

סטודנטים יקרים

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

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

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

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

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



אנו מאחלים לכם הצלחה מלאה בבחינות

צוות האתר GooL

גול זה בול. בשבילך!

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

כתב ופתר - ברק קנדל ©

תוכן

פרק 1 - הסקה סטטיסטית – הקדמה	5
פרק 2 - מושגים בסיסיים באמידה	8
פרק 3 - אמידה נקודתית	15
אומד חסר הטיה	15
אומד נראות מקסימלית	23
קריטריון MSE - תוחלת ריבוע הטעות	35
שאלות מסכמות באמידה נקודתית	38
פרק 4 - סטטיסטיקה תיאורית - הקדמה	44
פרק 5 - סטטיסטיקה תיאורית - הצגה של נתונים	47
פרק 6 - סטטיסטיקה תיאורית - מדדי מיקום מרכזי	55
פרק 7 - סטטיסטיקה תיאורית - מדדי פיזור: השווא, השונויות וסטיית התקן	65
פרק 8 - סטטיסטיקה תיאורית - מדדי מיקום יחסי - ציון תקן	70
פרק 9 - סטטיסטיקה תיאורית - מדדי מיקום יחסי - אחוזונים במחלקות	73
פרק 10 - סטטיסטיקה תיאורית - טרנספורמציה לינארית	77
פרק 11 - סטטיסטיקה תיאורית - שאלות מסכמות	81
פרק 12 - רווח סמך לתוחלת (ממוצע האוכלוסייה)	90
רווח סמך כששונויות האוכלוסייה ידועה	90
קביעת גודל מדגם באמידת תוחלת עם שונויות אוכלוסייה ידועה	97
רווח סמך לתוחלת (ממוצע האוכלוסייה) כששונויות האוכלוסייה אינה ידועה	100
פרק 13 - רווח סמך לפרופורציה	106
קביעת גודל מדגם באמידת פרופורציה	109
פרק 14 - רווח סמך להפרש פרופורציות	113
פרק 15 - רווח סמך להפרש תוחלות ממדגמים בלתי תלויים	116
כששונויות האוכלוסייה ידועות	116
כששונויות האוכלוסייה אינן ידועות אך שוות והמדגמים בלתי תלויים	119
פרק 16 - רווח סמך לתוחלת ההפרש במדגם מזווג	122
פרק 17 - רווח סמך לשונויות וסטיית תקן	125
פרק 18 - רווח סמך ליחס שונויות	130
פרק 19 - תרגול מסכם ברווחי סמך	135
פרק 20 - בדיקת השערות כללית	140

146	פרק 21 - בדיקת השערות על פרמטרים
146	הקדמה
149	טעויות בבדיקת השערות
151	פרק 22 - בדיקת השערות על תוחלת (ממוצע)
151	כאשר שונות האוכלוסייה ידועה
156	סיכוי לטעויות ועוצמה כאשר שונות האוכלוסייה ידועה
163	קביעת גודל מדגם כששונות האוכלוסייה ידועה
166	מובהקות התוצאה (p-value) בבדיקת השערות על תוחלת עם שונות ידועה
171	בדיקת השערות על תוחלת (ממוצע) כאשר שונות האוכלוסייה אינה ידועה
176	מובהקות התוצאה (p-value) כאשר שונות האוכלוסייה לא ידועה
180	ניתוח פלטי SPSS
185	הקשר בין רווח סמך לבדיקת השערות על תוחלת
188	פרק 23 - בדיקת השערות על פרופורציה
188	התהליך
192	סיכוי לטעויות ועוצמה
197	קביעת גודל מדגם
200	מובהקות התוצאה
204	פרק 24 - בדיקת השערות על הפרש פרופורציות
208	פרק 25 - בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים
208	כשהשונות של האוכלוסייה ידועות
212	כששונות האוכלוסייה לא ידועות ומניחים שהן שוות
216	ניתוח פלטים
223	פרק 26 - בדיקת השערות על תוחלת ההפרשים במדגמים מזווגים (תלויים)
223	בדיקת השערות למדגמים מזווגים
229	מדגמים מזווגים - ניתוח פלטים
235	פרק 27 - הקשר בין רווח סמך לבדיקת השערות על הפרש תוחלות
239	פרק 28- בדיקת השערות על שונות
239	בדיקת השערות על שונות האוכלוסייה כאשר התוחלת לא ידועה
244	בדיקת השערות על שתי שונות
251	פרק 29-שאלות מסכמות בבדיקת השערות על פרמטרים
251	שאלות מסכמות בבדיקת השערות על פרמטרים
258	שאלות מסכמות בסגנון רב ברירה (אמריקאיות) על בדיקת השערות

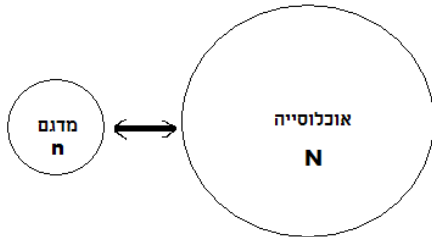
פרק 30 - מבחני חי בריבוע	274
מבחן טיב התאמה	274
הקשר בין מבחן טיב התאמה לבדיקת השערות על הפרופורציה	281
מבחן חי בריבוע לאי תלות בין משתנים	284
הקשר בין מבחן לאי תלות ובדיקת השערות להפרש פרופורציות	289
פרק 31 - ניתוח שונות חד כיוונית	292
פרק 32 - מדדי קשר - מדד הקשר הלינארי (פירסון)	304
פרק 33 - מדדי קשר - רגרסיה ליניארית	312

פרק 1 - הסקה סטטיסטית – הקדמה

רקע:

אוכלוסייה – קבוצה שאליה מפנים שאלה מחקרית.

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



מדגם – חלק מתוך האוכלוסייה.

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

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

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

סטטיסטי – גודל המחושב על המדגם.

פרמטר – גודל המתאר את האוכלוסייה.

הסימונים לפרמטר וסטטיסטי הם שונים

למשל:

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

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

דוגמה (פתרון בהקלטה):

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

א. מי האוכלוסייה?

ב. מה המשתנה?

ג. מה הפרמטרים?

ד. מהו גודל המדגם?

ה. מהו הסטטיסטי שמתכננים להוציא מהמדגם?

ו. האם הפרמטר או הסטטיסטי הוא משתנה מקרי?

תרגילים :

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

א. מי האוכלוסייה?

ב. מה המשתנה?

ג. מהם הפרמטרים?

ד. מהו גודל המדגם?

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

נגדיר את x להיות מספר המקלטים של משפחה אקראית.

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

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

א. מיהי האוכלוסייה ומהו המשתנה הנחקר?

ב. מהו הסטטיסטי שיילקח מהמדגם ומה סימונו?

3. נתון כי 20% מהשכירים במדינה הם אקדמאיים. נבחרו באקראי 10 שכירים באותה אוכלוסייה ומתכננים לפרסם את מספר האקדמאיים שנדגמו.

א. מהי האוכלוסייה ?

ב. מה המשתנה באוכלוסייה?

ג. מהם הפרמטרים?

ד. מהו הסטטיסטי?

פרק 2 - מושגים בסיסיים 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- $\hat{\theta}$. $\hat{\theta}$ 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 : $\sigma(\hat{\theta}) = S.E$

להלן פרמטרים מרכזיים והאומדים שלהם:

ממוצע האוכלוסייה: μ

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

$E(\bar{x}) = \mu$ לכן $\bar{x}$ הינו אומד חסר הטיות 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 אומד חסר הטיות 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 הטיות ל σ^2 .

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

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

דוגמה: (פתרון בהקלטה)

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

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

אמדו באמצעות אומדים חסרי הטיה את הפרמטרים הבאים :

1. ממוצע מספר הילדים למשפחה בתל אביב.

2. שונות מספר הילדים למשפחה בתל אביב.

3. פרופורציית המשפחות בנות שני ילדים.

תרגילים:

1. מתוך 500 טירונים נמצאו 120 בעלי שברי הליכה. נתון שהסיכוי שטירון יהיה עם שבר הליכה הוא 0.25.
 - א. מהי האוכלוסייה המוצגת בשאלה? מהם הפרמטרים שלה?
 - ב. מהי טעות התקן של האומדן כשהמדגם בגודל 500?
 - ג. מהו האומדן לפרמטר?
 - ד. מהי טעות האמידה?

2. לפי נתוני היצרן מקרר צורך בממוצע 2400 וואט לשעה עם סטיית תקן של 500 וואט לשעה. במדגם של 25 מקררים של היצרן התקבל ממוצע של 2342 וואט לשעה.
 - א. מהי האוכלוסייה המוצגת בשאלה? מהם הפרמטרים שלה?
 - ב. מהי טעות התקן של האומדן?
 - ג. מהו האומדן לפרמטר?
 - ד. מהי טעות האמידה?

3. נדגמו עשרה מתגייסים לצה"ל. גובהם נמדד בס"מ. להלן התוצאות שהתקבלו:

$$175, 168, 184, 192, 171, 180, 177, 187, 168, 177$$
 - א. מצא אומדן חסר הטיה לגובה הממוצע של מתגייסי צה"ל.
 - ב. מצא אומדן חסר הטיה לשונות הגבהים של מתגייסי צה"ל.
 - ג. מצא אומדן חסר הטיה לפרופורציות המתגייסים בגובה של לפחות 180 ס"מ.

4. נדגמו 20 שכירים באקראי. עבור כל שכיר נמדד השכר באלפי שקלים. להלן התוצאות שהתקבלו:

$$\sum_{i=1}^{20} X_i^2 = 1502.2 \quad \sum_{i=1}^{20} X_i = 162$$
 - א. אמדו את השכר הממוצע של השכירים במשק.
 - ב. אמדו את סטיית התקן של שכר השכירים במשק.

