

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

4.....	פרק 1 - הסקה סטטיסטית - הקדמה
7.....	פרק 2 - התפלגות הדגימה
7.....	ממוצע המדגם ומשפט הגבול המרכזי
15.....	התפלגות סכום תצפיות המדגם ומשפט הגבול המרכזי
18.....	פרק 3 - מושגים בסיסיים באמידה
25.....	פרק 4 - אמידה נקודתית
25.....	אומד חסר הטיה
32.....	אומד חסר הטיה יעיל ביותר - MVUE
35.....	פרק 5 - רווח סמך לתוחלת (ממוצע האוכלוסייה)
35.....	רווח סמך כששונות האוכלוסייה ידועה
41.....	קביעת גודל מדגם באמידת תוחלת עם שונות אוכלוסייה ידועה
44.....	רווח סמך לתוחלת (ממוצע האוכלוסייה) כששונות האוכלוסייה אינה ידועה
50.....	פרק 6 - רווח סמך להפרש תוחלות ממדגמים בלתי תלויים
50.....	כששונות האוכלוסייה אינן ידועות אך שוות והמדגמים בלתי תלויים
53.....	פרק 7 - רווח סמך לתוחלת ההפרש במדגם מזווג
56.....	פרק 8 - בדיקת השערות על פרמטרים
56.....	הקדמה
59.....	טעויות בבדיקת השערות
61.....	פרק 9 - בדיקת השערות על תוחלת (ממוצע)
61.....	כאשר שונות האוכלוסייה ידועה
66.....	מובהקות התוצאה (p-value) בבדיקת השערות על תוחלת עם שונות ידועה
71.....	בדיקת השערות על תוחלת (ממוצע) כאשר שונות האוכלוסייה אינה ידועה
76.....	מובהקות התוצאה (p-value) כאשר שונות האוכלוסייה לא ידועה
80.....	ניתוח פלטי SPSS
85.....	הקשר בין רווח סמך לבדיקת השערות על תוחלת
88.....	פרק 10 - בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים
88.....	כששונות האוכלוסייה לא ידועות ומניחים שהן שוות
92.....	ניתוח פלטים
99.....	פרק 11 - בדיקת השערות על תוחלת ההפרשים במדגמים מזווגים (תלויים)
99.....	בדיקת השערות למדגמים מזווגים
105.....	מדגמים מזווגים - ניתוח פלטים
111.....	פרק 12 - הקשר בין רווח סמך לבדיקת השערות על הפרש תוחלות
115.....	פרק 13 - בדיקת השערות על שונות
115.....	בדיקת השערות על שתי שונות

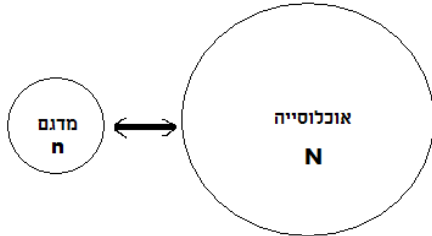
- פרק 14-שאלות מסכמות בבדיקת השערות על פרמטרים 121
- שאלות מסכמות בבדיקת השערות על פרמטרים 121
- שאלות מסכמות בסגנון רב ברירה (אמריקאיות) על בדיקת השערות 125
- פרק 15 - מבחני חי בריבוע 139
- מבחן חי בריבוע לאי תלות בין משתנים 139
- פליטים על מבחן לאי תלות 144

פרק 1 - הסקה סטטיסטית - הקדמה

רקע:

אוכלוסייה – קבוצה שאליה מפנים שאלה מחקרית.

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



מדגם – חלק מתוך האוכלוסייה.

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

במקרים רבים אין אפשרות לחקור את כל האוכלוסייה כיוון שאין גישה לכולה, היא גדולה מידי, אנו מוגבלים בזמן ובאמצעים טכניים ולכן מבצעים מדגם במטרה לבצע הסקה סטטיסטית מהמדגם לאוכלוסייה.

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

סטטיסטי – גודל המחושב על המדגם.

פרמטר – גודל המתאר את האוכלוסייה.

הסימונים לפרמטר וסטטיסטי הם שונים

למשל:

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

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

דוגמה (פתרון בהקלטה):

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

א. מי האוכלוסייה?

ב. מה המשתנה?

ג. מה הפרמטרים?

ד. מהו גודל המדגם?

ה. מהו הסטטיסטי שמתכננים להוציא מהמדגם?

ו. האם הפרמטר או הסטטיסטי הוא משתנה מקרי?

תרגילים :

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

א. מי האוכלוסייה?

ב. מה המשתנה?

ג. מהם הפרמטרים?

ד. מהו גודל המדגם?

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

נגדיר את x להיות מספר המקלטים של משפחה אקראית.

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

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

א. מיהי האוכלוסייה ומהו המשתנה הנחקר?

ב. מהו הסטטיסטי שיילקח מהמדגם ומה סימונו?

3. נתון כי 20% מהשכירים במדינה הם אקדמאיים. נבחרו באקראי 10 שכירים באותה אוכלוסייה

ומתכננים לפרסם את מספר האקדמאיים שנדגמו.

א. מהי האוכלוסייה ?

ב. מה המשתנה באוכלוסייה?

ג. מהם הפרמטרים?

ד. מהו הסטטיסטי?

פרק 2 - התפלגות הדגימה

ממוצע המדגם ומשפט הגבול המרכזי

רקע:

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

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

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

ממוצע האוכלוסייה נסמן ב μ (נקרא גם תוחלת).

שונות אוכלוסייה נסמן ב- σ^2 .

סטיית תקן של אוכלוסייה: σ .

א. תכונות התפלגות

ממוצע כל ממוצעי המדגם האפשריים שווה לממוצע האוכלוסייה:

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

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

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

יש יחס הפוך בין גודל המדגם לבין שונות ממוצעי המדגם.

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

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

דוגמה: (פתרון בהקלטה)

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

א. מי אוכלוסיית המחקר? מהו המשתנה הנחקר?

ב. מהם הפרמטרים של האוכלוסייה?

ג. מה התוחלת ומהי סטיית התקן של ממוצע המדגם?

ב. דגימה מהתפלגות נורמאלית

אם נדגום מתוך אוכלוסייה שהמשתנה בה מתפלג נורמאלית עם ממוצע μ ושונות σ^2 ממוצע המדגם גם יתפלג נורמאלית:

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

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

דוגמה: (פתרון בהקלטה)

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

ג. משפט הגבול המרכזי

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

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

דוגמה: (פתרון בהקלטה)

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

תרגילים :

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

- א. מי האוכלוסייה?
- ב. מה המשתנה?
- ג. מהם הפרמטרים?
- ד. מהו גודל המדגם?
- ה. מהו תוחלת ממוצע המדגם?
- ו. מהי טעות התקן?

2. להלן התפלגות מספר מקלטי הטלויזיה למשפחה בישוב מסוים :

מספר מקלטים	מספר המשפחות
0	500
1	2500
2	3500
3	3000
4	500
	סך הכול $N = 10000$

נגדיר את x להיות מספר המקלטים של משפחה אקראית.

א. בנו את פונקציית ההסתברות של x .

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

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

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

4. משקל תינוק ביום היוולדו מתפלג נורמאלית עם ממוצע 3400 גרם וסטיית תקן של 400 גרם
א. מה ההסתברות שתינוק אקראי בעת הלידה ישקול פחות מ-3800 גרם?

נתון כי ביום מסוים נולדו 4 תינוקות.

ב. מה ההסתברות שהמשקל הממוצע שלהם יעלה על 4 ק"ג?

ג. מה ההסתברות שהמשקל הממוצע של התינוקות יהיה מתחת ל-2.5 ק"ג?

ד. מה ההסתברות שהמשקל הממוצע של התינוקות יהיה רחוק מהתוחלת בלא יותר מ-50 גרם?

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

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

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

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

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

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

6. הזמן הממוצע שלוקח לאדם להגיע לעבודתו 30 דקות עם שונות של 16 דקות רבועות. האדם נוסע לעבודה במשך שבוע 5 פעמים. לצורך פתרון הניחו שזמן הנסיעה לעבודה מתפלג נורמאלית.

א. מה ההסתברות שבמשך שבוע משך הנסיעה הממוצע יהיה מעל 33 דקות?

ב. מהו הזמן שבהסתברות של 90% ממוצע משך הנסיעה השבועי יהיה גבוה ממנו?

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

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

7. נפח היין בבקבוק מתפלג נורמאלית עם תוחלת של 750 סמ"ק וסטיית תקן של 10 סמ"ק.

א. בארגז 4 בקבוקי יין. מה ההסתברות שהנפח הממוצע של הבקבוקים בארגז יהיה בדיוק 755 סמ"ק?

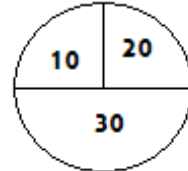
ב. בארגז 4 בקבוקי יין. מה ההסתברות שהנפח הממוצע של הבקבוקים בארגז יהיה יותר מ-755 סמ"ק?

ג. בארגז 4 בקבוקי יין. מה ההסתברות שהנפח הממוצע של הבקבוקים בארגז יהיה לפחות 755 סמ"ק?

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

8. משתנה מתפלג נורמאלית עם תוחלת 80 וסטיית תקן 4 .
 א. מה ההסתברות שממוצע המדגם יסטה מתוחלתו בלא יותר מיחידה כאשר גודל המדגם הוא 9?
 ב. מה ההסתברות שממוצע המדגם יסטה מתוחלתו בלא יותר מיחידה שגודל המדגם הוא 16?
 ג. הסבר את ההבדל בתשובות של שני הסעיפים.

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



- אדם מסובב את הרולטה וזוכה בסכום הרשום על הרולטה.
 א. בנו את פונקציית ההסתברות של סכום הזכייה במשחק בודד.
 ב. מה התוחלת ומה השונות של סכום הזכייה?
 ג. אם האדם ישחק את המשחק 5 פעמים מה התוחלת ומה השונות של ממוצע סכום הזכייה בחמשת המשחקים?
 ד. אם האדם משחק את המשחק 50 פעם מה ההסתברות שבסה"כ יזכה ב-1050 ₪ ומעלה?
 10. לפי הערכות הלשכה המרכזית לסטטיסטיקה השכר הממוצע במשק הוא 8000 ₪ עם סטיית תקן של 3000 ₪. מה ההסתברות שבמדגם מקרי של 100 עובדים השכר הממוצע יהיה יותר מ-8500 ₪?

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

12. אורך צינור שמפעל מייצר הינו עם ממוצע של 70 ס"מ וסטיית תקן של 10 ס"מ .
 א. נלקחו באקראי 100 מוטות, מה ההסתברות שממוצע אורך המוטות יהיה בין 68 ל 78 ס"מ?
 ב. יש לחבר 2 בניינים באמצעות מוטות. המרחק בין שני הבניינים הינו 7200 ס"מ. מה ההסתברות ש 100 המוטות יספיקו למלאכה?
 ג. מה צריך להיות גודל המדגם המינימאלי, כדי שבהסתברות של 5% ממוצע המדגם יהיה קטן מ-69 ס"מ. העזר במשפט הגבול המרכזי.

13. נתון משתנה מקרי בדיד בעל פונקציית ההסתברות הבאה :

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

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

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

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

א. 0

ב. 0.5

ג. 1

ד. לא ניתן לדעת.

15. נתון ש X מתפלג כלשהו עם תוחלת μ ושונות σ^2 .

החליטו לבצע מדגם בגודל 200 מתוך ההפלגות הנתונה לפי משפט הגבול המרכזי מתקיים ש :

(בחר בתשובה הנכונה)

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

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

ג. $\bar{X} \sim N(\mu, \sigma^2)$

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

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

(בחר בתשובה הנכונה)

א. μ ו- \bar{X} יהיו משתנים מקריים.

ב. μ יהיה משתנה מקרי ו \bar{X} קבוע.

ג. \bar{X} יהיה משתנה מקרי ו μ קבוע.

ד. μ ו \bar{X} יהיו קבועים.

פתרונות:שאלה 2

.א

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

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

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

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

שאלה 3

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

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

שאלה 4

.א 0.8413

.ב 0.0013

.ג 0

.ד 0.1974

שאלה 6

.א 0.0465

.ב 27.71

.ג 0.2628

שאלה 7

.א 0

.ב 0.1587

.ג 0.1587

.ד 0.5

שאלה 8

א. 0.5468

ב. 0.6826

שאלה 9

א.

30	20	10	
0.5	0.25	0.25	P(x)

ב. התוחלת: 22.5

השונות: 68.75

ג. התוחלת: 22.5

השונות: 13.75

ד. 0.8997

שאלה 10

0.0475

שאלה 11

0.1814

שאלה 12

א. 0.9772

ב. 0.0228

ג. 271

שאלה 14

התשובה ב

שאלה 15

התשובה ד

שאלה 16

התשובה ג

התפלגות סכום תצפיות המדגם ומשפט הגבול המרכזי

רקע:

$$T = \sum_{i=1}^n X_i \quad \text{כעת נדון בסטטיסטי המבטא את סכום התצפיות במדגם}$$

כאשר כל התצפיות נדגמו באקראי מאותה אוכלוסייה.

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

א. התוחלת והשונות של סכום התצפיות:

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

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

ב. דגימה מתוך התפלגות נורמלית:

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

$$Z = \frac{T - n\mu}{\sqrt{n\sigma^2}} \quad \text{אזי} \quad \text{אם } X \sim N(\mu, \sigma^2)$$

ג. משפט הגבול המרכזי:

$$E(X) = \mu$$

$$V(X) = \sigma^2 \quad \text{אם } X \text{ מתפלג כלשהו וידוע}$$

אזי עבור מדגם מספיק גדול (לפחות 30)

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

דוגמה: (פתרון בהקלטה)

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

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

תרגילים:

1. המשקל באוכלוסייה מסוימת מתפלג נורמאלית עם תוחלת של 60 ק"ג וסטיית תקן של 10 ק"ג.
 - א. מה הסיכוי שאדם אקראי מהאוכלוסייה ישקול מתחת ל-65 ק"ג?
 - ב. מה הסיכוי שהמשקל הממוצע של 4 אנשים אקראיים יהיה מתחת ל-65 ק"ג?
 - ג. מה הסיכוי שהמשקל הכולל של 4 אנשים אקראיים יהיה מתחת ל-240 ק"ג?

