

מבוא לסטטיסטיקה ב 55003

פרק 1

הסקה סטטיסטית - הקדמה

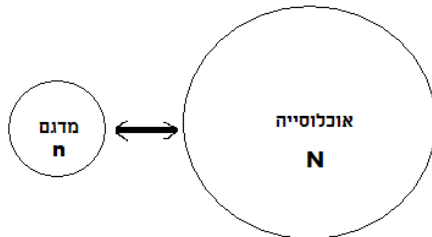
1 כללי

הסקה סטטיסטית – הקדמה:

רקע:

אוכלוסייה:

קבוצה שאליה מפנים שאלה מחקרית. למשל, חברת תרופות שמעוניינת לפתח תרופה למחלת הסוכרת מתעניינת באוכלוסיית חולי הסוכרת בעולם.



מדגם:

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

במקרים רבים אין אפשרות לחקור את כל האוכלוסייה כיוון שאין גישה לכולה, היא גדולה מידי, או מוגבלים בזמן ובאמצעים טכניים ולכן מבצעים מדגם במטרה לבצע הסקה סטטיסטית מהמדגם לאוכלוסייה. הדגימה בקורס תהיה דגימה מקרית - הכוונה לדגימה שבה לכל תצפית באוכלוסייה יש את אותו סיכוי להיכלל במדגם.

סטטיסטי:

גודל המחושב על המדגם.

פרמטר:

גודל המתאר את האוכלוסייה.

הסימונים לפרמטר וסטטיסטי הם שונים:

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

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

דוגמה (פתרון בהקלטה):

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

- א. מי האוכלוסייה?
- ב. מה המשתנה?
- ג. מה הפרמטרים?
- ד. מהו גודל המדגם?
- ה. מהו הסטטיסטי שמתכננים להוציא מהמדגם?
- ו. האם הפרמטר או הסטטיסטי הוא משתנה מקרי?

שאלות:

- (1) מתוך כלל הסטודנטים במכללה שסיימו סטטיסטיקה א נדגמו שני סטודנטים. נתון שממוצע הציונים של כלל הסטודנטים היה 78 עם סטיית תקן של 15.
- מי האוכלוסייה?
 - מה המשתנה?
 - מהם הפרמטרים?
 - מהו גודל המדגם?
- (2) להלן התפלגות מספר מקלטי 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.
- מיהי האוכלוסייה ומהו המשתנה הנחקר?
 - מהו הסטטיסטי שיילקח מהמדגם ומה סימונו?

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

- (3) נתון כי 20% מהשכירים במדינה הם אקדמאיים. נבחרו באקראי 10 שכירים באותה אוכלוסייה ומתכננים לפרסם את מספר האקדמאיים שנדגמו.
- מיהי האוכלוסייה?
 - מה המשתנה באוכלוסייה?
 - מהם הפרמטרים?
 - מהו הסטטיסטי?

תשובות סופיות:

- (1) א. כלל הסטודנטים במכללה שסיימו סטטיסטיקה א. ב. ציון. ג. ממוצע: 78, סטיית תקן: 15. ד. 2.
- (2) א. האוכלוסייה: 1000 משפחות בישוב העוגן, המשתנה הנחקר: מס' מקלטים. ב. \bar{X} = ממוצע מדגם.
- (3) א. השכירים במדינה. ב. השכלה: אקדמאי, לא אקדמאי. ג. שיעור ההצלחות באוכלוסייה: 0.2. ד. מס' האקדמאים במדגם.

מבוא לסטטיסטיקה ב 55003

פרק 2

התפלגות הדגימה ומשפט הגבול המרכזי

1. התפלגות ממוצע המדגם ומשפט הגבול המרכזי

התפלגות 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 .

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. מיהי אוכלוסיית המחקר? מהו המשתנה הנחקר?

ב. מהם הפרמטרים של האוכלוסייה?

ג. מה התוחלת ומהי סטיית התקן של ממוצע המדגם?

דגימה מהתפלגות נורמאלית:

אם נדגום מתוך אוכלוסייה שהמשתנה בה מתפלג נורמאלית עם ממוצע μ ושונות σ^2 .

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

ממוצע המדגם גם יתפלג נורמאלית:

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?

משפט הגבול המרכזי:

אם אוכלוסייה מתפלגת כלשהו עם ממוצע μ ושונות σ^2 אזי עבור מדגם מספיק

$$\bar{x} \rightsquigarrow N\left(\mu, \frac{\sigma^2}{n}\right) \quad (n \geq 30)$$

ממוצע המדגם מתפלג בקירוב נורמאלי:

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?

שאלות:

- (1) מתוך כלל הסטודנטים במכללה שסיימו סטטיסטיקה א נדגמו שני סטודנטים. נתון שממוצע הציונים של כלל הסטודנטים היה 78 עם סטיית תקן של 15.
- מיהי האוכלוסייה?
 - מה המשתנה?
 - מהם הפרמטרים?
 - מהו גודל המדגם?
 - מהו תוחלת ממוצע המדגם?
 - מהי טעות התקן?

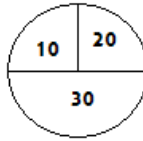
- (2) להלן התפלגות מספר מקלטי הטלוויזיה למשפחה בישוב מסוים:

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

- בנו את פונקציית ההסתברות של X .
 - חשבו את התוחלת, השונות וסטיית התקן של X .
 - 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.
- מה ההסתברות שתינוק אקראי בעת הלידה ישקול פחות מ-3800 גרם? נתון כי ביום מסוים נולדו 4 תינוקות.
 - מה ההסתברות שהמשקל הממוצע שלהם יעלה על 4 ק"ג?
 - מה ההסתברות שהמשקל הממוצע של התינוקות יהיה מתחת ל-2.5 ק"ג?
 - מה ההסתברות שהמשקל הממוצע של התינוקות יהיה רחוק מהתוחלת בלא יותר מ-50 גרם?
 - 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.
- מה ההסתברות שהגובה הממוצע שלהם יהיה לפחות 190 ס"מ?
 - מה ההסתברות שהגובה הממוצע שלהם יהיה בדיוק 180 ס"מ?
 - מה ההסתברות שהגובה הממוצע שלהם יסטה מתחולת הגבהים בפחות מ-5 ס"מ?
 - 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 הפתרון הניחו שזמן הנסיעה לעבודה מתפלג נורמאלית.
- מה ההסתברות שבמשך שבוע משך הנסיעה הממוצע יהיה מעל 33 דקות?
 - מהו הזמן שבהסתברות של 90% ממוצע משך הנסיעה השבועי יהיה גבוה ממנו?
 - מה ההסתברות שממוצע משך הנסיעה השבועי יהיה מרוחק מ-30 דקות בלפחות 2 דקות?
 - 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.
- בארגו 4 בקבוקי יין. מה ההסתברות שהנפח הממוצע של הבקבוקים בארגו יהיה בדיוק 755 סמ"ק?
 - בארגו 4 בקבוקי יין. מה ההסתברות שהנפח הממוצע של הבקבוקים בארגו יהיה יותר מ-755 סמ"ק?
 - בארגו 4 בקבוקי יין. מה ההסתברות שהנפח הממוצע של הבקבוקים בארגו יהיה לפחות 755 סמ"ק?
 - בקבוקי היין שבארגו נמזגים לקערה עם קיבולת של שלושה ליטר. מה ההסתברות שהיין יגלוש מהקערה?
- 8) משתנה מתפלג נורמאלית עם תוחלת של 80 וסטיית תקן 4.
- מה ההסתברות שממוצע המדגם יסטה מתוחלתו בלא יותר מיחידה כאשר גודל המדגם הוא 9?
 - מה ההסתברות שממוצע המדגם יסטה מתוחלתו בלא יותר מיחידה שגודל המדגם הוא 16?
 - 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.
- בנו את פונקציית ההסתברות של סכום הזכייה במשחק בודד.
 - מה התוחלת ומה השונות של סכום הזכייה?
 - אם האדם ישחק את המשחק 5 פעמים מה התוחלת ומה השונות של ממוצע סכום הזכייה בחמשת המשחקים?
 - 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) קובייה הוטלה 50 פעמים. מה ההסתברות שהממוצע של התוצאות יהיה לפחות 3.72?

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

- א. 0.
- ב. 0.5.
- ג. 1.
- ד. לא ניתן לדעת.

15) נתון ש- X מתפלג כלשהו עם תוחלת μ ושונויות σ^2 . החליטו לבצע מדגם בגודל 200 מתוך ההפלגות הנתונה לפי משפט הגבול המרכזי מתקיים (בחרו בתשובה הנכונה):

- א. $X \sim N\left(\mu, \frac{\sigma^2}{200}\right)$.
- ב. $\mu \sim N\left(\mu, \frac{\sigma^2}{200}\right)$.
- ג. $\bar{X} \sim N(\mu, \sigma^2)$.
- ד. $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{200}\right)$.

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 (בחרו 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.

תשובות סופיות:

- (1) א. כלל הסטודנטים במכללה שסיימו סטטיסטיקה א. ב. ציון.
ג. ממוצע: 78, סטיית תקן: 15.
ה. 78.

(2) א. להלן טבלה:

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

ב. $\sigma = 0.973$, $\sigma^2 = 0.9475$, $\mu = 2.05$

ג. $\sigma(\bar{X}) = 0.486$, $\sigma_{\bar{x}}^2 = 0.2369$, $\mu_{\bar{x}} = 2.05$

(3) $\sigma(\bar{X}) = 1.21$, $\mu_{\bar{x}} = 3.5$

- (4) א. 0.8413 ב. 0.0013 ג. 0 ד. 0.1974
- (5) א. 0 ב. 0 ג. 0.9544 ד. 0.178205
- (6) א. 0.0465 ב. 27.71 ג. 0.2628 ד. התשובה הייתה קטנה.
- (7) א. 0 ב. 0.1587 ג. 0.1587 ד. 0.5
- (8) א. 0.5468 ב. 0.6826
- (9) א. להלן טבלה:

30	20	10	X
0.5	0.25	0.25	P(X)

- א. התוחלת: 22.5, השונות: 68.75.
- ב. התוחלת: 22.5, השונות: 13.75.
- ד. 0.8997
- (10) 0.0475
- (11) 0.1814
- (12) א. 0.9772 ב. 0.0228 ג. 0.271
- (13) 0.5
- (14) ב'
- (15) ד'
- (16) ג'

מבוא לסטטיסטיקה ב 55003

פרק 3

מושגי יסוד באמידה

1 כללי

מושגי יסוד באמידה:

רקע:

כזכור מהמפגש הקודם, פרמטר הוא גודל המתאר את האוכלוסייה או התפלגות מסוימת. כמו ממוצע הגבהים בקרב מתגייסים 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- $\hat{\theta}$. 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 הטיה ל- θ אם התוחלת של $\hat{\theta}$ תהיה שווה ל- θ : $E(\hat{\theta}) = \theta$.
 - טעות התקן של אומד היא סטיית התקן שלו, כלומר: $\sigma(\hat{\theta}) = S.E$.

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: $\sigma(\bar{x}) = \frac{\sigma}{\sqrt{n}} = SE$

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.

שאלות:

- (1) מתוך 500 טירונים, נמצאו 120 בעלי שברי הליכה. נתון שהסיכוי שטירון יהיה עם שבר הליכה הוא 0.25.
- מהי האוכלוסייה המוצגת בשאלה? מהם הפרמטרים שלה?
 - מהי טעות התקן של האומדן כשהמדגם בגודל 500?
 - מהו האומדן לפרמטר?
 - 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.
- מהי האוכלוסייה המוצגת בשאלה? מהם הפרמטרים שלה?
 - מהי טעות התקן של האומדן?
 - מהו האומדן לפרמטר?
 - 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.
- מצאו אומדן חסר הטיה לגובה הממוצע של מתגייסי צה"ל.
 - מצאו אומדן חסר הטיה לשונות הגבהים של מתגייסי צה"ל.
 - מצאו 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: $\sum_{i=1}^{20} X_i = 162$, $\sum_{i=1}^{20} X_i^2 = 1502.2$.
- אמדו את השכר הממוצע של השכירים במשק.
 - אמדו את סטיית התקן של שכר השכירים במשק.
- (5) במטרה לאמוד את ממוצע האוכלוסייה, דגמו תצפיות בלתי תלויות מהאוכלוסייה וחישבו את הממוצע שלהם. מהי טעות התקן?
- סטיית התקן של האוכלוסייה.
 - סטיית התקן של ממוצע האוכלוסייה.
 - סטיית התקן של המדגם.
 - סטיית התקן של ממוצע המדגם.

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

- א. 3.
- ב. 2.5.
- ג. 1.581.
- ד. אין מספיק נתונים לדעת.

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

- א. כאשר n קטן.
- ב. כאשר תצפיות המדגם אינן בלתי תלויות.
- ג. כאשר האוכלוסייה אינה מתפלגת נורמאלית.
- ד. כאשר מעוניינים באומד 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 האומד 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.

תשובות סופיות:

- (1) א. 0.25 ב. 0.019 ג. 0.24 ד. 0.01
- (2) א. אוכלוסייה: מקררים של יצרן, תוחלת: 2400, סטיית תקן: 500.
 ב. 100 ג. 2342 ד. 58
- (3) א. 177.9 ב. 64.1 ג. 0.4
- (4) א. 8.1 ב. 3.16
- (5) ד.
- (6) ג.
- (7) ד.
- (8) ד.
- (9) ג.

מבוא לסטטיסטיקה ב 55003

פרק 4

רווח סמך לתוחלת ~ סוגריים 2 ~ ממוצע ~ סוגריים 1 ~

- 1 רווח סמך כששונות האוכלוסיה ידועה
- 6 קביעת גודל מדגם
- 8 רווח סמך כששונות האוכלוסיה לא ידועה.

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 האמיתי הוא אפסי.

מה שנהוג לעשות כדי לאמוד את ממוצע האוכלוסייה, זה לבנות רווח סמך.

בנה 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%.

1. מהי אוכלוסיית המחקר?

2. מה המשתנה באוכלוסייה?

3. מה הפרמטר שהחוקר רצה לאמוד?

4. מהו רווח הסמך?

5. מה אורך רווח הסמך?

6. מהי רמת הביטחון של רווח הסמך?

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

פרמטר אותו נרצה לאמוד: μ .

אומד נקודתי: \bar{x} .

תנאים לבניית רווח הסמך: $X \sim N$ או $n \geq 30$.

σ^2 (שוונות האוכלוסייה) ידועה.

$$\text{נוסחה לרווח הסמך: } \bar{x} \pm Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

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.

$$\text{שגיאת האמידה המקסימלית: } \varepsilon = Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

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 האומד יותר מדויק, ולכן נקבל רווח סמך יותר קצר.
- ככל שרמת הביטחון ($1-\alpha$) גבוהה יותר, כך $z_{1-\frac{\alpha}{2}}$ יותר גבוה, ורווח הסמך יותר ארוך.

שאלות:

- 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 ₪.
- מי האוכלוסייה במחקר?
 - מה המשתנה הנחקר?
 - מה הפרמטר שאותו רוצים לאמוד?
 - מה רווח הסמך לפרמטר?
 - מהי רמת הסמך לפרמטר?
 - מה אורך רווח הסמך?
 - 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. בנו את רווח הסמך.
- 3) מעוניינים לאמוד את ממוצע אורך החיים של מכשיר. מנתוני היצרן ידוע שאורך החיים מתפלג נורמאלית עם סטיית תקן של 20 שעות. נדגמו 25 מכשירים ונמצא כי ממוצע אורך החיים שלהם היה 230 שעות.
- בנו רווח סמך ברמת סמך של 90% לאורך החיים הממוצע של מכשיר.
 - בנו רווח סמך ברמת סמך של 95% לאורך החיים הממוצע של מכשיר.
 - הסבירו כיצד ומדוע השתנה רווח הסמך.
- 4) דגמו 200 עובדים מהמשק הישראלי. השכר הממוצע שלהם היה 9700 ₪. נניח שסטיית התקן של השכר במשק היא 3000 ₪.
- בנו רווח סמך ברמת סמך של 95% לתוחלת השכר במשק.
 - מה ניתן לומר בביטחון של 95% על הסטייה המרבית בין ממוצע המדגם לתוחלת השכר?
 - מה היה צריך להיות גודל המדגם אם הינו רוצים להקטין את רווח הסמך ב-50%?
 - 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 (105,99). שחזרו את:
- ממוצע המדגם.
 - שגיאת האמידה המקסימאלית.
 - 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.
- בנו רווח סמך לתוחלת זמן ההחלמה תחת האנטיביוטיקה החדשה ברמת סמך של 90%.
 - מה היה קורה לאורך רווח הסמך אם היה תקציב להגדלת גודל המדגם פי 4? הסבירו.
 - 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.
- מהו ממוצע המדגם?
 - מהי רמת הסמך של רווח הסמך שנבנה?
 - 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%, איזה מהמשפטים הבאים לא יהיה נכון.
- אורך רווח הסמך החדש יהיה קטן יותר.
 - גודל המדגם יהיה כעת קטן יותר.
 - המרחק בין ממוצע המדגם לקצות רווח הסמך יהיו קטנים יותר ברווח הסמך החדש.
 - 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 (בחרו בתשובה הנכונה):

א. רמת הביטחון.

