hat{S}_X^2 = 0$$

תכונה מס' 2: קשר בין משתנה לקבוע יהיה שווה תמיד 0.

1.4 קיצוץ תחום והשפעתו על המתאם

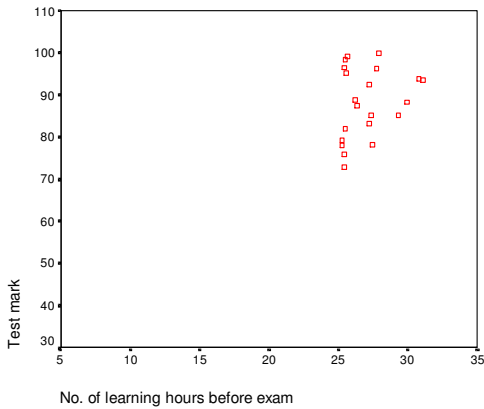
קיצוץ תחום משמעו התייחסות לתחום צר יותר של אחד המשתנים (בדר"כ המשתנה ה'ב'') באופן המקטין את השונות של המשתנים. הקטנת השונות של המשתנים גוררת הקטנת המתאם ביניהם. המתאם בין המשתנים אחרי הקיצוץ יהיה נמוך יותר מן המתאם לפני הקיצוץ.

בדוגמא 1- מתואר קשר בין שעות לימוד לפני המבחן לציון במבחן. הקשר נבדק כעת על קבוצה בעלת תחום צר יותר בשעות לימוד (מעל - 25). **לפני הקיצוץ: אחרי הקיצוץ:**

	Mean	Std. Deviation	N
No. of learning hours before exam	27.0251	1.87348	21
Test mark	88.0571	8.28331	21

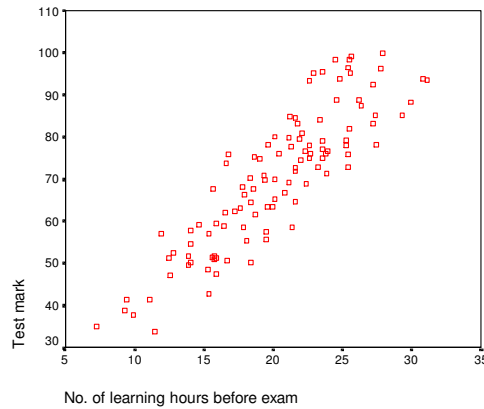
	Mean	Std. Deviation	N
HOURS No. of learn hours before exam	20.1754	5.03946	109
MARK Test mark	70.1530	16.22455	109

דיאגרמת הפיזור לאחר הקיצוץ:



r=0.264

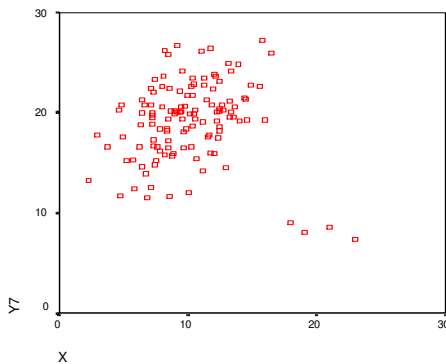
דיאגרמת הפיזור המקורית:



r=0.88

1.5 השפעת ערכים קיצוניים/חריגים על המתאם

• מקרה ראשון שבו הקיצוניים מקטינים קשר:

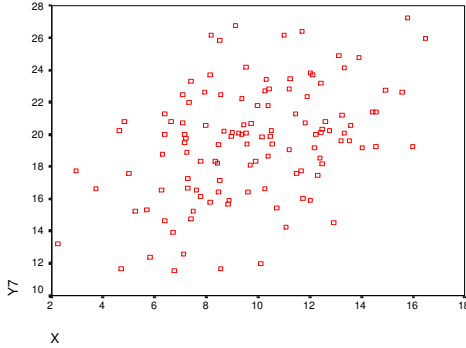


		X	Y7
X	Pearson Correlation	1	.023
	Sig. (2-tailed)	.	.799
Y7	Pearson Correlation	.023	1
	Sig. (2-tailed)	.799	.

a. Listwise N=124

כתב ופ

אותם נתונים לאחר השמטת 4 תצפיות שבניגוד למגמה:



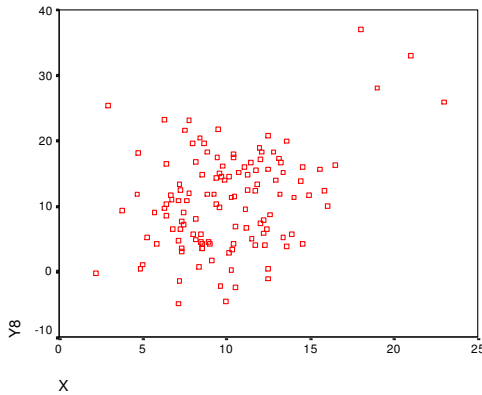
Correlations^a

		X	Y7
X	Pearson Correlation	1	.413**
	Sig. (2-tailed)	.	.000
Y7	Pearson Correlation	.413**	1
	Sig. (2-tailed)	.000	.

** .Correlation is significant at the 0.01 level

a. Listwise N=120

מקרה אחר שבו הקיצוניים מחזקים קשר:
אותם נתונים לאחר השמטת 4 תצפיות שגורמות



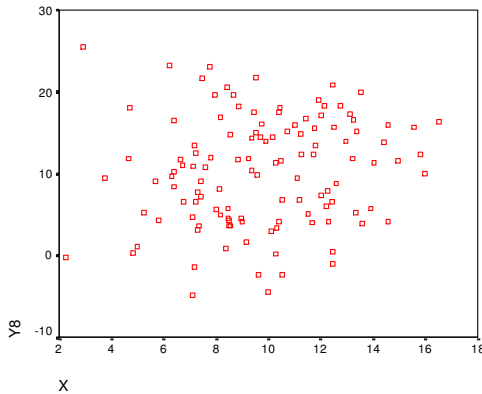
Correlations^b

		X	Y8
X	Pearson Correlation	1	.350**
	Sig. (2-tailed)	.	.000
Y8	Pearson Correlation	.350**	1
	Sig. (2-tailed)	.000	.

** .Correlation is significant at the 0.01 level

a. Listwise N=124

למגמה:



Correlations^a

		X	Y8
X	Pearson Correlation	1	.128
	Sig. (2-tailed)	.	.165
Y8	Pearson Correlation	.128	1
	Sig. (2-tailed)	.165	.

a. Listwise N=120

לסיכום: מקדם המתאם מושפע מאוד מערכים חריגים.

כאשר הערכים החריגים הם בניגוד למגמה הליניארית של מרבית הנתונים- הם יפגעו בעוצמת המתאם

כאשר הערכים החריגים מצויים בהתאם למגמה הליניארית של מרבית הנתונים- הם "ינפחו" את המתאם.