2. נפח יין בבקבוק מתפלג נורמאלית עם תוחלת של 750 מ"ל וסטיית תקן של 20 מ"ל. אדם קנה מארז של 4 בקבוקי יין.
 - א. מהי התוחלת ומהי סטיית התקן של נפח היין במארז?
 - ב. את היין שבמארז האדם מזג לכלי שקיבולתו 3.1 ליטר. מה ההסתברות שהיין יגלוש מהכלי?
 - ג. אם לא היה נתון שנפח היין מתפלג נורמאלית. האם התשובה לסעיף א הייתה משתנה? האם התשובה לסעיף ב הייתה משתנה?

3. בספר כלשהו 500 עמודים. קצב הקריאה הממוצע הוא עמוד אחד ב 4 דקות עם סטיית תקן של 1 דקות.
 - א. מה ההסתברות לסיים את הפרק הראשון (40 עמודים) תוך שעתיים וחצי?
 - ב. מהו האחוזון ה-95 לזמן סיום קריאת הספר?

פתרונות:**שאלה 1**

א. 0.6915

ב. 0.8413

ג. 0.5

שאלה 2

א. תוחלת 3000 מ"ל וסטיית תקן 40 מ"ל

ב. 0.0062.

פרק 3 - מושגים בסיסיים באמידה

רקע:

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

כמו ממוצע הגבהים בקרב מתגייעים לצה"ל- μ .

כמו פרופורציית התומכים בממשלה בקרב אזרחי המדינה - p .

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

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

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

דוגמה: (פתרון בהקלטה)

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

מה הפרמטר בדוגמה זו?

מהי טעות האמידה של ערוץ 10?

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

להלן פרמטרים מרכזיים והאומדים שלהם:

ממוצע האוכלוסייה: μ

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

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

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

פרופורציה באוכלוסייה: p

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

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

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

שונות האוכלוסייה: σ^2

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

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

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

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

דוגמה: (פתרון בהקלטה)

נדגמו 10 משפחות בתל אביב ונבדק עבור כל משפחה מספר הילדים שלה. להלן התוצאות שהתקבלו:

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

אמדו באמצעות אומדים חסרי הטיה את הפרמטרים הבאים:

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

תרגילים:

1. מתוך 500 טירונים נמצאו 120 בעלי שברי הליכה. נתון שהסיכוי שטירון יהיה עם שבר הליכה הוא 0.25.
- א. מהי האוכלוסייה המוצגת בשאלה? מהם הפרמטרים שלה?
- ב. מהי טעות התקן של האומדן כשהמדגם בגודל 500?
- ג. מהו האומדן לפרמטר?
- ד. מהי טעות האמידה?
2. לפי נתוני היצרן מקרר צורך בממוצע 2400 וואט לשעה עם סטיית תקן של 500 וואט לשעה. במדגם של 25 מקררים של היצרן התקבל ממוצע של 2342 וואט לשעה.
- א. מהי האוכלוסייה המוצגת בשאלה? מהם הפרמטרים שלה?
- ב. מהי טעות התקן של האומדן?
- ג. מהו האומדן לפרמטר?
- ד. מהי טעות האמידה?
3. נדגמו עשרה מתגייסים לצה"ל. גובהם נמדד בס"מ. להלן התוצאות שהתקבלו:
- 168, 184, 192, 171, 180, 177, 187, 168, 177 ו-175.
- א. מצא אומדן חסר הטיה לגובה הממוצע של מתגייסי צה"ל.
- ב. מצא אומדן חסר הטיה לשונות הגבהים של מתגייסי צה"ל.
- ג. מצא אומדן חסר הטיה לפרופורציות המתגייסים בגובה של לפחות 180 ס"מ.
4. נדגמו 20 שכירים באקראי. עבור כל שכיר נמדד השכר באלפי שקלים. להלן התוצאות שהתקבלו:
- $$\sum_{i=1}^{20} X_i^2 = 1502.2 \quad \sum_{i=1}^{20} X_i = 162$$
- א. אמדו את השכר הממוצע של השכירים במשק.
- ב. אמדו את סטיית התקן של שכר השכירים במשק.

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

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

א. 3

ב. 2.5

ג. 1.581

ד. אין מספיק נתונים לדעת.

7. במדגם מקרי, מתי סכום ריבועי הסטיות מהממוצע, $\sum_{i=1}^n (x_i - \bar{x})^2$, מחולק ב- $n-1$?

א. כאשר n קטן.

ב. כאשר תצפיות המדגם אינן בלתי תלויות.

ג. כאשר האוכלוסייה אינה מתפלגת נורמאלית.

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

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

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

$\sigma^2 = 64$. טעות התקן של האומדן ל- μ היא:

א. 16

ב. 8

ג. 4

ד. 2

9. מהו אומד חסר הטיה?

- א. אומד שערכו שווה לממוצע התפלגות הדגימה שלו.
- ב. אומד שערכו שווה לערך הפרמטר באוכלוסייה.
- ג. אומד שממוצע התפלגות הדגימה שלו שווה לערך הפרמטר באוכלוסייה.
- ד. אומד שהסיכוי שערכו יהיה גבוה מערך הפרמטר באוכלוסייה שווה לסיכוי שיהיה נמוך ממנו.

פתרונות:**שאלה 3**

א. 177.9

ב. 64.1

ג. 0.4

שאלה 4

א. 8.1

ב. 3.16

שאלה 5

התשובה היא ד.

שאלה 6

התשובה היא ג.

שאלה 7

התשובה היא ד.

שאלה 8

התשובה היא ד.

שאלה 9

התשובה היא ג.

פרק 4 - אמידה נקודתית

אומד חסר הטיה

רקע:

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

דוגמה: (פתרון בהקלטה)

המשתנה X הוא בעל פונקציית ההסתברות הבאה:

	3	2	1	X
הסתברות	4θ	$1-6\theta$	2θ	

מעוניינים לאמוד את θ על סמך שתי תצפיות מההתפלגות: X_1 ו- X_2

א. הראו שהאומד $T_1 = \frac{2X_1 + X_2}{2}$ הוא אומד מוטה ל- θ .

- הטיה של אומד היא: $E(\hat{\theta}) - \theta$, כמובן שלאומד חסר הטיה אין הטיה.

ב. מהי ההטיה של האומד T_1 .

ג. תקנו את T_1 כך שיהיה אומד חסר הטיה.

- אם יש שני אומדים חסרי הטיה עדיף זה עם השונות היותר קטנה.

ד. מוצא האומד הבא: $T_3 = 1.5X_1 - X_2 - 1$. האם הוא עדיף על האומד שהצעת בסעיף ג?

- אם $\hat{\theta}$ אומד חסר הטיה ל- θ אז $g(\hat{\theta})$ יהיה אומד חסר הטיה עבור $g(\theta)$ רק אם g תהיה לינארית.

ה. מצאו אומד חסר הטיה ל: $P(X=3)$.

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{\sum x_i^2 - n\bar{x}^2}{n-1} : \sigma^2 \quad \bullet \quad \text{אומד חסר הטיה לשונות האוכלוסייה}$$

ו. מצאו אומד חסר הטיה לשונות של X .

תזכורות חשובות:

• אם $Y = aX + b$ אזי:

$$\sigma_Y = |a| \sigma_X \quad V(Y) = a^2 \cdot V(X) \quad E(Y) = aE(X) + b$$

• אם X_1, X_2, \dots, X_n משתנים מקרים אזי:

$$E(T) = E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n)$$

אם X_1, X_2, \dots, X_n משתנים מקריים בלתי תלויים בזוגות, אזי:

$$V(T) = V(X_1 + X_2 + \dots + X_n) = V(X_1) + V(X_2) + \dots + V(X_n)$$

תרגילים:

1. הציון במבחן מסוים של תלמידי כיתה ח' הנו משתנה מקרי בעל תוחלת μ וסטטיית תקן 10. כדי לאמוד את התוחלת - μ , נלקח מדגם של 5 ציונים X_1, \dots, X_5 . שלושה חוקרים הציעו אומדים לתוחלת על סמך מדגם זה:

$$T_1 = \frac{X_1 + \dots + X_5}{5} \quad \text{חוקר א' הציע:}$$

$$T_2 = \frac{2X_1 - X_3 + X_4}{2} \quad \text{חוקר ב' הציע:}$$

$$T_3 = \frac{2X_1 + X_3}{2} \quad \text{חוקר ג' הציע:}$$

- א. איזה מן האומדים הוא חסר הטיה?
 ב. הצע תיקון לאומד המוטה כך שיהיה חסר הטיה.
 ג. במדגם התקבלו הציונים הבאים: 65, 78, 58, 82, 100. חשבו את האומדנים המתקבלים עבור האומדים חסרי ההטיה.
 ד. איזה מבין שני האומדים חסרי ההטיה עדיף? נמקו.

2. כדי לאמוד את המשקל הממוצע של הנשים בארה"ב, נבחר מדגם של $2n$ נשים. נסמן את שונות הגובה ב- σ^2 . הוצעו שני אומדים לממוצע המשקל על סמך מדגם זה:

$$T_1 = \frac{1}{n} \sum_{i=1}^n X_i \quad T_2 = \frac{1}{2n} \sum_{i=1}^{2n} X_i$$

- א. בדקו לגבי כל אומד אם הוא בלתי מוטה.
 ב. איזה אומד עדיף? נמקו.

3. $X \sim B(n, p)$ כלומר X הינו משתנה מקרי המתפלג בינומית עם פרמטר P (סיכוי להצלחה בניסיון בודד) במדגם בגודל n .
 א. פתחו אומד חסר הטיה ל- P .
 ב. מהו אומד חסר הטיה לסיכוי לכישלון בניסיון בודד.
 ג. מהו אומד חסר הטיה ל- $E(X)$.
 ד. מצאו אומד חסר הטיה ל- $E(X^2)$.

4. בתיק מניות שתי מניות . מספר המניות שיעלו ביום מסוים הוא משתנה מקרי התלוי בפרמטר פרמטר לא ידוע θ , $0 \leq \theta \leq 2$.

פונקציית ההסתברות של X – מספר המניות שיעלו ביום מסוים :

$$P(X=0) = 1 - \frac{\theta}{2} \quad P(X=1) = \frac{\theta}{3} \quad P(X=2) = \frac{\theta}{6}$$

- א. מצאו אומד בלתי מוטה ל- θ שמתבסס על מספר המניות שיעלו ביום מסוים.
 ב. מצאו אומד בלתי מוטה ל- θ שמתבסס על מספר המניות שעלו ביום במשך שלושה ימים X_1, X_2, X_3 (לכל אחד מהם אותה התפלגות כנ"ל והם בלתי תלויים).

5. בקרב המטפלות בת"א מספר התינוקות שבטיפולן הוא משתנה מיקרי בעל התפלגות התלויה

בפרמטר θ באופן הבא :

הסיכוי שמטפלת תטפל בתינוק אחד בלבד הוא 3θ ,

הסיכוי שמטפלת תטפל ב-2 תינוקות הוא $1 - 4\theta$,

הסיכוי שמטפלת תטפל ב-3 תינוקות הוא θ .

- במדגם מיקרי של 4 מטפלות מת"א, נמצא כי שתיים מהם מטפלות בתינוק אחד בלבד, אחת מהן בשנים ואחת השלושה תינוקות.
 א. מצא אומד חסר הטיה לפרמטר θ על סמך תצפית בודדת.
 ב. מצאו אומד חסר הטיה לפרמטר θ על סמך 4 תצפיות.
 ג. מהו האומדן לפרמטר θ על סמך תוצאות המדגם.
 ד. מצאו אומד חסר הטיה לסיכוי שלמטפלת בת"א תטפל בתינוק בודד אחד.
 ה. מצאו אומדים חסרי הטיה לתוחלת ולשונות של מספר התינוקות בטיפול אצל מטפלת מת"א. חשבו אומדנים.

6. קבע אילו מהטענות הבאות נכונות :

- א. אם T הוא אומד בלתי מוטה עבור פרמטר θ , אז $5T$ אומד בלתי מוטה עבור הפרמטר 5θ .
 ב. אם T הוא אומד בלתי מוטה עבור פרמטר θ , אז T^2 אומד בלתי מוטה עבור הפרמטר θ^2 .

7. במפעל שתי מכונות המייצרות מוצרים. במכונה הראשונה ההסתברות שמכשיר תקין היא p , מכונה השנייה ההסתברות שמכשיר תקין היא $2p$. דוגמים 20 מכשירים מהייצור של כל מכונה. נסמן ב- X את מספר המכשירים התקינים שיוצרו על ידי המכונה הראשונה, Y - מספר המכשירים התקינים שיוצרו על ידי המכונה השנייה. איזה מבין האומדים הבאים אינו אומד חסר הטיה ל- p ?

א. $\frac{X}{20}$

ב. $\frac{Y}{20}$

ג. $\frac{X+Y}{60}$

ד. $\frac{2X+Y}{80}$

8. יהי T_1 ו- T_2 אומדים חסרי הטיה ובלתי תלויים לפרמטר θ .

א. מצא אומד חסר הטיה ל- θ^2 המתבסס על T_1 ו- T_2 .

ב. מצא אומד חסר הטיה ל- $\theta(1-\theta)$ המתבסס על T_1 ו- T_2 .

9. נתון ש X הינו משתנה מקרי עם תוחלת μ ושוונות σ^2 . נדגמו n תצפיות בלתי תלויים

מאותה

אוכלוסייה.

א. הראה ש $\sum_{i=1}^n p_i x_i$ אומד חסר הטיה ל μ כאשר $\sum_{i=1}^n p_i = 1$.

ב. נתבונן במכפלת שתי התצפיות הראשונות $X_1 \cdot X_2$ הראה שהוא אומד חסרי הטיה ל-

μ^2 .