5. במטרה לאמוד את ממוצע האוכלוסייה. דגמו תצפיות בלתי תלויות מהאוכלוסייה וחישבו את

הממוצע שלהם. מהי טעות התקן?

- א. סטיית התקן של האוכלוסייה.
- ב. סטיית התקן של ממוצע האוכלוסייה.
- ג. סטיית התקן של המדגם.
- ד. סטיית התקן של ממוצע המדגם.

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

א. 3

ב. 2.5

ג. 1.581

ד. אין מספיק נתונים לדעת.

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

א. כאשר n קטן.

ב. כאשר תצפיות המדגם אינן בלתי תלויות.

ג. כאשר האוכלוסייה אינה מתפלגת נורמאלית.

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

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

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

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

א. 16

ב. 8

ג. 4

ד. 2

9. מהו אומד חסר הטיה?

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

פתרונות:**שאלה 3**

א. 177.9

ב. 64.1

ג. 0.4

שאלה 4

א. 8.1

ב. 3.16

שאלה 5

התשובה היא ד.

שאלה 6

התשובה היא ג.

שאלה 7

התשובה היא ד.

שאלה 8

התשובה היא ד.

שאלה 9

התשובה היא ג.

פרק 3 - אמידה נקודתית

אומד חסר הטיה

רקע:

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

דוגמה: (פתרון בהקלטה)

המשתנה X הוא בעל פונקציית ההסתברות הבאה:

3	2	1	X
4θ	$1-6\theta$	2θ	הסתברות

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

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

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

ב. מהי ההטיה של האומד T_1 .

ג. תקנו את T_1 כך שיהיה אומד חסר הטיה.

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

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

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

ה. מצאו אומד חסר הטיה ל: $P(X = 3)$.

• אומדן חסר הטיה לשונות האוכלוסייה: σ^2 : $S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{\sum x_i^2 - n\bar{x}^2}{n-1}$

ו. מצאו אומדן חסר הטיה לשונות של X .

תזכורות חשובות:

• אם $Y = aX + b$ אזי:

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

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

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

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

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

תרגילים:

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

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

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

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

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

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

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

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

3. $X \sim B(n, p)$ כלומר X הינו משתנה מקרי המתפלג בינומית עם פרמטר P (סיכוי להצלחה

- בניסיון בודד) במדגם בגודל n .
 א. פתחו אומד חסר הטיה ל- P .
 ב. מהו אומד חסר הטיה לסיכוי לכישלון בניסיון בודד.
 ג. מהו אומד חסר הטיה ל- $E(X)$.
 ד. מצאו אומד חסר הטיה ל- $E(X^2)$.

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

פונקציית ההסתברות של X – מספר המניות שיעלו ביום מסוים :

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

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

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

בפרמטר θ באופן הבא :

הסיכוי שמטפלת תטפל בתינוק אחד בלבד הוא 3θ ,
 הסיכוי שמטפלת תטפל ב-2 תינוקות הוא $1-4\theta$,
 הסיכוי שמטפלת תטפל ב-3 תינוקות הוא θ .
 במדגם מיקרי של 4 מטפלות מת"א, נמצא כי שתיים מהם מטפלות בתינוק אחד בלבד, אחת מהן בשנים ואחת השלושה תינוקות.
 א. מצא אומד חסר הטיה לפרמטר θ על סמך תצפית בודדת.
 ב. מצאו אומד חסר הטיה לפרמטר θ על סמך 4 תצפיות.
 ג. מהו האומדן לפרמטר θ על סמך תוצאות המדגם.
 ד. מצאו אומד חסר הטיה לסיכוי שלמטפלת בת"א תטפל בתינוק בודד אחד.
 ה. מצאו אומדים חסרי הטיה לתוחלת ולשונות של מספר התינוקות בטיפול אצל מטפלת מת"א. חשבו אומדנים.

6. קבע אילו מהטענות הבאות נכונות :

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

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

א. $\frac{X}{20}$

ב. $\frac{Y}{20}$

ג. $\frac{X+Y}{60}$

ד. $\frac{2X+Y}{80}$

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

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

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

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

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

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

μ^2 .

10. $X_i \sim N(\mu, 1)$ כאשר $i = 1, 2, \dots, n$

נתון שהתצפיות הינן בלתי תלויות זו בזו. מצא אומד חסר הטיה ל- μ^2 .

11. נתונות n תצפיות בלתי תלויות מתוך התפלגות בעלת הצפיפות הבאה :

$$f(x) = \begin{cases} \frac{1+\beta x}{2} & -1 < x < 1 \\ 0 & \text{אחר} \end{cases}$$

א. הראה כי האומד $3\bar{X}$ הנו אומד בלתי מוטה ל β .

ב. מצא את השונות של האומד מהסעיף הקודם.

12. $X_1, X_2, \dots, X_n$ הינם משתנים מקריים רציפים בלתי תלויים בעל פונקצית הצפיפות הבאה :

$$f(x) = \begin{cases} X \cdot A & 0 \leq x \leq \theta \\ 0 & \text{אחר} \end{cases}$$

א. בטא את ערכו של A באמצעות θ כדי שפונקצית הצפיפות תהיה לגיטימית.

ב. מצא אומד חסר הטיה ל- θ על סמך n התצפיות.

פתרונות:**שאלה 1**

א. T_2 ו- T_1

ב. $\frac{2}{3}T_3$

ג. $T_2 = 110 T_1 = 76.6$

ד. T_1

שאלה 2

ב. T_2

שאלה 3

א. $\frac{x}{n}$

ב. $1 - \frac{x}{n}$

ג. x

שאלה 4

א. $\frac{3x}{2}$

ב. $\frac{3\bar{x}}{2}$

שאלה 5

א. $1 - \frac{x}{2}$

ג. 0.125

ה. לשונות 0.917

שאלה 6

א. נכון.

ב. לא נכון.

שאלה 7

תשובה : ב

שאלה 8

א. $T_1 \cdot T_2$

ב. $T_1 - T_1 \cdot T_2$

שאלה 9

הוכחה

שאלה 10

$$\overline{X}^2 - \frac{1}{n}$$

שאלה 11

ב. $V(3\overline{X}) = \frac{3 - \beta^2}{n}$

שאלה 12

א. $A = \frac{2}{\theta^2}$

ב. $\theta = \frac{3}{2} \overline{X}$

אומד נראות מקסימלית

רקע

להלן נלמד את שיטת הנראות המקסימלית למציאת אומדים.

נניח ש X משתנה מקרי בדיד עם פונקצית הסתברות $P(x, \theta)$, כאשר θ הפרמטר הבלתי ידוע.

יהי $X_1, X_2, \dots, X_n$ תוצאות מדגם מקרי בגודל n הנלקח מאוכלוסייה זו.

נבנה את פונקצית ההסתברות המשותפת (פונקצית הדגימה).

אם אנו יודעים את תוצאות המדגם ולא את הפרמטר קוראים לפונקציית הנראות שהיא פונקציה של הפרמטר.

נגדיר את פונקציית הנראות :

$$L(\theta) = P(x_1, \theta) \cdot P(x_2, \theta) \cdot \dots \cdot P(x_n, \theta) = \prod_{i=1}^n P(x_i, \theta)$$

פונקצית הנראות היא ההסתברות לקבל את התצפית הראשונה (כפונקציה של θ) כפול ההסתברות לקבל את התצפית השנייה, וכולי, כלומר המשמעות של פונקציית הנראות היא ההסתברות לקבל את המדגם שהתקבל, כפונקציה של הפרמטר המבוקש θ .

אם מדובר במשתנה רציף נכפיל את פונקציות הצפיפות ולא את פונקציות ההסתברות:

$$L(\theta) = f(x_1, \theta) \cdot f(x_2, \theta) \cdot \dots \cdot f(x_n, \theta) = \prod_{i=1}^n f(x_i, \theta)$$

דוגמה : (פתרון בהקלטה)

הסיכוי של שחקן כדורסל לקלוע לסל הוא p (לא ידוע). השחקן זורק כדורים לסל עד שהוא קולע בפעם הראשונה. נניח כי הזריקות בלתי תלויות זו בזו. הכדור נכנס לסל לראשונה בניסיון השלישי. השחקן חוזר על התהליך שוב והפעם הכדור נכנס לסל בניסיון החמישי.

מצאו את פונקצית הנראות של p .

אומד נראות מקסימלית עבור θ הוא האומד $\hat{\theta}$ שממקסם את פונקציית הנראות $L(\theta)$, כלומר, אנו מחפשים את האומד שיגרום לכך שהמדגם המקרי שקבלנו יהיה כמה שיותר סביר.

שלבים למציאת אומד נראות מקסימלית:

- לוקחים את פונקציית ההסתברות המשותפת של המדגם (או צפיפות משותפת אם המשתנה רציף).
- מציבים את תוצאות המדגם ומקבלים את פונקציית הנראות (פונקציה של הפרמטר הנחקר).
- מוצאים מקסימום לפונקציית הנראות (לעיתים כדאי להוסיף $\ln$ כדי להקל על המלאכה).

המשך דוגמה:

חשבו את אומדן הנראות המקסימלית עבור p .

משפט: אם $\hat{\theta}$ הוא אומד נראות מקסימלית עבור θ , אזי $g(\hat{\theta})$ הוא אומד נראות מקסימלית עבור $g(\theta)$ בהנחה והפונקציה היא חד-חד ערכית (אינווריאנטיות).

המשך דוגמה:

מצאו את אומדן נראות מקסימלית לסיכוי של שחקן הכדורסל לקלוע לסל פעמיים ברצף.

תרגילים:

1. הסיכוי של שחקן לנצח במשחק הוא p (לא ידוע). השחקן משחק במשחק עד אשר הוא מנצח בפעם הראשונה. נתון שהשחקן ניצח לראשונה רק במשחק השני.
 - א. חשבו את פונקציית הנראות של p וציירו גרף שלה.
 - ב. מצאו אומדן נראות מקסימלית עבור p .
 - ג. מצאו אומדן נראות מקסימלית ל- p אם ביום אחד הוא נאלץ לשחק 4 פעמים וביום אחר הוא נאלץ לשחק 5 פעמים עד אשר ניצח.