ב. סטיית התקן באוכלוסייה.

ג. מספר המשתתפים.

ד. סטיית התקן במדגם.

תשובות סופיות:

1) א. העובדים במשק. ב. שכר ב-ש. ג. μ . ד. $9200 < \mu < 9800$.

ה. 0.95. ו. 600. ז. 0.05.

2) $4920.6 < \mu < 4979.4$

3) א. $223.42 < \mu < 236.58$. ב. $222.16 < \mu < 237.84$.

ג. ראה סרטון.

4) א. $10,116 < \mu > 9284$. ב. הסטיה המירבית בין \bar{x} ל- μ היא 416 שם בבטחון של 95%.

ג. 800. ד. לא.

5) א. 102. ב. 3. ג. 0.9544.

6) א. $4.42 < \mu < 83.5$. ב. יקטן פי 2. ג. גדל.

7) א. 87. ב. 5. ג. 0.9544.

8) ב'.

9) ג'.

10) ד'.

קביעת גודל מדגם:

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

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

בנוסחה הבאה:

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 האמתי לא יעלה על 0.5 ק"ג? (תשובה: 87).

שאלות:

- (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 מתגיסיים לצבא. מעוניינים שבביטחון של 95% שגיאת האמידה המרבית תהיה 0.5. נניח שהדופק מתפלג נורמאלית על סטיית תקן של 3 פעימות לדקה.
 א. כמה מתגיסיים 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 0.5σ . מהו גודל המדגם הנדרש?

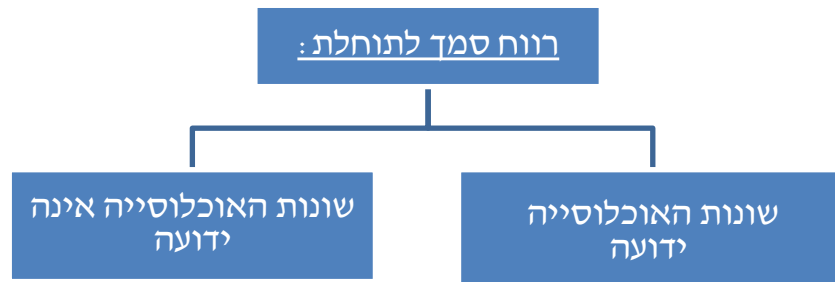
תשובות סופיות:

- (1) .780
 (2) א. 139. ב. הדבר יקטין את ε פי 2.
 (3) $n = 62$.

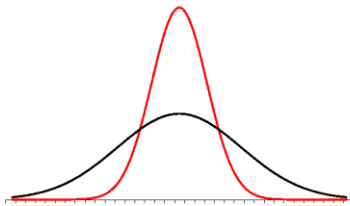
רווח סמך כששונות האוכלוסייה לא ידועה:

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.

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

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

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 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: 5.3, 4.6, 5.2, 4.7.

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

שאלות:

- (1) מחקר מעוניין לדעת כיצד תרופה מסוימת משפיעה על קצב פעימות הלב. ל-5 אנשים שנטלו את התרופה מדדו את הדופק והתקבל מספר פעימות לדקה: 84, 88, 84, 79, 89. הערה: לצורך פתרון הנח שקצב פעימות הלב מתפלג נורמאלית בקירוב.
- א. בנו רווח סמך ברמת סמך של 95% לתוחלת הדופק של נוטלי התרופה הנ"ל.
 ב. נתון שהדופק הממוצע ללא לקיחת התרופה הינו 70. לאור זאת, האם בביטחון של 95% התרופה משפיעה על הדופק?
 ג. בהמשך לסעיף א', אם היינו בונים את רווח הסמך ברמת ביטחון של 99%, כיצד הדבר היה משפיע על רווח הסמך?
- (2) במדגם שנעשה על 25 מתגייסים לצבא האמריקאי התקבל כי גובה ממוצע של חייל הינו 178 ס"מ עם סטיית תקן $S = 13$ 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: 30, 40, 32, 34, 27.
 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 האמיתית 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?
- א. סטטיסטיקאי א'.
 ב. סטטיסטיקאי ב'.
 ג. אותו אורך רווח סמך לשני הסטטיסטיקאים.
 ד. תלוי בתוצאות המדגם של כל סטטיסטיקאי.

תשובות סופיות:

- (1) א. $79.88 < \mu < 89.72$. ב. כן. ג. הוא היה גדל.
- (2) ראה בסרטון.
- (3) א. צריך להניח שהמשתנה מתפלג נורמלית. ב. לא ניתן לדעת.
- (4) א. $96.63 < \mu < 107.37$. ב. $96.90 < \mu < 107.10$. ג. ראה בסרטון.
- (5) $3.149 < \mu < 3.351$.
- (6) $13.42 < \mu < 14.18$.
- (7) ב'.

מבוא לסטטיסטיקה ב 55003

פרק 5

מבוא לבדיקת השערות על פרמטרים

1	הקדמה
5	סוגי טעויות

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 שנקרא כלל הכרעה. הכלל יוצר אזורי שנקראים:

1. **אזור דחייה:**

דחייה של השערת האפס כלומר קבלה של האלטרנטיבה.

2. **אזור 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 לידתם בישראל 3300 גרם. 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?

שאלות:**בשאלות הבאות, ענו על הסעיפים הבאים:**

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

- 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.
- 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.
- 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.
- 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.

תשובות סופיות:

- (1) א. נבחנים בבגרות באנגלית.
 ב. ציון.
 ג. ממוצע הציונים בשיטת לימוד חדשה.
 ד. $H_0: \mu = 72$
 $H_1: \mu > 72$
- (2) א. משקאות בבקבוק של חברה מסוימת.
 ב. נפח משקה בסמ"ק.
 ג. ממוצע נפח המשקה בבקבוק.
 ד. $H_0: \mu = 500$
 $H_1: \mu < 500$
- (3) א. מועמדים לפקולטה למשפטים.
 ב. משתנה דיכוטומי (התקבל, לא התקבל).
 ג. אחוז הקבלה.
 ד. $H_0: p = 0.25$
 $H_1: p < 0.25$
- (4) א. אזרחים בוגרים במשק.
 ב. משתנה דיכוטומי (מובטל, עובד).
 ג. אחוז האבטלה כיום.
 ד. $H_0: p = 0.08$
 $H_1: p \neq 0.08$

סוגי טעויות:

רקע:

בתהליך של בדיקת השערות יוצרים כלל שניקרא כלל הכרעה. הכלל יוצר אזורים שנקראים:

1. אזור דחייה – דחייה של השערת האפס כלומר קבלה של האלטרנטיבה.
2. אזור 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. אם נשנה את כלל ההכרעה אז אנחנו יכולים לקבל מסקנה אחרת. אם נבצע מדגם חדש אז אנחנו עלולים לקבל תוצאה אחרת. לכן יתכנו טעויות במסקנות שלנו:

		הכרעה	
		H_0	H_1
מציאות	H_0	אין טעות	טעות מסוג 1
	H_1	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?

שאלות:

- (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.
- רשמו את השערות המחקר.
 - מה מסקנת המחקר?
 - איזו סוג טעות יתכן וביצעו במחקר?
- (2) במחקר על פרמטר מסוים הוחלט בסופו של דבר לדחות את השערת האפס.
- האם ניתן לדעת אם בוצע טעות במחקר?
 - 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.
- מהי אוכלוסיית המחקר?
 - מה המשתנה הנחקר?
 - מה הפרמטר הנחקר?
 - מה השערות המחקר?
 - מה מסקנת המחקר?
 - מהי סוג הטעות האפשרית במחקר?

תשובות סופיות:

- (1) א. $H_0: \mu = 500$
ב. $H_1: \mu < 500$
ג. טעות מסוג שני.
- (2) א. לא ניתן לדעת.
ב. טעות מסוג ראשון.
ג. משפחות כיום.
- (3) א. תוחלת מספר הילדים למשפחה כיום.
ב. לא לדחות את H_0 .
ג. טעות מסוג שני.
- ד. $H_0: \mu = 2.3$
ה. $H_1: \mu < 2.3$

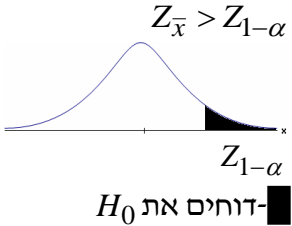
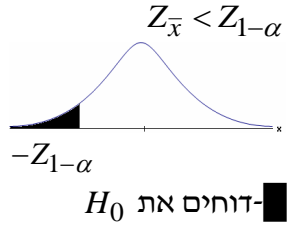
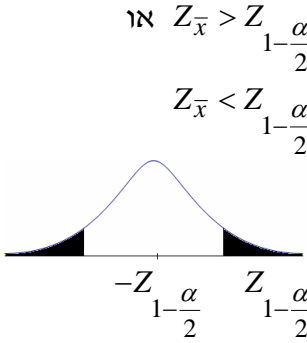
תוכן העניינים:

2	בדיקת השערות על תוחלת (ממוצע).....
2	כאשר שונות האוכלוסייה ידועה – רקע.....
4	שאלות:.....
5	תשובות סופיות:.....
6	סיכוי לטעויות ועוצמה כאשר שונות האוכלוסייה ידועה – רקע.....
8	שאלות:.....
10	תשובות סופיות:.....
11	קביעת גודל מדגם כששונות האוכלוסייה ידועה – רקע.....
12	שאלות:.....
12	תשובות סופיות:.....
13	מובהקות התוצאה (value-p) בבדיקת השערות על תוחלת עם שונות ידועה – רקע.....
14	שאלות:.....
15	תשובות סופיות:.....
16	בדיקת השערות על תוחלת (ממוצע) כאשר שונות האוכלוסייה אינה ידועה – רקע.....
18	שאלות:.....
19	תשובות סופיות:.....
20	מובהקות התוצאה (value-p) כאשר שונות האוכלוסייה לא ידועה – רקע.....
21	שאלות:.....
21	תשובות סופיות:.....
22	ניתוח פלטי SPSS – רקע.....
24	שאלות:.....
26	תשובות סופיות:.....
27	הקשר בין רווח סמך לבדיקת השערות על תוחלת – רקע.....
28	שאלות:.....
28	תשובות סופיות:.....

לתשומת לבך, יש ללמוד לפי הסרטונים באתר. ייתכנו שאלות בספר הפרק אשר אינן חלק מחומר הלימוד שלך.

בדיקת השערות על תוחלת (ממוצע)

כאשר שונות האוכלוסייה 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			
 <p>$Z_{\bar{x}} > Z_{1-\alpha}$</p> <p>$Z_{1-\alpha}$</p> <p>דוחים את H_0</p>	 <p>$Z_{\bar{x}} < -Z_{1-\alpha}$</p> <p>$-Z_{1-\alpha}$</p> <p>דוחים את H_0</p>	 <p>או $Z_{\bar{x}} > Z_{1-\frac{\alpha}{2}}$ $Z_{\bar{x}} < -Z_{1-\frac{\alpha}{2}}$</p> <p>$-Z_{1-\frac{\alpha}{2}}$ $Z_{1-\frac{\alpha}{2}}$</p> <p>דוחים את H_0</p>	כלל ההכרעה: אזור הדחייה של H_0:

$$Z_{\bar{x}} = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \quad \text{סטטיסטי המבחן:}$$

חלופה אחרת לכלל הכרעה:

$\bar{X} > \mu_0 + Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$	$\bar{X} < \mu_0 + Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$	$\bar{X} > \mu_0 + Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$ או $\bar{X} < \mu_0 + Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$	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. בדקו את ההשערה ברמת מובהקות של 1%.

שאלות:

- (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 תהיה 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 מכיילת (מאופסת). המכונה כוונה לחתוך מוטות באורך 50 ס"מ. לפי נתוני היצרן סטיית התקן בחיתוך המוטות היא 0.5 ס"מ. במדגם של 50 מוטות התקבל ממוצע אורך המוט 50.93 ס"מ. מה מסקנתכם ברמת מובהקות של 5%?
- (4) המשקל הממוצע של הספורטאים בתחום ספורט מסוים הוא 90 ק"ג, עם סטיית תקן 8 ק"ג. לפי דעת מומחים בתחום יש צורך בהורדת המשקל ובשימוש בדיאטה מסוימת שצריכה להביא להורדת המשקל. לשם בדיקת יעילות הדיאטה נלקח מדגם מקרי של 50 ספורטאים ובתום שנה של שימוש בדיאטה התברר שהמשקל הממוצע במדגם זה היה 84 ק"ג.
יש לבדוק בר"מ של 10%, האם הדיאטה גורמת להורדת המשקל.
- (5) לפי מפרט נתון, על עובי בורג להיות 4 מ"מ עם סטיית תקן של 0.2 מ"מ. במדגם של 25 ברגים העובי הממוצע היה 4.07 מ"מ.
קבעו ברמת מובהקות 0.05, האם עובי הברגים מתאים למפרט.
הניחו כי עובי של בורג מתפלג נורמלית וסטיית התקן של עובי בורג היא אכן 0.2 מ"מ.
- (6) במחקר נמצא שתוצאה היא מובהקת ברמת מובהקות של 5% מה תמיד נכון? בחרו בתשובה הנכונה.
א. הגדלת רמת המובהקות לא תשתנה את מסקנת המחקר.
ב. הגדלת רמת המובהקות תשנה את מסקנת המחקר.
ג. הקטנת רמת המובהקות לא תשנה את מסקנת המחקר.
ד. הקטנת רמת המובהקות תשנה את מסקנת המחקר.

(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:

- א. השערת האפס הייתה נדחית.
- ב. השערת האפס הייתה לא נדחית.
- ג. 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.

- א. אם חוקר א' החליט לדחות את H_0 , מה יחליט חוקר ב'? נמקו.
- ב. אם חוקר א' יחליט לא לדחות את H_0 , מה יחליט חוקר ב'? נמקו.

תשובות סופיות:

- (1) נקבל H_0 .
- (2) נדחה H_0 .
- (3) נדחה H_0 .
- (4) נדחה H_0 .
- (5) נקבל H_0 .
- (6) א
- (7) ג
- (8) א. אותה מסקנה. ב. לא ניתן לדעת.

סיכוי לטעויות ועוצמה כאשר שונות האוכלוסייה ידועה - רקע

מציאות	הכרעה	
	H_0	H_1
H_0	אין טעות	טעות מסוג 1
H_1	טעות מסוג 2	אין טעות

הגדרת הסתברויות

הסיכוי לבצע טעות מסוג 1 (רמת מובהקות)

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

הסיכוי לבצע טעות מסוג 2:

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

רמת בטחון:

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

עוצמה:

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

התהליך לחישוב סיכוי לטעות מסוג שני:

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

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

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

התקנון:

דוגמה (פתרון בהקלטה):

- בתחילת השנה חשבון הטלפון הסלולארי הממוצע לאדם היה ש200 עם סטיית תקן של ש80 לחודש. בעקבות כניסתן של חברות טלפון סלולארית חדשות מעוניינים לבדוק האם כיום ממוצע חשבון הטלפון הסלולארי פחת. לצורך בדיקה דגמו באקראי 36 אנשים וחשבון הטלפון הסלולארי שלהם היה ש150 בממוצע לחודש.
- א. רשמו את השערות המחקר ובנו כלל הכרעה במונחי חשבון ממוצע מדגמי ברמת מובהקות של 5%.
 - ב. מה מסקנתכם? איזה סוג טעות אפשרית במסקנה?
 - ג. נניח שבמציאות כיום החשבון הממוצע הוא ש160. מה הסיכוי לבצע טעות מסוג שני?
 - ד. אם נקטין את רמת המובהקות מסעיף א', כיצד הדבר ישפיע על התשובה מסעיף ג'?