פתרונות:**שאלה 1**

א. T_2 ו- T_1

ב. $\frac{2}{3}T_3$

ג. $T_2 = 110 T_1 = 76.6$

ד. T_1

שאלה 2

ב. T_2

שאלה 3

א. $\frac{x}{n}$

ב. $1 - \frac{x}{n}$

ג. x

שאלה 4

א. $\frac{3x}{2}$

ב. $\frac{3\bar{x}}{2}$

שאלה 5

א. $1 - \frac{x}{2}$

ג. 0.125

ה. לשונות 0.917

שאלה 6

א. נכון.

ב. לא נכון.

שאלה 7

תשובה: ב

שאלה 8

א. $T_1 \cdot T_2$

ב. $T_1 - T_1 \cdot T_2$

שאלה 9

הוכחה

אומד חסר הטיה יעיל ביותר - MVUE
(Minimum- variance unbiased estimator)

רקע:

T יהיה MVUE אם מתקיים ש-T אומד חסר הטיה ל- θ , ובנוסף מתקיים ש: $V(T) \leq V(\hat{\theta})$ לכל $\hat{\theta}$ חסר הטיה אחר.

דוגמה: (פתרון בהקלטה)

לרשת חנויות ישנם שני סניפים. מספר הלקוחות הנכנסים לכל סניף ביום מתפלג פואסונית עם קצב של λ בסניף A וקצב של 2λ בסניף B. נדגמו n ימים מכל סניף ונבדק בכל יום:

X_i - מספר הלקוחות שנכנסו לסניף A ביום i .

Y_j - מספר הלקוחות שנכנסו לסניף B ביום j .

על מנת לאמוד את λ מוצע האומד: $\alpha \bar{X} + \beta \bar{Y}$.

א. מה התנאי שצריך להתקיים על α ו- β כדי שהאומד יהיה חסר הטיה?

ב. מה צריך להיות α ו- β כדי שהאומד יהיה גם בעל שונות מינימלית?

תרגילים:

1. T_1 ו- T_2 הינם אומדים חסרי הטיה ובלתי תלויים לפרמטר θ .

$$T = aT_1 + bT_2 : \text{ כמו כן נגדיר}$$

א. מה צריך להיות התנאי על a ו- b כדי ש- T יהיה אומד חסר הטיה?

ב. σ_1^2 ו- σ_2^2 הם השונות של T_1 ו- T_2 בהתאמה. מצאו את a ו- b כך ש- T יהיה אומד חסר הטיה ל θ ובעל שונות מינימלית.

2. במפעל 3 מכונות המייצרות את אותו חלק. תוחלת הקוטר של החלקים המיוצרים בכל מכונה זהה אומנם השונות של כל מכונה שונות ומקיימות:

$$\sigma_2^2 = 2\sigma_1^2 \quad \sigma_3^2 = 3\sigma_1^2$$

הוחלט לדגום n חלקים מכל מכונה ולחשב את ממוצע הקוטר המתקבל.

\bar{X}_i - יהיה הממוצע המתקבל במכונה i .

יהי $W = \sum_{i=1}^3 a_i \bar{X}_i$ האומד לתוחלת קוטר החלקים המיוצרים על ידי מכונה כלשהי.

א. מה התנאי שצריך להתקיים על המשקלים a_i כדי שהאומד המוצע יהיה בלתי מוטה?

ב. נניח ש $a_1 = a_2$. מה במקרה זה המשקלים המביאים את האומד להיות MVUE?

פתרונות :**שאלה 1:**

$$a + b = 1. \text{א.}$$

$$b = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}, \quad a = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}. \text{ב.}$$

שאלה 2:

$$\sum_{i=1}^3 a_i = 1. \text{א.}$$

$$a_1 = a_2 = 0.4 \quad \text{ב.}$$

$$a_3 = 0.2$$

פרק 5 - רווח סמך לתוחלת (ממוצע האוכלוסייה)

רווח סמך כששונות האוכלוסייה ידועה

רקע:

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

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

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

A - גבול התחתון של רווח הסמך

B - הגבול העליון של רווח הסמך

$L = B - A$ - אורך רווח הסמך

דוגמה : (פתרון בהקלטה)

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

מהי אוכלוסיית המחקר?

מה המשתנה באוכלוסייה?

מה הפרמטר שהחוקר רצה לאמוד?

מהו רווח הסמך?

מה אורך רווח הסמך?

מהי רמת הביטחון של רווח הסמך?

בפרק זה נרצה לבנות רווח סמך לתוחלת (μ) במקרה ש σ^2 (שוונות האוכלוסייה) ידועה

הפרמטר שנרצה לאמוד : μ

האומד נקודתי : \bar{x}

התנאים לבניית רווח הסמך :

1 $X \sim N$ או $n \geq 30$

2 σ^2 (שוונות האוכלוסייה) ידועה

הנוסחה לרווח הסמך :

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

דוגמה : (פתרון בהקלטה)

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

מעוניינים לאמוד את תוחלת חיי סוללה.

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

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

שגיאת האמידה המקסימלית:

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

ε -נותן את שגיאת האמידה המקסימלית, דבר שנקרא גם טעות סטטיסטית, טעות דגימה.

דוגמה: (פתרון בהקלטה)

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

קשרים מתמטיים ברווח הסמך:

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

תרגילים :

1. חוקר התעניין לאמוד את השכר הממוצע במשק. על סמך מדגם הוא קבע שבביטחון של 95% כי השכר הממוצע במשק נע בין 9200 ל-9800.
 - א. מי האוכלוסייה במחקר?
 - ב. מה המשתנה הנחקר?
 - ג. מה הפרמטר שאותו רוצים לאמוד?
 - ד. מה רווח הסמך לפרמטר?
 - ה. מהי רמת הסמך לפרמטר?
 - ו. מה אורך רווח הסמך?
 - ז. מה הסיכוי שטעות הדגימה תעלה על 300 ₪?

2. מעוניינים לאמוד את התפוקה היומית הממוצעת של מפעל מסוים ברמת סמך של 95%. במדגם אקראי של 100 ימים התקבלה תפוקה ממוצעת 4950 מוצרים ביום. לצורך פתרון הנח שסטיית התקן האמתית ידועה ושווה 150 מוצרים ביום. בנה את רווח הסמך.
 - א. בנו רווח סמך ברמת סמך של 90% לאורך החיים הממוצע של מכשיר.
 - ב. בנו רווח סמך ברמת סמך של 95% לאורך החיים הממוצע של מכשיר.
 - ג. הסבר כיצד ומדוע השתנה רווח הסמך.

3. מעוניינים לאמוד את ממוצע אורך החיים של מכשיר. מנתוני היצרן ידוע שאורך החיים מתפלג נורמאלי עם סטיית תקן של 20 שעות. נדגמו 25 מכשירים ונמצא כי ממוצע אורך החיים שלהם היה 230 שעות.
 - א. בנו רווח סמך ברמת סמך של 90% לאורך החיים הממוצע של מכשיר.
 - ב. בנו רווח סמך ברמת סמך של 95% לאורך החיים הממוצע של מכשיר.
 - ג. הסבר כיצד ומדוע השתנה רווח הסמך.

4. דגמו 200 עובדים מהמשק הישראלי. השכר הממוצע שלהם היה 9700 ₪. נניח שסטיית התקן של השכר במשק היא 3000 ₪.
 - א. בנו רווח סמך ברמת סמך של 95% לתוחלת השכר במשק.
 - ב. מה ניתן לומר בביטחון של 95% על הסטייה המרבית בין ממוצע המדגם לתוחלת השכר?
 - ג. מה היה צריך להיות גודל המדגם אם הינו רוצים להקטין את רווח הסמך ב-50%?
 - ד. אם היינו מגדילים את גודל המדגם ובונים רווח סמך באותה רמת סמך האם היה ניתן לטעון בביטחון רב יותר שרווח הסמך מכיל את הפרמטר?

5. בנו רווח סמך לממוצע הציונים של מבחן אינטליגנציה. ידוע שסטיית התקן היא 15 והמדגם מתבסס על 100 תצפיות. רווח הסמך שהתקבל הוא (99,105). שחזרו את :
 - א. ממוצע המדגם.
 - ב. שגיאת האמידה המקסימאלית.
 - ג. רמת הסמך.

6. זמן החלמה מאנגינה מתפלג עם סטיית תקן של יומיים. חברת תרופות מעוניינת לחקור אנטיביוטיקה חדשה שהיא פיתחה. במחקר השתתפו 60 אנשים שחלו באנגינה וקיבלו את האנטיביוטיקה החדשה. בממוצע הם החלימו לאחר 4 ימים.
- א. בנו רווח סמך לתוחלת זמן ההחלמה תחת האנטיביוטיקה החדשה ברמת סמך של 90%.
- ב. מה היה קורה לאורך רווח הסמך אם היה תקציב להגדלת גודל המדגם פי 4? הסבירו.
- ג. מה היה קורה לאורך רווח הסמך אם היינו בונים את רווח הסמך ברמת סמך גדולה יותר? הסבירו.
7. חוקר בנה רווח סמך לממוצע וקיבל את רווח הסמך הבא: $82 < \mu < 92$.
- נתון שסטיית התקן בהתפלגות שווה ל-10 ושהמדגם מתבסס על 16 תצפיות. התפלגות המשתנה היא נורמאלית.
- א. מהו ממוצע המדגם?
- ב. מהי רמת הסמך של רווח הסמך שנבנה?
- ג. מה הסיכוי ששגיאת האמידה באמידת ממוצע האוכלוסייה תעלה על 5 ?
8. חוקר בנה רווח סמך לתוחלת כאשר השונות בהתפלגות ידועה ברמת סמך של 95%. אם החוקר כעת יבנה על סמך אותם נתונים רווח סמך ברמת סמך קטנה מ-95%, מי מהמשפטים הבאים אינו יהיה נכון.
- א. אורך רווח הסמך החדש יהיה קטן יותר.
- ב. גודל המדגם יהיה כעת קטן יותר.
- ג. המרחק בין ממוצע המדגם לקצות רווח הסמך יהיו קטנים יותר ברווח הסמך החדש.
- ד. רמת הביטחון לבנות רווח הסמך החדש תהיה קטנה יותר.
9. חוקר בנה רווח סמך ל- μ וקיבל $48 < \mu < 54$ מה נכון בהכרח:
- א. $\mu = 51$
- ב. $\bar{X} = 6$
- ג. $\bar{X} = 51$
- ד. אורך רווח הסמך הינו 3.
10. איזה מהגורמים הבאים אינו משפיע על גודלו של רווח בר סמך, כאשר שונות האוכלוסייה ידועה? (בחר בתשובה הנכונה)
- א. רמת הביטחון.
- ב. סטיית התקן באוכלוסייה.
- ג. מספר המשתתפים.
- ד. סטיית התקן במדגם.

פתרונות :**שאלה 2**

$$4920.6 < \mu < 4979.4$$

שאלה 3

$$א. \quad 223.42 < \mu < 236.58$$

$$ב. \quad 222.16 < \mu < 237.84$$

שאלה 5

$$א. \quad 102$$

$$ב. \quad 3$$

$$ג. \quad 0.9544$$

שאלה 6

$$א. \quad 83.5 < \mu < 4.42$$

$$ב. \quad \text{יקטן פי 2}$$

$$ג. \quad \text{גדל}$$

שאלה 7

$$א. \quad 87$$

$$ב. \quad 5$$

$$ג. \quad 0.9544$$

שאלה 8

$$א. \quad 139$$

$$ב. \quad 21 < \mu < 25$$

שאלה 9

התשובה היא : ב

שאלה 10

התשובה היא : ג

קביעת גודל מדגם באמידת תוחלת עם שונות אוכלוסייה ידועה

רקע:

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

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

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

דוגמה: (פתרון בהקלטה)

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

תרגילים:

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

פתרונות :**שאלה 1**

780

שאלה 2

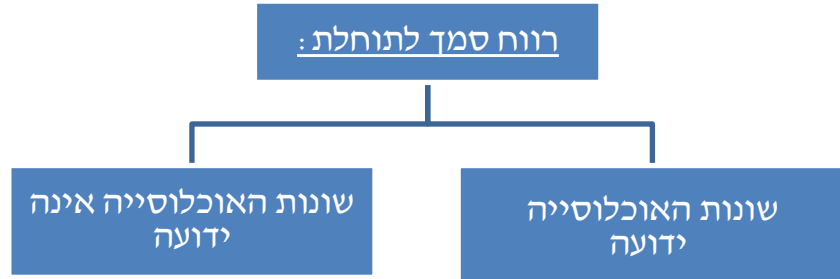
א. 139

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

רווח סמך לתוחלת (ממוצע האוכלוסייה) כששונות האוכלוסייה אינה ידועה

רקע:

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



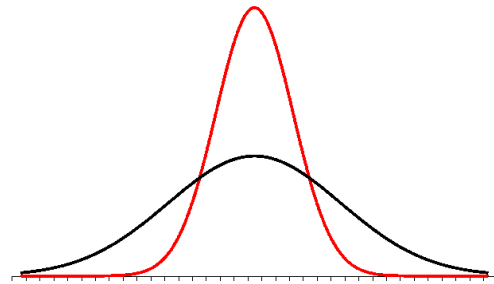
בפרק זה נעסוק במקרה ששונות האוכלוסייה (σ^2) אינה ידועה לנו. מקרה יותר פרקטי.