1.6 השפעת טרנספורמציה ליניארית על המתאם

טרנספורמציה ליניארית איננה משנה את ערכו של המתאם. הדגמה על דוגמא מס' 1 (הקשר בין שעות לימוד לבחינה לציון בבחינה): הוחלט לתת פקטור לציון בבחינה-להוסיף לכל ציון 10 נקודות. המתאם בין הציון לאחר הטרנספורמציה לבין שעות הלימוד יהיה שווה למתאם לפני הטרנספורמציה:

$$r' = r = 0.88$$

1.7 נוסחת המתאם הליניארי

$$r = \frac{\text{COV}(x, y)}{\hat{S}_x^2 \cdot \hat{S}_y^2}$$

הנוסחה מבוססת על ההשתנות המשותפת של X ושל Y:

$$\text{cov}(x, y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

ההשתנות המשותפת של משתנה עם עצמו מהווה את האומדן לשונות של המשתנה:

$$\text{cov}(x, x) = \frac{\sum (x_i - \bar{x})^2}{n-1} = \hat{S}_x^2$$

$$\text{cov}(y, y) = \frac{\sum (y_i - \bar{y})^2}{n-1} = \hat{S}_y^2$$

כך שניתן לרשום את הנוסחה של המתאם גם כך:

$$r = \frac{\text{cov}(x, y)}{\sqrt{\text{cov}(x, x)} \cdot \sqrt{\text{cov}(y, y)}}$$

פלט ה-correlation מתאר את המתאם ואת אומדני ה-cov השונים:

Correlations^a

		HOURS No. of learning hours before exam	MARK Test mark
HOURS No. of learning hours before exam	Pearson Correlation	1	.880**
	Sig. (2-tailed)	.	.000
	Sum of Squares and Cross-products	2742.785	7766.854
	Covariance	25.396	71.915
MARK Test mark	Pearson Correlation	.880**	1
	Sig. (2-tailed)	.000	.
	Sum of Squares and Cross-products	7766.854	28429.479
	Covariance	71.915	263.236

** . Correlation is significant at the 0.01 level (2-tailed).

a. Listwise N=109

1.8 מובהקות המתאם

את מובהקות המתאם נבדוק על סמך הפלט בלבד (ללא חישוב). הפלט הרלוונטי למובהקות המתאם הוא פלט ה-correlations. נדגים את בדיקת מובהקות המתאם (ברמת מובהקות של 0.05) על נתוני דוגמא מס' 1 (הקשר בין שעות לימוד לבחינה לציון בבחינה):

Correlations^a

		HOURS No. of learning hours before exam	MARK Test mark
HOURS No. of learning hours before exam	Pearson Correlation	1	.880**
	Sig. (2-tailed)	.	.000
	Sum of Squares and Cross-products	2742.785	7766.854
	Covariance	25.396	71.915
MARK Test mark	Pearson Correlation	.880**	1
	Sig. (2-tailed)	.000	.
	Sum of Squares and Cross-products	7766.854	28429.479
	Covariance	71.915	263.236

** . Correlation is significant at the 0.01 level (2-tailed).

a. Listwise N=109

השערות:

$$H_0: \eta = 0$$

$$H_1: \eta \neq 0$$

כלל הכרעה (על סמך ה-sig שבפלט ה-correlations):

$$\text{Sig}=0.00 < \alpha = 0.05$$



יש סיבה מספקת לדחות את H_0 ברמת מובהקות של 0.05

מסקנה:

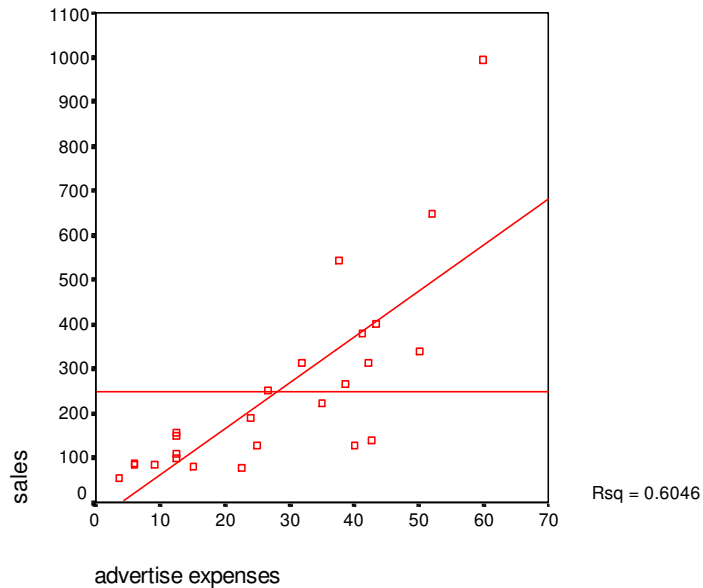
יש עדות למתאם מובהק באוכלוסייה בין שעות לימוד לבחינה לבין הציונים בבחינה ברמת מובהקות של 0.05.

פרק 16 - קו הרגרסיה במדגם

1.9 בניית קו הרגרסיה במדגם

1.9.1 עיקרון הריבועים הפחותים (L.S.E)

לאחר שנוכחנו שיש מתאם ליניארי בין שני המשתנים אנו מעוניינים לבנות את קו הניבוי שיאפשר לנו לנבא משתנה אחד על סמך השני.
 נשאלת השאלה: מהו הקו (מבין קווים רבים אפשריים) שיהיה המתאים ביותר לנתוני המדגם?



תשובה: קו הרגרסיה אשר יביא למינימום את הטעויות בניבוי ממנו.

את הניבוי למשתנה y במדגם נסמן: \hat{y}_i

את הערך האמיתי של y במדגם נסמן: y_i

הטעות בניבוי במדגם מהווה את ההפרש בין שני הערכים: $e_i = y_i - \hat{y}_i$

הקו הטוב ביותר בנתוני המדגם הוא הקו שמבטיח כי סכום כל הטעויות הריבועיות הוא המינימאלי ביותר, או במילים אחרות קו הרגרסיה אשר נבנה על בסיס "עיקרון הריבועים הפחותים" (Least

:Square Estimation)

$$\sum e_i^2 = \sum (y_i - \hat{y}_i)^2 = \sum [y_i - (\hat{\alpha}_i + \hat{\beta}_i x_i)]^2 = MIN$$

באמצעות פונקציה זו (L.S.E) נוכל למצוא את מקדמי הקו הישר $(\hat{\alpha}, \hat{\beta})$ אשר יביאו למינימום את הטעויות מהקו.

לאחר פיתרון המשוואה (אין צורך לדעת לפתור) מתקבלות הנוסחאות הבאות עבור $\hat{\alpha}$ ו- $\hat{\beta}$:

$$\hat{\beta} = r \cdot \frac{\hat{s}_y}{\hat{s}_x} = \frac{\cos(x, y)}{\hat{s}_x^2}$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

1.9.2 מקדמי הקו

א. משמעות המקדמים:

(1) $\hat{\beta}$ - מקדם השיפוע של הקו.