שאלות:

- (1) נתון ש $X \sim N(\mu, \sigma^2 = 1)$.
להלן השערות של חוקר לגבי הפרמטר μ : $H_0: \mu = 5$, $H_1: \mu = 7$.
מעוניינים ליצור כלל הכרעה המתבסס על הסמך תצפית בודדת כך שרמת המובהקות תהיה 5%.
א. עבור אילו ערכים של X שידגם נדחית השערת H_0 ?
ב. מה הסיכוי לבצע טעות מסוג שני?
ג. אם במדגם התקבל ש $X = 6.9$ מה תהיה המסקנה ומה הטעות האפשרית?
- (2) לפי נתוני משרד הפנים בשנת 1980 למשפחה ממוצעת היה 2.3 ילדים למשפחה עם סטיית תקן 0.4. מעוניינים לבדוק אם כיום ממוצע מספר הילדים למשפחה קטן יותר. לצורך כך הוחלט לדגום 121 משפחות. במדגם התקבל ממוצע 2.17 ילדים למשפחה.
א. רשמו כלל הכרעה במונחי ממוצע מדגם קריטי ברמת מובהקות של 5%.
ב. בהמשך לסעיף א' מה תהיה המסקנה ומהי הטעות האפשרית במסקנה?
ג. אם באמת ממוצע מספר הילדים במשפחה פחת לכדי 2.1 מהי העצמה של הכלל מסעיף א'?
- (3) להלן נתונים על תהליך של בדיקת השערות על תוחלת:
 $H_0: \mu = 200$ $H_1: \mu \neq 200$ $\sigma = 30$ $n = 225$
א. רשמו כלל הכרעה במונחי ממוצע מדגם קריטי וברמת מובהקות של 10%.
ב. בהמשך לסעיף א', מהי העצמה אם התוחלת שווה ל-195?
ג. הסבירו, ללא חישוב, איך העצמה תשתנה אם רמת המובהקות תהיה 5%?
- (4) מפעל לייצור צינורות מייצר צינור שקוטרו מתפלג נורמלית עם תוחלת של 50 מ"מ וסטית תקן של 6 מ"מ. במחלקת ביקורת האיכות דוגמים בכל יום 81 צינורות ומוודדים את קוטרים, בכדי לבדוק, בעזרת מבחן סטטיסטי, האם מכונת הייצור מכוילת כנדרש או שקוטר הצינורות קטן מהדרוש.
א. רשמו את ההשערות ואת כלל ההכרעה ברמת מובהקות של 5%.
ב. אם ביום כלשהו מכונת הייצור התקלקלה והיא מייצרת את הצינורות בקוטר שתוחלתו 48 מ"מ בלבד (סטית התקן לא השתנתה), מה ההסתברות שהתקלה לא תתגלה בביקורת האיכות? כיצד נקראת הסתברות זו?
ג. הסבירו ללא חישוב כיצד התשובה לסעיף ב' תשתנה אם רמת המובהקות תגדל.
ד. הסבירו ללא חישוב כיצד התשובה לסעיף ב' תשתנה אם התוחלת האמיתית היא 47 ולא 48 מ"מ.

- (5) להלן השערות של מחקר: $H_0: \mu = 50$, $H_1: \mu = 58$. ידוע שסטיית התקן של ההתפלגות הינה 20.
 א. בנו כלל הכרעה שהסיכוי לטעות מסוג שני בו הוא 10%. מהי רמת המובהקות?
 ב. כיצד הייתה משתנה רמת המובהקות אם (כל סעיף בפני עצמו)?
 i. סטיית התקן הייתה יותר גדולה.
 ii. הסיכוי לטעות מסוג שני גדול יותר.

השאלות שלהלן הן שאלות רב-ברירה, בחרו בתשובה הנכונה ביותר:

- (6) אם חוקר החליט להגדיל את רמת המובהקות במחקר שלו אזי:
 א. הסיכוי לטעות מסוג ראשון גדל.
 ב. העוצמה של המבחן גדלה.
 ג. הסיכוי לטעות מסוג שני גדל.
 ד. תשובות א ו-ב נכונות.

- (7) חוקר ביצע מחקר ובו עשה טעות מסוג שני לכן:
 א. השערת האפס נכונה.
 ב. השערת האפס נדחתה.
 ג. השערת האפס לא נדחתה.
 ד. אף אחת מהתשובות לא נכונה בהכרח.

- (8) מה המצב הרצוי לחוקר המבצע בדיקת השערה:

$$1 - \beta \quad \alpha$$

- א. גדולה גדולה
 ב. גדולה קטנה
 ג. קטנה גדולה
 ד. קטנה קטנה

- (9) נערך שינוי בכלל ההחלטה של בדיקת השערה מסוימת ובעקבותיו אזור דחיית H_0 קטן. כל שאר הגורמים נשארו ללא שינוי. כתוצאה מכך:

- א. הן 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.

- 10) ידוע כי לחץ דם תקין באוכלוסייה הוא 120. רופא מניח שלחץ הדם בקרב עיתונאים גבוה יותר מהממוצע באוכלוסייה. הוא לקח מדגם של 60 עיתונאים וקיבל ממוצע 137. על סמך המדגם, הוא בודק טענתו ברמת מובהקות 0.02 ומסיק שלחץ הדם בקרב העיתונאים אינו גבוה יותר. מה הטעות האפשרית שהרופא עושה?
- טעות מסוג ראשון.
 - טעות מסוג שני.
 - טעות מסוג שלישי.
 - אין טעות במסקנתו.

תשובות סופיות:

- א. מעל 6.645 ב. 0.3632
- א. נדחה H_0 אם $\bar{X} < 2.24$ ב. נדחה H_0 ג. 1
- א. נדחה H_0 אם $\bar{X} > 203.29$ או $\bar{X} < 196.71$ ב. 0.8051 ג. תקטן
- א. נדחה H_0 אם $\bar{X} < 48.9$ ב. 0.0885 ג. תקטן ד. תקטן
- ד
- ג
- ג
- א
- ב

קביעת גודל מדגם כששונות האוכלוסייה ידועה – רקע

השערות המחקר הן: $H_0: \mu = \mu_0$ $H_1: \mu = \mu_1$

סטיית התקן של האוכלוסייה ידועה σ ומעוניינים לבצע מחקר שרמת המובהקות לא תעלה על α והסיכוי לטעות מסוג שני לא יעלה על β .

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

הנוסחה הבאה נותנת את גודל המדגם הרצוי:

דוגמה (פתרון בהקלטה):

משרד החינוך מפעיל בגן חובה שיטת חינוך שפותחה בשנת 1995. לפי שיטת חינוך זו תוחלת הציון במבחן אוצר מילים לגיל הרך הוא 70. אנשי חינוך החליטו לבדוק שיטת חינוך שפותחה בהולנד הנותנת שם תוחלת ציון אוצר מילים של 80. נניח שציוני מבחן זה מתפלגים נורמאלית עם $\sigma = 17$. כדי לבדוק האם גם בישראל הפעלת שיטת החינוך ההולנדית תעבוד בגנים, רוצים לבנות מחקר ברמת מובהקות של 5%. כמו כן, מעוניינים שאם בהפעלת השיטה ההולנדית תוחלת הציונים תעלה לכדי 80, המחקר יגלה זאת בסיכוי של 90%. כמה ילדי גן חובה דרושים למחקר?

שאלות:

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

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

(3) השערות המחקר הן: $H_0: \mu = \mu_0$ $H_1: \mu = \mu_1$.

כמו כן נתון שהמשתנה מתפלג נורמלית עם סטיית התקן ידועה σ מעוניינים לבצע מחקר שרמת המובהקות לא תעלה על α והסיכוי לטעות מסוג שני לא

יעלה על β . הוכיחו שגודל המדגם הרצוי לכך יהיה: $n \geq \left(\frac{(Z_{1-\alpha} + Z_{1-\beta}) \times \sigma}{\mu_0 - \mu_1} \right)^2$.

תשובות סופיות:

- (1) 41
(2) 78
(3) הוכחה

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

דרך נוספת להגיע להכרעות שלא דרך כלל הכרעה, היא דרך חישוב מובהקות התוצאה:

באמצעות תוצאות המדגם מחשבים את מובהקות התוצאה שמסומן ב- p_v . את רמת המובהקות החוקר קובע מראש לעומת זאת, את מובהקות התוצאה החוקר יוכל לחשב רק אחרי שיהיו לו את התוצאות.

המסקנה של המחקר תקבע לפי העיקרון הבא: אם $p_v \leq \alpha$, דוחים את H_0 . מובהקות התוצאה זה הסיכוי לקבלת תוצאות המדגם וקיצוני מתוצאות אלה בהנחת השערת האפס.

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

אם ההשערה היא דו צדדית:

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

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: תנאים:
			σ ידועה $X \sim N$ 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

$$Z_{\bar{x}} = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \quad \bar{X} \sim N(\mu_0, \frac{\sigma^2}{n}) \text{ כאשר בהנחת השערת האפס:}$$

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 נורמאלי 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%?

שאלות:

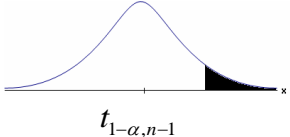
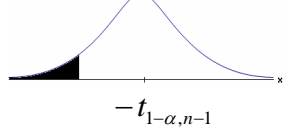
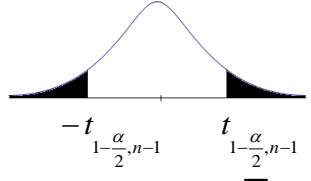
- (1) להלן השערות של מחקר: $H_0: \mu = 70$ $H_1: \mu > 70$. המשתנה הנחקר מתפלג נורמלית עם סטיית תקן 20. במדגם מאותה אוכלוסייה התקבלו התוצאות הבאות: $\bar{x} = 74$ $n = 100$. מהי מובהקות התוצאה?
- (2) השכר הממוצע במשק בשנת 2012 היה ₪8800 עם סטיית תקן 2000. במדגם שנעשה אתמול על 100 עובדים התקבל שכר ממוצע ₪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 נוסיף עוד תצפית שערכה יהיה 82 ס"מ, כיצד הדבר ישפיע על התשובה של הסעיף הקודם?
 ג. הכרע ברמת מובהקות של 5% האם המכונה מכוילת.
 ד. אם מקבלים בחישובים אלפא מינימלית (p-value) קטנה מאוד, סביר להניח כי החוקר ידחה את השערת האפס בקלות. נכון? לא נכון? נמקו.

- (5) בבדיקת השערות התקבל שה- $p\text{-value} = 0.02$.
מה תהיה מסקנת חוקר המשתמש ברמת מובהקות 1%?
בחרו 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.
- (6) מובהקות התוצאה היא גם (בחרו בתשובה הנכונה):
א. רמת המובהקות המינימאלית לדחות השערת האפס.
ב. רמת המובהקות המקסימאלית לדחיית השערת האפס.
ג. רמת המובהקות שנקבעת מראש על ידי החוקר.
ד. רמת המובהקות המינימאלית לאי דחיית השערת האפס.
- (7) בבדיקת השערות מסוימת התקבל $p\text{ value} = 0.0254$ לכן (בחרו 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 .

תשובות סופיות:

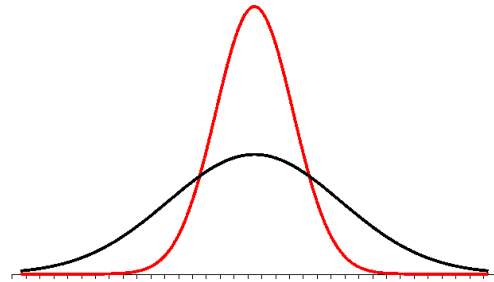
- (8) 0.0228
- (9) עבור כל רמת מובהקות סבירה
- (10) א. 0.1056 ב. 0.1056 ג. 0.1056 ד. נכריע שיש עמידה בהתחייבות של החברה
- (11) א. 0.0006 ב. יקטן ג. נכריע שאין כיול
- (12) נכון
- (13) א'
- (14) א'
- (15) ג'

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: תנאים:
σ אינה ידועה $X \sim N$ 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 ■	או $t_{\bar{x}} > t_{1-\frac{\alpha}{2}}^{(n-1)}$ 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
$\bar{X} > \mu_0 + t_{1-\alpha}^{n-1} \cdot \frac{S}{\sqrt{n}}$	$\bar{X} > \mu_0 - t_{1-\alpha}^{n-1} \cdot \frac{S}{\sqrt{n}}$	$\bar{X} > \mu_0 + t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S}{\sqrt{n}}$ או $\bar{X} < \mu_0 - t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S}{\sqrt{n}}$	חלופה לכלל הכרעה: נדחה H_0 אם מתקיים:

$$t_{\bar{x}} = \frac{\bar{x} - \mu_0}{S / \sqrt{n}}, \quad S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} = \frac{\sum_{i=1}^n X_i^2 - n\bar{X}^2}{n-1} : \text{סטטיסטי המבחן}$$

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):

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. בדקו ברמת מובהקות של 5%.

שאלות:

(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: 125, 80, 100, 95, 90. 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:

$$n = 20 \quad \bar{x} = 3120 \quad S = 280$$
 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:

$$\sum_{i=1}^{10} X_i = 750, \quad \sum_{i=1}^{10} (X_i - \bar{X})^2 = 900$$

נתון שההתפלגות היא נורמלית. בדקו 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? בחרו 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: $H_0: \mu = \mu_0$, $H_0: \mu < \mu_0$.
 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. בחרו בתשובה הנכונה:
 א. כעת בבירור הוא ידחה את השערת האפס.
 ב. כעת הוא דווקא יקבל את השערת האפס.
 ג. כעת לא ניתן לדעת מה תהיה מסקנתו.

תשובות סופיות:

- (1) נדחה H_0 .
- (2) נדחה H_0 .
- (3) נקבל H_0 .
- (4) נקבל H_0 .
- (5) ב'
- (6) ג'

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

נזכיר שהמסקנה של המחקר תיקבע לפי העיקרון הבא: אם $p_v \leq \alpha$ דוחים את H_0 . מובהקות התוצאה היא 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:

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

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: תנאים:
- σ אינה ידועה - $X \sim N$ או מדגם מספיק גדול			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: 34, 40, 30, 32, 27. הניחו שזמן הנסיעה מתפלג נורמלית.

- א. רשמו את השערות המחקר.
- ב. מצאו חסמים למובהקות התוצאה.
- ג. מה המסקנה ברמת מובהקות של 5%?

שאלות:

(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: 1024, 1008, 1005, 996, 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?

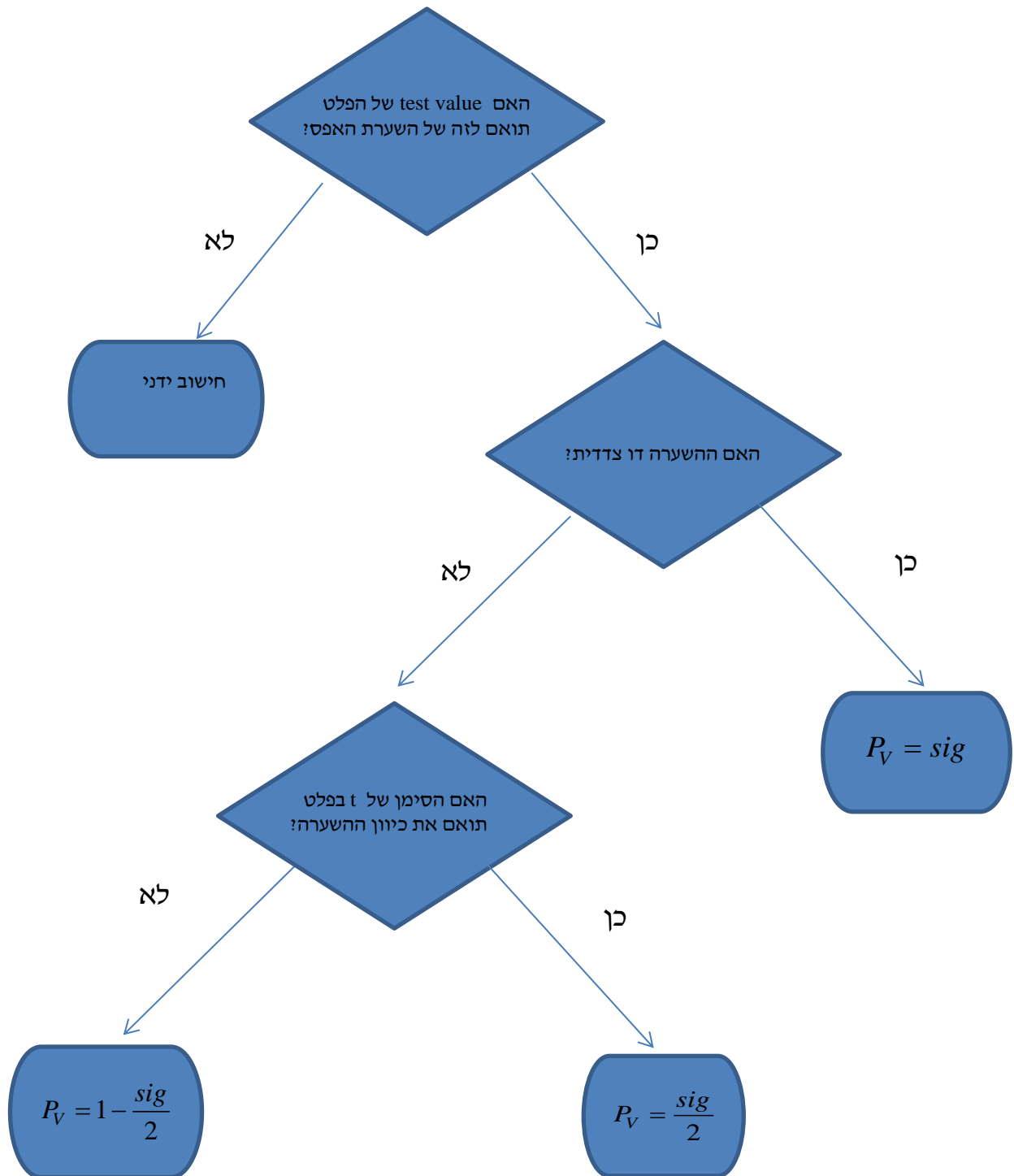
(3) הגובה של מתגייסים לצה"ל מתפלג נורמלית. במדגם של 25 מתגייסים מדדו את הגבהים שלהם בס"מ והתקבלו התוצאות הבאות: $\sum (x_i - \bar{x})^2 = 2832$, $\bar{x} = 176.2$. מטרת המחקר היא לבדוק האם תוחלת הגבהים של המתגייסים גבוה מ-174 ס"מ באופן מובהק. מהי בקרוב מובהקות התוצאה ועל פיה מה תהיה המסקנה ברמת מובהקות של 6%?