התנאי: $X \sim N$ או שהמדגם גדול

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

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

התפלגות T:



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

דוגמה : (פתרון בהקלטה)

הזמן שלוקח לפתור שאלה מסוימת בחשבון מתפלג אצל תלמידי 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: 89, 79, 84, 88, 84. הערה: לצורך פתרון הנח שקצב פעימות הלב מתפלג נורמאלית בקירוב.
 - א. בנו רווח סמך ברמת סמך של 95% לתוחלת הדופק של נוטלי התרופה הנ"ל.
 - ב. נתון שהדופק הממוצע ללא לקיחת התרופה הינו 70. לאור זאת, האם בביטחון של 95% התרופה משפיעה על הדופק?
 - ג. בהמשך לסעיף א, אם היינו בונים את רווח הסמך ברמת ביטחון של 99% כיצד הדבר היה משפיע על רווח הסמך?
2. במדגם שנעשה על 25 מתגייסים לצבא האמריקאי התקבל כי: גובה ממוצע של חייל הינו 178 ס"מ עם סטיית תקן $S=13$ ס"מ. בנו רווח סמך ברמת סמך של 90% לתוחלת גובה המתגייסים לצבא האמריקאי. מה יש להניח לצורך פתרון?
3. אדם מעוניין לאמוד את זמן הנסיעה הממוצע שלו לעבודה. לצורך כך הוא דוגם 5 ימים שזמן הנסיעה בהם בדקות הוא: 27, 34, 32, 40, 30.
 - א. ברמת ביטחון של 95% אמוד את זמן הנסיעה הממוצע. מהי ההנחה הדרושה לצורך פתרון?
 - ב. איך גודל רווח הסמך היה משתנה אם היו דוגמים עוד ימים?
4. ציוני מבחן אינטליגנציה מתפלגים נורמאלית. נדגמו 25 מבחנים והתקבל ממוצע ציונים 102 וסטיית תקן מדגמית 13.
 - א. בנו רווח סמך לממוצע הציונים באוכלוסייה ברמת ביטחון של 95%.
 - ב. חזרו על סעיף א' אם סטיית התקן הינה סטיית התקן האמתית של כלל הנבחנים.
 - ג. הסבירו את ההבדלים בין שני הסעיפים הנ"ל.
5. נשקלו 60 תינוקות אשר נולדו בשבוע ה-40 של ההיריון. המשקל נמדד בקילוגרמים. להלן התוצאות שהתקבלו: $\sum_{i=1}^{60} X_i = 195$, $\sum_{i=1}^{60} X_i^2 = 643.19$. בנו רווח סמך ברמת סמך של 95% לתוחלת משקל תינוק ביום היוולדו.
6. נדגמו 120 אנשים אקראיים מעל גיל 50. עבור כל אדם נבדק מספר שנות השכלתו. להלן התוצאות שהתקבלו: $\bar{x} = 13.8$, $S = 2$. בנו רווח סמך ברמת סמך של 96% לממוצע ההשכלה של אזרחים מעל גיל 50.

7. שני סטטיסטיקאים בנו רווח בר-סמך לאותו פרמטר 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 $S = 20$

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

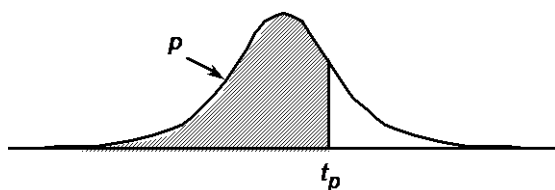
א. סטטיסטיקאי א

ב. סטטיסטיקאי ב

ג. אותו אורך רווח סמך לשני הסטטיסטיקאים.

ד. תלוי בתוצאות המדגם של כל סטטיסטיקאי.

8. נתון ש $X \sim N(\mu, \sigma^2)$ ביצעו מדגם בגודל 16 וקיבלו סטיית תקן מדגמית 10. אורך רווח הסמך שהתקבל הוא : 8.765. מהי רמת הביטחון של רווח הסמך?



P

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

פתרונות:**שאלה 1**

$$א. \quad 79.88 < \mu < 89.72$$

שאלה 4

$$א. \quad 96.63 < \mu < 107.37$$

$$ב. \quad 96.90 < \mu < 107.10$$

שאלה 5

$$3.149 < \mu < 3.351$$

שאלה 8

90%

פרק 6 - רווח סמך להפרש תוחלות ממדגמים בלתי תלויים

כששונות האוכלוסייה אינן ידועות אך שוות והמדגמים בלתי תלויים

רקע:

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

האומד נקודתי: $\bar{x}_1 - \bar{x}_2$

התנאים לבניית רווח הסמך:

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

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

3. מדגמים בלתי תלויים.

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

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

דרגות החופש: $d.f = n_1 + n_2 - 2$

רווח סמך:

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

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

דוגמה: (פתרון בהקלטה)

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

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

בנו רווח סמך ברמת ביטחון של 90% להפרש תוחלות ההכנסה בשני האזורים.
 הניחו שהשכר מתפלג נורמלית עם אותה שונות בכל אחד מהאזורים.
 פתרון : (1357,1643)

תרגילים:

1. נדגמו 15 ישראלים ו-15 אמריקאים.

כל הנדגמים נגשו למבחן IQ. להלן תוצאות המדגם:

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

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

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

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 - רווח סמך לתוחלת ההפרש במדגם מזווג

רקע:

מדגם מזווג: מדגם אחד שבו יש n צמדדים.

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:

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

כאשר דרגות החופש: $d.f = n - 1$

דוגמה: (פתרון בהקלטה)

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

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

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

תרגילים:

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

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

נניח שהציונים מתפלגים נורמאלית.

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

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

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

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

פרק 8 - בדיקת השערות על פרמטרים

הקדמה

רקע:

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

שלב א: נזהה את הפרמטר הנחקר.

שלב ב: נרשום את השערות המחקר.

השערת האפס המסומנות ב- H_0

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

השערה אלטרנטיבית (השערת המחקר) המסומנת ב- H_1 .

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

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

שלב ד: נרשום את כלל ההכרעה.

בתהליך של בדיקת השערות יוצרים כלל שניקרא כלל הכרעה :

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

אזור הדחייה מוכתב על ידי סיכון שלוקח החוקר מראש שנקרא רמת מובהקות ומסומן ב- α .

שלב ה:

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

שלב ו:

להסיק מסקנה בהתאם לתוצאות המדגם.

דוגמה: (פתרון בהקלטה)

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

$$n = 20$$

$$\bar{X} = 3120$$

$$S = 280$$

א. מהי אוכלוסיית המחקר?

ב. מה המשתנה הנחקר?

ג. מה הפרמטר הנחקר?

ד. מהן השערות המחקר?

תרגילים:

1. ממוצע הציונים בבחינת הבגרות באנגלית הנו 72 עם סטיית תקן 15 נקודות. מורה טוען שפיתח שיטת לימוד חדשה שתעלה את ממוצע הציונים. משרד החינוך החליט לתת למורה 36 תלמידים אקראיים. ממוצע הציונים של אותם תלמידים לאחר שלמדו בשיטתו היה 75.5.

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

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

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

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

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

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

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

טעויות בבדיקת השערות

רקע:

בתהליך של בדיקת השערות יוצרים כלל שניקרא כלל הכרעה :

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

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

לכן יתכנו טעויות במסקנות שלנו :

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

הגדרת הטעויות:

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

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

דוגמה: (פתרון בהקלטה)

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

תרגילים:

1. לפי הצהרת היצרן של חברת משקאות מסוימת נפח הנוזל בבקבוק מתפלג נורמלית עם תוחלת 500 סמ"ק וסטיית תקן 20 סמ"ק. אגודת הצרכנים מתלוננת על הפחתת נפח המשקה בבקבוק מהכמות המוצהרת. במדגם שעשתה אגודת הצרכנים התקבל נפח ממוצע של 492 סמ"ק במדגם בגודל 25. בסופו של דבר הוחלט להכריע לטובת חברת המשקאות.
 - א. רשמו את השערות המחקר.
 - ב. מה מסקנת המחקר?
 - ג. איזו סוג טעות יתכן וביצעו במחקר?

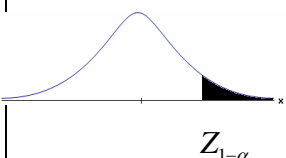
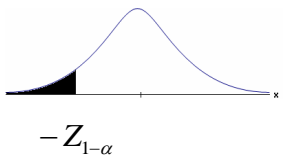
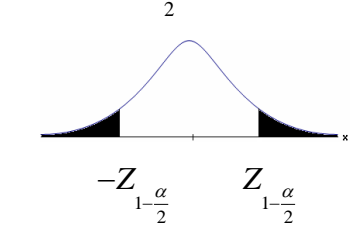
2. במחקר על פרמטר מסוים הוחלט בסופו של דבר לדחות את השערת האפס.
 - א. האם ניתן לדעת אם בוצע טעות במחקר?
 - ב. מה סוג הטעות האפשרית?

3. לפי נתוני משרד הפנים בשנת 1980 למשפחה ממוצעת היה 2.3 ילדים למשפחה עם סטיית תקן 0.4. ישנה טענה שכיום ממוצע מספר הילדים במשפחה קטן יותר. לצורך כך הוחלט לדגום 121 משפחות. במדגם התקבל ממוצע 2.17 ילדים למשפחה. על סמך תוצאות המדגם נקבע שלא ניתן לקבוע שבאופן מובהק תוחלת מספר הילדים למשפחה קטנה כיום.
 - א. מהי אוכלוסיית המחקר?
 - ב. מה המשתנה הנחקר?
 - ג. מה הפרמטר הנחקר?
 - ד. מה השערות המחקר?
 - ה. מה מסקנת המחקר?
 - ו. מהי סוג הטעות האפשרית במחקר?

פרק 9 - בדיקת השערות על תוחלת (ממוצע)

כאשר שונות האוכלוסיה ידועה

רקע:

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

סטטיסטי המבחן:

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

חלופה אחרת לכלל הכרעה:

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

דוגמה : (פתרון בהקלטה)

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

תרגילים:

1. ממוצע הציונים בבחינת הבגרות באנגלית הנו 72 עם סטיית תקן 15 נקודות. מורה טוען שפיתח שיטת לימוד חדשה שתעלה את ממוצע הציונים. משרד החינוך החליט לתת למורה 36 תלמידים אקראיים. ממוצע הציונים של אותם תלמידים לאחר שלמדו בשיטתו היה 75.5. בהנחה שגם בשיטתו סטיית התקן תהייה 15 מה מסקנתכם ברמת מובהקות של 5%?

2. לפי הצהרת היצרן של חברת משקאות מסוימת נפח הנוזל בבקבוק מתפלג נורמלית עם תוחלת 500 סמ"ק וסטיית תקן 20 סמ"ק. אגודת הצרכנים מתלוננת על הפחתת נפח המשקה בבקבוק מהכמות המוצהרת. במדגם שעשתה אגודת הצרכנים התקבל נפח ממוצע של 492 סמ"ק במדגם בגודל 25.
 - א. מה מסקנתכם ברמת מובהקות של 2.5%?
 - ב. האם ניתן לדעת מה תהיה המסקנה עבור רמת מובהקות הגבוהה מ-5%?

3. מהנדס האיכות מעוניין לבדוק אם מכונה מכוילת (מאופסת). המכונה כוונה לחתוך מוטות באורך 50 ס"מ. לפי נתוני היצרן סטיית התקן בחיתוך המוטות היא 0.5 ס"מ. במדגם של 50 מוטות התקבל ממוצע אורך המוט 50.93 ס"מ. מה מסקנתכם ברמת מובהקות של 5%?

4. המשקל הממוצע של הספורטאים בתחום ספורט מסוים הוא 90 ק"ג, עם סטיית תקן 8 ק"ג. לפי דעת מומחים בתחום יש צורך בהורדת המשקל ושימוש בדיאטה מסוימת שצריכה להביא להורדת המשקל. לשם בדיקת יעילות הדיאטה נלקח מדגם מקרי של 50 ספורטאים ובתום שנה של שימוש בדיאטה התברר שהמשקל הממוצע במדגם זה היה 84 ק"ג. יש לבדוק בר"מ של 10%, האם הדיאטה גורמת להורדת המשקל.

5. לפי מפרט נתון, על עובי בורג להיות 4 מ"מ עם סטיית תקן של 0.2 מ"מ. במדגם של 25 ברגים העובי הממוצע היה 4.07 מ"מ.
 - א. קבעו ברמת מובהקות 0.05, האם עובי הברגים מתאים למפרט. הניחו כי עובי של בורג מתפלג נורמלית וסטיית התקן של עובי בורג היא אכן 0.2 מ"מ.

6. במחקר נמצא שתוצאה היא מובהקת ברמת מובהקות של 5% מה תמיד נכון? בחר בתשובה הנכונה.
- הגדלת רמת המובהקות לא תשתנה את מסקנת המחקר.
 - הגדלת רמת המובהקות תשנה את מסקנת המחקר.
 - הקטנת רמת המובהקות לא תשנה את מסקנת המחקר.
 - הקטנת רמת המובהקות תשנה את מסקנת המחקר.
7. חוקר ערך מבחן דו צדדי ברמת מובהקות של α והחליט לדחות את השערת האפס.
- אם החוקר היה עורך מבחן צדדי ברמת מובהקות של $\frac{\alpha}{2}$ אזי בהכרח: (בחר בתשובה הנכונה)
- השערת האפס הייתה נדחית.
 - השערת האפס הייתה לא נדחית.
 - לא ניתן לדעת מה תהיה מסקנתו במקרה זה.
8. שני סטטיסטיקאים בדקו השערות $H_0: \mu = \mu_0$ כנגד $H_1: \mu > \mu_0$ עבור שונות ידועה ובאותה רמת מובהקות. שני החוקרים קבלו אותו ממוצע במדגם אך לחוקר א' היה מדגם בגודל 100 ולחוקר ב' מדגם בגודל 200.
- אם חוקר א' החליט לדחות את H_0 , מה יחליט חוקר ב'? נמקו.
 - אם חוקר א' יחליט לא לדחות את H_0 , מה יחליט חוקר ב'? נמקו.

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

א

שאלה 7:

ג

שאלה 8:

א. אותה מסקנה

ב. לא ניתן לדעת.

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 :
.3 σ ידועה			תנאים :
.4 $X \square N$ או מדגם מספיק גדול			
$P_{H_0}(\bar{X} \geq \bar{x})$	$P_{H_0}(\bar{X} \leq \bar{x})$	אם $2 \cdot P_{H_0}(\bar{X} \geq \bar{x}) \leftarrow \bar{x} > \mu_0$ אם $2 \cdot P_{H_0}(\bar{X} \leq \bar{x}) \leftarrow \bar{x} < \mu_0$	p-value

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

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

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 סטיית 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 נורמלית עם תוחלת אליה כוונה המכונה וסטיית תקן 2 ס"מ. ביום מסוים כוונה המכונה לחתוך מוטות באורך 80 ס"מ. אחראי האיכות מעוניין לבדוק האם המכונה מכוילת. לצורך כך נדגמו מקו הייצור 16 מוטות שנחתכו אורכן הממוצע היה 81.7 ס"מ.