קצב ההשתנות של \hat{y}_i כפונקציה של עליה ביחידה אחת של x_i .

עבור עליה ביחידה אחת של x_i , \hat{y}_i עולה פי $\hat{\beta}$.

יש קשר ישיר בין $\hat{\beta}$ ל- r : ה- $\hat{\beta}$ מקבלת את הסימן של ה- r .

כאשר אין קשר בין X ל-Y ($r=0$), $\hat{\beta} = 0$.

(2) $\hat{\alpha}$ - מקדם החיתוך של הקו.

נקודת החיתוך של קו הרגרסיה עם ציר ה-Y.

כאשר $x_i = 0$, $\hat{y}_i = \hat{\alpha}$.

כאשר $r=0$, $\hat{y}_i = \hat{\alpha} = \bar{y}$.

ב. חישוב המקדמים:

? דוגמא מס' 2: ניבוי המכירות באמצעות ההוצאות על פרסום (אלפי ש"ח).

להלן הנתונים המתארים את הקשר בין המכירות לבין ההוצאות על הפרסום:

Case Summaries

	N	Mean	Std. Deviation	Minimum	Maximum
X advertise expenses	25	28.1040	16.36932	3.70	60.00
Y sales	25	249.2400	217.85283	55.00	994.00

Correlations^a

		X advertise expenses	Y sales
X advertise expenses	Pearson Correlation	1	.778**
	Sig. (2-tailed)	.	.000
	Sum of Squares and Cross-products	6430.910	66548.276
	Covariance	267.955	2772.845
Y sales	Pearson Correlation	.778**	1
	Sig. (2-tailed)	.000	.
	Sum of Squares and Cross-products	66548.276	1139037
	Covariance	2772.845	47459.857

** . Correlation is significant at the 0.01 level (2-tailed).

a. Listwise N=25

א. חשב את משוואת הרגרסיה לניבוי המכירות על סמך הוצאות הפרסום.

ב. מה יהיה ניבוי המכירות עבור הוצאות פרסום של 50 אלף ש"ח?

ג. מה יהיה ניבוי הוצאות הפרסום עבור מכירות של 267 אלף ש"ח?

1.10 תכונות קו הרגרסיה במדגם

(1) סכום הטעויות בניבוי ($\sum e_i$) וממוצע הטעויות בניבוי (\bar{e}_i) שווים ל-0:

$$\bar{e}_i = \sum e_i = 0$$

(2) נקודת הממוצעים (\bar{x}, \bar{y}) נמצאת תמיד על קו הרגרסיה.

(3) ממוצע הניבויים/האומדנים (\hat{y}) = ממוצע ציונים אמיתיים/מנובאים (\bar{y}):

$$\hat{y} = \bar{y}$$

(4) ככל שהציון המנבא (X_i) קרוב יותר לממוצע שלו (\bar{X}) כך הניבוי באמצעותו (\hat{Y}_i) יהיה מדויק יותר.

(5) ניתן להשתמש בקו הרגרסיה רק עבור תחום ציוני המנבא-X עליו הוא נבנה.

(6) עבור ערך x_i מסוים מתקיים כי:

טעות בניבוי + אומדן לניבוי = ציון אמיתי

$$y_i = \hat{y}_i + e_i$$

עבור כלל התצפיות מתקיים כי:

סכום ריבועי הטעויות + סכום ריבועי הניבויים = סכום ריבועי הציונים האמיתיים

$$ssy = ssreg + ssres$$

$$\sum (y_i - \bar{y})^2 = \sum (\hat{y}_i - \bar{y})^2 + \sum (y_i - \hat{y}_i)^2$$

1.11 אומדנים חשובים בגרסיה פשוטה

1.11.1 פרופורציית השונות המוסברת

$$r^2 = \frac{ssreg}{ssy}$$

פרופורציית השונות הלא-מוסברת:

$$1 - r^2 = \frac{ssres}{ssy}$$

1.11.2 שונות הטעויות

$$MSRES = \frac{SSRES}{n-2} = \frac{(1-r^2) \cdot (n-1) \cdot \hat{\sigma}_y^2}{n-2}$$

MSRES מהווה אומדן חסר הטיה לשונות הלא מוסברת: $E(MSRES) \rightarrow \sigma_e^2$

ניסוחים נוספים לשונות הלא מוסברת:

- אומדן לשונות של Y לאחר ניכוי הקשר הקווי בין X ל-Y.
- אומדן לשונות של Y עבור ערך מסוים של X.

? על סמך נתוני דוגמא מס' 2 :

Case Summaries

	N	Mean	Std. Deviation	Minimum	Maximum
X advertise expenses	25	28.1040	16.36932	3.70	60.00
Y sales	25	249.2400	217.85283	55.00	994.00

כאשר נתון בנוסף כי $r_{xy} = 0.778$

חשב את האומדנים הבאים:

- סכום הסטיות הריבועיות של המכירות
- אומדן לשונות הכללית של המכירות
- אומדן לשונות הכללית של המכירות לאחר ניכוי הקשר הקווי עם הוצאות הפרסום

לפתרון מלא בסרטון וידאו היכנסו ל- www.Gool.co.il

- ד. אומד לשונות של המכירות עבור הוצאות פרסום של 52 אלף ש"ח אומד לשונות הלא מוסברת של המכירות
- ה. אומד לסטיית התקן של הטעויות בניבוי המכירות
- ו. אומד לפרופורציית השונות הלא מוסברת של המכירות על ידי הוצאות הפרסום
- ז. אומד לפרופורציית שונות הטעויות של הוצאות הפרסום על ידי המכירות סכום ריבועי הסטיות של ניבויי המכירות
- ח. סכום ריבועי הסטיות של הטעויות בניבוי המכירות
- ט. סכום הסטיות הריבועיות של הוצאות הפרסום

1.12 מתאמים בין משתני הרגרסיה

	y ערכי המשתנה התלוי/מנובא	\hat{y} ניבויים/אומדנים/PRED	E טעויות בניבוי/ שאריות/RES
X ערכי המשתנה הב"ת/מנובא	r_{xy}	$r_{x\hat{y}} = 1$ כי המתאם בין משתנה (x) לטרנספורמציה הליניארית של אותו משתנה (\hat{y}) שווה תמיד ל-1.	$r_{xe} = 0$ כי מתאם בין קבוע (x) למשתנה (e) שווה תמיד ל-0.
y		$r_{y\hat{y}} = r_{xy}$ כי טרנספורמציה ליניארית (מ-x ל- \hat{y}) לא משנה את ערך המתאם.	$r_{ye} = \sqrt{1 - r_{xy}^2}$ כי r_{ye}^2 מהווה את פרופורציית שונות הטעויות ולכן שווה ל- $1 - r_{xy}^2$ **יש יחס הפוך בין r_{ye} לבין r_{xy} או $r_{y\hat{y}}$
\hat{y}			$r_{\hat{y}e} = 0$ כי מתאם בין קבוע (\hat{y}) למשתנה (e) שווה תמיד ל-0.