2. מספר הלקוחות שנכנסים לחנות מסוימת, מתפלג פואסונית עם תוחלת של λ לקוחות ביום.
 - א. מצאו אומדן נראות מקסימלית ל- λ על סמך מספר הלקוחות שנכנסים ביום מסוים.
 - ב. מצאו אומדן נראות מקסימלית ל- λ על סמך מספר הלקוחות שנכנסים ב- n ימים מסוימים.

3. הזמן שלוקח לאדם לחכות בתור מתפלג מעריכית עם פרמטר λ . דגמו 4 אנשים מקריים שחיכו בתור ומדדו את זמני ההמתנה שלהם. התוצאות שהתקבלו בדקות הן: 3, 7, 5 ו-3.
 - א. פתחו אומדן נראות מקסימלית לפרמטר זה על סמך n תצפיות כלשהן.
 - ב. מהו האומדן לפרמטר?

4. משך זמן הכנת שיעורי הבית (בשעות) של בני נוער ביום אחד מתפלג אחיד $U(0, \theta)$. כדי לאמוד את θ , נשאלו ביום מסוים מספר בני נוער כמה שעות הם הכינו שיעורי בית באותו יום.
 - א. אלעד הכין ביום מסוים שיעורי בית במשך שעה שלמה. חשבו את פונקציית הנראות של θ המתבססת על תצפית זו, וציירו את הגרף שלה.
 - ב. מצאו אומדן נראות מקסימלית ל- θ על סמך התצפית.
 - ג. משכי הכנת שיעורי בית (שעות) של 3 בני נוער היו 1.5, 3, 1. מצאו אומדן נראות מקסימלית ל- θ על סמך המדגם הזה.
 - ד. מצאו באופן כללי אומדן נראות מקסימלית ל- θ על סמך מדגם של n בני נוער - $X_1, \dots, X_n$.

5. הגובה של אוכלוסייה מסוימת מתפלג נורמאלית עם תוחלת ידועה של 170 ס"מ ושונות σ^2 לא ידועה.

א. מצאו אומד נראות מקסימלית עבור השונות על סמך מדגם $X_1, \dots, X_n$ מ תצפיות מהאוכלוסייה.

ב. נדגמו 5 אנשים בלתי תלויים בעלי הגבהים: 170, 182, 174, 165, 174. מהו האומדן לשונות הגבהים באוכלוסייה?

6. פתחו אומד נראות מקסימלית לפרמטר P בהתפלגות הבינומית על סמך מדגם בגודל n בו X הוא מספר ההצלחות במדגם.

7. X הוא משתנה מקרי בעל פונקציית הצפיפות:

$$f(x) = \begin{cases} 2\theta x e^{-\theta x^2}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

א. מצאו אומד נראות מקסימלית ל- θ על סמך n תצפיות בלתי תלויות: $X_1, \dots, X_n$.

ב. מצאו אומד נראות מקסימלית ל- θ^2 .

8. בכד א 10 כדורים שחורים ו 10 לבנים בכד ב 5 כדורים שחורים ו- 15 לבנים. דוגמים באקראי כדור אך אינך יודע מאיזה כד.

א. מצא אומד נראות מקסימלית לכד שממנו הוצא הכדור על סמך הצבע של הכדור.

ב. מהו האומדן אם הצבע הוא שחור?

9. הזמן שלוקח ליוסי לפתור תשבץ מתפלג מעריכית עם תוחלת לא ידועה. נתנו ליוסי לפתור חמישה תשבצים ובממוצע לקח לו 32 דקות לפתור אותם.

א. מה אומדן הנראות המקסימלית לתוחלת זמן הפתרון של תשבץ על ידי יוסי (אין חובה לפתח).

ב. מה אומדן הנראות המקסימלית לסיכוי שייקח לו לפחות חצי שעה לפתור את התשבץ הבא?

10. מספר הלקוחות הממתינים בתור במוקד טלפוני הוא משתנה מיקרי X בעל התפלגות התלויה בפרמטר θ באופן הבא:

2	1	0	X
$1 - 4\theta + 4\theta^2$	$4\theta - 8\theta^2$	$4\theta^2$	$P(X)$

- בחישה זמנים שונים שנבחרו באקראי נמצאו: 0, 1, 0, 0, 0 לקוחות ממתינים בתור.
 א. מצאו אומדן בשיטת הנראות המקסימלית עבור הפרמטר θ על-סמך המדגם הנתון.
 ב. מצאו אומדן בשיטת הנראות המקסימלית לסיכוי שלא יהיו לקוחות בתור.

11. אדם מחזיק בידו שני מטבעות: מטבע הוגן ומטבע שאינו הוגן שהסיכוי בו לתוצאה עץ הוא 0.2. האדם מטיל את אחד המטבעות פעמיים ומודיע לך כמה פעמים הוא קיבל עץ. אתה צריך לנחש איזה מטבע הוא הטיל: את ההוגן או זה שאינו הוגן.

- א. מצא אומדן בשיטת הנראות המקסימלית לסוג המטבע שהוטל.
 ב. מהו האומדן אם האדם קיבל פעמיים עץ?

12. מעוניינים לאמוד את אחוז המובטלים באוכלוסייה. דוגמים 50 אנשים אקראיים ומתקבל ש 4 מהם מובטלים.

- א. מצא אומדן נראות מקסימלית לשיעור המובטלים באוכלוסייה.
 ב. מצא אומדן לשיעור העובדים באוכלוסייה.
 ג. מצא אומדן ליחס בין שיעור העובדים לשיעור המובטלים באוכלוסייה.

13. במשחק מחשב שלוש רמות משחק :

ברמה 1 הסיכוי של יוסי לסיים את המשחק הוא 0.9.

ברמה 2 הסיכוי של יוסי לסיים את המשחק הוא 0.7.

ברמה 3 הסיכוי של יוסי לסיים את המשחק הוא 0.4.

יוסי בחר ברמה מסוימת אך אינו יודע איזו רמה הוא בחר. הוא משחק במשחק ברמה שבחר פעמיים.

א. הציעו א.נ.מ. לרמה של המשחק שיוסי שיחק על סמך מספר הפעמים שסיים את משחק.

ב. אם יוסי סיים את שני המשחקים, מה יהיה האומדן לרמה?

ג. מהו א.נ.מ. לסיכוי שמתוך שני משחקים הוא יצליח בדיוק משחק אחד?

14. $X_1, X_2, \dots, X_n$ מתפלגים אחיד בקטע $[-\theta, \theta]$ מצא אומד נראות מקסימלית לפרמטר θ .

15. $X_1, X_2, \dots, X_n$ מתפלגים בדיד לפי פונקציה ההסתברות הבאה :

$$P(X = k) = \frac{\binom{2}{k} \cdot P^k \cdot (1-P)^{2-k}}{1 - (1-P)^2} \quad K = 1, 2$$

הוכח שא.נ.מ. ל- P הינו: $2 - \frac{2}{\bar{X}}$

16. במכשיר חשמלי יש 2 סוללות שפועלות באופן ב"ת זו בזו והוא מפסיק לפעול ברגע שאחת

הסוללות מפסיקה לעבוד. הסיכוי של סוללה לתפקד לפחות חודש הוא P . כאשר המכשיר

מפסיק לפעול מחליפים את שתי הסוללות שלו. בתחילת הניסוי נלקחו 80 מכשירים

כאלה עם סוללות חדשות ולאחר חודש נמצא של 30 מהם עדיין פועלים.

א. מצא אומדן נראות מקסימלית עבור P .

ב. רשמו את האומדן שבו השתמשותם בחלק א' באופן כללי, עבור מדגם של n מכשירים

שמתוכם נמצאו Y מכשירים שעדיין פועלים לאחר חודש אחד.

ג. בהנחה שאורך החיים (בחודשים) של סוללה בודדת הוא מעריכי עם פי צפיפות

$$f(t) = \theta e^{-\theta t} \text{ עבור } t > 0.$$

מצא א.נ.מ. עבור θ המבוסס על Y . מהו האומדן המתאים מן המדגם הנתון?

17. חיוג אוטומטי של מכשיר טלפון משדר אות אחת לשתי דקות. אם לאחר 20 דקות (10

אותות חיוג) המספר שאליו מטלפנים עדיין תפוס-החיוג האוטומטי נפסק.

א. רשמו את פונקציית ההסתברות של המשתנה X - מספר הפעמים שהחייג האוטומטי

מחייג למספר הטלפון המבוקש, אם ההסתברות לקבלת צליל "פנוי" בשידור אחד של

אות חיוג הוא P .

ב. מתוך 12 ניסיונות חיוג אוטומטי למשרד הרישוי בזמנים שונים במשך 5 ימים, התקבלו

התוצאות הבאות : בשני ניסיונות הופסק החיוג האוטומטי ובשאר הניסיונות שבהם

הצליח המטלפן להשיג את המספר המבוקש, מספר החיוגים האוטומטיים עד לקבל צליל

"פנוי" היו :

5,1,2,2,8,3,7,2,6,1

מצאו אומדן נראות מקסימלית עבור P על סמך התוצאות שהתקבלו.