א. מהי רמת המובהקות המינימלית עבורה נכריע שהמכונה לא מכוילת?

ב. אם נוסף עוד תצפית שערכה יהיה 82 ס"מ, כיצד הדבר ישפיע על התשובה של הסעיף הקודם?

ג. הכרע ברמת מובהקות של 5% האם המכונה מכוילת.

5. אם מקבלים בחישובים אלפא מינימלית (P value) קטנה מאוד, סביר להניח כי החוקר ידחה את השערת האפס בקלות. נכון? לא נכון? נמק.
6. בבדיקת השערות התקבל שה- $p\text{-value}=0.02$.
 מה תהיה מסקנת חוקר המשתמש ברמת מובהקות 1%? בחר בתשובה הנכונה.
 א. יקבל את השערת האפס בכל מקרה.
 ב. ידחה את השערת האפס מקרה.
 ג. ידחה את השערת האפס רק אם המבחן הנו דו צדדי.
 ד. לא ניתן לדעת כי אין מספיק נתונים.
7. מובהקות התוצאה (PV) היא גם : (בחר בתשובה הנכונה)
 א. רמת המובהקות המינימלית לדחות השערת האפס.
 ב. רמת המובהקות המקסימלית לדחיית השערת האפס.
 ג. רמת המובהקות שנקבעת מראש על ידי החוקר טרם קיבל את תוצאות המחקר.
 ד. רמת המובהקות המינימלית לאי דחיית השערת האפס.
8. בבדיקת השערות מסוימת התקבל $p\text{ value}=0.0254$ לכן (בחר בתשובה הנכונה):
 א. ברמת מובהקות של 0.01 אך לא של 0.05 נדחה את H_0 .
 ב. ברמת מובהקות של 0.01 ושל 0.05 לא נדחה את H_0 .
 ג. ברמת מובהקות של 0.05 אך לא של 0.01 נדחה את H_0 .
 ד. ברמת מובהקות של 0.01 ושל 0.05 נדחה את H_0 .

פתרונות :**שאלה 1:**

0.0228

שאלה 2:

עבור כל רמת מובהקות סבירה.

שאלה 3:

ב. 0.1056

ג. 0.1056

ד. נכריע שיש עמידה בהתחייבות של החברה.

שאלה 4:

א. 0.0006

ב. יקטן.

ג. נכריע שאין כיוול.

שאלה 5:

נכון

שאלה 6:

תשובה א:

שאלה 7:

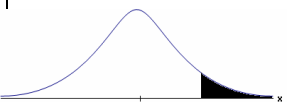

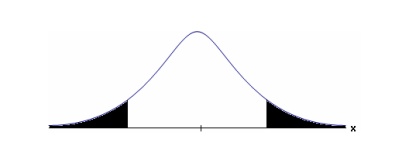
תשובה א :

שאלה 8:

תשובה : ג

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

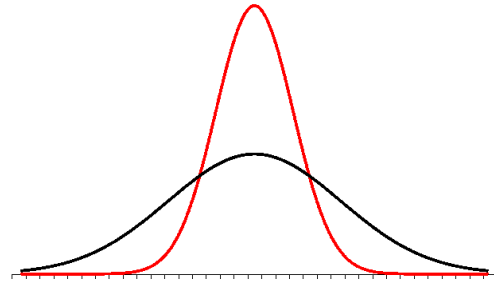
רקע:

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

סטטיסטי המבחן:

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

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

התפלגות T:

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

דוגמה: (פתרון בהקלטה)

מפעל קיבל הזמנה לייצור משטחים בעובי של 0.1 ס"מ.
 כדי לבדוק האם המפעל עומד בדרישה נדגמו 10 משטחים ונמצא שהעובי הממוצע הוא 0.104 עם אומדן לסטיית תקן 0.002 ס"מ.
 א. מהן השערות המחקר?
 ב. מה ההנחה הדרושה לצורך פתרון?
 ג. בדוק ברמת מובהקות של 5%.

תרגילים:

1. משך זמן ההחלמה בלקיחת אנטיביוטיקה מסוימת הוא 120 שעות בממוצע עם סטיית תקן לא ידועה. מעוניינים לבדוק האם אנטיביוטיקה אחרת מקטינה את משך זמן ההחלמה. במדגם של 5 חולים שלקחו את האנטיביוטיקה האחרת התקבלו זמני ההחלמה הבאים: 90, 95, 100, 80, 125 שעות. מה מסקנתכם ברמת מובהקות של 5%. מהי ההנחה הדרושה לצורך הפתרון?

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

$$n = 20$$

$$\bar{x} = 3120$$

$$S = 280$$

מה מסקנתכם ברמת מובהקות של 5% מה יש להניח לצורך פתרון?

3. ציוני מבחן אינטליגנציה מתפלגים נורמלית. בארה"ב ממוצע הציונים הוא 100. במדגם שנעשה על 23 נבחנים ישראלים, התקבל ממוצע ציונים 104.5 וסטיית התקן המדגמית 16. האם בישראל ממוצע הציונים שונה מבארה"ב? הסיקו ברמת מובהקות של 5%.

4. באוכלוסייה מסוימת נדגמו 10 תצפיות והתקבלו התוצאות הבאות:

$$\sum_{i=1}^{10} X_i = 750$$

$$\sum_{i=1}^{10} (X_i - \bar{X})^2 = 900$$

נתון שההתפלגות היא נורמלית.

בדוק ברמת מובהקות של 5% האם התוחלת של ההתפלגות שונה מ-80.

5. ליאור ורוני העלו את אותן השערות על ממוצע האוכלוסייה. כמו כן הם התבססו על אותן תוצאות של מדגם.

ליאור השתמש בטבלה של התפלגות Z .

רוני השתמשה בטבלה של התפלגות t .

מה נוכל לומר בנוגע להחלטת המחקר שלהם? בחר בתשובה הנכונה.

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

ב. אם רוני תדחה את השערת האפס אז גם בהכרח ליאור.

ג. שני החוקרים בהכרח יגיעו לאותה מסקנה.

ד. לא ניתן לדעת על היחס בין דחיית השערת האפס של שני החוקרים.

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

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu < \mu_0$$

חוקר בדק את ההשערות הללו על סמך מדגם שכלל 10 תצפיות. 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.

בחר בתשובה הנכונה :

א. כעת בברור הוא ידחה את השערת האפס.

ב. כעת הוא דווקא יקבל את השערת האפס.

ג. כעת לא ניתן לדעת מה תהיה מסקנתו.

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

התשובה היא : ב

שאלה 6:

התשובה היא : ג

מובהקות התוצאה (p-value) כאשר שונות האוכלוסייה לא ידועה

רקע:

נזכיר שהמסקנה של המחקר תקבע לפי העיקרון הבא :

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

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

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

אם ההשערה היא דו צדדית :

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

מובהקות התוצאה היא גם האלפא המינימלית לדחיית השערת האפס.

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

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

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

$$d.f = n - 1$$

דוגמה : (פתרון בהקלטה)

- ממוצע זמן הנסיעה של אדם לעבודה הינו 40 דקות. הוא מעוניין לבדוק דרך חלופית שאמורה להיות יותר מהירה. לצורך כך הוא דוגם 5 ימים שבהם הוא נוסע בדרך החלופית. זמני הנסיעה שקיבל בדקות הם : 27, 34, 32, 40, 30 . הנח שזמן הנסיעה מתפלג נורמלית.
- א. רשום את השערות המחקר.
- ב. מצא חסמים למובהקות התוצאה.
- ג. מה המסקנה ברמת מובהקות של 5% ?

תרגילים :

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

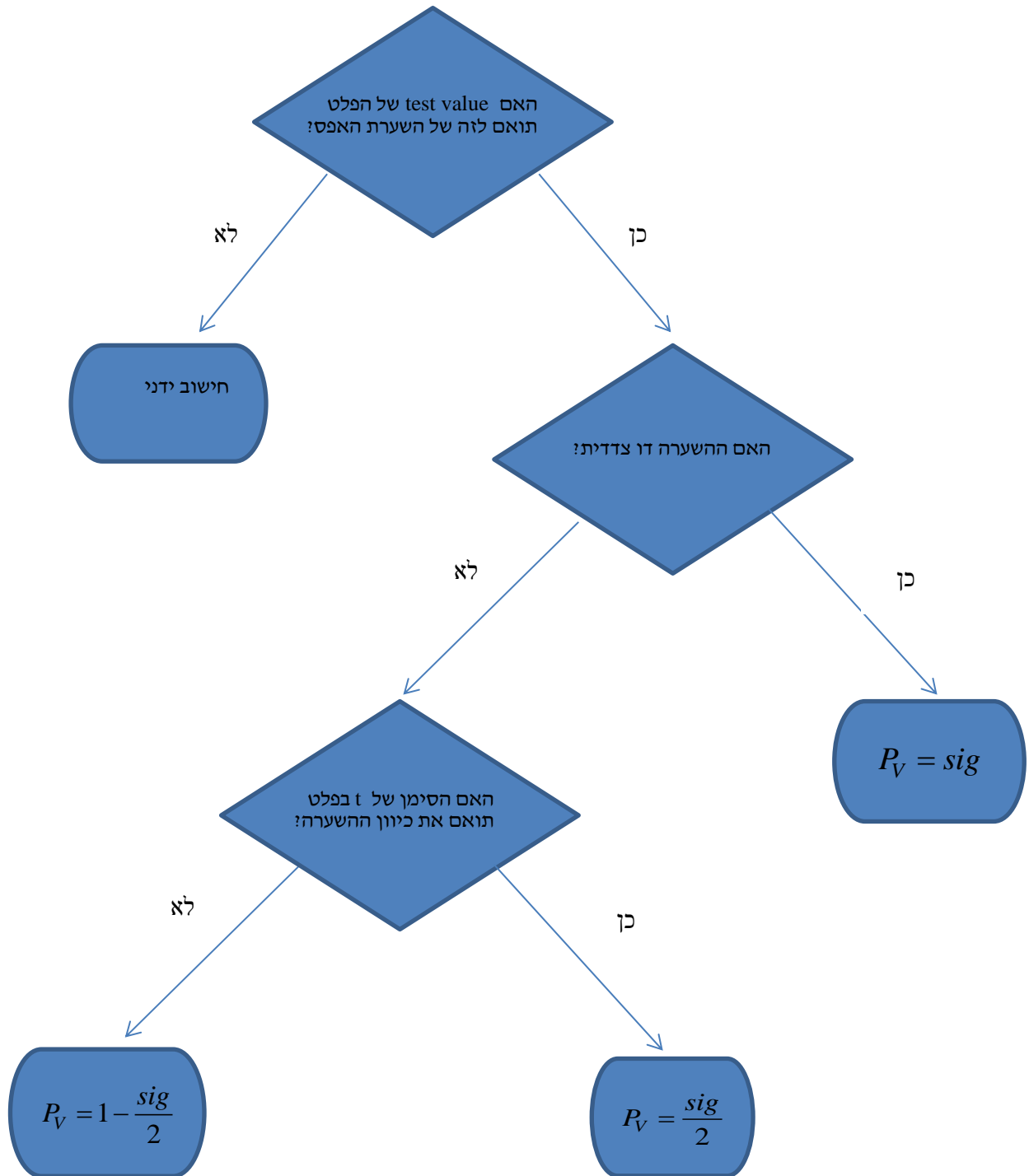
שאלה 3:

נקבל H_0

ניתוח פלטי SPSS

רקע:

חישוב מובהקות התוצאה באמצעות פלט תוכנת SPSS



דוגמה: (פתרון בהקלטה)

One-Sample Statistics

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

One-Sample Test

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

ממוצע הציונים במבחן המיצב בחשבון הוא 60. הוחלט לדגום כיתה אקראית של 25 תלמידים וללמד אותם בשיטת לימוד חדשה.

א. מהו רווח הסמך לממוצע הציונים בחשבון אם יוחלט ליישם את שיטת הלימוד החדשה?

ב. מהו P_V לבדיקת יעילותה של שיטת הלימוד החדשה?

ג. מה יוכרע ברמת מובהקות של 5% לגבי יעילותה של שיטת הלימוד החדשה?

תרגילים :

1. באוניברסיטה גדולה גיל הסטודנטים לתואר ראשון מתפלג נורמאלי. בעבר פורסם שהגיל הממוצע של הסטודנטים הינו 23. להלן פלט תוכנת SPSS על מדגם של 16 סטודנטים אקראיים מתואר ראשון:

One-Sample Statistics

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

One-Sample Test

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

- א. מהו המבחן הסטטיסטי שנעשה כאן?
 ב. מה ערכו של הפרמטר לפי השערת האפס?
 ג. רשום רווח סמך ברמת סמך של 95% לתוחלת גיל הסטודנטים באוניברסיטה לתואר ראשון.
 ד. בדוק ברמת מובהקות של 5% האם הגיל הממוצע כיום שונה מבעבר?

2. קבוצת ילדים בגיל 6 קיבלה משימה לביצוע. עבור כל ילד בדקו כמה זמן לקח לו לסיים את המשימה בדקות. להלן תוצאות הניתוח הסטטיסטי:

One-Sample Test

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

- א. כמה ילדים השתתפו במחקר?
 ב. מצא רווח סמך ברמת סמך של 95% לתוחלת זמן ביצוע המשימה עבור ילדים בני 6.
 ג. מה יש להניח כדי שרווח הסמך מסעיף א' יהיה תקף?
 ד. בדוק ברמת מובהקות של 5% שזמן ביצוע המשימה הממוצע נמוך מ-4.5 דקות.

3. לפניך פלט מחשב עבור ניתוח סטטיסטי שנעשה בתוכנת SPSS. הניתוח הוא עבור מדגם אקראי של קבוצת נבחנים בבגרות באנגלית.

One-Sample Statistics

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

One-Sample Test

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

- א. השלם את הגדלים החסרים המסומנים בסמני שאלה בפלט.
 ב. מהי מובהקות התוצאה לבדיקת ההשערה שהתוחלת של הציונים שונה מ-75?
 ג. מהי מובהקות התוצאה לבדיקת ההשערה שהתוחלת של הציונים קטנה מ-75?
 ד. מהי מובהקות התוצאה לבדיקת ההשערה שהתוחלת של הציונים גדולה מ-75?