תשובות סופיות:

(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		
X	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
X	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?

- 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) להלן 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. השלימו 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, 20, 24, 22 מ"ג. הניחו 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 היינו מוסיפים עוד תצפית שערכה 20. כיצד הדבר היה משפיע על הערך Sig ועל המסקנה?
- ג. בדקו האם ניתן להגיד שתוחלת רמת הניקוטין שונה מ-26 ברמת מובהקות של 5%.

תשובות סופיות:

- (1) ב. 23 ג. (22.104, 24.771) ד. נקבל H_0 .
- (2) א. 25 ב. (4.3056, 4.5104) ד. נדחה H_0 .
- (3) א. $n = 44$, $\bar{X} = 67.66$, $t = -2.48$, $upper = -1.3735$.
ב. 0.017 ג. 0.0085 ד. 0.9915
- (4) א. נכריע שיש אמת בפרסום.
ב. מסקנה לא תשתנה.
ג. נכריע שהתוחלת שונה מ-26.

הקשר בין רווח סמך לבדיקת השערות על תוחלת – רקע

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

$$H_0: \mu = \mu_0, \quad H_1: \mu \neq \mu_0$$

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 :

$$H_0: \mu = 80, \quad H_1: \mu \neq 80, \quad \alpha = 5\%$$

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

האם אפשר לדעת מה מסקנתו, ואם כן מהי?

שאלות:

- (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? הסבירו.
- (3) יצרן אנטיביוטיקה רושם על גבי התרופות שכמות הפנצלין היא 200 מ"ג לקפסולה. משרד הבריאות ביצע מדגם של 8 קפסולות אקראיות מקו הייצור ומצא שבממוצע יש 196 מ"ג פנצלין לקפסולה עם סטיית תקן מדגמית של של 5 מ"ג. בהנחה וכמות הפנצלין בקפסולה מתפלגת נורמלית.
א. בנו רווח סמך ברמת סמך של 95% לממוצע כמות הפנצלין לקפסולה המיוצרת על ידי יצרן האנטיביוטיקה.
ב. בדקו ברמת מובהקות של 5% האם יש אמת באינפורמציה המסופקת על ידי היצרן.

תשובות סופיות:

- (1) נקבל השערת.
- (2) א. $112.87 \leq \mu \leq 118.13$
ב. נכריע שהדיאטה משפיעה על תוחלת רמת הסוכר בדם.
- (3) א. $191.8 \leq \mu \leq 200.2$ ב. נכריע שיש אמת בפרסום.

מבוא לסטטיסטיקה ב 55003

פרק 7

רווח סמך להפרש תוחלות ~ סוגריים 2 ~ ממוצעים ~ סוגריים 1 ~ במדגמים בלתי תלויים

כששוננויות האוכלוסיה לא ידועות ובהנחת שוויון שוננויות.

1.....

כששונויות האוכלוסייה לא ידועות ובהנחת שוויון שונויות:

רקע:

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

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

התנאים לבניית רווח הסמך:

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

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

3. מדגמים בלתי תלויים.

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

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

הנוסחה הבאה:

$$d.f = n_1 + n_2 - 2$$

דרגות החופש:

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

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:

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

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

שאלות:

- (1) נדגמו 15 ישראלים ו-15 אמריקאים. כל הנדגמים נגשו למבחן IQ. להלן תוצאות המדגם:

מדינה	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.

תשובות סופיות:

- (1) הנחות:
- השונות שווה.
 - שהציונים מתפלגים נורמלית.
 - המדגמים אינם תלויים זה בזה.
- $$-5.52 < \mu_1 - \mu_2 < 17.52$$
- (2)
- $$-9.6 < \mu_y - \mu_x < 1.6$$


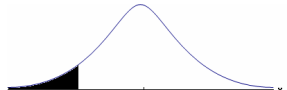

תוכן העניינים:

2.....	בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים
2.....	כשהשונויות של האוכלוסייה ידועות – רקע
4.....	תרגילים:
5.....	תשובות סופיות:
6.....	כשהשונויות של האוכלוסייה לא ידועות והמדגמים גדולים – רקע
8.....	שאלות:
8.....	תשובות סופיות:
9.....	כשהשונויות האוכלוסייה לא ידועות ומניחים שהן שוות – רקע
11.....	שאלות:
12.....	תשובות סופיות:
13.....	ניתוח פלטים – רקע
15.....	שאלות:
19.....	תשובות סופיות:

לתשומת לבך, יש ללמוד לפי הסרטונים באתר. ייתכנו שאלות בספר הפרק אשר אינן חלק מחומר הלימוד שלך.

בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

כשהשונויות של האוכלוסייה ידועות – 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:
מדגמים בלתי תלויים σ_1, σ_2 ידועות $X_1, X_2 \sim N$ או מדגמים מספיק גדולים			תנאים:
$Z_{\bar{x}_1 - \bar{x}_2} > Z_{1-\alpha}$  $Z_{1-\alpha}$ דוחים את H_0	$Z_{\bar{x}_1 - \bar{x}_2} < -Z_{1-\alpha}$  $-Z_{1-\alpha}$ דוחים את H_0	או $Z_{\bar{x}_1 - \bar{x}_2} < -Z_{1-\frac{\alpha}{2}}$ $Z_{\bar{x}_1 - \bar{x}_2} > Z_{1-\frac{\alpha}{2}}$  $-Z_{1-\frac{\alpha}{2}} \quad Z_{1-\frac{\alpha}{2}}$ דוחים את H_0	כלל ההכרעה: אזור הדחייה של H_0 $-Z_{1-\frac{\alpha}{2}} \quad Z_{1-\frac{\alpha}{2}}$

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

סטטיסטי המבחן:

חלופה אחרת לכלל הכרעה:

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

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

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

דוגמה (פתרון בהקלטה):

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

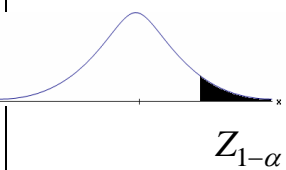
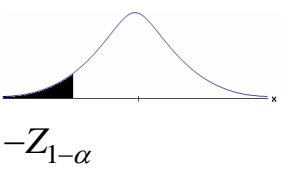
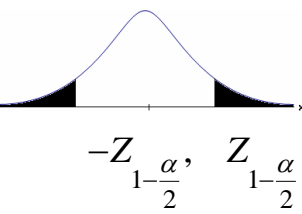
תרגילים:

- (1) מחקר טוען שאנשים החיים במרכז הארץ צופים בממוצע בטלוויזיה יותר מאנשים שלא חיים במרכז. נדגמו 100 אנשים מהמרכז ו-107 אנשים לא מהמרכז. אנשים אלה נשאלו כמה שעות ביום הם נוהגים לצפות בטלוויזיה. במדגם של מרכז הארץ התקבל ממוצע 2.7 שעות. במדגם של מחוץ למרכז הארץ התקבל ממוצע 1.8 שעות. לצורך פתרון הניחו שבכל אזור, סטיית התקן היא שעה 1 ביום. בדקו את טענת המחקר ברמת מובהקות של 1%.
- (2) ציוני פסיכומטרי מתפלגים נורמלית עם סטיית תקן 100. מכון ללימוד פסיכומטרי טוען שהוא יכול לשפר את ממוצע הציונים ביותר מ-30 נקודות. במדגם של 20 נבחנים שניגשו למבחן ללא הכנה במכון התקבל ממוצע 508. במדגם של 25 נבחנים שעברו הכנה במכון התקבל ממוצע ציונים 561. מה מסקנתכם ברמת מובהקות של 5%.
- (3) במדגם אקראי של 20 ימים נבדקה התפוקה של מפעל ביום. התפוקה הממוצעת הייתה של 340 מוצרים ליום. במדגם אקראי של 20 ימים אחרים נבדקה התפוקה של המפעל בלילה והתפוקה הממוצעת הייתה 295. לצורך פתרון נניח שסטיית התקן של התפוקה ביום היא 40 מוצרים ובלילה 30 מוצרים.
א. מהי מובהקות התוצאה לבדיקה האם התפוקה הממוצעת היומית גבוהה מהתפוקה הממוצעת הלילית.
ב. מה תהיה המסקנה ברמת מובהקות של 8%?
- (4) במחקר מקיף שנעשה באירופה נקבע שגברים גבוהים מנשים ב-8 ס"מ בממוצע. מחקר ישראלי מתעניין לבדוק האם בישראל הפער גדול יותר. לצורך המחקר נדגמו 40 גברים ו 40 נשים באקראי. כמו כן, נניח שסטיות התקן של הגברים והנשים ידועות ושוות ל-6 ס"מ אצל הנשים ו-12 ס"מ אצל הגברים.
א. מהן השערות המחקר ומהו כלל ההכרעה ברמת מובהקות של 10%?
ב. אם בישראל הפער בין גברים לנשים מבחינת הגובה הממוצע הוא 11 ס"מ, מה ההסתברות שהמחקר לא יגלה זאת? איך קוראים להסתברות הזאת?

תשובות סופיות:

- (5) נדחה H_0 .
- (6) לא נדחה את H_0 .
- (7) א. 0 ב. נדחה את H_0 .
- (8) א. נדחה את H_0 , אם במדגם הגברים יהיו גבוהים בממוצע מהנשים ביותר מ-10.72 ס"מ.
ב. 0.6331

כשהשונויות של האוכלוסייה לא ידועות והמדגמים גדולים – רקע

$H_0 \mu_1 - \mu_2 = c$ $H_1 \mu_1 - \mu_2 > c$	$H_0 \mu_1 - \mu_2 = c$ $H_1 \mu_1 - \mu_2 < c$	$H_0 \mu_1 - \mu_2 = c$ $H_1 \mu_1 - \mu_2 \neq c$	<p>השערת האפס: השערה אלטרנטיבית:</p>
<p>1. מדגמים בלתי תלויים 2. σ_1, σ_2 לא ידועות 3. מדגמים מספיק גדולים</p>			<p>תנאים:</p>
$Z_{\bar{x}_1 - \bar{x}_2} > Z_{1-\alpha}$  $Z_{1-\alpha}$ דוחים את H_0	$Z_{\bar{x}_1 - \bar{x}_2} < -Z_{1-\alpha}$  $-Z_{1-\alpha}$ דוחים את H_0	<p>או $Z_{\bar{x}_1 - \bar{x}_2} < -Z_{1-\frac{\alpha}{2}}$ $Z_{\bar{x}_1 - \bar{x}_2} > Z_{1-\frac{\alpha}{2}}$</p>  $-Z_{1-\frac{\alpha}{2}}, Z_{1-\frac{\alpha}{2}}$ דוחים את H_0	<p>כלל ההכרעה: אזור הדחייה של H_0:</p>

סטטיסטי המבחן:
$$Z_{\bar{x}_1 - \bar{x}_2} = \frac{\bar{x}_1 - \bar{x}_2 - c}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

חלופה אחרת לכלל הכרעה:

אם מתקיים: H_0 נדחה	
$\bar{x}_1 - \bar{x}_2 > c + Z_{1-\alpha} \cdot \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$	$\bar{x}_1 - \bar{x}_2 > c + Z_{1-\alpha/2} \cdot \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$ <p style="text-align: center;">או</p> $\bar{x}_1 - \bar{x}_2 < c - Z_{1-\alpha/2} \cdot \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$
$\bar{x}_1 - \bar{x}_2 < c - Z_{1-\alpha} \cdot \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$	

דוגמה (פתרון בהקלטה) :

נרצה לבדוק האם קיים הבדל בין ממוצע ציוני הפסיכומטרי של חיילים לממוצע ציוני הפסיכומטרי של תלמידי תיכון. במדגם של 46 נבחנים חיילים התקבל ממוצע 543 וסטיית תקן 123. במדגם של 50 תלמידי תיכון התקבל ממוצע 508 וסטיית תקן 178. מה המסקנה ברמת מובהקות 5%?

שאלות:

(1) חברה להנדסת בניין מעוניינת להשוות ברמת הקשיות של שני סוגי ברגים. במדגם של 35 ברגים מסוג א' התקבל רמת קשיות ממוצעת של 28 יחידות וסטיית תקן 4, ובמדגם של 45 ברגים מסוג ב' התקבל רמת קשיות ממוצעת של 25 וסטיית תקן 6. האם על סמך תוצאות המדגם יש הבדל בין סוגי הברגים מבחינת רמת הקשיות שלהם? בדקו ברמת מובהקות של 5%.

(2) כדי לבדוק האם נהגים השותים תחת השפעת אלכוהול נוהגים מהר יותר מאלו שאינם שותים בוצע מדגם שבו בדקו את המהירות המקסימאלית של כל נהג בקמ"ש. להלן התוצאות:


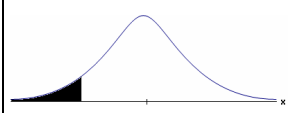
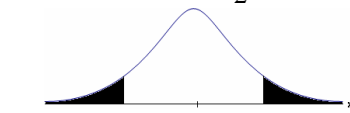
S	\bar{X}	גדול מדגם	
20	80	70	נהגים השותים אלכוהול
15	60	100	נהגים שאינם שותים אלכוהול

- א. מהי מובהקות התוצאה?
 ב. מה המסקנה ברמת מובהקות של 5%?

תשובות סופיות:

- (3) נדחה את H_0 .
 (4) א. 0. ב. נדחה את H_0 .

כששונויות האוכלוסייה לא ידועות ומניחים שהן שוות – רקע

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

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

סטטיסטי המבחן:

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

השונויות המשוקללת:

חלופה אחרת לכלל הכרעה:

אם מתקיים: H_0 נדחה	
$\bar{x}_1 - \bar{x}_2 < c - t_{1-\alpha}^{(n_1+n_2-2)} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$	$\bar{x}_1 - \bar{x}_2 > c + t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$ <p style="text-align: center;">או</p> $\bar{x}_1 - \bar{x}_2 < c - t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$
$\bar{x}_1 - \bar{x}_2 > c + t_{1-\alpha}^{(n_1+n_2-2)} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$	

דוגמה (פתרון בהקלטה):

חברה המייצרת מוצרי בנייה טוענת שפיתחה סגסוגת (תערובת מתכות) שטמפרטורת ההתכה שלה גבוהה משמעותית מטמפרטורת ההתכה של הסגסוגת לבנייה שמתמשים 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%.

שאלות:

(1) להלן נתונים של שטחי דירות מתוך דירות שנבנו בשנת 2012 ובשנת 2013 (במ"ר):

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:

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

בדקו ברמת מובהקות של 5% האם קיים הבדל של נקודה בין ישראלים לאמריקאים מבחינת ממוצע הציונים במבחן ה-IQ לטובת ישראל. רשמו את כל ההנחות הדרושות לצורך פתרון התרגיל.

(3) להלן תוצאות מדגם הבדק אורך חיים של נורות מסוג W60 ומסוג W100. אורך החיים נמדד בשעות.