פתרונות :**שאלה 1**

ב. 0.5

ג. $\frac{2}{9}$ **שאלה 2**א. X ב. $\bar{X}$ **שאלה 3**א. $\frac{1}{\bar{X}}$ ב. $\frac{2}{9}$ **שאלה 4**

א. 1

ג. 3

ד. $X_{\max}$ **שאלה 5**א. $\frac{\sum_{i=1}^n (x_i - 170)^2}{n}$

ב. 40.2

שאלה 6 $\frac{x}{n}$ **שאלה 7**א. $\frac{n}{\sum X_i^2}$ ב. $(\frac{n}{\sum X_i^2})^2$

שאלה 8

ב. כד א

שאלה 9

א. 32

ב. 0.3916

שאלה 10

א. 0.45

ב. 0.81

שאלה 11

ב. הוגן

שאלה 12

א. 0.08

ב. 0.92

ג. 11.5

שאלה 13

$$\hat{\theta} = \begin{cases} 3 & X = 0,1 \\ 1 & X = 2 \end{cases} \quad \text{א.}$$

ב. 1

$$\hat{p} = \begin{cases} 2 \cdot 0.4 \cdot 0.6 & X = 0,1 \\ 2 \cdot 0.9 \cdot 0.1 & X = 2 \end{cases} \quad \text{ג.}$$

שאלה 14 $\max |X_i|$ **שאלה 15**

הוכחה

שאלה 16

א. 0.6124

$$\hat{p} = \sqrt{\frac{y}{n}} \quad \text{ב.}$$

ג. 0.49

שאלה 17

ב. 0.1818

נספח

התפלגויות רציפות

ההתפלגות	פונקציית הצפיפות	פונקציית ההתפלגות המצטברת	תוחלת	שונות	הערות	אנ"מ
$X \sim U(a, b)$	$f(x) = \frac{1}{b-a}$ $a \leq x \leq b$	$\frac{t-a}{b-a}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$		$b = \max(X_i)$ $a = \min(X_i)$
$X \sim \exp(\lambda)$	$f(x) = \lambda e^{-\lambda x}$	$1 - e^{-\lambda t}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	הזמן עד להתרחשות מאורע מסוים. λ - הוא ממוצע האירועים ביחידת זמן.	$\hat{\lambda} = \frac{1}{\bar{X}}$
$X \sim N(\mu, \sigma^2)$	$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$\Phi(t)$	μ	σ^2	$Z = \frac{X - \mu}{\sigma}$	$\bar{\mu} = \bar{X}$ $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$

התפלגויות בדידות

ההתפלגות	פונקציית ההסתברות	תוחלת	שונות	הערות	אנ"מ
בינומית	$P(X=k)$ $\binom{n}{k} p^k (1-p)^{n-k}$ $k = 0, 1, \dots, n$ $B(n, p)$ $0 \leq p \leq 1$	np	$np(1-p)$	מספר ההצלחות ב- n ניסויי ברנולי ב"ת. p - ההסתברות להצלחה	$\hat{p} = \frac{Y}{n}$
גיאומטרית	$(1-p)^{k-1} p$ $k = 1, 2, \dots, \infty$ $G(p)$ $0 < p \leq 1$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	מספר הניסויים עד להצלחה הראשונה בסדרת ניסויי ברנולי ב"ת, p - ההסתברות להצלחה	$\hat{p} = \frac{1}{\bar{X}}$
אחידה	$\frac{1}{b-a+1}$ $K=a, \dots, b$ $U(a, b)$	$\frac{a+b}{2}$	$\frac{(b-a+1)^2 - 1}{12}$	בחירה אקראית של מספר בין a ו- b .	$b = \max(X_i)$ $a = \min(X_i)$
פואסונית	$\frac{e^{-\lambda} \lambda^k}{k!}$ $k = 0, 1, \dots, \infty$ $P(\lambda)$ $\lambda > 0$	λ	λ	מספר אירועים ביחידת זמן λ - קצב האירועים	$\hat{\lambda} = \bar{X}$

קריטריון MSE - תוחלת ריבוע הטעות

רקע

הקריטריון הנפוץ ביותר כדי לבדוק את טיב האומד הוא קריטריון MSE. תוחלת ריבוע טעות האמידה.

$$MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2 = V(\hat{\theta}) + (E(\hat{\theta}) - \theta)^2$$

$V(\hat{\theta})$ - הינה שונות האומד.

$E(\hat{\theta}) - \theta$ - הינה ההטיה של האומד.

אם T_1 ו- T_2 הינם אומדים לפרמטר θ . האומד העדיף יהיה זה עם MSE קטן יותר כלומר, אם

$$MSE(T_1) > MSE(T_2) \text{ עדיף על } T_1.$$

דוגמה: (הפתרון בהקלטה)

נתון משתנה X המתפלג אחיד רציף באופן הבא: $X \sim U(3, \theta)$. מוצעים שני אומדים לפרמטר θ

$$T_1 = 2X - 3 \text{ ו- } T_2 = \frac{3X - 3}{2}$$

איזה אומד עדיף לאמידת הפרמטר θ ?

תרגילים:

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

כתב ופתר - ברק קנדל ©

1. מעוניינים לאמוד את התוחלת של התפלגות מסוימת. מוצעים שני אומדים אפשריים ממוצע של שתי תצפיות וממוצע של שלוש תצפיות. לפי קריטריון תוחלת ריבוע הטעות (MSE) איזה אומד עדיף? הסבירו.

2. בעיר מסוימת בשוויץ בכל θ דקות רכבת מגיעה לתחנה מסוימת. דוד מגיע לתחנה בזמן אקראי ומודד את זמן ההמתנה לרכבת X .

א. הצע אומד חסר הטיה ל- θ על סמך X .

ב. סטטיסטיקאי הציע לאמוד את θ על סמך האומד: $1.5X$ האם האומד הנ"ל מוטה?

ג. איזה אומד מבין האומדים של סעיף א או ב עדיף?

3. חוקר מעוניין לאמוד את הסיכוי לחלות במחלת השפעת בחורף (להלן הפרמטר P). הוא דוגם חמישה אנשים בריאים ומתבונן בסטטיסטי X מספר האנשים שחלו בשפעת בחורף. הוא מתלבט

$$\text{בין שני אומדים: } T_1 = \frac{X}{5} \text{ ו- } T_2 = \frac{X+1}{7}$$

א. מי מבין האומדים הללו הוא חסר הטיה?

ב. מי מבין האומדים עדיף אם $P=0.5$?

ג. מי מבין האומדים עדיף אם $P=0.1$?

4. מספר השריפות המתרחשות בחודש אוקטובר בארץ מתפלג פואסונית עם תוחלת λ . נלקח מדגם של 10 חודשי אוקטובר. להלן שני אומדים אפשריים:

$$\hat{\lambda}_1 = \frac{\sum_{i=1}^{10} X_i}{10}$$

$$\hat{\lambda}_2 = \frac{2 \cdot \sum_{i=1}^5 X_i + 2 \cdot \sum_{i=6}^{10} X_i}{10}$$

$$X_i = \text{מספר השריפות בחודש אוקטובר ה- } i.$$

איזה מהאומדים עדיף לצורך אמידת הפרמטר λ ?

$$5. \text{ הוכח ש: } E(\hat{\theta} - \theta)^2 = V(\hat{\theta}) + (E(\hat{\theta}) - \theta)^2$$

פתרונות:

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

כתב ופתר - ברק קנדל ©

שאלה 1

זה עם השלוש תצפיות.

שאלה 2

א. $2x$

ג. סעיף ב

שאלה 3

א. T_1

ב. T_2

ג. T_1

שאלה 5

הוכחה

שאלות מסכמות באמידה נקודתית

תרגילים:

1. במפעל מייצרים מוצרים בשלוש מכונות שונות ובלתי תלויות. במכונה הראשונה הסיכוי שמוצר יהיה תקין הוא P , במכונה השנייה ההסתברות שמוצר יהיה תקין הוא P^2 ובמכונה השלישית הסיכוי הוא $2P$. דוגמים 20 מוצרים מכל מכונה. נסמן ב- X את מספר המוצרים התקינים שיוצרו במכונה א'. נסמן ב- Y את מספר המוצרים התקינים שיוצרו במכונה השנייה וב- Z את מספר המוצרים התקינים שיוצרו במכונה השלישית.
 - א. מהם הערכים האפשריים של הפרמטר P ?
 - ב. מצאו אומד בלתי מוטה עבור הפרמטר P על סמך X ו- Z .
 - ג. אם התקבל ש- $X=6, Y=3$ מהו אומדן נראות מקסימלית ל- P ?
2. מספר תאונות הדרכים בקטע כביש א' מתפלג פואסונית עם קצב של λ תאונות בחודש. מספר תאונות הדרכים בקטע כביש ב' מתפלג פואסונית עם קצב של 2λ תאונות בחודש. הוחלט לספור את מספר התאונות בחודש בכל אחד מקטעי הכביש.
 - א. מצאו אומד נראות מקסימלית לפרמטר λ על סמך X ו- Y .
 - ב. מצאו אומד נראות מקסימלית לסיכוי שבקטע כביש א' תהיה לפחות תאונה אחת בחודש?
 - ג. האם האומד שמצאת בסעיף א' הוא חסר הטיה ל- λ ?
3. זמן הייצור של מוצר מסוים בתהליך ייצור מתפלג נורמאלי עם תוחלת ושונות שאינן ידועות.
 - א. הציעו אומדים חסרי הטיה לתוחלת והשונות של זמן הייצור של המוצר.
 - ב. הציעו אומדי נראות מקסימלית לתוחלת ולשונות של זמן הייצור של המוצר.
 - ג. הציעו אומד נראות מקסימלית לריבוע התוחלת של זמן הייצור.
 - ד. האם האומד מהסעיף הקודם הוא גם חסר הטיה?

4. בקזינו משחק בו 4 תאים ממוספרים מ 1 עד 4. מפעיל המשחק שם כסף באחד מארבעת התאים והאדם המשתתף צריך לנחש באיזה תא הכסף מוחבא. מפעיל הקזינו מודיע שהסיכוי להחביא את הכסף בכל אחד משלושת התאים הראשונים שווה אך לא בהכרח שווה לסיכוי להחביא אותו בתא הרביעי.

יש לאמוד את הסיכוי להחביא את הכסף בתא הראשון : P .
א. מצא את תחום ההגדרה של הפרמטר P .