4. יצרן סיגריות מפרסם כי תוחלת הניקוטין בסיגריות שהוא מיצר קטנה מ- 27 מ"ג. בבדיקה מקרית של 5 סיגריות מתוצרתו נמצאו כמויות הניקוטין הבאות:
 21, 21, 20, 24, 22 מ"ג. הנח כי כמות הניקוטין בסיגריות מפולג נורמאלי.

One-Sample Statistics

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

One-Sample Test

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

- א. האם ברמת מובהקות של 5% ניתן להסיק שיש אמת בפרסום?
 ב. אם הינו מוסיפים עוד תצפית שערכה 20. כיצד הדבר היה משפיע על הערך Sig ועל המסקנה?
 ג. בדוק האם ניתן להגיד שתוחלת רמת הניקוטין שונה מ- 26 ברמת מובהקות של 5%.

פתרונות :**שאלה 1:**

ב. 23

ג. (22.104, 24.771)

ד. נקבל H_0 **שאלה 2:**

א. 25

ב. (4.3056, 4.5104)

ד. נדחה H_0 **שאלה 3:**א. $n = 44$

$$\bar{X} = 67.66$$

$$t = -2.48$$

$$upper = -1.3735$$

ב. 0.017

ג. 0.0085

ד. 0.9915

שאלה 4:

א. נכריע שיש אמת בפרסום.

ב. מסקנה לא תשתנה.

ג. נכריע שהתוחלת שונה מ-26.

הקשר בין רווח סמך לבדיקת השערות על תוחלת

רקע:

ניתן לבצע בדיקת השערות זו צדדית ברמת מובהקות α על μ :

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

על ידי בניית רווח סמך ברמת סמך של $1 - \alpha$ ל μ :

אם μ_0 נופל ברווח \leftarrow נקבל את H_0

אם μ_0 לא נופל ברווח \leftarrow נדחה את H_0

דוגמה: (פתרון בהקלטה)

חוקר ביצע בדיקת השערות לתוחלת. להלן השערותיו:

$$H_0 : \mu = 80$$

$$H_1 : \mu \neq 80$$

$$\alpha = 5\%$$

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

האם אפשר לדעת מה מסקנתו, ואם כן מהי?

תרגילים :

1. חוקר רצה לבדוק את ההשערות הבאות :

$$H_0 : \mu = 90$$

$$H_1 : \mu \neq 90$$

החוקר בנה רווח סמך לתוחלת ברמת סמך של 95% וקיבל את רווח הסמך הבא : (87,97).
אם החוקר מעוניין לבצע בדיקת השערות ברמת מובהקות של 1% האם ניתן להגיע למסקנה ע"ס רווח הסמך? נמקו.

2. חוקר מעוניין לבדוק השפעת דיאטה חדשה על רמת הסוכר בדם. ידוע כי מספר מיליגרם הסוכר בסמ"ק דם הוא משתנה מקרי שמתפלג נורמלית עם סטיית תקן 10.4 מ"ג. נלקח מדגם של 60 נבדקים שניזונו מדיאטה זו. נמצא כי ממוצע מספר המיליגרם סוכר היה 115.5 מ"ג לסמ"ק.

א. בנה רווח סמך ברמת סמך 95% לתוחלת רמת הסוכר בדם אצל הניזונים מדיאטה זו.
ב. ידוע שתוחלת רמת הסוכר בדם באוכלוסיה היא 90 מ"ג לסמ"ק. האם לדעתך ניתן להסיק על סמך תוצאת סעיף א שהדיאטה משפיעה על רמת הסוכר בדם? הסבר.

3. יצרן אנטיביוטיקה רושם על גבי התרופות שכמות הפנצלין היא 200 מ"ג לקפסולה.
משרד הבריאות ביצע מדגם של 8 קפסולות אקראיות מקו הייצור ומצא שבממוצע יש 196 מ"ג פנצילין לקפסולה עם סטיית תקן מדגמית של 5 מ"ג. בהנחה וכמות הפנצלין בקפסולה מתפלגת נורמלית.

א. בנה רווח סמך ברמת סמך של 95% לממוצע כמות הפנצלין לקפסולה המיוצרת על ידי יצרן האנטיביוטיקה.

ב. בדוק ברמת מובהקות של 5% האם יש אמת באינפורמציה המסופקת על ידי היצרן.

פתרונות :**שאלה 1:**

1. נקבל השערת H_0

שאלה 2:

א. $112.87 \leq \mu \leq 118.13$

ב. נכריע שהדיאטה משפיעה על תוחלת רמת הסוכר בדם.

שאלה 3:

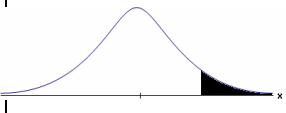
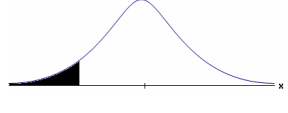
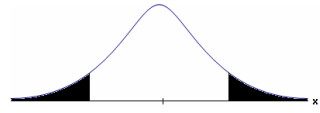
א. $191.8 \leq \mu \leq 200.2$

ב. נכריע שיש אמת בפרסום.

פרק 10 - בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

כששונות האוכלוסיה לא ידועות ומניחים שהן שוות

רקע:

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

סטטיסטי המבחן :

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

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

חלופה אחרת לכלל הכרעה:

$\bar{x}_1 - \bar{x}_2 > c + t_{1-\alpha}^{(n_1+n_2-2)} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$	$\bar{x}_1 - \bar{x}_2 < c - t_{1-\alpha}^{(n_1+n_2-2)} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$	או $\bar{x}_1 - \bar{x}_2 > c + t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$ 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
הנדגמים נגשו למבחן IQ. להלן תוצאות המדגם:

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

בדקו ברמת מובהקות של 5% האם קיים הבדל של נקודה בין ישראלים לאמריקאים מבחינת ממוצע הציונים במבחן 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 שנוורות 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:**

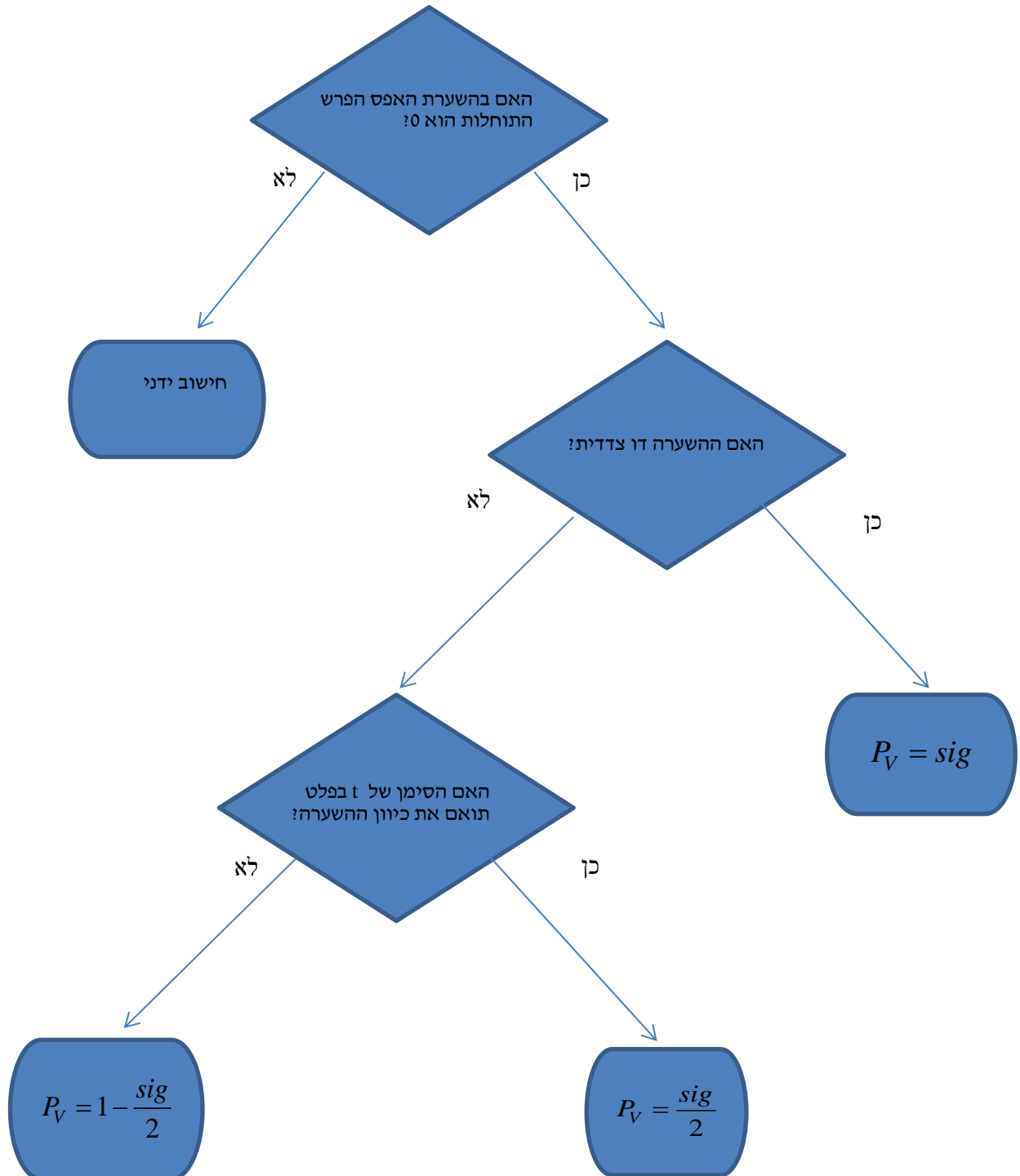
א. נדחה H_0

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	1190.977	.053	.05500	.02838	-.00068	.11067

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

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

T-Test

Group Statistic:

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

Independent Samples Test

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

מקרא:

A = נגשו פעם אחת.

B = נגשו יותר מפעם אחת.

- רשמו את השערות המחקר והסבירו מהו המבחן המתאים כאן.
 - כיצד הייתה משתנה התשובה לסעיף הקודם אם היה מדובר על אותם אנשים שציונם נבדק פעם אחרי המבחן הראשון שעשו ופעם אחרי המבחן השני?
 - האם ניתן לומר כי מידת הפיזור של ציוני אנשים הנבחנים בפעם הראשונה שונה ממידת הפיזור של ציוני האנשים אשר נבחנים בפעם השנייה. בדוק ברמת מובהקות של $\alpha=0.05$.
 - האם נכונה טענת הפסיכולוגים ברמת מובהקות של $\alpha=0.01$.
3. כחלק ממחקר בנושא הנישואין בישראל, אחד החוקרים העלה השערה שיש הבדל 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.

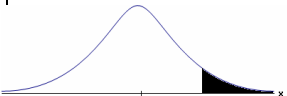

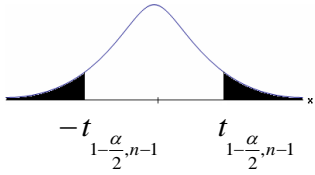
שאלה 4:

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

**פרק 11 - 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:
.4 σ_D אינה ידועה .5 $D \square N$ או מדגם מספיק גדול			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 הכרעה: אזור הדחייה של H_0 :
$\bar{D} > C + t_{1-\alpha}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$	$\bar{D} < C - t_{1-\alpha}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$	$\bar{D} > C + t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$ ו $\bar{D} < C - t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$	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 :

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

בהנחה והמחירים מתפלגים נורמאלית בדקו ברמת מובהקות של 5% את טענת רשת "מגה בעיר".

תרגילים:

1. במטרה לבדוק האם קיים הבדל בין חברת X לחברת Y מבחינת המחירים לשיחות בינ"ל. נגדמו באקראי 7 מדינות ועבור כל מדינה נבדקה עלות דקת שיחה. להלן התוצאות:

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

- בהנחה והמחירים מתפלגים נורמלית בכל חברה, בדקו ברמת מובהקות של 5% האם קיים הבדל בין החברות מבחינת המחירים במוצעי?

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

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

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

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

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

- פורסם שתלמידים שמסיימים את סמסטר ב משפרים 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:
- מבחן Z למדגם יחיד.
 - מבחן T למדגם יחיד.
 - מבחן T למדגמים בלתי תלויים.
 - מבחן T למדגמים מזווגים.
6. כדי להשוות בין שני אצנים נדגמו 5 תוצאות מריצת 100 מטר של כל אצן. זמני הריצה נרשמו ויש להניח שמתפלגים נורמלית. המטרה להשוות בין האצנים.
 המבחן שיש לבצע כאן הוא:
- מבחן Z למדגם יחיד.
 - מבחן T למדגם יחיד.
 - מבחן T למדגמים בלתי תלויים.
 - מבחן T למדגמים מזווגים.

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

ב. 0.5

ג. לא נדחה H_0 **שאלה 4:**

התשובה היא ד.