100W	60W	הקבוצה
956	1007	\bar{x}
72	80	S
15	13	n

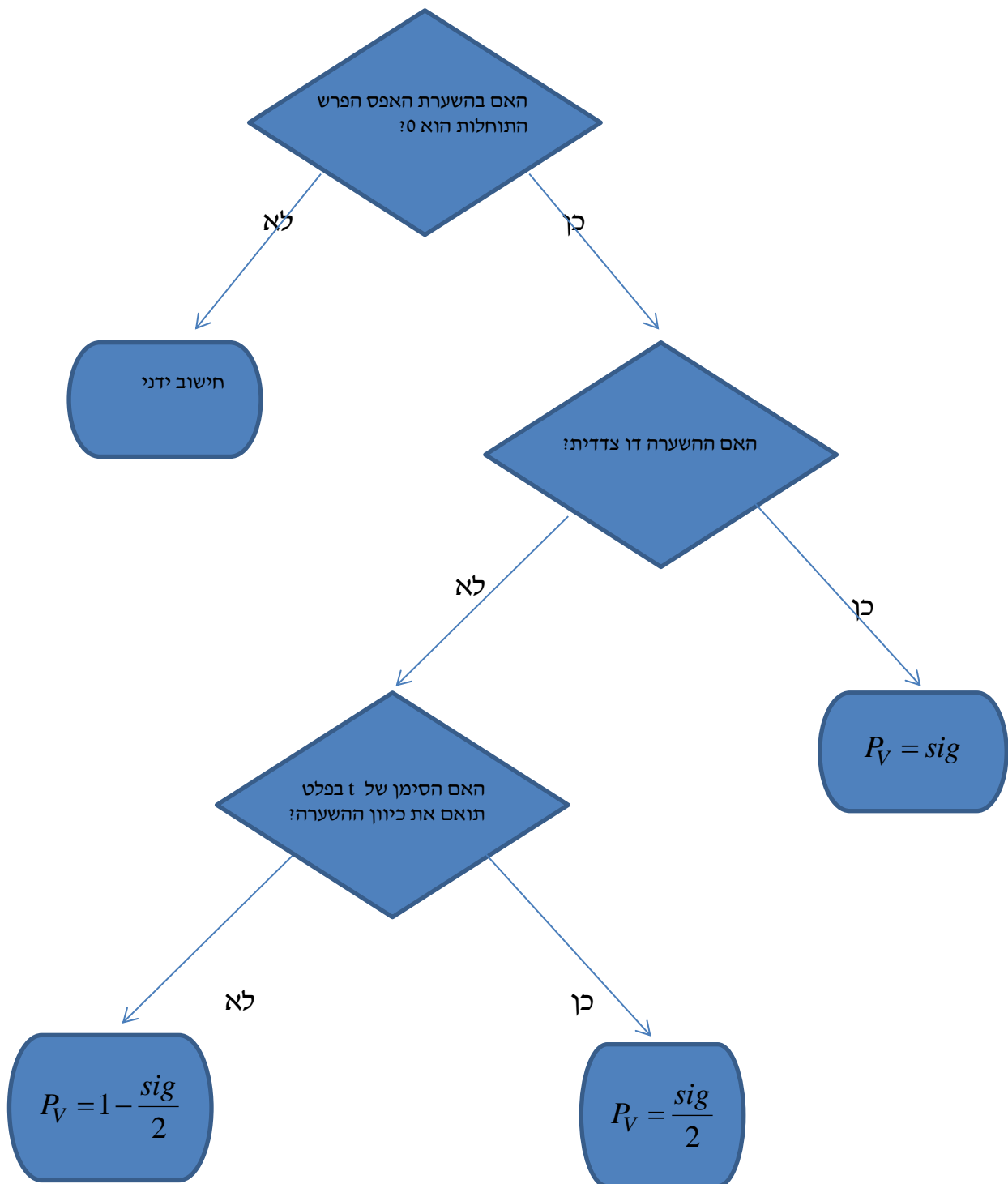
- א. בדקו ברמת מובהקות של 5% האם נורות מסוג W60 דולקות בממוצע יותר מאשר נורות מסוג W100. רשמו את כל ההנחות הדרושות לפתרון.
- ב. עבור איזו רמת מובהקות ניתן לקבוע שנורות מסוג W60 דולקות בממוצע יותר מאשר נורות מסוג 100?
- ג. בדקו ברמת מובהקות של 5% האם נורות מסוג W60 דולקות יותר מ 1000 שעות. רשמו את כל ההנחות הדרושות.

תשובות סופיות:

- (1) לא נדחה את H_0 .
- (2) א. נדחה את H_0 .
ב. רמת מובהקות של לפחות 5%.
- (3) ג. לא נדחה את 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?

1) להלן פלט מתוכנת SPSS מתוך מחקר שבחן את רמת האופטימיות של גברים ונשים. רמת האופטימיות נמדדה בסולם 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:

Group Statistics

		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

T-Test

מקרא:

A = נגשו פעם אחת.

B = נגשו יותר מפעם אחת.

א. רשמו את השערות המחקר והסבירו מהו המבחן המתאים כאן.

ב. כיצד הייתה משתנה התשובה לסעיף הקודם אם היה מדובר על אותם אנשים

שציונם נבדק פעם אחרי המבחן הראשון שעשו ופעם אחרי המבחן השני?

ג. האם ניתן לומר כי מידת הפיזור של ציוני אנשים הנבחנים בפעם הראשונה

שונה ממידת הפיזור של ציוני האנשים אשר נבחנים בפעם השנייה. בדוק

ברמת מובהקות של $\alpha = 0.05$.

ד. האם נכונה טענת הפסיכולוגים ברמת מובהקות של $\alpha = 0.01$.

3) כחלק ממחקר בנושא הנישואין בישראל, אחד החוקרים העלה השערה שיש הבדל 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. השלימו 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?

תשובות סופיות:

- (1) א. נכריע שוויון שוניות.
 ב. נקבע שלא קיים הבדל בין נשים לגברים מבחינת האופטימיות הממוצעת.
 ג. 0.11076
 ד. $2.5665 \leq \mu \leq 2.6441$
- (2) א. מבחן T למדגמים בלתי תלויים.
 ב. מבחן T למדגמים מזווגים.
 ג. נקבע לקיים שוויון שוניות.
 ד. לא נקבל את טענת הפסיכולוגים.
- (3) א. מבחן T להשוואת תוחלת במדגמים בלתי תלויים.
 ב. $1.19497 \leq \mu_1 - \mu_2 \leq 2.39332$ ג. כן.
- (4) א.-48, 4.69, 10.15 ג. 0.978
 ב. 0.04
 ד. 0.022

מבוא לסטטיסטיקה ב 55003

פרק 9

רווח סמך לתוחלת ~ סוגריים 2 ~ ממוצע ~ סוגריים 1 ~ הפרשים במדגמים מזווגים

רווח סמך לתוחלת (ממוצע) הפרשים במדגמים מזווגים

1.....

רווח סמך לתוחלת (ממוצע) הפרשים במדגמים מזווגים:

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:

1. $x, y \sim N$.

2. המדגם מזווג.

נוסחת רווח הסמך: $\bar{D} \pm t_{\frac{1-\alpha}{2}}^{n-1} \frac{S_D}{\sqrt{n}}$.

כאשר דרגות החופש: $df = n - 1$.

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.

שאלות:

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

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

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

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

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

תשובות סופיות:

- (1) א. $-19 < \mu_0 < 38$. ב. בביטחון של 95% לא קיים הבדל. ג. ראה הסבר בסרטון.
- (2) $-0.013 < \mu < 0.185$.

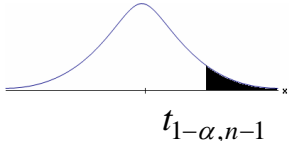
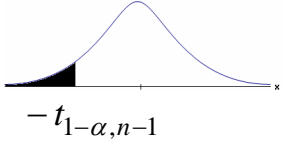
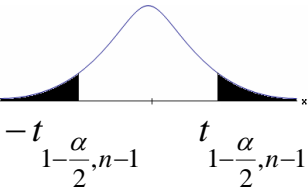
תוכן העניינים:

- 1 בדיקת השערות על תוחלת ההפרשים במדגמים מזווגים (תלויים)
- 2 בדיקת השערות למדגמים מזווגים – רקע
- 4 שאלות:
- 5 תשובות סופיות:
- 6 מדגמים מזווגים – ניתוח פלטים – רקע
- 8 שאלות:
- 11 תשובות סופיות:

לתשומת לבך, יש ללמוד לפי הסרטונים באתר. ייתכנו שאלות בספר הפרק אשר אינן חלק מחומר הלימוד שלך.

בדיקת השערות על תוחלת ההפרשים במדגמים מזווגים (תלויים)

בדיקת השערות למדגמים מזווגים – רקע

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

$$S_D^2 = \frac{\sum_{i=1}^n (D_i - \bar{D})^2}{n-1} = \frac{\sum_{i=1}^n D_i^2 - n\bar{D}^2}{n-1}, \quad t_{\bar{D}} = \frac{\bar{D} - \mu_D}{S_D / \sqrt{n}}$$

סטטיסטי המבחן:

דוגמה (פתרון בהקלטה):

חברה שיווקית מעוניינת לבדוק את טענת רשת השיווק "מגה בעיר" הטוענת שמחיריה נמוכים מהמחירים מרשת השיווק "שופרסל". לצורך הבדיקה נבחרו באקראי 4 מוצרים שונים. המחירים נבדקו בשתי הרשתות. להלן המחירים:

המוצר / רשת	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".

שאלות:

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

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

- בהנחה והמחירים מתפלגים נורמלית בכל חברה, בדקו ברמת מובהקות של 5% האם קיים הבדל בין החברות מבחינת המחירים במוצע?
 (2) מכון המכין לפסיכומטרי טוען שהוא מעלה את ממוצע הציונים ביותר מ-30 נקודות. 8 נבחנים נבדקו לפני ואחרי שהם למדו במכון. להלן התוצאות שהתקבלו:

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

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

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

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

- פורסם שתלמידים שמסיימים את סמסטר ב משפרים במוצע את הציונים ב-5 נקודות לעומת סמסטר א'. הניחו 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 נורמלית ללא ידיעת השונות האמתית. המבחן שיש לבצע כאן הוא:
- מבחן Z למדגם יחיד.
 - מבחן T למדגם יחיד.
 - מבחן T למדגמים בלתי תלויים.
 - מבחן T למדגמים מזווגים.

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

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

נניח שהמשקלים מתפלגים נורמלית, המבחן שיש לבצע כאן הוא:

- א. מבחן Z למדגם יחיד.
- ב. מבחן T למדגם יחיד.
- ג. מבחן T למדגמים בלתי תלויים.
- ד. מבחן T למדגמים מזווגים.

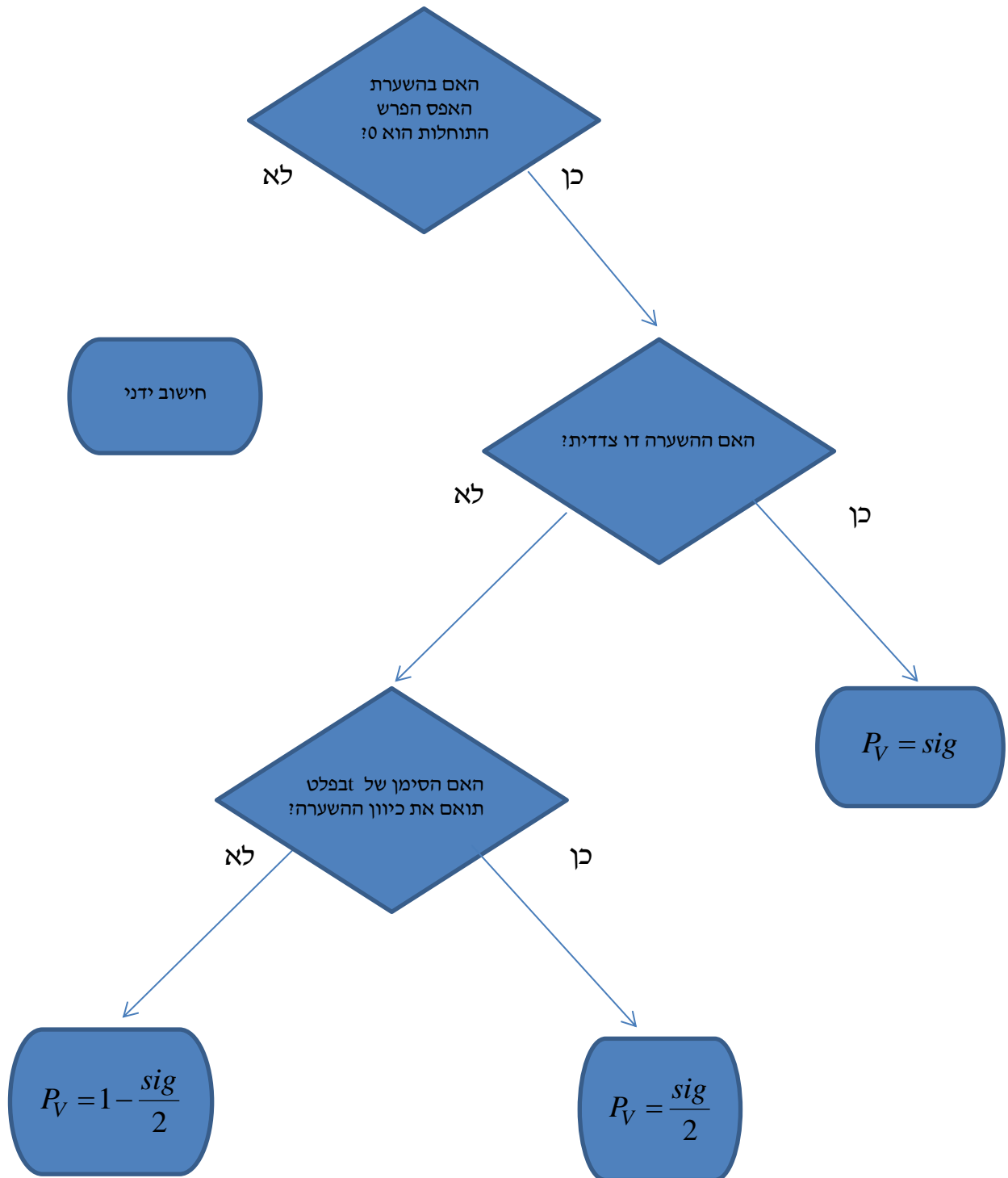
6) כדי להשוות בין שני אצנים נדגמו 5 תוצאות מריצת 100 מטר של כל אצן. זמני הריצה נרשמו ויש להניח שמתפלגים נורמלית. המטרה להשוות בין האצנים. המבחן שיש לבצע כאן הוא:

- א. מבחן Z למדגם יחיד.
- ב. מבחן T למדגם יחיד.
- ג. מבחן T למדגמים בלתי תלויים.
- ד. מבחן T למדגמים מזווגים.

תשובות סופיות:

- (1) לא נדחה H_0 .
- (2) לא נדחה H_0 .
- (3) א. $0.25 \leq p \leq 0.5$ ב. 0.5 ג. לא נדחה H_0 .
- (4) ד'
- (5) ד'
- (6) 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 :

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

Paired Samples Statistics

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

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	90% Confidence Interval of the Difference				
				Lower	Upper			
Pair 1 BEFORE - AFTER	11.20000	8.19756	3.66606	3.38452	19.01548	3.055	4	.038

בדקו ברמת מובהקות של 5% האם הקורס יעיל.

שאלות:

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

- תנו אומדן להפרש הממוצעים.
- תנו אומדן לטעות התקן של הפרש הממוצעים.
- 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 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	???	???

- השלימו את החלקים החסרים בפלט (מסומנים בסימני שאלה).
- בדקו את הטענה שהשופט הדרום קוריאני נותן בממוצע 0.2 נקודות יותר מאשר השופט האיטלקי ברמת מובהקות של 5%.
- מהו רווח הסמך ברמת סמך של 95% לתוחלת פער הציונים בין השופטים?
- בנו 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. רשמו מהו המבחן הסטטיסטי ונסח את ההשערות אליהם מתייחס הפלט.
- ב. בדקו את ההשערה שממוצע ציונים משתפרים לאחר ההכנה ברמת מובהקות של 5%.
- ג. בדקו את ההשערה שממוצע ציונים משתפרים לאחר ההכנה ביותר מ-5 נקודות ברמת מובהקות של 5%.
- ד. מצאו רווח סמך לתוחלת שיפור ממוצע הציונים לאחר ההכנה ברמת ביטחון של 95%.