? על סמך נתוני דוגמא מס' 2, כאשר נתון כי: $r_{xy} = 0.778$

נא השלם את המתאמים בין משתני הרגרסיה:

	X	Y	PRED	RES
X				
Y	$Z_i = \frac{x_i - \bar{x}}{\hat{s}_x}$			
PRED				
RES				

1.13 משוואת הרגרסיה המתוקנת

עד עתה בנינו את משוואת הרגרסיה בציוני גלם. כעת נבנה את משוואת הרגרסיה בציוני תקן. להזכירכם - ציון תקן הוא מדד למיקום יחסי; מדד זה נותן אינפורמציה על מיקומה של תצפית מסוימת ביחס להתפלגות כלל התצפיות:

בדוגמא א' נתון קשר בין שעות לימוד לפני המבחן לציון במבחן ערכו טרנספורמציה ליניארית לשני המשתנים והעבירו אותם לציוני תקן. דוגמא לטרנספורמציה על 5 תצפיות ראשונות בקובץ:

Case Summaries^a

	HOURS No. of learning hours before exam	MARK Test mark	ZHOURS Zscore: No. of learning hours before ex	ZMARK Zscore: Test mark
1	21.12	69.24	.18838	-.05646
2	15.75	51.00	-.87911	-1.18054
3	14.01	50.21	-1.22270	-1.22902
4	20.13	70.11	-.00817	-.00283
5	23.51	77.07	.66221	.42633
Total N	5	5	5	5

a. Limited to first 5 cases.

Descriptive Statistics

	Mean	Std. Deviation	N
ZHOURS Zscore: No. of learning hours before ex	.000	1.000	109
ZMARK Zscore: Test mark	.000	1.000	109

$$r_{xy} = r_{zxzy} = 0.88$$

1.13.1 מקדמי המשוואה המתוקנת

א. מקדם השיפוע המתוקנן:

$$z\hat{\beta} = r$$

$$\text{הסבר: } z\hat{\beta} = r_{zxy} \cdot \frac{\hat{s}_{zy}}{\hat{s}_{zx}} = r_{xy} \cdot \frac{1}{1} = r_{xy}$$

מקדם השיפוע במשוואה המתוקנת שווה למתאם בין X ל-Y

ב. מקדם החיתוך המתוקנן:

$$z\hat{\alpha} = 0$$

$$\text{הסבר: } z\hat{\alpha} = z\bar{y} - z\hat{\beta} \cdot z\bar{x} = 0 - r \cdot 0 = 0$$

משוואת הרגרסיה המתוקנת יוצאת תמיד מראשית הצירים.

ג. משוואת הרגרסיה המתוקנת:

$$z\hat{y}_i = r \cdot zx_i$$

? חוקר מנבא את Y ע"י X דרך משוואת הרגרסיה. הנתונים שהיו בפניו הינם:

$$\text{א.ח.ה לשונות של } X = 16$$

$$\text{א.ח.ה לשונות של } Y = 100$$

$$\text{הממוצע של } X = 20$$

$$\text{ממוצע ערכי הניבויים} = 45$$

$$\text{המתאם בין } X \text{ ל- } Y = 0$$

א. חשב את משוואת הניבוי של Y ע"י X בציוני תקן.

ב. מהו הציון הגולמי שנבא עבור $X = 25$?

ג. חווה דעתך על הטענה הבאה: ברגרסיה בציוני תקן, ציון התקן המנובא יהיה קטן

יותר בערכו המוחלט מציון התקן המנבא.

2.6 השוואה בין מודל רגרסיה חד משתנית לבין מודל ניתוח שונות חד כיווניתרגרסיה חד משתנית היא מקרה פרטי של ניתוח שונות חד כיוונית: $eta^2 \geq r^2$ א. כאשר בין ממוצעי הקבוצות קיים פער קבוע ($\hat{\beta}$)-הנתונים ליניאריים $\Leftarrow eta^2 = r^2$,

מודל הרגרסיה מתאים יותר לניתוח הנתונים מניתוח השונות כי הוא ממוקד ויעיל יותר.

ב. כאשר הפער בין ממוצעי הקבוצות איננו קבוע- הנתונים אינם ליניאריים \Leftarrow $eta^2 > r^2$, מודל הרגרסיה איננו מתאים לנתונים (אינם ליניאריים) ויש לבצע ניתוח שונות.

? חברה יצרה מוצר חדש ומתלבטת ביחס למחיר המכירה של מוצר זה. מנהל השיווק יודע שיש לקבוע מחיר שהינו בסביבות 10 ₪ ליחידה אולם הוא איננו בטוח אם קביעת המחיר כ- 9 ₪ ליחידה או 11 ₪ ליחידה ישנה באופן משמעותי את רמת המכירות. לצורך מתן מענה לשאלה זו בוצע הניסוי הבא: המוצר החדש שווק למדגם מקרי של 60 חנויות הנמצאות באזורים בעלי מאפיינים דומים. ב-20 חנויות שנבחרו באופן מקרי מבין ה-60 נמכר המוצר במחיר של 9 ₪ ליחידה, ב-20 חנויות מקריות אחרות הוא נמכר ב-10 ₪ ליחידה וביתר במחיר של 11 ₪ ליחידה. בתום תקופת הניסוי נרשם סך היחידות שנמכרו בכל אחת מהחנויות. לפניך פלט ובו סכום חלק מהממצאים:

Level	Number	Mean
9 ₪	20	153.6
10 ₪	20	151.5
11 ₪	20	123.2

- שני סטודנטים שהסתכלו על הנתונים, הציעו לנתחם באמצעות 2 מודלים שונים: סטודנט א' הציע לנתח את הנתונים באמצעות מודל ניתוח שונות חד כיווני. סטודנט ב' הציע לנתח את הנתונים באמצעות מודל הרגרסיה הפשוטה כאשר X הינו המחיר בו נמכר המוצר בחנות ואילו ה- Y הוא מספר האריזות הנמכרות בחנות (במחיר זה).
- סטודנט ג' שהתבקש לחוות את דעתו טען שהיות ובשני המודלים נבדקת השפעה של המחיר על המכירות לא חשוב באמצעות איזה מודל ינותחו הנתונים. התוצאות תהינה זהות בשני המודלים.
- א. חווה דעתך על טענתו של הסטודנט השלישי. במידה ואתה סבור שיש הבדל בין שני המודלים, באיזה מודל היית בוחר לצורך ניתוח הנתונים? נמק.
- ב. במידה והיית מחשב את הקשר בין המחיר למס' האריזות הנמכר בשני המודלים- מה להערכתך היה היחס בין שני מדדי הקשר- r ו- η (שווים או שאחד גדול יותר מהשני)? נמק.