שאלה 5:

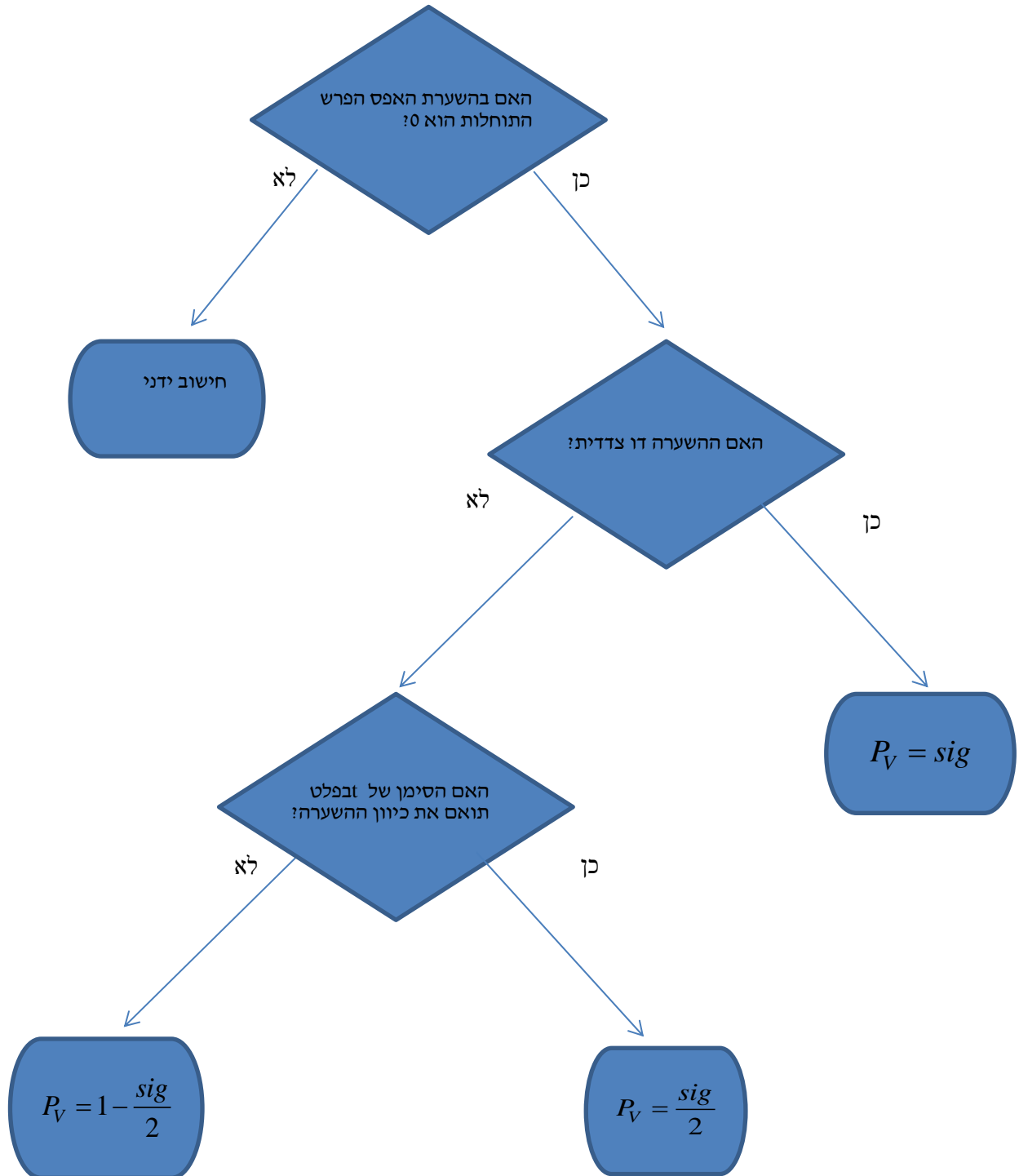
התשובה היא ד.

שאלה 6:

התשובה היא ג.

מדגמים מזווגים - ניתוח פלטים

רקע:



דוגמה : (פתרון בהקלטה)

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

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

Paired Samples Statistics

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

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	90% Confidence Interval of the Difference				
				Lower	Upper			
Pair 1 BEFORE - AFTER	11.20000	8.19756	3.66406	3.38452	19.01548	3.055	4	.038

בדוק ברמת מובהקות של 5% האם הקורס יעיל .

תרגילים :

1. בסקר שנערך בארה"ב בשנת 1993 נשאלו נסקרים על השכלת הוריהם, להלן הפלט שהתקבל:

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

א. תן אומדן להפרש הממוצעים.

ב. תן אומדן לטעות התקן של הפרש הממוצעים.

ג. האם קיים הבדל מובהק בין השכלת האבות להשכלת האימהות ברמת מובהקות של 5%?

2. בתחרות קפיצה למים שופטים באופן קבוע שופט איטלקי ושופט דרום קוריאני. להלן פלט

המנתח את הציונים ששופטים אלה נתנו בתחרויות השונות:

Paired Samples Statistics

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

Paired Samples Test

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

א. השלימו את החלקים החסרים בפלט. (מסומנים בסימני שאלה).

ב. בדוק את הטענה שהשופט הדרום קוריאני נותן בממוצע 0.2 נקודות יותר מאשר השופט האיטלקי ברמת מובהקות של 5%.

ג. מהו רווח הסמך ברמת סמך של 95% לתוחלת פער הציונים בין השופטים.

ד. בנה את הרווח כעת ברמת סמך של 98% לתוחלת פער בציונים בין השופטים.

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

Paired Samples Test

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

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

T-Test

Paired Samples Statistics

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

Paired Samples Test

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

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

פתרונות:שאלה 1:א. -0.007 ב. 0.1

ג. אין הבדל מובהק

שאלה 2:א. $d.f = 299$ $n = 300$ $\bar{X} = 8.496$

Sig=0

ב. נדחה H_0 ג. $(0.46341, 0.38126)$ ד. $(0.4708, 0.3738)$ שאלה 3:ב. נדחה H_0 ג. לא נדחה את H_0 ד. $(1.592, 13.317)$ שאלה 4:א. $1.5876, 11.5715$ ב. 0.173

ג. יגדל

ד. 0.0865 ה. 0.9135 ו. $49.18 < \mu < 58.82$

פרק 12 - 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 :
- $$-0.0293 < \mu_b < 0.2145$$
- רווח הסמך הוא ברמת סמך של 95%.
- לכן מסקנת המחקר היא :
- א. ברמת מובהקות של 5% לא נוכל לקבוע שקיים הבדל בין החברות.
- ב. ברמת מובהקות של 5% נקבע שקיים הבדל מובהק בין החברות.
- ג. לא ניתן לדעת מה המסקנה ברמת מובהקות של 5% כיוון שלא נאמר מה ההגדרה של D.

פתרונות:**שאלה 1:**

א. $-3.8 \leq \mu_D \leq 19$

ב. נכריע שיש אמת בפרסום.

שאלה 2:

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

ב. נכריע שאין הבדל.

שאלה 3:

התשובה היא ג.

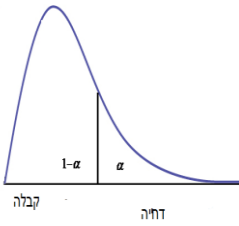
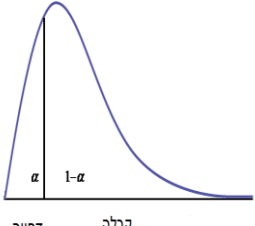
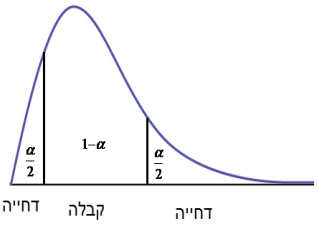
שאלה 4:

התשובה היא א.

פרק 13 - בדיקת השערות על שונויות

בדיקת השערות על שתי שונויות

רקע:

$H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1$ $H_1 : \frac{\sigma_1^2}{\sigma_2^2} > 1$	$H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1$ $H_1 : \frac{\sigma_1^2}{\sigma_2^2} < 1$	$H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1$ $H_1 : \frac{\sigma_1^2}{\sigma_2^2} \neq 1$	השערת האפס : השערה אלטרנטיבית :
$X_1, X_2 \sim N .2$		1. מדגמים בלתי תלויים	תנאים :
 $F \geq f_{1-\alpha}^{(n_1-1, n_2-1)}$	 $F \leq \frac{1}{f_{1-\alpha}^{(n_2-1, n_1-1)}}$	 $F \geq f_{1-\alpha/2}^{(n_1-1, n_2-1)}$ או $F \leq \frac{1}{f_{1-\alpha/2}^{(n_2-1, n_1-1)}}$	נדחה את השערת האפס אם :

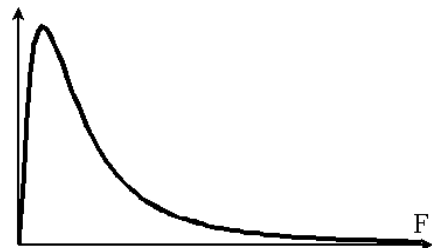
$$F = \frac{S_1^2}{S_2^2} : \text{סטטיסטי המבחן}$$

התפלגות F:

אם $X_1 \sim N(\mu_1, \sigma^2)$ ו- $X_2 \sim N(\mu_2, \sigma^2)$ אזי : $F(n_1-1, n_2-1) \sim \frac{S_1^2}{S_2^2}$

התפלגות F הינה התפלגות אסימטרית חיובית התלויה בדרגות חופש של המונה ושל המכנה.

כמו כן בהתפלגות F מתקיימת התכונה הבאה : $F_\alpha(n_1-1, n_2-1) = \frac{1}{F_{1-\alpha}(n_2-1, n_1-1)}$



$$df_1 = n_1 - 1$$

$$df_2 = n_2 - 1$$

דוגמה: (פתרון בהקלטה)

מעוניינים להשוות בין נשים וגברים מבחינת השונות בזמנים שלהם לבצע משימה מסוימת. במדגם של 10 גברים התקבלו התוצאות הבאות לגבי זמני ביצוע המשימה:

$$\sum (y_i - \bar{y})^2 = 204$$

במדגם של 13 נשים התקבלו התוצאות הבאות:

$$\sum (x_i - \bar{x})^2 = 200$$

בדקו ברמת מובהקות של 2% האם קיים הבדל בין השונות? מה יש להניח?

טבלת התפלגות F . ערכי החלוקה F_p של התפלגות $F(m, n)$
 m — דרגות חופש המונה, n — דרגות חופש המכנה

		m											
p	n	1	2	3	4	5	6	7	8	9	10	12	15
.95	1	161	200	216	225	230	234	237	239	241	242	244	246
.975		648	800	864	900	922	937	948	957	963	969	977	985
.99		4052	5000	5403	5625	5764	5859	5928	5981	6022	6056	6106	6157
.995		16211	20000	21615	22500	23056	23437	23715	23925	24091	24224	24426	24630
.95	2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43
.975		38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41	39.43
.99		98.50	99.00	99.16	99.25	99.30	99.33	99.36	99.38	99.39	99.40	99.42	99.43
.995		198.50	199.01	199.16	199.24	199.30	199.33	199.36	199.38	199.39	199.39	199.42	199.43
.95	3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70
.975		17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.34	14.25
.99		34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.34	27.23	27.05	26.87
.995		55.55	49.80	47.47	46.20	45.39	44.84	44.43	44.13	43.88	43.68	43.39	43.08
.95	4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86
.975		12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.75	8.66
.99		21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20
.995		31.33	26.28	24.26	23.15	22.46	21.98	21.62	21.35	21.14	20.97	20.70	20.44
.95	5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62
.975		10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.43
.99		16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72
.995		22.78	18.31	16.53	15.56	14.94	14.51	14.20	13.96	13.77	13.62	13.38	13.15
.95	6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94
.975		8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.27
.99		13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56
.995		18.63	14.54	12.92	12.03	11.46	11.07	10.79	10.57	10.39	10.25	10.03	9.81
.95	7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51
.975		8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.57
.99		12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31
.995		16.24	12.40	10.88	10.05	9.52	9.16	8.89	8.68	8.51	8.38	8.18	7.97
.95	8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22
.975		7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.10
.990		11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52
.995		14.69	11.04	9.60	8.81	8.30	7.95	7.69	7.50	7.34	7.21	7.01	6.81

m

p	n	1	2	3	4	5	6	7	8	9	10	12	15
.95	9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01
.975		7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87	3.77
.99		10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96
.995		13.61	10.11	8.72	7.96	7.47	7.13	6.88	6.69	6.54	6.42	6.23	6.03
.95	10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85
.975		6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.52
.99		10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56
.995		12.83	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.97	5.85	5.66	5.47
.95	12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62
.975		6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.18
.99		9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01
.995		11.75	8.51	7.23	6.52	6.07	5.76	5.52	5.35	5.20	5.09	4.91	4.72
.95	15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40
.975		6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.96	2.86
.99		8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52
.995		10.80	7.70	6.48	5.80	5.37	5.07	4.85	4.67	4.54	4.42	4.25	4.07
.95	20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20
.975		5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.57
.99		8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09
.995		9.94	6.99	5.82	5.17	4.76	4.47	4.26	4.09	3.96	3.85	3.68	3.50
.95	30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01
.975		5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.41	2.31
.99		7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70
.995		9.18	6.35	5.24	4.62	4.23	3.95	3.74	3.58	3.45	3.34	3.18	3.01
.95	60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84
.975		5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.17	2.06
.99		7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35
.995		8.49	5.79	4.73	4.14	3.76	3.49	3.29	3.13	3.01	2.90	2.74	2.57
.95	120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75
.975		5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	2.05	1.94
.99		6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19
.995		8.18	5.54	4.50	3.92	3.55	3.28	3.09	2.93	2.81	2.71	2.54	2.37
.95	∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67
.975		5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.94	1.83
.99		6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04
.995		7.88	5.30	4.28	3.72	3.35	3.09	2.90	2.74	2.62	2.52	2.36	2.19

תרגילים :

1. להלן נתונים על שטחי דירות במ"ר עבור דירות חדשות שנבנו בשנת 2012 ובשנת 2013:

120	94	90	130	95	112	120	2012
	69	74	105	91	82	100	2013

א. בדוק ברמת מובהקות של 10% את ההשערה ששונויות שטחי הדירות החדשות בשנת 2012

ובשנת 2013 שוות. מה הן ההנחות הדרושות לביצוע הבדיקה?

ב. האם וכיצד הייתה משתנה המסקנה מהסעיף הקודם אם מסתבר שחלה טעות ברישום ויש

להפחית 10 מ"ר מכל הדירות שמופיעות במדגם?

2. בתחום הבינוי משתמשים בשני סוגי מתכות: מתכת A ומתכת B. מחקר מעוניין לבדוק האם קיים

הבדל בין שני סוגי המתכות מבחינת החוזק שלהן. דגמו מספר יחידות מתכת מכל סוג והתקבלו

התוצאות הבאות:

סוג המתכת	A	B
n	8	10
$\sum X_i$	16	30
$\sum X_i^2$	60	198

יש להניח שרמת החוזק של המתכות מתפלגת נורמאלית.

א. האם קיים הבדל בין שונויות החוזק של מתכות?

ב. האם קיים הבדל בין תוחלות החוזק של מתכות?

בכל סעיף רמת מובהקות של 10%.