יעל שיחק את המשחק 3 פעמים וקיבלה שפעם אחת הכסף הוחבא בתא מספר 1 ובפעמים האחרות בתא מספר 2.

ב. מצאו אומדן ל- P על סמך התוצאות הללו בשיטת הנראות המקסימלית.

ג. מצאו אומדן חסר הטיה ל- P מהו האומדן לפי התוצאות של יעל?

ד. מצאו אומדן חסר הטיה ונראות מקסימלית לסיכוי שהכסף יוחבא בתא מספר 4 על סמך התוצאות של יעל.

5. יהי $X_1, X_2, \dots, X_n$ מדגם מקרי מתוך ההתפלגות הבאה :

$$f(x) = \begin{cases} \frac{\theta}{\lambda} \left(\frac{x}{\lambda}\right)^{\theta-1}, & 0 < x < \lambda, \theta > 0 \\ 0, & \text{else} \end{cases}$$

א. מצא אחי"ה ל- λ (כאשר θ קבוע ידוע).

ב. מצא אחי"מ ל- θ (כאשר λ קבוע ידוע).

ג. מצא אחי"מ ל- λ (כאשר θ קבוע ידוע).

6. X -משך זמן הפרסומות בערוץ 2 מתפלג אחיד רציף בתחום $(0, \theta)$

Y -משך זמן הפרסומות בערוץ 10 מתפלג אחיד רציף בתחום $(0, 2\theta)$

א. מצא אומדן חסר הטיה ל- θ המשתמש במשך זמן אקראי של פרסומת בודדת בערוץ 2 ופרסומת בודדת בערוץ 10.

ב. מוצע האומדן $T_2 = X + 0.5Y$, האם האומדן הנ"ל הוא חסר הטיה?

ג. איזה אומדן יותר עדיף זה של סעיף א או זה של סעיף ב?

ד. מצא אומדן נראות מקסימלית ל- θ על סמך X -ו- Y .

פתרונות:שאלה 1:

א. $0 \leq P \leq 0.5$

ג. 0.345

שאלה 2:

א. $\frac{x+y}{3}$

ג. כן

שאלה 4:

א. $0 \leq P \leq \frac{1}{3}$

ב. $\frac{1}{3}$

ג. 0.389

שאלה 5:

א. אח"ה יהיה $\hat{\lambda} = \frac{\theta+1}{\theta} \bar{x}$

ב.

$$\hat{\theta} = \frac{n}{n \ln \lambda - \sum_{i=1}^n \ln x_i}$$

ג. $\hat{\lambda} = X_{\max}$

נספח : אומדי נראות מכסימלית ואומדים חסרי הטיה בהתפלגויות השונות

מודל בינומי

נתון מדגם של משתנה בינומי $X \sim B(n, p)$.

א.נ.מ עבור p הוא $\hat{p} = \frac{X}{n}$ והוא גם א.ח.ה.

מודל אחיד (בדיד)

נתון מדגם $X_1, X_2, \dots, X_n$ של משתנים אחידים $X_i \sim U(1, N)$ בלתי-תלויים בזוגות.

א.נ.מ עבור N הוא $\hat{N} = \max\{X_1, \dots, X_n\}$ ואינו א.ח.ה.

מודל פואסוני

נתון מדגם $X_1, X_2, \dots, X_n$ של משתנים פואסוניים $X_i \sim P(\lambda)$ בלתי-תלויים בזוגות.

א.נ.מ עבור λ הוא $\hat{\lambda} = \bar{X}$ וגם א.ח.ה.

מודל גיאומטרי

נתון מדגם $X_1, X_2, \dots, X_n$ של משתנים גיאומטריים $X_i \sim G(p)$ בלתי-תלויים בזוגות.

א.נ.מ עבור p הוא $\hat{p} = \frac{1}{\bar{X}}$ אינו א.ח.ה. וא.נ.מ עבור התוחלת $\frac{1}{p}$ הוא $\bar{X}$ והנו א.ח.ה.

מודל נורמלי

נתון מדגם $X_1, X_2, \dots, X_n$ של משתנים נורמליים $X_i \sim N(\mu, \sigma^2)$ בלתי-תלויים בזוגות.

א.נ.מ עבור μ הוא $\hat{\mu} = \bar{X}$

כאשר μ ידוע א.נ.מ עבור σ^2 הוא $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$ (אומד חסר-הטייה)

כאשר μ לא-ידוע א.נ.מ עבור σ^2 הוא $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ (אומד מוטה!!!)

אומד חסר-הטיה עבור σ^2 :

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 \text{ כאשר } \mu \text{ ידוע}$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \text{ כאשר } \mu \text{ לא-ידוע}$$

מודל מעריכי

נתון מדגם $X_1, X_2, \dots, X_n$ של משתנים מעריכיים $X_i \sim \exp(\theta)$ בלתי-תלויים בזוגות.

א.נ.מ עבור θ הוא $\hat{\theta} = \frac{1}{\bar{X}}$ -מהווה אומד מוטה. וא.נ.מ עבור התוחלת $\frac{1}{\theta}$ הוא $\bar{X}$ א.נ.ה.

מודל אחיד (רציף)

נתון מדגם $X_1, X_2, \dots, X_n$ של משתנים אחידים $X_i \sim U(0, \theta)$ בלתי-תלויים בזוגות.

א.נ.מ עבור θ הוא $\hat{\theta} = \max\{X_1, \dots, X_n\}$ אינו א.נ.ה

בכל התפלגות:

א.ח.ה עבור μ הוא $\hat{\mu} = \bar{X}$

אומד חסר-הטיה עבור σ^2 :

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 \quad \text{כאשר } \mu \text{ ידוע}$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad \text{כאשר } \mu \text{ לא-ידוע}$$

פרק 4 - סטטיסטיקה תיאורית - הקדמה

רקע:

בסטטיסטיקה תיאורית אנו חוקרים קבוצה מסוימת. הקבוצה יכולה להיות קבוצת ילדים בגן, קבוצת מניות בתיק, כלל התושבים בעיר מסוימת וכולי. בין ישות לישות בקבוצה ישנם גורמים היכולים לקבל מספר ערכים. גורמים אלה נקראים משתנים. למשל, בין מניה למניה בתיק משתנה התשואה היומית של המניה, הוותק של המניה, תחום המניה וכדומה.

בסטטיסטיקה תיאורית אנחנו נתבונן בקבוצה מסוימת ובתוך הקבוצה הזו נאסוף נתונים לגבי משתנה מסוים ונלמד להציג את הנתונים ולנתח אותם מכל מיני אספקטים.

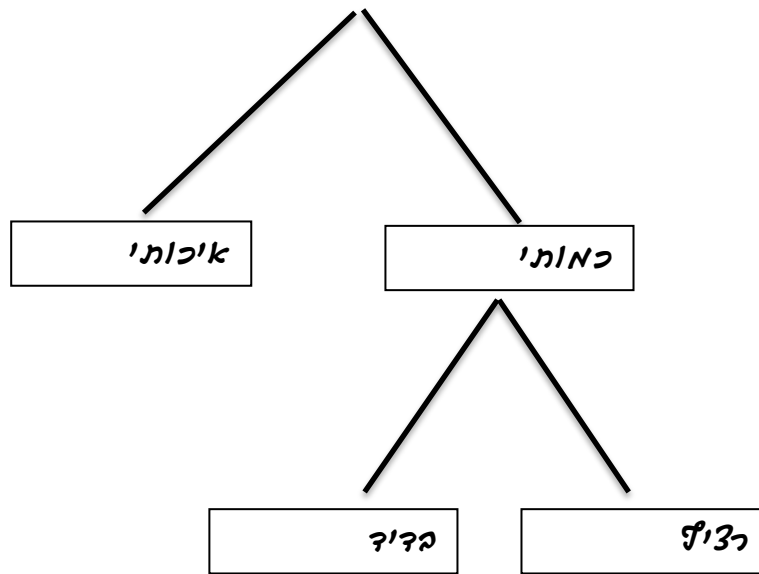
דוגמה:

בתיק מניות 10 מניות. מנהל התיק פרסם את התשואה של כל מניה בשנת 2011.

מי הקבוצה הנחקרת?

מה גודל הקבוצה?

מה המשתנה הנחקר?

סוגי משתנים:

משתנה איכותי הוא משתנה שלערכיו אין משמעות של יותר או פחות, אין עניין כמותי לערכים המתקבלים.

כמו : מקום מגורים של אדם (רעננה, תל אביב, אשדוד..)

מין האדם (זכר, נקבה)

מצב משפחתי (רווק, נשוי, גרוש, אלמן)

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

את המשתנה הכמותי נסווג לשני סוגים :

משתנה בדיד : משתנה שערכיו מתקבלים מתוך סידרה של ערכים אפשריים. כמו : מספר ילדים למשפחה (1,2,3..)

ציון בבחינה (מ 0 ועד 100 בקפיצות של 1)

הערה :

משתנה רציף : משתנה שערכיו מתקבלים מתוך אינסוף ערכים בתחום מסוים, הערכים מתקבלים ברצף וללא קפיצות של ערכים.

כמו : גובה בס"מ – אם למשל, הגובה הנמוך ביותר הוא 150 ועד 190 ס"מ בקבוצה הגבהים הם ברצף. גם בין 160 ל 161 ס"מ יש רצף אינסופי של ערכים אפשריים לגובה (160.33 ס"מ הוא גם גובה אפשרי)

משקל בק"ג, מהירות בקמ"ש וכולי.

תרגילים:

1. סווג את המשתנים הבאים לפי: איכותי / כמותי בדיד / כמותי רציף:
 - א. מספר הדירות בבניין.
 - ב. גיל אדם בשנים.
 - ג. אחוז האבטלה בעיר.
 - ד. מקצוע לימוד מועדף.