פרק 17 - מובהקות הרגרסיה באוכלוסיה

1.14 ההנחות הנדרשות למודל הרגרסיה הפשוטה

3.1.1 ההנחות ומשמעותן

למודל הרגרסיה החד משתנית נדרשות שתי ההנחות הבאות:
 (1) דגימה מקרית של התצפיות; אי תלות בין הטעויות

$$\text{cov}(\varepsilon_i, \varepsilon_j) = 0 \quad i \neq j$$

הסבר: ההנחה הראשונה טוענת שאם התצפיות נדגמו באופן בלתי תלוי אחת בשנייה, הדבר מבטיח שהטעויות בניבוי שלהן יהיו אף הן בלתי תלויות.

$$(2) \quad \varepsilon_i \approx N(0, \sigma_e^2) \quad \text{עבור כל ערך קבוע של } X_i$$

הסבר: ההנחה השנייה מגלמת בתוכה 3 הנחות שונות המתייחסות להתפלגות הטעויות א. תוחלת התפלגות הטעויות שווה לאפס ($E(\varepsilon_i) = 0$) עבור כל ערך של x_i .

הנחה זו מבטיחה שקו הרגרסיה באוכלוסיה הוא קו התוחלות – הציונים המנובאים שווים למומצע ערכי Y עבור כל ערך קבוע של x_i ($\mu_{y_i/x=x_i} = \alpha + \beta x_i$). במילים אחרות הנחה זו מבטיחה כי מדובר במודל ליניארי ולא אחר.

רק כאשר הנחה זו מתקיימת – מקדמי קו הרגרסיה במדגם $\hat{\alpha}$ ו- $\hat{\beta}$ הם אומדנים חסרי הטיה למקדמי הקו באוכלוסיה α ו- β .

ב. שונות הטעויות (σ_e^2) שווה וקבועה עבור כל ערך של x_i .

הנחה זו מבטיחה כי שונות הטעויות תהיה שווה סביב קו הרגרסיה ומינימאלית. תחת הנחה זו, האומד לשונות הטעויות MSRES מהווה אומד לשונות ערכי Y עבור כל אחד מערכי x_i ($MSRES \rightarrow \sigma_{y_1}^2 = \sigma_{y_2}^2 = \dots = \sigma_e^2$).

רק כאשר הנחה זו מתקיימת האומדנים $\hat{\alpha}$ ו- $\hat{\beta}$ יהיו אומדנים יעילים לפרמטרים α ו- β באוכלוסיה.

ג. הטעויות מתפלגות נורמאלית עבור כל ערך של x_i .

הנחה זו מבטיחה התפלגות נורמאלית של הטעויות (או ערכי Y) עבור כל ערך קבוע של x_i .

לסיכום ההנחות של מודל הרגרסיה:

1. דגימה מקרית של תצפיות/ טעויות ב"ת.

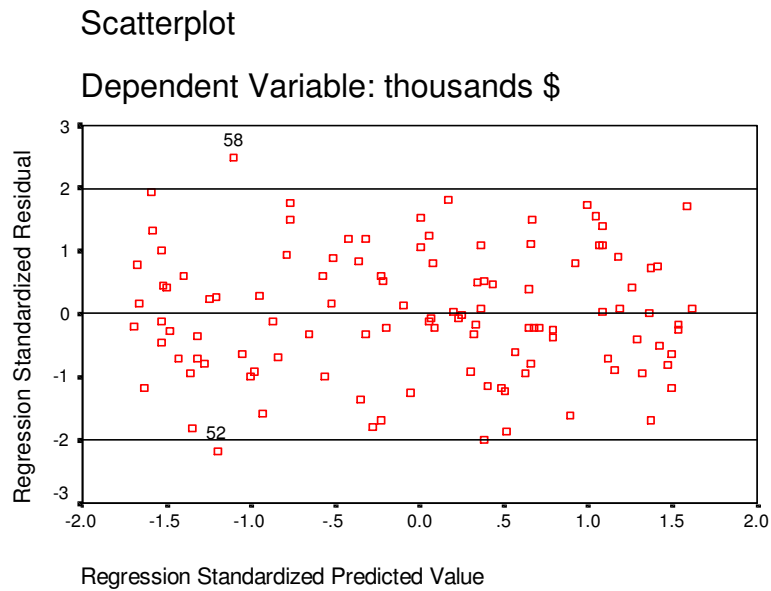
$$\text{cov}(\varepsilon_i, \varepsilon_j) = 0 \quad i \neq j$$

2. $\varepsilon_i \approx N(0, \sigma_e^2)$ עבור כל ערך קבוע של X_i .

שונות קבועה ומינימאלית
 תוחלת התפלגות הטעויות = 0
 האומדנים $\hat{\alpha}, \hat{\beta}$ יעילים לפרמטרים α, β באוכלוסיה
 האומדנים $\hat{\alpha}, \hat{\beta}$ חסרי הטיה לפרמטרים α, β באוכלוסיה

3.1.2 בדיקת ההנחות

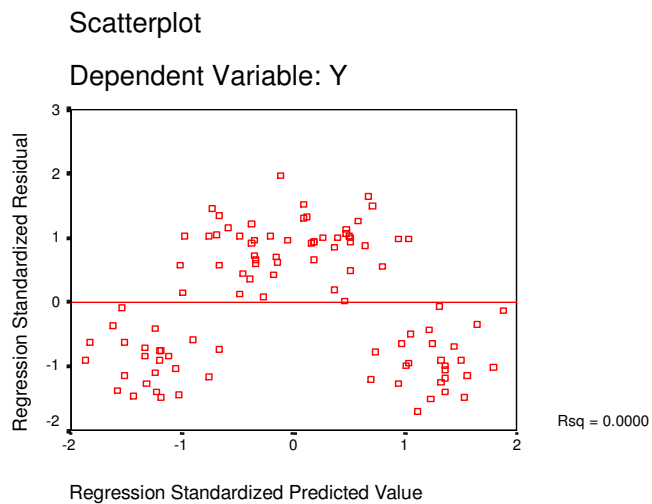
את ההנחות נהוג לבדוק על ידי דיאגרמת פיזור של ניבויים $(z\hat{y}_i)$ * שאריות (ze_i) בציוני תקן כמו לדוגמא:



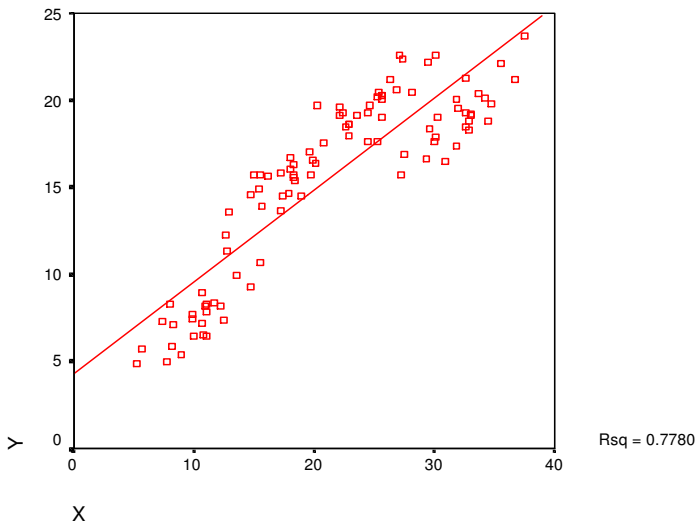
על דיאגרמה מעין זו ניתן לבצע הערכה (שאיננה מבחן סטטיסטי) האם מתקיימות כל אחת משלושת ההנחות. דיאגרמה זו מאפשרת בנוסף איתור תצפיות חריגות (הסושות יותר משתי סטיות תקן מעל או מתחת לממוצע).