3. מחקר סוציולוגי מעוניין לחקור את הרגלי הבילויים בקבוצות גיל שונות. ידוע כי בקרב

האוכלוסייה הבוגרת (מעל 18) ההוצאה החודשית על בילויים מתפלגת נורמאלית עם תוחלת של

500 ₪ וסטיית תקן של 300 ₪.

במדגם שנעשה על סטודנטים בגילאי 21-26 התקבל אומד חוסר הטיה לשוונות ההוצאה החודשית

על בילויים 10,000. כמות הסטודנטים שנדגמה 16.

במדגם שנעשה על 11 מבוגרים בשנות השלושים התקבל אומד חסר הטיה לשוונות ההוצאה

החודשית על בילויים 490,000.

א. בדקו ברמת מובהקות של 5% האם שוונות ההוצאה על בילויים בקרב סטודנטים בקבוצת

גילאי 21-26 נמוכה מהשוונות אצל כלל המבוגרים.

ב. בדקו ברמת מובהקות של 1% האם הפיזור של ההוצאה החודשית לבילויים גדולה יותר

בקבוצת גיל ה-30 מאשר בקבוצת גיל 21-26.

פתרונות:**שאלה 1**

- א. לא נדחה H_0
ב. מסקנה לא תשתנה

שאלה 2

- א. לא נדחה H_0
ב. לא נדחה H_0

שאלה 3

- א. נדחה H_0
ב. נדחה H_0

פרק 14- שאלות מסכמות בבדיקת השערות על פרמטרים

שאלות מסכמות בבדיקת השערות על פרמטרים

1. שני חוקרים נתבקשו לבדוק את ההשערות הבאות:

$$H_0 : \mu = 520$$

$$H_1 : \mu > 520$$

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

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

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

א. למי מהחוקרים הסתברות לטעות מסוג ראשון קטנה יותר?

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

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

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

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

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

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

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

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

מוצר	A	B
1	5	5
2	4	5
3	5	3
4	7	4

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

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

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

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

.5 נתון כי:

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

מעוניינים לבדוק את ההשערות:

$$H_0 : \mu = 40$$

$$H_1 : \mu > 40$$

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

א. חשבו את P-value (מובהקות התוצאה).

ב. חזור על סעיף א אם ההשערה האלטרנטיבית הייתה:

$$H_1 : \mu < 40$$

ג. חזור על סעיף א אם ההשערה האלטרנטיבית הייתה:

$$H_1 : \mu \neq 40$$

פתרונות:**שאלה 1:**

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

שאלה 2:

- א. לפחות 0.0808
- ב. 0.7995

שאלה 3:

- א. לא נדחה H_0
- ב. לא נדחה H_0

שאלה 4:

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

שאלה 5:

- א. 0.0062
- ב. 0.9938
- ג. 0.0124

שאלות מסכמות בסגנון רב ברירה (אמריקאיות) על בדיקת השערות

1. בבדיקת השערה חד-צדדית ימנית ברמת מובהקות $\alpha=0.01$, נדחתה השערת האפס. מה הייתה המסקנה לו נבדקה אותה ההשערה באמצעות אותם נתונים ברמת מובהקות $\alpha=0.05$?
- השערת האפס הייתה נדחית.
 - השערת האפס לא הייתה נדחית.
 - ההשערה המחקרית הייתה נדחית.
 - בהעדר נתונים נוספים, לא ניתן לדעת.
2. לצורך בדיקת השפעת היפנוזה על לימוד אנגלית, נבחרו 10 זוגות תאומים זהים. אחד התאומים למד אנגלית בהשפעת היפנוזה, והשני ללא היפנוזה. לאחר מכן נערך לשניהם מבחן באנגלית. נניח שציוני המבחן מתפלגים נורמאלית ללא ידיעת השונות האמיתית.
- המבחן שיש לבצע כאן הוא :
- מבחן Z למדגם יחיד.
 - מבחן T למדגם יחיד.
 - מבחן T למדגמים בלתי תלויים.
 - מבחן T למדגמים מזווגים.
3. כדי לבדוק את הטענה שגברים רווקים שוקלים פחות מגברים נשואים לקח חוקר מדגם מקרי של 4 גברים ומדד את משקלם לפני נישואיהם ולאחר נישואיהם. הנה התוצאות :

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

מהן ההשערות הנבדקות: (ההפרש חושב $X-Y$)

א. $H_1 : \mu_d < 0, H_0 : \mu_d = 0$

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

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

ד. $H_1 : \mu_d > 0, H_0 : \mu_d = 0$

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

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

5. 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.

- מבחן Z למדגם יחיד ; מבחן T למדגם יחיד.
- מבחן T למדגם יחיד ; מבחן T למדגמים תלויים.
- מבחן T למדגם יחיד ; מבחן T למדגמים בלתי תלויים.
- מבחן T למדגמים בלתי תלויים ; מבחן T לממוצע יחיד.

6. מובהקות התוצאה (PV) היא גם :

- רמת המובהקות המינימאלית לדחות השערת האפס.
- רמת המובהקות המקסימאלית לדחיית השערת האפס.
- רמת המובהקות שנקבעת מראש על ידי החוקר טרם קיבל את תוצאות המחקר.
- רמת המובהקות המינימאלית לאי דחיית השערת האפס.

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

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

באיזה התפלגות משתמשים לבדיקת ההשערות, ובכמה דרגות חופש :

- ההתפלגות Z ללא דרגות חופש.
- ההתפלגות T ו-3 דרגות חופש.
- ההתפלגות T ו-6 דרגות חופש.
- ההתפלגות χ^2 ו-3 דרגות חופש.

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

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

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

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

מה יחליט סטטיסטיקאי ב'?

א. לדחות את השערת האפס.

ב. לא לדחות את השערת האפס.

ג. ללא נתונים נוספים אי אפשר לדעת מה יחליט.

9. חוקר בדק השערה מסוימת והחליט לדחות את השערת האפס ברמת מובהקות 5%. מה נכון לומר?

א. הוא בוודאות ידחה את השערת האפס ברמת מובהקות 9% ואילו ברמת מובהקות 2% יש לבדוק מחדש.

ב. הוא בוודאות לא ידחה את השערת האפס ברמת מובהקות 9% ואילו ברמת מובהקות 2% יש לבדוק מחדש.

ג. הוא בוודאות ידחה את השערת האפס ברמת מובהקות 9% וברמת מובהקות 2%.

ד. הוא בוודאות לא ידחה את השערת האפס ברמת מובהקות 9% ואילו ברמת מובהקות 2% יש לבדוק מחדש.

10. 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.

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

12. ידוע מנתוני העבר כי תוחלת הציונים בבחינה בפסיכולוגיה היא 79. הועלתה השערה כי תוחלת הציונים בקרב העולים החדשים נמוכה יותר. לצורך בדיקת הטענה נלקח מדגם מקרי של 47 סטודנטים עולים ונמצא ממוצע של 75. מה משמעות הפרמטר בניסוח ההשערות?
- א. תוחלת ציוני העולים באוכלוסייה.
 ב. ממוצע ציוני העולים במדגם.
 ג. תוחלת ציוני האוכלוסייה מנתוני העבר.
 ד. ממוצע ציוני שאר האוכלוסייה במדגם.

13. חוקר ביצע מחקר וידוע כי עשה טעות מסוג 1. מה מהבאים נכון?
- א. החוקר דחה את השערת H_0 כאשר היא הייתה נכונה.
 ב. החוקר דחה את השערת H_1 כאשר היא הייתה נכונה.
 ג. החוקר לא דחה את השערת H_0 כאשר היא הייתה לא נכונה.
 ד. המדגם של החוקר שייך בפועל להתפלגות הדגימה של H_1 .

14. חוקר ביקש לבחון האם תאומים זהים אשר הופרדו בילדותם שונים מתאומים זהים אשר גדלו יחדיו מבחינת מידת הפער בין התאומים בלחץ הדם. הוא דגם 20 זוגות תאומים מכל אוכלוסייה ומדד את הפרש בין לחץ הדם בכל זוג תאומים. מהו המבחן הסטטיסטי המתאים?
- א. מבחן t למדגמים בלתי תלויים עם 38 דרגות חופש.
 ב. מבחן t למדגמים מזווגים, עם 39 דרגות חופש.
 ג. מבחן t למדגמים בלתי תלויים עם 39 דרגות חופש.
 ד. מבחן t למדגמים מזווגים עם 38 דרגות חופש.

15. שלושה חוקרים רצו לבדוק את השפעתו של שידור פרסומות נגד תאונות דרכים על מהירות הנהיגה של נהגים בישראל (השונות של מהירות הנהיגה בישראל אינה ידועה). עידו השווה את מהירות הנהיגה של קבוצת נהגים אחת, 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.
- שני החוקרים בהכרח יגיעו לאותה מסקנה.
- 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.
- בסיכוי של 2% נדחה את השערת האפס.
 - בסיכוי של 2% לא נדחה את השערת האפס.
 - בסיכוי של 2% השערת האפס לא נכונה.
 - 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:
- כעת בברור הוא ידחה את השערת האפס.
 - כעת הוא דווקא יקבל את השערת האפס.
 - 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:

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

ה. לדחות את השערת האפס.

ו. לא לדחות את השערת האפס.

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

ח. שלא נתונות בשאלה סטיות התקן של האוכלוסיות, ולכן לא ניתן להסיק דבר.

38. במטרה לבדוק האם קיים הבדל בין קווי זהב לבזק מבחינת ממוצע המחירים

לשיחות בינ"ל. נגדמו באקראי 7 מדינות ועבור כל מדינה נבדקה עלות דקת שיחה.

בהנחה והמחירים מתפללים נורמלית בנו רווח סמך לממוצע ההפרשים וקיבלו:

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

לכן מסקנת המחקר היא:

ד. ברמת מובהקות של 5% לא נוכל לקבוע שקיים הבדל בין החברות.

ה. ברמת מובהקות של 5% נקבע שקיים הבדל מובהק בין החברות.

ו. לא ניתן לדעת מה המסקנה ברמת מובהקות של 5% כיוון שלא נאמר מה ההגדרה

של D.

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

א. תמיד נדחה H_0 כאשר היא נכונה אך לא תמיד נדחה אותה כאשר היא לא נכונה.

ב. לא נדחה את H_0 אף פעם.

ג. לא נדחה את H_0 כאשר היא נכונה אך יתכן ונדחה אותה כאשר היא לא נכונה.

ד. כל התשובות לא נכונות.

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

מדגם מקרי בגודל 5 מתוך אוכלוסייה המתפלגת נורמאלית עם שונות לא ידועה. על סמך

תוצאות המדגם הוא חישב וקיבל: $t_x = -2.63$. לכן המסקנה היא:

- הוא ידחה H_0 ברמת מובהקות 0.1 אך לא כן ברמת מובהקות 0.05.
- הוא ידחה H_0 ברמת מובהקות 0.05 אך לא כן ברמת מובהקות 0.025.
- הוא ידחה H_0 ברמת מובהקות 0.025 אך לא כן ברמת מובהקות 0.01.
- הוא לא ידחה H_0 ברמת מובהקות 0.1.

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

מה הסיכוי לבצע טעות מסוג ראשון במחקר?

- 0.005
- 0.03
- 0.0287
- 0.05

42. חוקרת שיערה, כי משקלן של נשים כשנה לאחר החתונה גבוה ממשקלן בעת החתונה. החוקרת דגמה 15 נשים, ובדקה את משקלן בשתי נקודות הזמן (בעת החתונה, ושנה לאחריה), אך לא מצאה הבדל מובהק ברמת מובהקות 0.01 בהנחה, כי במציאנות השערותה 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?

- כל התצפיות במחקר אינן תלויות זו בזו.
- כל התצפיות במחקר תלויות זו בזו.
- כל הצמדים של תצפיות במחקר אינם תלויים זה בזה.
- 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:

$$H_1 : \mu \neq 10$$

- ניתן להכריע בין ההשערות רק אם שונות האוכלוסייה נתונה.
- מקבלים את השערת האפס.
- דוחים את השערת האפס.
- 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.

- שלושת הטענות אינן נכונות.
- טענות 2 ו-3 אינן נכונות וטענה 1 נכונה.
- טענות 1 ו-2 נכונות וטענה 3 אינה נכונה.
- טענות 1 ו-3 אינן נכונות וטענה 2 נכונה.

פתרונות :

א	26	א	1
ב	27		
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		
א	38	א	13
ג	39	א	14
א	40	א	15
א	41	א	16
א	42		
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 15 - 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

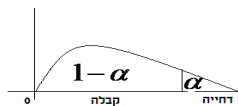
H_1 : יש תלות בין המשתנים

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

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

סטטיסטי המבחן:

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

O_i - השכיחות נצפית במדגם בתא i .

E_i - שכיחות צפויה במדגם בתא i בהנחת השערת האפס.

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

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

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

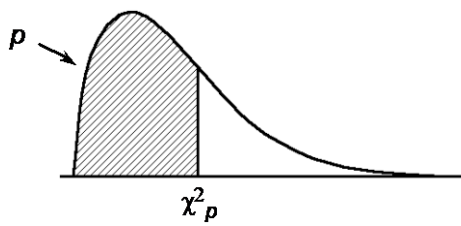
מהתאים E קטן מ-5.

דוגמה : (הפתרון בהקלטה)

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

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

טבלת התפלגות חי-בריבוע – ערכי החלוקה χ^2_p



p

df	.005	.01	.025	.05	.10	.25	.50	.75	.90	.95	.975	.99	.995
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 :

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

מה המסקנה ברמת מובהקות של 2.5%?

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

	יום	ערב	לילה
פגומים	50	60	70
תקינים	600	700	800

האם יש הבדל בין שיעורי הפגומים

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

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

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.

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		
hair_color	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 הביילוי 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%
		% within	21.6%	53.3%
		% of Total	15.4%	15.4%
חום	Count			
		% within	83.3%	16.7%
		% within	27.0%	13.3%
		% of Total	19.2%	3.8%
ירוק	Count			
		% within	79.2%	20.8%
		% within	51.4%	33.3%
		% of Total	36.5%	9.6%
Total	Count			
		% within	71.2%	28.8%
		% within	100.0%	100.0%
		% of Total	71.2%	28.8%

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