2. להלן התפלגות מספר האיחורים לעבודה בחודש של העובדים בחברת "סטאר".
בחברה 200 עובדים.

מספר העובדים	מספר האיחורים
17	0
23	1
85	2
50	3
25	4

- א. מהו המשתנה הנחקר כאן?
- ב. האם מדובר במשתנה איכותי או כמותי? אם הוא כמותי האם הוא בדיד או רציף?
3. לפניכם רשימה של משתנים כמותיים. ציין ליד כל משתנה אם הוא רציף או בדיד.
 - א. שכר עובד בש"ח.
 - ב. ציון בחינת בגרות.
 - ג. תוצאה בהטלת קובייה.
 - ד. מהירות ריצה בתחרות.
 - ה. שיעור התמיכה בממשלה.

פרק 5 - סטטיסטיקה תיאורית - הצגה של נתונים

רקע:

דרכים להצגת נתונים שנאספו :

א. רשימה של תצפיות:

התצפית היא הערך שנצפה עבור ישות מסוימת בקבוצה.
 רושמים את התצפיות שהתקבלו כרשומה, יעיל שיש מספר מועט של תצפיות.
 ההצגה הזו רלבנטית לכל סוגי המשתנים.
 למשל, להלן מספר החדרים בבניין בן 5 דירות :
 3 4 3 5 4

ב. טבלת שכיחויות בדידה :

שם המשתנה - X	שכיחות - $f(X)$	שכיחות יחסית באחוזים
X_1	f_1	$\frac{f_1}{N} \times 100$
X_2	f_2	$\frac{f_2}{N} \times 100$
X_3	f_3	$\frac{f_3}{N} \times 100$
$\vdots$	$\vdots$	$\vdots$
X_k	f_k	$\frac{f_k}{N} \times 100$
סה"כ	$N = \sum_{i=1}^k f_i$	100%

רושמים את התצפיות בטבלה שבה עמודה אחת מבטאת את ערכי המשתנה והשנייה את השכיחות. יעיל עבור משתנה איכותי וכמותי בדיד וכשיש מספר רב של תצפיות.
 לא יעיל למשתנה כמותי רציף.

למשל, להלן התפלגות הציונים בכיתה מסוימת:

הציון X	מספר התלמידים – השכיחות f	F_i	$\frac{f_i}{n}$
5	2	2	$0.08=2/25$
6	4	6	$0.16=4/25$
7	8	14	$0.32=8/25$
8	5	19	$0.2=5/25$
9	4	23	$0.16=4/25$
10	2	25	$0.08=2/25$

שכיחות מצטברת – צבירה של השכיחויות: F_i - השכיחות המצטברת נותנת כמה תצפיות קטנות או שוות לערך.

שכיחות יחסית (פרופורציה) – השכיחות מחולקת לכמות התצפיות הכללי: $\frac{f_i}{n}$ - איזה חלק מהתצפיות בקבוצה שוות לערך.

ג. טבלת שכיחויות במחלקות:

משתמשים שהמשתנה כמותי רציף או כאשר יש מספר ערכים רב במשתנה הבדיד וטבלת שכיחויות תהיה ארוכה מידי.
למשל, נתנו לקבוצת ילדים לבצע משימה מסוימת ובדקו את התפלגות זמן ביצוע המשימה בדקות.
להלן ההתפלגות שהתקבלה:

מספר הילדים	זמן בדקות
20	0.5-3.5
18	3.5-9.5
14	9.5-19.5
8	19.5-29.5

ד. דיאגרמת עוגה :

זהו התיאור הגרפי של משתנה איכותי. בדיאגרמת עוגה כל ערך במשתנה מקבל "נתח" יחסי מהעוגה. הנתח בעוגה פרופורציוני לשכיחות היחסית של ערך המשתנה בנתונים.

ה. דיאגרמת מקלות :

הציר האופקי הוא הציר של המשתנה הציר האנכי של השכיחות – הגובה של המקל מעיד על השכיחות .
רלבנטי למשתנה כמותי בדיד. לא נהוג להשתמש בתיאור למשתנה איכותי וכמו כן לא למשתנה כמותי רציף .
כמו כן בסולמות מדידה עבור משתנה מסולם סדר.

1. **היסטוגרמה:**

ההיסטוגרמה היא הדרך הגרפית כדי לתאר טבלת שכיחויות במחלקות.

רלבנטית למשתנה כמותי רציף.

בהיסטוגרמה ציר האופקי הוא הציר של המשתנה וציר האנכי הוא הציר של הצפיפות.

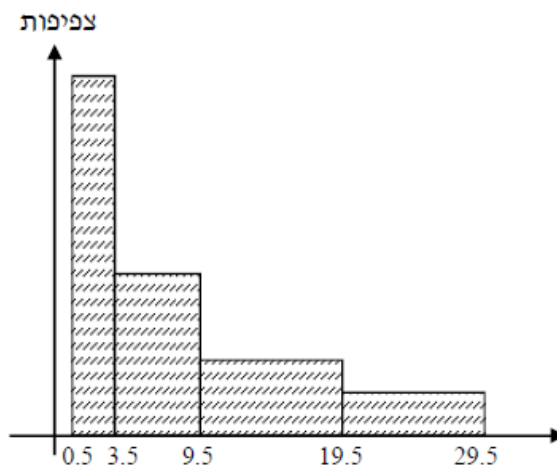
הצפיפות מחושבת בכל מחלקה על ידי חלוקת השכיחות ברוחב של כל המחלקה והיא נותנת

את מספר התצפיות הממוצע בכל מחלקה ליחידה.

אם המחלקות הן שוות ברוחב, ניתן לשרטט את ההיסטוגרמה לפי השכיחות ואין צורך

בצפיפות.

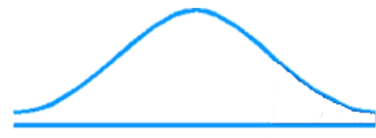
			X		
צפיפות	מצטברת	שכיחות	אמצע	רוחב	
6.6667	20	20	2	3	0.5 - 3.5
3	38	18	6.5	6	3.5 - 9.5
1.4	52	14	14.5	10	9.5 - 19.5
0.8	60	8	24.5	10	19.5 - 29.5



פוליגון-מצולעון: אם נחבר את אמצע קצה כל מלבן בקווים ישרים. נותן מראה חזותי לצורה של התפלגות המשתנה.

צורות התפלגות נפוצות

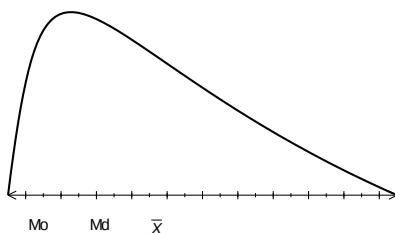
התפלגות סימטרית פעמונית- רוב התצפיות במרכז וככל שנתרחק מהמרכז יהיו פחות תצפיות באופן סימטרי. למשל, ציוני IQ.



ישנן התפלגויות סימטריות שאינן פעמוניות:

התפלגות אסימטרית ימנית (חיובית) – רוב התצפיות מקבלות ערכים נמוכים ויש מיעוט הולך וקטן של תצפיות שמקבלות ערכים גבוהים קיצוניים. למשל, שכר במשק.

התפלגות א-סימטרית ימנית או חיובית



התפלגות אסימטרית שמאלית (שלילית) רוב התצפיות מקבלות ערכים גבוהים ויש מיעוט הולך וקטן של תצפיות שמקבלות ערכים נמוכים קיצוניים. למשל, אורך חיים.

התפלגות א-סימטרית שמאלית או שלילית



תרגילים:

1. בסקר צפייה בטלוויזיה התקבלו התוצאות הבאות: 25 צפו בערוץ הראשון, 25 צפו בערוץ 10, 75 צפו בערוץ השני, 50 צפו באחד מערוצי הכבלים ו- 25 לא צפו בטלוויזיה בזמן הסקר.

א. רשמו את טבלת השכיחות ואת השכיחות היחסית.

ב. תארו את הנתונים באופן גרפי.

2. להלן נתונים על התפלגות המקצוע המועדף של תלמידי שכבה ו' בבית הספר "מעוף":

מספר התלמידים	המקצוע
44	מתמטיקה
20	תנ"ך
12	אנגלית
26	היסטוריה

א. מהו המשתנה הנחקר?

ב. מהי פרופורציית התלמידים שמעדיפים תנ"ך?

3. להלן התפלגות ההשכלה במקום עבודה מסוים:

מספר העובדים	השכלה
60	נמוכה
120	תיכונית
20	אקדמאית

א. מהו המשתנה הנחקר? מאיזה סולם הוא?

ב. תארו את הנתונים באופן גרפי.

4. להלן רשימת הציונים של 20 תלמידים שנבחנו במבחן הבנת הנקרא:

7, 6, 8, 9, 10, 6, 4, 5, 8, 7, 6, 7, 6, 8, 9, 6, 7, 8, 5, 6

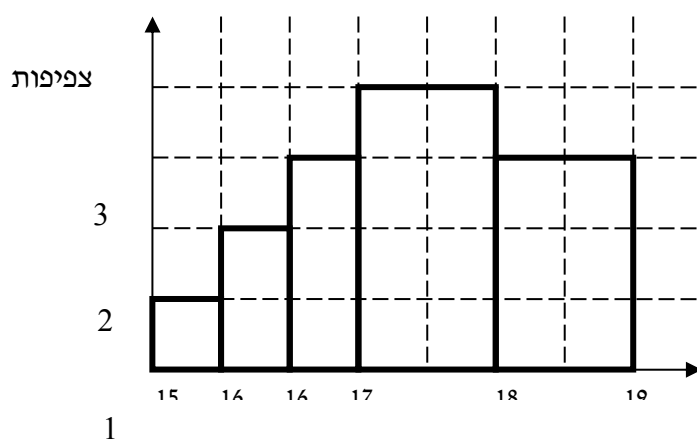
א. מהו המשתנה? האם הוא בדיד או רציף?