4) להלן פלט של תכנת SPSS :

T-Test

Paired Samples Statistics

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

Paired Samples Test

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

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

תשובות סופיות:

- (1) א. -0.007 ב. 0.1 ג. אין הבדל מובהק.
- (2) א. $d.f = 299$, ב. $n = 300$, ג. $\bar{X} = 8.496$, ד. $Sig = 0$
- ב. נדחה את H_0 .
- ג. (0.46341, 0.38126)
- ד. (0.4708, 0.3738)
- (3) א. נדחה את H_0 . ב. לא נדחה את H_0 . ג. (1.592, 13.317) . ד.
- (4) א. 1.5876 , 11.5715
- ב. 0.173 ג. יגדל ד. 0.0865
- ה. 0.9135 ו. $49.18 < \mu < 58.82$

תוכן העניינים:

- 2 הקשר בין רווח סמך לבדיקת השערות על הפרש תוחלות
- 2.....:רקע
- 3.....:שאלות
- 4.....:תשובות סופיות :

לתשומת לבך, יש ללמוד לפי הסרטונים באתר. ייתכנו שאלות בספר הפרק אשר אינן חלק מחומר הלימוד שלך.

הקשר בין רווח סמך לבדיקת השערות על הפרש תוחלות

רקע:

ניתן לבצע בדיקת השערות זו צדדית ברמת מובהקות α על $\mu_1 - \mu_2$:

$$. H_0 : \mu_1 - \mu_2 = C, \quad H_1 : \mu_1 - \mu_2 \neq C$$

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: $\alpha = 5\%$, $H_0 : \mu_D = 80$, $H_1 : \mu_D \neq 80$.

החוקר בנה רווח סמך ברמה של 90%, $78 < \mu_D < 83$.

האם אפשר לדעת מה מסקנתו, ואם כן מהי?

שאלות:

(1) נדגמו 5 סטודנטים שסיימו את הקורס סטטיסטיקה ב'.

להלן ציוניהם בסמסטר א' ו- ב' :

סמסטר א'	סמסטר 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 :

מרצה X	82	90	68
מרצה Y	68	81	64

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 רב-ברירה:

(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 :

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

א. לדחות את השערת האפס.

ב. לא לדחות את השערת האפס.

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

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

- (4) במטרה לבדוק האם קיים הבדל בין קווי זהב לבזק מבחינת ממוצע המחירים לשיחות בינ"ל. נגדמו באקראי 7 מדינות ועבור כל מדינה נבדקה עלות דקת שיחה. בהנחה והמחירים מתפלים נורמלית בנו רווח סמך לממוצע ההפרשים וקיבלו : $-0.0293 < \mu_D < 0.2145$, רווח הסמך הוא ברמת סמך של 95%.
לכן מסקנת המחקר היא :
- 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 .

תשובות סופיות:

- (1) א. $-3.8 \leq \mu_D \leq 19$ ב. נכריע שיש אמת בפרסום.
- (2) א. $-8.5 \leq \mu_X - \mu_Y \leq 26.5$ ב. נכריע שאין הבדל.
- (3) ג'
- (4) א'

תוכן העניינים:

2	מבחן חי בריבוע
2	מבחן טיב התאמה – רקע
5	שאלות:
6	תשובות סופיות:
7	הקשר בין מבחן טיב התאמה לבדיקת השערות על הפרופורציה – רקע
8	שאלות:
8	תשובות סופיות:
9	מבחן חי בריבוע לאי תלות בין משתנים – רקע
12	שאלות:
13	תשובות סופיות:
14	הקשר בין מבחן לאי תלות ובדיקת השערות להפרש פרופורציות – רקע
14	שאלות:
14	תשובות סופיות:
15	מדדי קשר-מדד הקשר של קרמר – רקע
16	שאלות:
16	תשובות סופיות:
17	פלטים על מבחן לאי תלות – רקע
20	שאלות:
27	תשובות סופיות:

לתשומת לבך, יש ללמוד לפי הסרטונים באתר. ייתכנו שאלות בספר הפרק אשר אינן חלק מחומר הלימוד שלך.

מבחני חי בריבוע

מבחן טיב התאמה – רקע

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

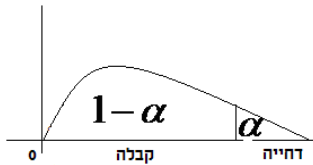
מבנה המבחן:

השערות:

- המשתנה מתפלג לפי התפלגות מסוימת - H_0 .
- אחרת - H_1 .

כלל הכרעה:

הערך הקריטי נקבע על סמך התפלגות חי בריבוע. התפלגות זו היא אסימטרית חיובית ותלויה בדרגות החופש. $d.f = K - 1$, כאשר K - מספר הקטגוריות.



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

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

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

סטטיסטי המבחן:

O_i - השכיחות שנצפתה במדגם בקטגוריה i .

p_i - הסתברות לקטגוריה i לפי השערת האפס.

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

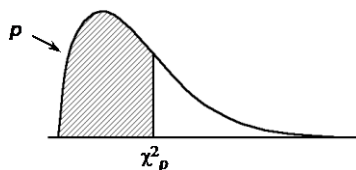
הערה:

תנאי כדי לבצע את המבחן הוא $E_i \geq 5$ לכל i . במידה ותנאי זה לא מתקיים יש אפשרות לאחד קטגוריות סמוכות עד שהתנאי יתקיים.

דוגמה (פתרון בהקלטה):

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

טבלת התפלגות חי-בריבוע – ערכי החלוקה



df	p												
	.005	.01	.025	.05	.10	.25	.50	.75	.90	.95	.975	.99	.995
1	0.04393	0.03157	0.03982	0.02393	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

שאלות:

- (1) במטרה לבדוק האם קובייה הוגנת, מטיילים אותה 120 פעמים. התקבלו 17 פעמים 1, 23 פעמים 2, 20 פעמים 3, 25 פעמים 4, 18 פעמים 5 ו-17 פעמים 6. מה מסקנתכם ברמת מובהקות של 5%?
- (2) מפעל מייצר סוכריות בצבעים כחול, אדום, ירוק וכתום. מעוניינים לבדוק שפרופורציית הסוכריות הכחולות גדולה פי 2 מכל צבע אחר. לצורך כך נדגמו באקראי 200 סוכריות והתקבל: 70 כחולות, 50 אדומות, 40 ירוקות והיתר כתומות. מה מסקנתכם ברמת מובהקות של 5%?
- (3) משרד החינוך טוען שבקרב השכירים במשק היחס בין השכירים בעלי השכלה נמוכה, תיכונית ואקדמאית הוא 1:2:1 בהתאמה. במדגם של 200 שכירים התקבלו 56 אנשים בעלי השכלה נמוכה, 105 בעלי השכלה תיכונית והיתר בעלי השכלה גבוהה.
- א. על סמך תוצאות המדגם, האם התפלגות ההשכלה היא כמו שמשדרד החינוך מפרסם? בדוק ברמת מובהקות של 5%.
- ב. בנו רווח סמך ברמת סמך של 95% לפרופורציית השכירים במשק בעלי השכלה אקדמאית.
- (4) 200 איש נתבקשו לבחור ספרה באקראי והנה התוצאות שהתקבלו: 18 איש בחרו בספרה 0, 24 איש בחרו בספרה 1, 17 איש בחרו בספרה 2, 19 איש בחרו בספרה 3, 20 איש בחרו בספרה 4, 18 איש בחרו בספרה 5, 22 איש בחרו בספרה 6 והיתר בחרו בספרות 7-9.
- א. על סמך התוצאות הללו האם בחירת הספרות אקראית? בדקו ברמת מובהקות של 2.5%.
- ב. תנו הערכה למובהקות התוצאה.
- ג. אם נגדיל את גודל המדגם פי 2 ונשמור על אותם יחסים של כמות האנשים במדגם שבחרו בספרות, כיצד הדבר ישפיע על ערכו של הסטטיסטי χ^2 ? מה תהיה המסקנה במקרה זה?
- (5) מעוניינים לבדוק האם קובייה היא הוגנת. הטיילו את הקובייה פעמיים והתבוננו בסכום הוצאות. חזרו על התהליך 72 פעמים. להלן התוצאות שהתקבלו במדגם:

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

מה המסקנה ברמת מובהקות של 5%?

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

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

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

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

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

א. בהנחה והמטבע הוגן, מהי ההתפלגות של X ?
 ב. בדקו האם המטבע הוגן, על סמך תוצאות המדגם ברמת מובהקות של 5%.

8) להלן השערות מחקר: $H_0: X \sim N(40, 2^2)$, $H_1: else$

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

תוצאות המדגם הן: מהו ערכו המקסימלי של A עבורו נקבל את H_0 ברמת מובהקות של 5%?

תשובות סופיות:

- 1) לא נדחה H_0 .
- 2) לא נדחה H_0 .
- 3) א. לא נדחה H_0 . ב. (0.14, 0.25).
- 4) א. לא נדחה H_0 . ב. בין 0.95 ל-0.975.
- 5) ג. יגדל פי 2; המסקנה לא תשתנה.
- 6) נכריע שהקובייה אינה הוגנת.
- 7) 0.005
- 8) א. $X \sim G(0.5)$. ב. נסיק שהמטבע לא הוגן.

14 (8)

הקשר בין מבחן טיב התאמה לבדיקת השערות על הפרופורציה – רקע

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

דוגמה (פתרון בהקלטה) :

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

א. באמצעות מבחן טיב התאמה.

ב. באמצעות מבחן Z לפרופורציה בודדת.

שאלות:

- (1) בסקר שנעשה על 320 נשאלים, 43.75% טענו שהחיה המועדפת עליהם היא כלב. עד היום היה נהוג לחשוב ש40% מהאנשים מעדיפים כלבים. בדקו ברמת מובהקות של 5% האם הסקר יישנה את הסברה שהייתה נהוגה עד היום לגבי העדפת כלב.
- א. באמצעות מבחן טיב התאמה.
ב. באמצעות מבחן על פרופורציה.

- (2) לסוכנות מכוניות שלושה סניפים ברחבי הארץ. המכוניות נמכרות בסניפים השונים. מתוך 100 מכוניות נמצא ש-65 נמכרו בסניף תל-אביב, 23 בסניף ירושלים והיתר בסניף חיפה.
- א. בדקו ברמת מובהקות של 5% האם שיעור המכוניות שנמכרות בסניף ת"א גדול פי 2 מכל סניף אחר.
- ב. בדקו באמצעות מבחן טיב התאמה האם 60% מהמכוניות נהוגות להימכר בסניף תל אביב. האם יש דרך אחרת לבדוק את ההשערה?

- (3) בתחרות ריצה בית ספרית שלושה מסלולי ריצה. ב-50 תחרויות בדקו באיזה מסלול היה הניצחון.

המסלול	1	2	3
מספר הניצחונות	20	15	15

התוצאות שהתקבלו מסוכמות בטבלה הבאה:

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

ב. בדקו ברמת מובהקות של 5% האם הסיכוי לנצח במסלול מספר 1

$$\text{גבוה מ-} \frac{1}{3}.$$

תשובות סופיות:

- (1) לא נדחה H_0 .
- (2) א. נדחה H_0 .
ב. לא נדחה H_0 .
- (3) א. לא נדחה H_0 .
ב. לא נדחה H_0 .

מבחן חי בריבוע לאי תלות בין משתנים – רקע

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

מבנה המבחן:

השערות:

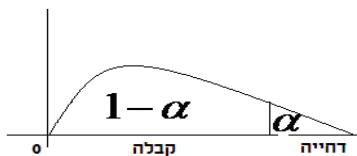
אין תלות בין המשתנים H_0 .

יש תלות בין המשתנים H_1 .

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.



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

כאשר:

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:

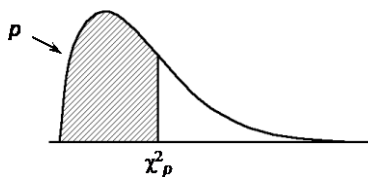
תנאי כדי לבצע את המבחן הוא $E_i \geq 5$ לכל i . במידה ותנאי זה לא מתקיים יש אפשרות לאחד קטגוריות סמוכות עד שהתנאי יתקיים.
תנאי חלופי: אין E קטן מ-1 וגם אין ביותר מ 20% מהתאים E קטן מ-5.

דוגמה (הפתרון בהקלטה):

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

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

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



df	p												
	.005	.01	.025	.05	.10	.25	.50	.75	.90	.95	.975	.99	.995
1	0.004393	0.004457	0.004982	0.005393	0.005988	0.006745	0.007781	0.009003	0.010421	0.012033	0.013858	0.015901	0.018176
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

שאלות:

1) נבדקה התלות בין גודל הארגון לבין שביעות הרצון של העובדים.

להלן התוצאות:

גודל המפעל	שביעות רצון	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:

	לילה	ערב	יום
פגומים	70	60	50
תקינים	800	700	600

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?

בדקו ברמת בטחון של 95%.

4) במטרה לבדוק האם השתנו דפוסי ההצבעה למפלגות השונות בין שבוע שעבר לשבוע נלקחו שני סקרים אחד מהשבוע שעבר והאחר מהשבוע.

להלן דפוסי ההצבעה שהתקבלו בסקרים אלה.

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

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

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

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

- (5) בחנות בגדים A בדקו את התפלגות הצבעים של הבגדים הנמכרים ביום מסוים. כמו כן בדקו את התפלגות הצבעים בחנות שכנה B:

מספר פריטים / צבע	שחור	לבן	אדום	כחול
חנות A	15	20	15	50
חנות B	60	20	10	20

- א. בדקו ברמת מובהקות של 5% האם התפלגות הצבעים בחנות A היא ביחס של 1:1:1:3 לטובת הכחול.
- ב. בדוק ברמת מובהקות של 2.5% האם קיים הבדל בין החנויות מבחינת התפלגות הצבעים של הפריטים הנמכרים.
- (6) סטודנט קיבל בבדיקת השערות ערך χ^2 (chi-square) השוו לאפס. הסטודנט הסיק כי לא קיימת תלות בין שני המשתנים שבדק, בכל רמת מובהקות. נכון / לא נכון? נמקו.

- (7) להלן טבלת O של שני משתנים שהתקבל במדגם כלשהו:

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

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

תשובות סופיות:

- (1) נסיק שיש קשר בין גודל הארגון לשביעות הרצון של העובדים.
- (2) נסיק שאין הבדל מובהק בין שיעור הפגומים במשמרות השונות.
- (3) נסיק שאין קשר בין היות מוצא לייצוא למועד שבו הוא יוצר.
- (4) א. 10% ב. קטן ג. (0.223, 0.297)
- (5) א. נסיק שהתפלגות הצבעים בחנות היא כמו שמצויין.
ב. נסיק שיש הבדל בין החנויות מבחינת התפלגות הצבעים.
- (6) נכון

הקשר בין מבחן לאי תלות ובדיקת השערות להפרש פרופורציות – רקע

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

דוגמה (פתרון בהקלטה) :

בקרב מדגם של 200 נשים 120 טענו שהן תצבענה למועמד R לראשות העיר. בקרב מדגם של 200 גברים 80 טענו שהם יצביעו למועמד R האם קיים הבדל בין דפוס ההצבעה של הנשים ושל הגברים? האם אפשר לבדוק זאת גם על ידי מבחן לאי תלות וגם על ידי בדיקת השערות לשתי פרופורציות?

שאלות:

- (1) בקרב מדגם של 200 נשים 120 טענו שהן תצבענה למועמד R לראשות העיר. בקרב מדגם של 200 גברים 80 טענו שהם יצביעו למועמד R האם קיים הבדל בין דפוס ההצבעה של הנשים ושל הגברים? בדוק ברמת מובהקות של 5%.
א. על ידי מבחן לאי תלות.
ב. על ידי בדיקת השערות לשתי פרופורציות.
- (2) נלקחו 200 אנשים שמתוכם 60 הצהירו שהם עוסקים בפעילות גופנית סדירה. מתוך אלו שעוסקים בפעילות גופנית סדירה 50 נמצאו במצב בריאותי תקין. מתוך אלו שלא עוסקים בפעילות גופנית סדירה 90 נמצאו במצב בריאותי תקין.
א. בנו טבלת שכיחות משותפת לנתונים שהוצגו בשאלה.
ב. האם ניתן להגיד שהסיכוי להימצא במצב בריאותי תקין גבוה יותר כאשר עוסקים בפעילות גופנית סדירה לעומת המצב שלא עוסקים בפעילות גופנית סדירה? בדוק ברמת בטחון של 90%.

תשובות סופיות:

- (1) נדחה את השערת האפס.
- (2) ב. נדחה את השערת האפס.

מדדי קשר-מדד הקשר של קרמר – רקע

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

דוגמה (פתרון בהקלטה) :

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

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

בהקשר של קרמר הטבלה נקראת טבלת O (observed)

X - מין (גבר/אישה) – סולם שמי.

Y - דעה (בעד/נמנע/נגד) – סולם שמי/סדר.

שלב ב' בחישוב r_c :

שלב א' :

נבנה את טבלת E (Expected)

נעתיק את המסגרת של טבלת O ואז כל $E_i = \frac{f(x) \cdot f(y)}{n}$.