דוגמאות לאי עמידה בהנחות המודל:

א. אי עמידה בהנחת תוחלת הטעויות $E(\varepsilon_i) = 0$ עבור כל ערך של X :



המשמעות של הפרת הנחה זו היא כי קו הרגרסיה איננו קו התוחלות ולמעשה המודל איננו ליניארי:

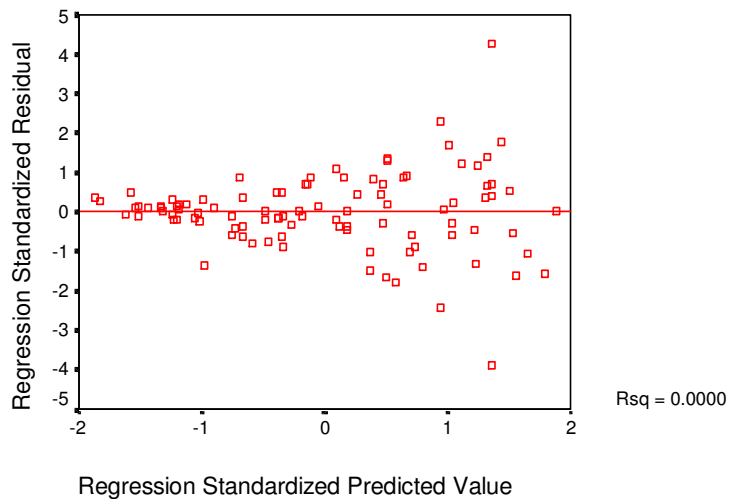


במקרה זה האומדנים $\hat{\alpha}$ ו- $\hat{\beta}$ יהיו אומדנים מוטים לפרמטרים α ו- β באוכלוסיה.

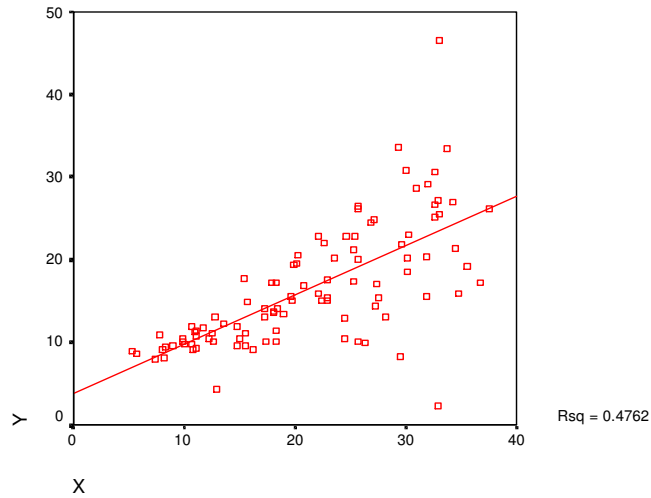
ב. אי עמידה בהנחת שוויון שוניות:

Scatterplot

Dependent Variable: Y



במקרה זה שונות ערכי Y איננה שווה לאורך קו הרגרסיה:



במקרה זה האומדנים $\hat{\alpha}$ ו- $\hat{\beta}$ יהיו אומדנים בלתי יעילים לפרמטרים α ו- β באוכלוסיה.

1.15 בדיקת השערות למובהקות משוואת הרגרסיה באוכלוסיה

מבחן סטטיסטי שמטרתו לבדוק האם קו הרגרסיה שבנינו במדגם מובהק באוכלוסיה.

3.2.1 השערות

$$H_0; \beta = 0$$

$$H_1; \beta \neq 0$$

ההשערות מנוסחות במונחי הפרמטר β , שיפוע הקו. האפקט במודל הרגרסיה תלוי במקדם השיפוע. כאשר הוא שונה מאפס יש אפקט של ניבוי לקו הרגרסיה ואילו כאשר השיפוע = 0 אין אפקט ניבוי לקו הרגרסיה.

3.2.2 סטטיסטי המבחן

$$F = \frac{MSREG}{MSRES}$$

רציונאל המבחן זהה לרציונאל של ניתוח שונות חד כיוונית:

$$E(MSREG) = \Delta + \sigma_e^2$$

$$E(MSRES) = \sigma_e^2$$

תחת H_0 ($\Delta = 0$):

$$E(MSREG) = \sigma_e^2 = E(MSRES)$$

3.2.3 חישוב הסטטיסטי

את סטטיסטי המבחן F נחשב באמצעות טבלת ה-ANOVA :

מקור השונות	df חופש	SS	MS=SS/df	$F = \frac{MS_{REG}}{MS_{RES}}$ סטטיסטי
רגרסיה reg	k=1	SSreg	MSreg	
טעויות res	n-k-1=n-2	SSres	MSres	
אמיתית y	n-1	SSy		קריטי

ניתן לחשב את סטטיסטי המבחן F גם באמצעות נוסחה מקוצרת (כאשר אנו לא נדרשים לבנות את טבלת ה-ANOVA):

$$F = \frac{r^2}{(1-r^2)/n-2}$$

3.2.4 כלל הכרעה ומסקנה

נדחה את H_0 אם F מחושב $F <$ קריטי

מסקנה: יש/אין עדות לכך שמשוואת הרגרסיה לניבוי Y על ידי X מובהקת באוכלוסיה ברמת מובהקות של α .

? מבוססת על דוגמא מס' 3: ניבוי מחיר הדירה על ידי שטחה.

בבדיקת הקשר בין המשתנים התקבלו הנתונים הבאים:

Case Summaries

	N	Mean	Std. Deviation	Minimum	Maximum
SIZE square meter	112	116.7404	39.13942	50.46	179.76
PRICE thousands \$	112	185.0664	44.45345	86.20	286.56

Correlations^a

		SIZE square meter	PRICE thousands \$
SIZE square meter	Pearson Correlation	1	.823**
	Sig. (2-tailed)	.	.000
PRICE thousands \$	Pearson Correlation	.823**	1
	Sig. (2-tailed)	.000	.

** . Correlation is significant at the 0.01 level (2-tailed).

a. Listwise N=112

א. בדוק האם המשוואה לניבוי מחיר הדירה (באלפי דולרים) על ידי שטח הדירה (במטר מרובע) מובהקת באוכלוסיה ברמת מובהקות של אלפא=0.01 (רשום הנחות, הצב השערות, רשום טבלת מקור שונות והסק מסקנה).