ב. תאר את הרשימה בטבלת שכיחות.

ג. הוסף שכיחות יחסית לטבלה.

ד. תאר את הנתונים באופן גרפי.

5. להלן היסטוגרמה המתארת את התפלגות הגבהים בס"מ של קבוצה מסוימת:



גובה

- מהו המשתנה הנחקר? האם הוא בדיד או רציף?
- תאר את הנתונים בטבלת שכיחויות במחלקות.
- הוסף שכיחות יחסית לטבלה.
- הוסף את הצפיפות של כל מחלקה לטבלה.
- מהי צורת ההתפלגות של הגבהים?

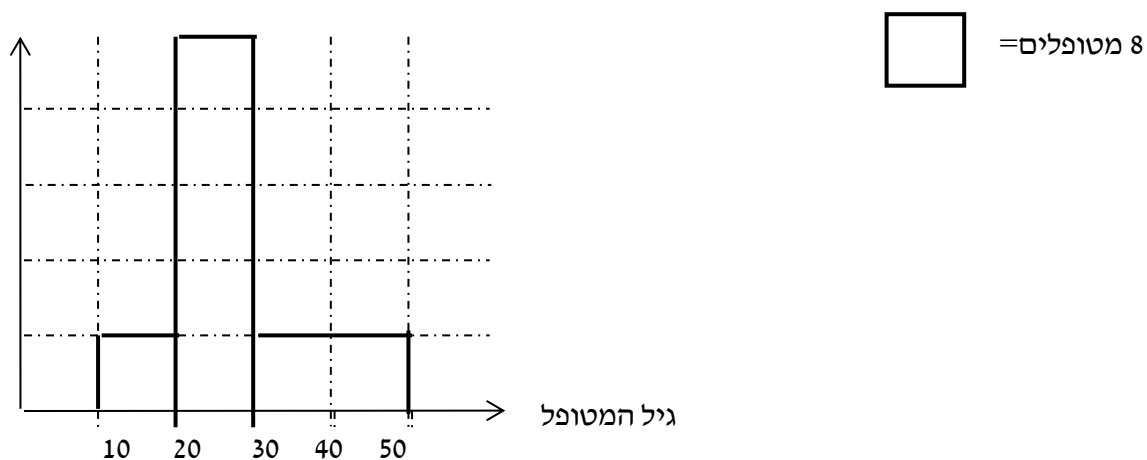
6. להלן התפלגות המשקל של קבוצה מסוימת בק"ג:

משקל	מספר מקרים
40-45	10
45-50	20
50-60	30
60-65	20
65-70	10

- תאר את ההתפלגות באופן גרפי.
- מה ניתן להגיד על צורת ההתפלגות?

7. להלן גיל המטופלים של ד"ר שוורץ בשנים :

קנה מידה :



א. מה המשתנה הנחקר? האם הוא בדיד או רציף?

ב. מהי הקבוצה הנחקרת?

ג. תרגמו את ההסיטוגרמה לטבלת שכיחות.

ד. מהי הפרופורציה של המטופלים של ד"ר שוורץ בגילאים 20-30?

פרק 6 - סטטיסטיקה תיאורית - מדדי מיקום מרכזי

רקע:

המטרה במדדי המיקום המרכזי למדוד את מרכז ההתפלגות של התצפיות.

השכיח – MODE

השכיח הוא הערך הנפוץ ביותר בהתפלגות.

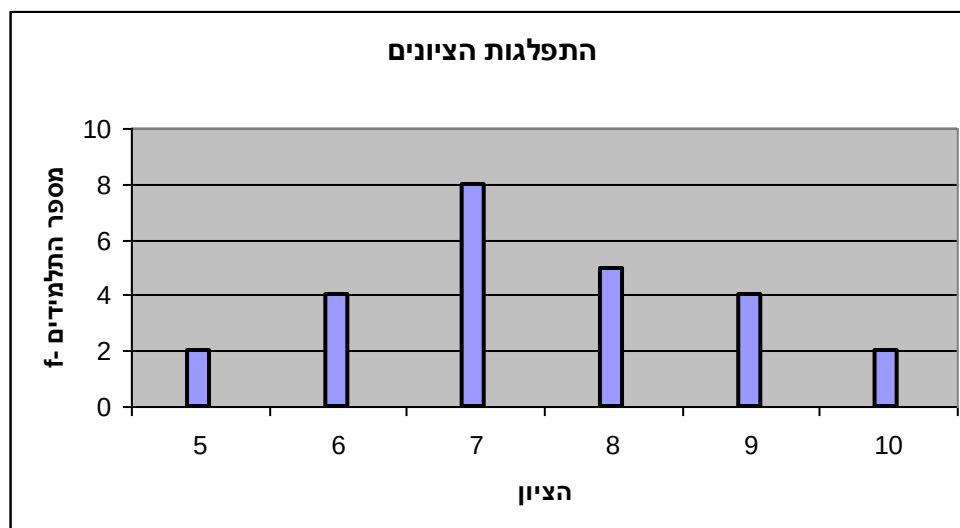
ברשימה : הערך החוזר על עצמו הכי הרבה פעמים .

7 9 4 8 4 10 6

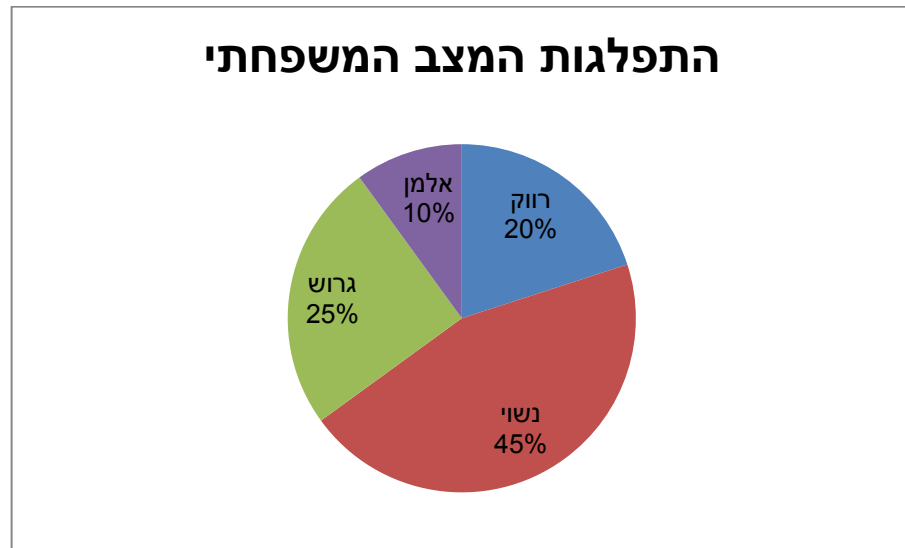
בטבלת שכיחויות בדידה : הערך שהשכיחות שלו היא הגבוהה ביותר.

# תכניות החיסכון	$f(x)$
0	100
1	75
2	25
3	25
4	25

בדיאגרמת מקלות : שיעור ה- X של המקל הגבוה ביותר.



בעוגה: הערך של הפלח הגדול ביותר.

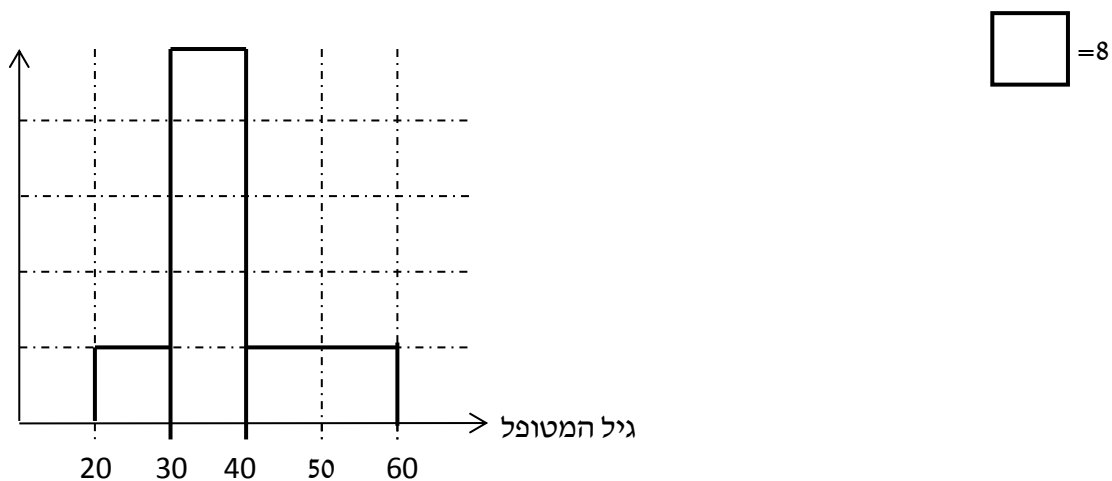


בטבלת שכיחויות במחלקות: אמצע המחלקה עם הצפיפות הגבוהה ביותר.
התפלגות הציונים בכיתה.

$f(x)$	X
20	0-60
10	60-70
18	70-80
15	80-90
15	90-100

בהיסטוגרמה: שיעור ה- X של אמצע המחלקה הגבוהה ביותר.

להלן גיל המטופלים של ד"ר שוורץ בשנים :



כללי: יתכן שלהתפלגות יותר משכיח אחד.

השכיח הוא מדד הרלבנטי לכל סוגי המשתנים.

MIDRANGE – (טווח) – אמצע תחום

הממוצע בין התצפית הגבוהה ביותר לתצפית הנמוכה ביותר.

$$MR = \frac{X_{\min} + X_{\max}}{2}$$

MEDIAN - החציון

החציון הוא ערך שמחצית מהתצפיות קטנות או שוות לו ומחצית מהתצפיות גדולות או שוות לו. ברשימה: נסדר את התצפיות בסדר עולה.