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

שלב ב' :

$$\text{נחשב } \chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

שלב ג' :

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

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

שאלות:

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

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

האם קיים קשר בין מין להשכלה? נמקו!

- (2) נלקחו 200 אנשים שמתוכם 60 הצהירו שהם עוסקים בפעילות גופנית סדירה. מתוך אלו שעוסקים בפעילות גופנית סדירה 50 נמצאו במצב בריאותי תקין. מתוך אלו שלא עוסקים בפעילות גופנית סדירה 90 נמצאו במצב בריאותי תקין.

א. בנו טבלת שכיחות משותפת לנתונים שהוצגו בשאלה.

ב. האם קיים קשר בין פעילות גופנית למצב בריאותי?

חשבו לפי מדד הקשר של קרמר.

תשובות סופיות:

(1) 0.595

(2) ב. 0.19

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

r : מספר הקטגוריות של המשתנה שבשורות.

c : מספר הקטגוריות של המשתנה שבעמודות.

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

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 צ 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%
		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%
		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%
	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.

שאלות:

- 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 פלשתינית 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.

b. The minimum expected count is 7.25.

gender * opinion Crosstabulation

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

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

Crosstab

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

Chi-Square Tests

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

a. 5 cells (55.6%) have expected count less than 5.

b. 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.
- מה תהיה מסקנת המחקר ברמת מובהקות של 1%?
 - מה ערכו של E עבור עיניים כחולות וצבע שיער כהה.
 - מה יהיה בקירוב ערכו של מקדם המתאם של קרמר?
 - 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 צבע עיניים	50.0%	50.0%	100.0%
		% within צבע שיער	21.6%	53.3%	30.8%
		% of Total	15.4%	15.4%	30.8%
חום		Count			
		% within צבע עיניים	83.3%	16.7%	100.0%
		% within צבע שיער	27.0%	13.3%	23.1%
		% of Total	19.2%	3.8%	23.1%
ירוק		Count			
		% within צבע עיניים	79.2%	20.8%	100.0%
		% within צבע שיער	51.4%	33.3%	46.2%
		% of Total	36.5%	9.6%	46.2%
Total		Count			
		% within צבע עיניים	71.2%	28.8%	100.0%
		% within צבע שיער	100.0%	100.0%	100.0%
		% of Total	71.2%	28.8%	100.0%

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

תשובות סופיות:

- (1) א. 1.973 ב. קטן ג. לא נדחה H_0 .
 ד. 34.5% ה. 38.5%
- (2) א. בילוי מועדף ומוזיקה מועדפת עם 3 קטגוריות לכל משתנה.
 ב. לא
 ג. 4
- (3) א. 6, 21.6%, 71.2% ב. 15.8 ג. קטן מ-0.005
- (4) א. לא נדחה H_0 . ב. 11.4 ג. 0.394
 ד. 20.8%
- (5) נקבל H_0 .
- (6) א
- (7) ג
- (8) א. אותה מסקנה. ב. לא ניתן לדעת.

תוכן העניינים:

2 ניתוח שונות חד כיוונית
2 רקע תיאורטי :
4 שאלות :
9 פתרונות סופיים חלקיים

לתשומת לבך, יש ללמוד לפי הסרטונים באתר. ייתכנו שאלות בספר הפרק אשר אינן חלק מחומר הלימוד שלך.

ניתוח שונות חד כיוונית

רקע תיאורטי:

ניתוח שונות (חד כיוונית) הוא מבחן להשוואת תוחלות (μ_1, \dots, μ_k) של k אוכלוסיות שונות. לכן, בנייתוח שונות, השערות המחקר הן:

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

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

ההנחות הדרושות לביצוע התהליך:

- 1) בכל אוכלוסייה מתוך k האוכלוסיות ההתפלגות נורמלית.
- 2) כל האוכלוסיות הן עם אותה שונות σ^2 .
- 3) המדגמים בלתי תלויים זה בזה.

ישנו משתנה המבדיל בין הקבוצות השונות, הוא המשתנה הבלתי תלוי הנקרא גורם (factor). משתנה זה הוא קטגוריאלי עם k רמות (levels). כדי לבצע את התהליך יש לבצע מדגם מכל אוכלוסייה: נסמן ב- n_i את גודל המדגם בקבוצה i .

$$n = \sum_{i=1}^k n_i \quad \text{- מספר התצפיות סך הכול (בכל המדגמים).}$$

$$\bar{X}_1 - \text{ממוצע המדגם הראשון, } \dots, \bar{X}_k - \text{ממוצע המדגם ה-} k \text{-י.}$$

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

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

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

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

$$SST = SSB + SSW$$

יש למלא את טבלת ניתוח השונות הבאה:

מקור השונות	סכום הריבועים SS	דרגות חופש df	ממוצע הריבועים MS	F
B - בין הקבוצות	SSB	$k - 1$	$\frac{SSB}{k - 1}$	$\frac{MSB}{MSW}$
W - בתוך הקבוצות	SSW	$n - k$	$\frac{SSW}{n - k}$	
T - סה"כ	SST	$n - 1$		

$$F = \frac{\frac{SSB}{k-1}}{\frac{SSW}{n-k}} \sim F(k-1, n-k)$$

אזור דחיית $H_0 : 1 - \alpha : F > F_{(k-1, n-k)}$

שאלות:

- (1) מחקר מעוניין להשוות בין שלוש תרופות לשיכוך כאבים במטרה לבדוק האם קיים הבדל בין התרופות מבחינת הזמן בדקות שלוקח עד שהתרופה משפיעה. לצורך הבדיקה נלקחו 15 אנשים שסובלים מכאבי ראש. אנשים אלה חולקו באקראי לשלוש קבוצה 1 קיבלה "אקמול" קבוצה 2 קיבלה "אופטלגין" קבוצה 3 קיבלה "נורופן". כל אדם במחקר מסר את מספר הדקות עד שהתרופה השפיעה עליו.
- א. מהו המשתנה התלוי ומהו המשתנה הבלתי תלוי במחקר?
מהו ה"גורם" וכמה רמות יש לו?
- ב. מהו המבחן הסטטיסטי המתאים כאן? רשמו את ההשערות.
- ג. מה הן ההנחות הדרושות כדי לבצע את המבחן הסטטיסטי שהצעת בסעיף הקודם?

- (2) בעיר מסוימת שלושה בתי ספר תיכון. ראש העיר התעניין לבדוק האם קיים הבדל בהצלחה של בתי הספר במקצוע מתמטיקה. לצורך כך הוא דגם מספר תלמידים שנבחנו במבחן הבגרות במתמטיקה ברמה של 3 יחידות בעירו ובדק עבור כל תלמיד מה ציון הבגרות שלו במתמטיקה. להלן הציונים שהתקבלו:

"הס"	"רביץ"	"המתמיד"
85	98	78
83	62	65
74	55	70
85	80	90
75		56

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

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

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

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

- 4) שלושה אופים נתבקשו להכין עוגת שוקולד. לכל אופה בדקו את משך זמן ההכנה בדקות. כל אופה נדרש לאפות בכל יום 4 עוגות. האם קיים הבדל בין האופים מבחינת תוחלת זמני ההכנה של העוגות? בדקו ברמת מובהקות של 5%.

האופה	ניר	מוזס	שלום
סכום הזמנים	206	212	182
סכום ריבועי הזמנים	10644	11250	8982

- 5) להלן טבלת ניתוח שונות חד כיוונית. במחקר בחנו 4 סוגי סוללות. רצו לבדוק האם לסוג הסוללה השפעה על תוחלת אורך החיים שלה. הפעילו את כל הסוללות על אותו מכשיר ובדקו את אורך החיים של כל סוללה בשעות. מה המסקנה ברמת מובהקות של 10%? רשמו את ההשערות וההנחות הדרושות.

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	10.317	3	3.439	1.361	.279
Within Groups	60.648	24	2.527		
Total	70.964	27			

- 6) להלן טבלת ANOVA בטבלה הושמטו חלקים. השלימו את החלקים בטבלה שהושמטו ומסומנים באותיות.

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	357.450	ב	ג	ה	.000
Within Groups	א	17	ד		
Total	522.950	19			

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

ANOVA						pulse			
pulse						Tukey HSD ^a			
	Sum of Squares	df	Mean Square	F	Sig.	dosage	N	Subset for alpha = 0.05	
								1	2
Between Groups	414.400	2	207.200	19.733	.000	30.00	5	71.0000	
Within Groups	126.000	12	10.500			20.00	5		80.2000
Total	540.400	14				10.00	5		83.4000
						Sig.		1.000	.299

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 5.000.

Post Hoc Tests

Multiple Comparisons

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

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

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

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

Oneway

ANOVA

grade

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

Post Hoc Tests

Multiple Comparisons

grade

Scheffe

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

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

Homogeneous Subsets

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

Means for groups in homogeneous subsets are displayed.
a. Uses Harmonic Mean Sample Size = 5.000.

פתרונות סופיים חלקיים

(2) אם חישבת נכון, ה F הסטטיסטי יוצא: 0.58.

(3) נדחה את השערת האפס.

(4) להלן טבלת הניתוח השונות המתקבלת:

	Sum of Squares	df	Mean Square	F
Between Groups	126.000	2	63.000	.756
Within Groups	750.000	9	83.333	
Total	876.000	11		

(5) נקבל את השערת האפס.

(6) א. 165.5 ב. 2 ג. 178.725 ד. 9.375

ה. 18.36

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

(8) א. 5 ב. 25 ג. כן

תוכן העניינים:

2	שאלות מסכמות בבדיקת השערות על פרמטרים
2	שאלות:
5	תשובות סופיות:
6	שאלות סיכום – שאלות רב ברירה על בדיקת השערות
19	תשובות סופיות:

לתשומת לבך, יש ללמוד לפי הסרטונים באתר. ייתכנו שאלות בספר הפרק אשר אינן חלק מחומר הלימוד שלך.

שאלות מסכמות בבדיקת השערות על פרמטרים

שאלות:

- (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 $\bar{X} > 522$.
- למי מהחוקרים הסתברות לטעות מסוג ראשון קטנה יותר?
 - מהי ההסתברות לטעות מסוג שני של חוקר ב' עבור $\mu_1 = 525$.
 - הסבר ללא חישוב נוסף, האם ההסתברות לטעות מסוג שני עבור $\mu_1 = 525$, של חוקר א' שווה/קטנה/גדולה לזו של חוקר ב'.
 - חוקר א' קיבל במדגם שלו $\bar{X} = 523$. מהי מסקנתו?

- (2) ידוע כי תוחלת מספר הלייקים היומי של דנה היא 12 עם סטיית תקן 5. דני טוען שהוא יותר פופולארי מדנה בכך שהוא מקבל יותר לייקים מדנה ביום. על-מנת לבדוק זאת ספר דני כמה לייקים הוא קיבל בכל יום במהלך 7 שבועות (כלומר, ב-49 ימים) וקיבל סך-הכול 637 לייקים. נניח כי סטיית התקן של מספר הלייקים שדני מקבל ביום זהה לסטיית התקן של דנה.
- מהי רמת המובהקות שכדאי לדני לדרוש, כדי שדנה תשתכנע בצדקת טענתו (שדני פופולרי יותר בכך שהוא מקבל יותר לייקים מדנה ביום).
 - אם דני משער שתוחלת מספר ה"לייקים" שהוא מקבל ביום היא 14 וקובע רמת מובהקות 2.5%, מהי עוצמת המבחן של דני?

B	A	מוצר / רשת
5	5	1
5	4	2
3	5	3
4	7	4

- (3) ברצוננו להשוות בין רשתות A לבין B. לשם כך בחרנו 4 מוצרים, ובדקנו את מחיריהם בשתי הרשתות. להלן התוצאות:
- הניחו כי המחירים מתפלגים נורמלית.
- אם יש הנחות נוספות כדי לבצע את המבחן הפרמטרי רשמו אותן.
- בדקו האם קיים הבדל בין הרשתות מבחינת תוחלת המחירים. רמת מובהקות של 5%.
 - חזרו על הסעיף הקודם בהנחה ונבחרו בכל רשת מוצרים באקראי ולא בהכרח אותם מוצרים.

4) במדגם של 10 ישראלים שנבחנו במבחן ה-IQ נתקבלו התוצאות הבאות:

$$n = 10, \quad \sum X_i = 1020, \quad \sum X_i^2 = 105120$$

במדגם של 14 אמריקאים שנבחנו במבחן ה-IQ נתקבלו התוצאות הבאות:

$$n = 14, \quad \sum X_i = 1386, \quad \sum X_i^2 = 138644$$

נתון שציוני הבחינה מתפלגים נורמלית בכל מדינה.

א. בדקו ברמת מובהקות של 10% האם קיים שוויון שונויות בין אוכלוסיית אמריקה לאוכלוסיית ישראל?

ב. בדקו האם קיים הבדל בממוצע הציונים בבחינת ה-IQ בין ישראל לארה"ב. ברמת מובהקות של 5%?

5) 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. נסחו את ההשערות ובדקו אותן ברמת מובהקות של 5%. רשום את כלל ההכרעה ואת ההנחות הדרושות לביצוע המבחן הפרמטרי.

ב. חשבו את p-value.

ג. ישנה טענה שממוצע זמן ההשקעה בלימודים במכללות הוא 3.5 שעות ביום. בדקו את הטענה כאשר רמת המובהקות הינה 5%.

6) במדינת טרפפו המשכורות במשק מתפלגות נורמלית עם ממוצע של 1 אלף דולר וסטטיית תקן של 0.2 אלף דולר. בוצע מדגם מקרי בו השתתפו 5 נשים ו 5 גברים במדינת שומקום שבה המשכורות מתפלגות נורמלית גם כן. להלן משכורותיהם באלפי דולר:

1.1	1.2	0.7	0.9	2	גברים
1.2	1.8	1.9	1.1	1.4	נשים

א. בדקו את הטענה שממוצע משכורותיהם של אזרחי שומקום גבוה מאשר ממוצע משכורותיהם של אזרחי טרפפו ברמת מובהקות של 5%.

בהנחה שסטטיית התקן זהה בשתי המדינות.

ב. חזרו על הסעיף הקודם ללא ההנחה הנ"ל.

ג. ישנה טענה שסטטיית התקן במדינת שומקום גבוהה מזו של טרפפו. בדקו ברמת מובהקות של 5%.

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

לא צופים	צופים	
42	320	נשים
120	72	גברים

- א. האם יש הבדל בין אחוזי הצפייה של גברים ונשים ברמת מובהקות של 1%?
 ב. עבור רמת מובהקות של 5% בדוק טענה שמבין הצופים בתוכנית הטלוויזיה אחוז הנשים גדול פי 2 מאחוז הגברים.

8) בשנת 2000 ל-60% היה מדיח כלים בבית. מחקר רוצה לבדוק האם כיום פרופורציית המשפחות עם מדיח כלים עלה. הוחלט לבצע מדגם אקראי של 150 משפחות.

- א. רשמו את השערות המחקר.
 ב. מה היא מסקנת המחקר ברמת מובהקות של 5% אם במדגם ל-102 משפחות היה מדיח כלים.
 ג. מהי הטעות האפשרית במסקנה מהסעיף הקודם. האם ניתן לדעת את הסתברותה?

9) נערך מחקר על הקשר בין עישון ויתר לחץ דם. נבדק מדגם מקרי של 200 מעשנים ונמצא כי 30 סבלו מיתר לחץ דם. ידוע שבאוכלוסייה 18% סובלים מיתר לחץ דם.

- א. בדקו ברמת מובהקות 0.1 את ההשערה כי אחוז הסובלים מיתר לחץ דם בקרב המעשנים גדול מאשר כלל האוכלוסייה.
 ב. מהי רמת המובהקות המינימלית לקבלת הטענה שאחוז הסובלים מיתר לחץ דם בקרב המעשנים גדול מאשר כלל האוכלוסייה.
 ג. מהי עצמת המבחן, אם אחוז הסובלים מיתר לחץ דם בקרב אוכלוסיית המעשנים היא בפועל 25%.

10) להלן התפלגות מספר הנסיעות לחופשה השנתית במדגם של משפחות ישראליות. בדקו ברמת מובהקות של 5%:

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

- א. באיטליה משפחות נוסעות בממוצע פעמיים בשנה לחופשה. האם בישראל משפחות נוסעות פחות מאשר באיטליה?
 ב. בהולנד 80% מהמשפחות נוסעות לפחות פעם אחת בשנה לחופשה, האם בישראל אחוז המשפחות שנוסעות לפחות פעם אחת בשנה לחופשה נמוך מאשר בהולנד?

11 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. חזרו על סעיף א אם ההשערה האלטרנטיבית הייתה: $H_1: \mu < 40$.

ג. חזרו על סעיף א אם ההשערה האלטרנטיבית הייתה: $H_1: \mu \neq 40$.