ב. בדוק את מובהקות המשוואה על סמך הפלט הבא:

Model Summary ^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.823 ^a	.677	.674	25.39155

a. Predictors: (Constant), SIZE square meter

b. Dependent Variable: PRICE thousands \$

כאשר נתון בנוסף ש: $n=112$

1. עם חישוב טבלת מקור שונות.

2. ללא צורך בחישוב טבלת מקור שונות.

ג. בנה טבלת מקור שונות על פי הפלט הבא:

Descriptive Statistics

	Mean	Std. Deviation	N
SIZE square meter	116.740	39.139	112
PRICE thousands \$	185.066	44.453	112
PRE_1 Unstandardized Predicted Value	185.066	36.568	112
RES_1 Unstandardized Residual	.000	25.277	112

1.16 בדיקת השערות למובהקות מקדם השיפוע באוכלוסייה

מבחן סטטיסטי שמטרתו לבדוק האם מקדם השיפוע מובהק באוכלוסייה.

1.16.1 השערות

$$H_0; \beta = 0$$

$$H_1; \beta \neq 0$$

הערה: שימו לב כי ניתן לשער גם השערות חד צדדיות וכמו כן השערות המייחסות ערך כלשהו (שאינו 0) לשיפוע.

1.16.2 סטטיסטי המבחן

$$t_{\hat{\beta}} = \frac{\hat{\beta} - \beta}{\sqrt{\frac{MSRES}{SSX}}}$$

1.16.3 כלל הכרעה ומסקנה

נדחה את H_0 אם:

$$t_{\hat{\beta}} > \pm t\left(\frac{\alpha}{2}\right)(n-2)$$

הערה: במבחן חד צדדי יש לקחת α שלמה.

מסקנה: יש/אין עדות לכך ששיפוע קו הרגרסיה לניבוי Y על ידי X מובהק באוכלוסייה ברמת מובהקות של α .

1.16.4 גורמים המשפיעים על טעות התקן של $\hat{\beta}$

$$\hat{s}_{\hat{\beta}} = \sqrt{\frac{MSRES}{SSX}}$$

טעות התקן של $\hat{\beta}$ ($\hat{s}_{\hat{\beta}}$) מושפעת משני גורמים:

א. במונה-האומד לשונות הטעויות-MSRES. קיים קשר ישיר בינו לבין $\hat{s}_{\hat{\beta}}$.

קיים קשר הפוך בין ה-MSRES לבין עוצמת הקשר הליניארי בין המשתנים)
 $|r_{xy}|$.

ב. במכנה-סכום הסטיות הריבועיות של X-SSX. קיים קשר הפוך בינו לבין $\hat{s}_{\hat{\beta}}$.

קשור לתופעת קיצוץ תחום; כאשר מקוצץ תחום ערכי X (הב"ת), ה-SSX קטן.

1.16.5 רב"ס ל- β

כאשר אנו מתבקשים לאמוד את שיפוע המשוואה, נבנה רווח בר סמך ל- β :

$$\hat{\beta} \pm t_{n-2}\left(\frac{\alpha}{2}\right)\sqrt{\frac{MSRES}{SSX}}$$

??

(1) על סמך נתוני דוגמא מס' 3 (ניבוי מחיר הדירה על סמך שטחה):

Case Summaries

	N	Mean	Std. Deviation	Minimum	Maximum
SIZE square meter	112	116.7404	39.13942	50.46	179.76
PRICE thousands \$	112	185.0664	44.45345	86.20	286.56

Correlations^a

		SIZE square meter	PRICE thousands \$
SIZE square meter	Pearson Correlation	1	.823**
	Sig. (2-tailed)	.	.000
PRICE thousands \$	Pearson Correlation	.823**	1
	Sig. (2-tailed)	.000	.

** Correlation is significant at the 0.01 level (2-tailed).

a. Listwise N=112

א. בדוק האם השיפוע לניבוי מחיר הדירה על ידי שטחה מובהק באוכלוסייה ברמת מובהקות של 0.05.

ב. אמוד את השיפוע באוכלוסייה ברמת ביטחון של 0.95.

(2) חווה דעתך על הטענות הבאות:

א. ברגרסיה חד משתנית, ככל שנדגום טווח ערכים גדול יותר במשתנה ה'ב', אזי נגדיל את הפיזור ועקב כך ברוב המקרים שונות הטעויות תגדל ויהיה קשה יותר לדחות את השערת האפס למובהקות מקדם בטא.

ב. השונות של התפלגות הדגימה של בטא גדלה ככל שהקשר בין המשתנה התלוי למשתנה הבלתי תלוי חלש יותר.

1.16.6 פלט המקדמים (coefficients)

להלן פלט המציג את מקדמי הרגרסיה ומובהקותם בציוני גלם ובציוני תקן:

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	75.997	7.578		10.028	.000	60.979	91.015
	SIZE square meter	.934	.062	.823	15.173	.000	.812	1.056

a. Dependent Variable: PRICE thousands \$

**שימו לב לקשר בין סטטיסטי המבחן T לערך המקדם וטעות התקן:

$$t = \frac{B}{Std.Error}$$

1.17 קשר בין התפלגויות T ו-F

ברגרסיה חד משתנית קיימת זהות בין שלושת המבחנים הסטטיסטיים הבאים:
 (1) מבחן F למובהקות משוואת הרגרסיה

(2) מבחן T למובהקות מקדם השיפוע (β)

(3) מבחן למובהקות מקדם המתאם (η)

כל אחד מן המבחנים מעיד על האחרים: ה-sig בשלושת שווה.
 בנוסף קיים קשר בין סטטיסטי המבחן T ו-F (המבחן הראשון והשני):

$$F(1, n-2) = t_{n-2}^2 \left(\frac{\alpha}{2} \right)$$

על פי נתוני דוגמא מס' 2 (ניבוי מחיר הדירה על פי שטחה) ניתן לראות כי:
 סטטיסטי המבחן F שחישבנו למובהקות משוואת הרגרסיה שווה לריבוע סטטיסטי המבחן
 T שחישבנו למובהקות מקדם השיפוע:

$$F = 230.9 = t^2 = 15.2^2$$

ניתן לראות בנוסף כי ה-sig של שלושת המבחנים שווה:
 ה-sig של מובהקות המשוואה (המופיע בפלט ה-ANOVA):

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	148427.7	1	148427.680	230.216	.000 ^a
	Residual	70920.396	110	644.731		
	Total	219348.1	111			

a. Predictors: (Constant), SIZE square meter

b. Dependent Variable: PRICE thousands \$

שווה ל-sig של מובהקות השיפוע (שורה שנייה בפלט ה-coefficients):

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	75.997	7.578		10.028	.000	60.979	91.015
	SIZE square meter	.934	.062	.823	15.173	.000	.812	1.056

a. Dependent Variable: PRICE thousands \$

ול-sig של מובהקות מקדם המתאם בין שני המשתנים (המופיע בפלט ה-correlations):

Correlations^a

		SIZE square meter	PRICE thousands \$
SIZE square meter	Pearson Correlation	1	.823**
	Sig. (2-tailed)	.	.000
PRICE thousands \$	Pearson Correlation	.823**	1
	Sig. (2-tailed)	.000	.