12 ציוני בחינת הבגרות במתמטיקה מתפלגים נורמלית עם שונות 150. במדגם של 16 נבחנים מתל אביב התקבלה שונות מדגמית-190. במדגם של 25 ירושלמים התקבלה שונות מדגמית 118.

א. בדקו ברמת מובהקות של 2.5% האם שונות הציונים במתמטיקה בקרב נבחני תל אביב גבוהה מהשונות בכלל הארץ.

ב. בדקו ברמת מובהקות של 5% האם שונות ציונים במתמטיקה בקרב תלמידי תל אביב גבוהה מאשר בקרב תלמידי ירושלים.

תשובות סופיות:

- | | | | |
|--------------------|-----------------------------------|-----------------|-----------------|
| א. חוקר א' | ב. 0.0122 | ג. גדלה | ד. נדחה H_0 . |
| א. לפחות 0.0808 | ב. 0.7995 | | |
| א. לא נדחה H_0 . | ב. לא נדחה H_0 . | | |
| א. לא נדחה H_0 . | ב. לא נדחה H_0 . | | |
| א. לא נדחה H_0 . | ב. בין 5% ל-10% | ג. נדחה H_0 . | |
| א. נדחה H_0 . | ב. נדחה H_0 . | ג. נדחה H_0 . | |
| א. נדחה H_0 . | ב. נדחה H_0 . | | |
| ב. נדחה H_0 . | ג. טעות מסוג ראשון בסיכוי של 0.05 | | |
| א. לא נדחה H_0 . | ב. 0.8643 | ג. 0.8749 | |
| א. נדחה H_0 . | ב. נדחה H_0 . | | |
| א. 0.0062 | ב. 0.9938 | ג. 0.0124 | |
| א. לא נדחה H_0 . | ב. לא נדחה H_0 . | | |

שאלות סיכום – שאלות רב ברירה על בדיקת השערות

(1) בבדיקת השערה חד-צדדית ימנית ברמת מובהקות $\alpha = 0.01$, נדחתה השערת האפס. מה הייתה המסקנה לו נבדקה אותה ההשערה באמצעות אותם נתונים ברמת מובהקות $\alpha = 0.05$?

- א. השערת האפס הייתה נדחית.
- ב. השערת האפס לא הייתה נדחית.
- ג. ההשערה המחקרית הייתה נדחית.
- ד. בהעדר נתונים נוספים, לא ניתן לדעת.

(2) על מנת לבדוק האם ההסתברות ללידת בן הינה חצי, נבחר מדגם מקרי של 200 ילדים, ונמצא שישנם 120 בנים. מהן ההשערות האלטרנטיביות להשערת האפס?

א. $H_1 : p = 0.5$

ב. $H_1 : p = 0.6$

ג. $H_1 : p > 0.5$

ד. $H_1 : p \neq 0.5$

(3) לצורך בדיקת השפעת היפנוזה על לימוד אנגלית, נבחרו 10 זוגות תאומים זהים. אחד התאומים למד אנגלית בהשפעת היפנוזה, והשני ללא היפנוזה. לאחר מכן נערך לשניהם מבחן באנגלית. נניח שציוני המבחן מתפלגים נורמאלית ללא ידיעת השונות האמיתית. המבחן שיש לבצע כאן הוא:

א. מבחן Z למדגם יחיד.

ב. מבחן T למדגם יחיד.

ג. מבחן T למדגמים בלתי תלויים.

ד. מבחן T למדגמים מזווגים.

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

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

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

5) 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.

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

- 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.

7) 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.

8) כדי לבדוק את הטענה שגברים רווקים שוקלים פחות מגברים נשואים לקח חוקר מדגם מקרי של 4 גברים ומדד את משקלם לפני נישואיהם ולאחר

נישואיהם. הנה התוצאות :

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

לבדיקת

באיזה התפלגות משתמשים

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.

- 9) שני סטטיסטיקאים בודקים השערות ברמת מובהקות $\alpha = 0.05$ על סמך אותו מדגם. סטטיסטיקאי א' בודק את ההשערה: $H_0: \mu = 20$ כנגד האלטרנטיבה $H_1: \mu \neq 20$ ומחליט לא לדחות את השערת האפס. סטטיסטיקאי ב' בודק את ההשערה $H_0: \mu \leq 20$ כנגד האלטרנטיבה $H_1: \mu > 20$ מה יחליט סטטיסטיקאי ב'?
- לדחות את השערת האפס.
 - לא לדחות את השערת האפס.
 - ללא נתונים נוספים אי אפשר לדעת מה יחליט.

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

- 11) רמת הכולסטרול בדמם של אנשים מתפלג נורמאלית עם תוחלת של 180 מ"ג (ל 100 סמ"ק דם). וסטיית תקן של 10 מ"ג. מעוניינים לבדוק את הטענה שצמחונים הם בעלי רמת כולסטרול נמוכה יותר. נניח שסטיית התקן אצל צמחונים זהה לסטיית התקן של כלל האנשים. במדגם של 20 צמחונים התקבל ממוצע רמת כולסטרול 174.5 מ"ג. אם הוחלט לקבל את הטענה שצמחונים הם בעלי רמת כולסטרול נמוכה יותר איזה סוג טעות אפשרית במסקנה?
- טעות מסוג ראשון.
 - טעות מסוג שני.
 - טעות מסוג שלישי.
 - לא ניתן לדעת כיוון שאנו לא יודעים מה התוחלת האמתית אצל הצמחוניים.

12 בסקר שנערך התקבל ש 60% מתוך 220 נשאלים מבקרים אצל השיננית לפחות פעם אחת בשנה. עבור אילו רמות מובהקות ניתן יהיה לקבוע שרוב האוכלוסייה מבקרת אצל השיננית לפחות פעם בשנה?

- א. רמת מובהקות הגדולה מ-5%.
- ב. רמת מובהקות הקטנה מ-5%.
- ג. רמת מובהקות הגדלה מ-0.0015.
- ד. רמת מובהקות הקטנה מ-0.0015.

13 שני חוקרים העוסקים בתחום מחקרי משותף החליטו להסתמך על נתונים של מדגם שפורסם על ידי הלשכה המרכזית לסטטיסטיקה.

חוקר א' ניסח השערה דו צדדית ואילו חוקר ב' ניסח השערה חד צדדית. מסקנתו של איזה מבין המשפטים הבאים הוא הנכון בנוגע למסקנות החוקרים?

- א. אם חוקר א' ידחה את השערת האפס לא ניתן לדעת מה יחליט חוקר ב' באותה רמת מובהקות.
- ב. אם חוקר א' יקבל את השערת האפס גם חוקר ב' יקבל את השערת האפס באותה רמת מובהקות.
- ג. אם חוקר ב' ידחה את השערת האפס גם חוקר א' ידחה את השערת האפס באותה רמת מובהקות.
- ד. אם חוקר א' ידחה את השערת האפס גם חוקר ב' ידחה את השערת האפס בתנאי שרמת המובהקות כפולה בגודלה.

14 ידוע מנתוני העבר כי תוחלת הציונים בבחינה בפסיכולוגיה היא 79. הועלתה השערה כי תוחלת הציונים בקרב העולים החדשים נמוכה יותר. לצורך בדיקת הטענה נלקח מדגם מקרי של 47 סטודנטים עולים ונמצא ממוצע של 75. מה משמעות הפרמטר בניסוח ההשערות?

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

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

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

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

- 17** בינואר השנה פורסם שהשכר הממוצע במשק הוא 8,900. במדגם שנעשה בחודש יוני על 60 עובדים נרשם עבור כל עובד במדגם האם השכר שלו נמוך או לא נמוך מהשכר הממוצע שפורסם בחודש ינואר. מהו המבחן המתאים כדי לבדוק שרוב העובדים בחודש יוני קיבלו שכר הנמוך מהשכר הממוצע שפורסם בחודש ינואר?
- א. מבחן Z על פרופורציה.
 - ב. מבחן T על תוחלת אחת.
 - ג. מבחן T על שתי תוחלות במדגמים בלתי תלויים.
 - ד. מבחן T על שתי תוחלות במדגמים תלויים.

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

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

- 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.

- כעת בברור הוא ידחה את השערת האפס.
- כעת הוא דווקא יקבל את השערת האפס.
- כעת לא ניתן לדעת מה תהיה מסקנתו.

23) אם חוקר החליט להגדיל את רמת המובהקות במחקר שלו אזי:

- הסיכוי לטעות מסוג ראשון גדל.
- העוצמה של המבחן גדלה.
- הסיכוי לטעות מסוג שני גדל.
- 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 |
| א. גדולה | גדולה |
| ב. גדולה | קטנה |
| ג. קטנה | גדולה |
| ד. קטנה | קטנה |

26) נערך שינוי בכלל ההחלטה של בדיקת השערה מסוימת ובעקבותיו אזור דחיית H_0 קטן. כל שאר הגורמים נשארו ללא שינוי. כתוצאה מכך :

- א. הן 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 :

- 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.

(29) מובהקות התוצאה (PV) היא גם :

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

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

- א. ברמת מובהקות של 0.01 אך לא של 0.05 נדחה את H_0 .
- ב. ברמת מובהקות של 0.01 ושל 0.05 לא נדחה את H_0 .
- ג. ברמת מובהקות של 0.05 אך לא של 0.01 נדחה את H_0 .
- ד. ברמת מובהקות של 0.01 ושל 0.05 נדחה את H_0 .

(31) רמת המובהקות במחקר הייתה 2% לכן.

- א. בסיכוי של 2% נדחה את השערת האפס.
- ב. בסיכוי של 2% לא נדחה את השערת האפס.
- ג. בסיכוי של 2% השערת האפס לא נכונה.
- ד. אף תשובה לא נכונה.

(32) נתון ש: $X \sim N(\mu, \sigma^2)$. כמו כן נתונות ההשערות הבאות: $H_0 : \mu = \mu_0$, $H_1 : \mu < \mu_0$.

- חוקר בדק את ההשערות הללו על סמך מדגם שכלל 10 תצפיות.
 σ^2 לא הייתה ידועה לחוקר. החוקר החליט לדחות את השערת האפס ברמת מובהקות של 5%. אם הוא היה מגדיל את רמת המובהקות ל-10% אזי :
- א. כעת בברור הוא ידחה את השערת האפס.
 - ב. כעת הוא דווקא יקבל את השערת האפס.
 - ג. כעת לא ניתן לדעת מה תהיה מסקנתו.

(33) לצורך בדיקת השפעת היפנוזה על לימוד אנגלית, נבחרו 10 זוגות תאומים

זהים. אחד התאומים למד אנגלית בהשפעת היפנוזה, והשני ללא היפנוזה. לאחר מכן נערך לשניהם מבחן באנגלית. נניח שציוני המבחן מתפלגים נורמאלית ללא ידיעת השונות האמתית. מספר דרגות החופש במבחן הוא :

- א. 9
- ב. 19
- ג. 18
- ד. 8

- 2 | משקל במכשיר | 5.0 | 7.1 | 6.9 | 5.3
- 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):
- 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.
- 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:
- מבחן Z למדגם יחיד.
 - מבחן T למדגם יחיד.
 - מבחן T למדגמים בלתי תלויים.
 - מבחן T למדגמים מזווגים.
- 36** סטטיסטיקאי ערך מבחן סטטיסטי. הוא חישב את עוצמת המבחן וקיבל 0. המשמעות של תוצאה זו היא:
- לעולם לא לדחות את השערת האפס כאשר היא לא נכונה.
 - תמיד לדחות את השערת האפס כאשר היא נכונה.
 - לעולם לא לדחות את השערת האפס כאשר היא נכונה.
 - 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:
- לדחות את השערת האפס.
 - לא לדחות את השערת האפס.
 - שלא ניתן לדעת את המסקנה עבור רמת מובהקות 0.05.
 - שלא נתונות בשאלה סטיות התקן של האוכלוסיות, ולכן לא ניתן להסיק דבר.

38 במטרה לבדוק האם קיים הבדל בין קווי זהב לבזק מבחינת ממוצע המחירים לשיחות בינ"ל. נגדמו באקראי 7 מדינות ועבור כל מדינה נבדקה עלות דקת שיחה. בהנחה והמחירים מתפלים נורמלית בנו רווח סמך למוצע הפרשים וקיבלו: $-0.0293 < \mu_D < 0.2145$ רווח הסמך הוא ברמת סמך של 95%.
לכן מסקנת המחקר היא:

- ברמת מובהקות של 5% לא נוכל לקבוע שקיים הבדל בין החברות.
- ברמת מובהקות של 5% נקבע שקיים הבדל מובהק בין החברות.
- לא ניתן לדעת מה המסקנה ברמת מובהקות של 5% כיוון שלא נאמר מה ההגדרה של D .

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

- תמיד נדחה H_0 כאשר היא נכונה, אך לא תמיד נדחה אותה כאשר היא לא נכונה.
- לא נדחה את H_0 אף פעם.
- לא נדחה את H_0 כאשר היא נכונה אך יתכן ונדחה אותה כאשר היא לא נכונה.
- כל התשובות לא נכונות.

40 חוקר ביצע ניסוי. הוא ניסח את ההשערות הבאות: $H_0: \mu = 10$, $H_1: \mu \neq 10$. לצורך בדיקה הוא לקח מדגם מקרי בגודל 5 מתוך אוכלוסייה המתפלגת נורמאלית עם שונות לא ידועה. על סמך תוצאות המדגם הוא חישב וקיבל:
 $t_{\bar{x}} = -2.63$. לכן המסקנה היא:

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

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

- 0.005
- 0.03
- 0.0287
- 0.05

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

א. יקטן הסיכוי לטעות מסוג שני (β).

ב. תגדל רמת הביטחון ($1-\alpha$).

ג. אף תשובה לא נכונה.

ד. כל התשובות נכונות.

43) איזה מהמשפטים הבאים נכון תמיד?

א. $POWER + \alpha + \beta = 1$

ב. $POWER = 0.5 - \beta$

ג. $POWER + \alpha = 1$

ד. $\beta + \alpha = 1$

ה. הכול לא נכון.

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

א. היא אומרת שהשונויות המדגמיות שוות.

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

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

ד. אף תשובה אינה נכונה.

45) חוקר החליט לא לדחות השערה ברמת מובהקות של α . במידה וחוקר זה היה בודק השערה זו ברמת מובהקות של 2α על סמך אותם נתונים, האם ההשערה תדחה?

א. ההשערה תדחה.

ב. ההשערה לא תדחה.

ג. התשובה תלויה בעוצמת המבחן.

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

46) חוקרת שיערה, כי בגילאי הגן בנות יותר תקשורתיות מבנים. אם החוקרת תדגום אקראית 30 בנים ו-30 בנות, ובמדגם יתקבל אותו ממוצע של ציון תקשורת. סטטיסטי המבחן יהיה:

א. אפס

ב. חיובי

ג. שלילי

ד. לא ניתן לדעת

(47) עוצמה שווה ל-1 פרושה :

- א. לעולם לא לדחות את השערת האפס כאשר היא נכונה.
- ב. תמיד לדחות את השערת האפס כאשר היא נכונה.
- ג. לעולם לא לדחות את השערת האפס כאשר היא לא נכונה.

(48) מה מהבאים **נכון** לגבי מבחן T מדגמים מזווגים?

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

(49) לבדיקת ההשערה החד צדדית על התוחלת של התפלגות נורמלית $H_0: \mu \geq 10$,

$H_1: \mu < 10$. נלקח מדגם והתקבלה רמת מובהקות מינימאלית לדחיית השערת האפס 0.058. לו רצינו לבדוק את ההשערה הדו צדדית $H_0: \mu = 10$, $H_1: \mu \neq 10$, אז על סמך תוצאת אותו המדגם ברמת מובהקות 0.05:

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

(50) לבדיקת ההשערה החד צדדית ימנית $H_0: \mu = 55$, $H_1: \mu = 65$

נלקח מדגם מקרי בגודל n מאוכלוסייה בעלת התפלגות נורמלית ושונות σ^2 . רמת המובהקות היא 5%. נמצא שהעוצמה היא 0.9. להלן 3 טענות:

- עבור מדגם בגודל n וברמת מובהקות 5% לבדיקת ההשערות: $H_0: \mu = 55$, $H_1: \mu = 60$ העוצמה תהיה גדולה מ-0.9.
- עבור מדגם בגודל $2n$ ורמת מובהקות 5% לבדיקת ההשערות: $H_0: \mu = 55$, $H_1: \mu = 65$ העוצמה תהיה גדולה מ-0.9.
- עבור מדגם בגודל n ורמת מובהקות 10% לבדיקת ההשערות: $H_0: \mu = 55$, $H_1: \mu = 65$ העוצמה תהיה קטנה מ-0.9.

א. שלושת הטענות אינן נכונות.

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

תשובות סופיות:

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