** . Correlation is significant at the 0.01 level (2-tailed).

a. Listwise N=112

1.18 רב"ס לתוחלת ערכי Y (μ_0) עבור ערך מסוים של X

רב"ס שמטרתו לאמוד את תוחלת ערכי המשתנה התלוי (μ_0) עבור ערך מסוים של המשתנה הבלתי תלוי (x_0). במילים אחרות אנו מתבקשים לאמוד את הניבוי באוכלוסייה עבור ערך מסוים של X .

האומד הנקודתי (הסטטיסטי) סביבו בנוי הרב"ס הוא הניבוי במדגם עבור אותו ה- X :

$$\hat{\mu}_0 = \hat{y}_0 = \hat{\alpha} + \hat{\beta}x_0$$

נוסחת הרב"ס:

$$\hat{\mu}_0 \pm t_{n-2} \left(\frac{\alpha}{2} \right) \cdot \sqrt{MSRES \cdot \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{SSX} \right)}$$

טעות התקן/גודל הרב"ס מושפעים מ-4 גורמים:

א. **MSRES** - האומד לשונות הטעויות. ככל שגדל, טעות התקן/ הרב"ס גדלים ולהפך.

ב. **n** - גודל המדגם. ככל שגדל, טעות התקן/ הרב"ס קטנים ולהפך.

ג. **SSX** - מונה השונות של X (קשור לתופעת קיצוץ תחום). ככל שגדל, טעות התקן/ הרב"ס קטנים ולהפך.

ד. $(x_0 - \bar{x})$ - הסטייה של ערך X המסוים מהממוצע של X . ככל שגדלה טעות התקן/ הרב"ס גדלים ולהפך.

? השאלה מבוססת על נתוני דוגמא מס' 2 (ניבוי הקשר בין מחיר הדירה על סמך שטחה) והפלטים הבאים:

Case Summaries

	N	Mean	Std. Deviation	Minimum	Maximum
SIZE square meter	112	116.7404	39.13942	50.46	179.76
PRICE thousands \$	112	185.0664	44.45345	86.20	286.56

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	75.997	7.578		10.028	.000	60.979	91.015
	SIZE square meter	.934	.062	.823	15.173	.000	.812	1.056

a. Dependent Variable: PRICE thousands \$

Descriptive Statistics

	Mean	Std. Deviation	N
SIZE square meter	116.740	39.139	112
PRICE thousands \$	185.066	44.453	112
PRE_1 Unstandardized Predicted Value	185.066	36.568	112
RES_1 Unstandardized Residual	.000	25.277	112

- א. חשב רב"ס ברמת סמך של 95% לתוחלת מחיר הדירה כאשר שטח הדירה הוא 100 מ"ר.
- ב. תן אומדן נקודתי לתוחלת מחיר הדירה עבור שטח הדירה 105 מ"ר ועבור שטח דירה של 70 מ"ר. מי משני האומדנים הללו יעיל ומדויק יותר בניבוי התוחלת של מחיר הדירה? נמק.
- ג. האם נוכל לנבא על סמך משוואת הרגרסיה הנ"ל את מחיר הדירה עבור שטח של 40 מ"ר?
- ד. מהו האומדן לשינוי בתוחלת מחיר הדירה עבור עליה ביחידה אחת של שטח הדירה?
- ה. מהו האומדן להפרש הצפוי בתוחלת מחיר הדירה בין שטח דירה ששווה ל-120 מ"ר לבין שטח דירה ששווה ל-90 מ"ר?
- ו. אמוד את שונות מחיר הדירה עבור שטח של 100 מ"ר. על איזה הנחה מהנחות המודל מבוסס אומדן זה?

1.19 רב"ס לערכי Y עבור ערך מסוים של X

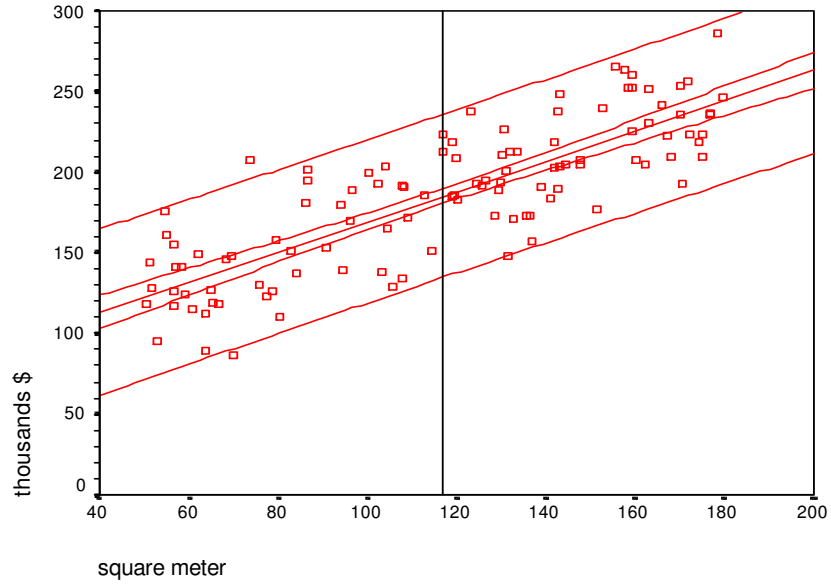
רב"ס שמטרתו לאמוד את כל טווח ערכי Y (y_0) עבור ערך מסוים (x_0).
נוסחת הרב"ס:

$$\hat{\mu}_0 \pm t_{n-2} \left(\frac{\alpha}{2} \right) \cdot \sqrt{MSRES \cdot \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{SSX} \right)}$$

ניתן לראות כי גם רב"ס זה בנוי סביב האומדן הנקודתי לתוחלת ערכי Y עבור ערך ה-X המסוים ($\hat{\mu}_0$).

ההבדל בין רב"ס לערכי Y לבין הרב"ס לתוחלת ערכי Y בא לידי ביטוי בטעות התקן. ניתן לראות כי טעות התקן של הרב"ס לערכי Y גדולה יותר מטעות התקן של הרב"ס לתוחלת ערכי Y. כאשר כל יתר הפרמטרים נשארים קבועים רב"ס זה יהיה רחב יותר מן הרב"ס לתוחלת.

התרשים הבא מתאר רב"ס לתוחלת ולערך המשתנה התלוי וממחיש זאת בבירור:



:

? חשב רב"ס ברמת ביטחון של 95% למחיר הדירה עבור שטח דירה של 100 מ"ר. מה ההבדל בין רב"ס זה לרב"ס הקודם?