אם יש מספר אי זוגי של איברים מקומו של החציון יהיה התצפית שמיקומה: $\frac{n+1}{2}$

אם יש מספר זוגי של איברים החציון יהיה הממוצע של האיבר ה- $\frac{n}{2}$ והאיבר ה- $\frac{n}{2} + 1$

כלומר שיש מספר אי-זוגי של תצפיות החציון יהיה: $md = X_{\frac{n+1}{2}}$

ושיש מספר זוגי של תצפיות החציון יהיה: $md = \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2}$

בטבלת שכיחויות בדידה: נעשה תהליך דומה אך נעזר בשכיחות המצטברת.

דיאגרמת מקלות: נמיר לטבלת שכיחויות בדידה במטרה למצוא את החציון.

בטבלת שכיחויות במחלקות:

שלב א: נימצא את המחלקה החציונית שמיקומה יהיה $\frac{n}{2}$.

שלב ב: נציב בנוסחה הבאה: $Md = L_0 + \frac{\frac{n}{2} - F(x_{m-1})}{f(x_m)} \cdot (L_1 - L_0)$

$F(x_{m-1})$ - שכיחות מצטברת של מחלקה אחת לפני המחלקה החציונית.

$f(x_m)$ - השכיחות של המחלקה החציונית.

L_0 - גבול התחתון של המחלקה.

L_1 - גבול העליון של המחלקה.

היסטוגרמה: החציון הוא הערך על ציר ה- X שמחלק את ההיסטוגרמה לשני חלקים שווים בשטח.

כללי: החציון אינו רלבנטי למשתנה מסולם שמי ולא רלבנטי למשתנה איכותי.

הממוצע

הנו מרכז הכובד של ההתפלגות.

ברשימה: $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$

בטבלת שכיחויות: $\bar{x} = \frac{\sum x \cdot f}{n}$

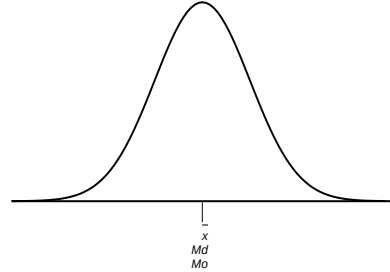
במחלקות: נשתמש באותה נוסחה רק נתייחס לאמצע המחלקה בתור ה- X . הממוצע הזה יהיה ממוצע מקורב.

כללי: הממוצע רלבנטי רק למשתנה כמותי.

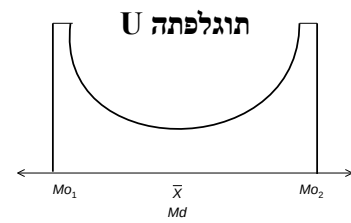
מדדי המיקום המרכזי בהתפלגויות המיוחדות:

בהתפלגות סימטרית פעמונית כל מדדי המרכז שווים זה לזה:

התפלגות סימטרית

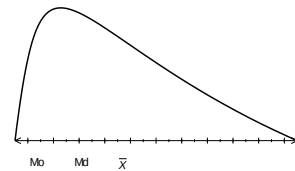


בהתפלגות סימטרית השכיח לא חייב להיות במרכז:



בהתפלגות אסימטרית

**התפלגות א-סימטרית
ימנית או חיובית**



**התפלגות
א-סימטרית
שמאלית או
שלילית**



תרגילים:

1. להלן רשימת הציונים של 20 תלמידים שנבחנו במבחן הבנת הנקרא :
 $7, 6, 8, 9, 10, 6, 4, 5, 8, 7, 6, 7, 6, 8, 9, 6, 7, 8, 5, 6$
 חשב את החציון, השכיח, והממוצע של הציונים.
2. בדקו את מספר החדרים לדירה בבניין בן 5 דירות והתקבל ממוצע 3.8
 לגבי 4 דירות נמצא מספר חדרים : 4, 3, 4, 5 .
 א. כמה חדרים יש בדירה החמישית?
 ב. מהו השכיח ומהו החציון?
3. להלן התפלגות מספר מקלטי הטלויזיה שנספרו עבור כל משפחה ביישוב מסוים :

מספר משפחות	מספר מקלטים
22	0
28	1
18	2
22	3
10	4

- א. חשב את הממוצע, החציון והשכיח של ההתפלגות.
- ב. הסבר ללא חישוב כיצד כל מדד שחישבת בסעיף א' היה משתנה אם חלק מהמשפחות (לא כולן) שלא היה להם עד היום טלויזיה היו רוכשים מקלט אחד.
4. להלן התפלגות מספר המכוניות למשפחה ביישוב "הגורן"

מספר מכוניות למשפחה	1	2	3	4	5
שכיחות	65	150	220	140	55

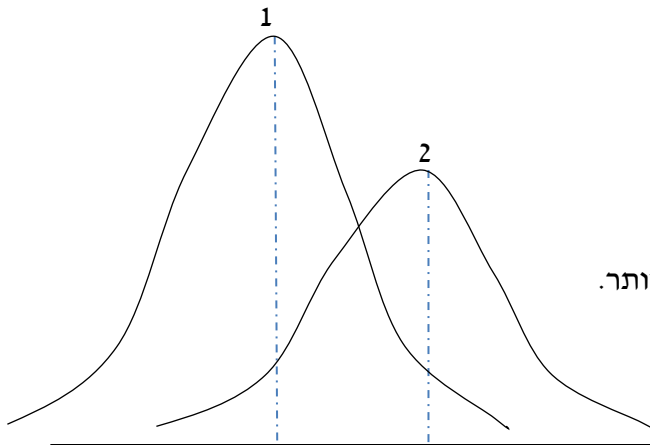
א. כמה משפחות יש

בישוב?

- ב. מה אחוז המשפחות ביישוב עם לכל היותר 2 מכוניות?
- ג. חשבו את הממוצע, החציון והשכיח.

הקפידו להסביר לגבי כל סעיף מה משמעות התוצאה שקיבלתם!

5. מורה לימד 2 כיתות, הוא תיאר באותה מערכת צירים את התפלגות הציונים בכל כיתה. בחר בתשובה הנכונה:



- א. בכיתה 1 השכיח גבוה יותר מכיתה 2.
 ב. בכיתה 2 השכיח גבוה יותר מכיתה 1.
 ג. בשתי הכיתות אותו שכיח.
 ד. לא ניתן לדעת באיזו כיתה השכיח גדול יותר.

6. ביישוב מסוים בדקו לכל משפחה את מספר הטלוויזיות שיש לה בבית. ביישוב גרות 200 משפחות. בממוצע יש למשפחה 1.5 טלוויזיות.

מספר טלוויזיות	מספר משפחות
0	28
1	62
2	
3	

- א. השלימו את הטבלה.
 ב. מהו השכיח, אמצע טווח והחציון.
 ג. חלק מהמשפחות להן הייתה טלוויזיה אחת בדיוק הוציאו את הטלוויזיה מביתם, כיצד כל מדד ישתנה (יגדל, יקטן או לא ישתנה) הסבירו ללא חישוב.

7. להלן התפלגות המשקל של קבוצה מסוימת בק"ג:

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

משקל	מספר מקרים
40-45	10
45-50	20
50-60	30
60-65	20
65-70	10

מה הממוצע והחציון של ההתפלגות?

8. להלן התפלגות הגבהים בס"מ בקבוצה מסוימת.

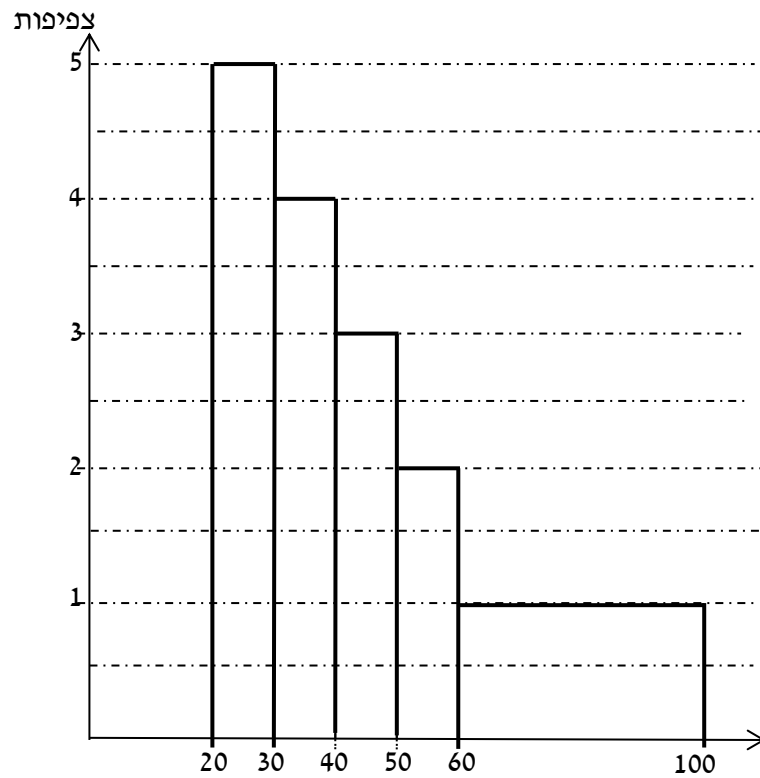
שכיחות	גובה בס"מ
30	150-160
40	160-170
60	170-175
70	175-180
40	180-190

חשב את הממוצע, החציון והשכיח של הגבהים בקבוצה זו.

9. בפקולטה מסוימת בדקו לסטודנטים העובדים בה את השכר לשעת עבודה. להלן התוצאות :

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

כתב ופתר - ברק קנדל ©



א. מצא את השכיח בהתפלגות.

ב. מצא את החציון בהתפלגות.

ג. הסבירו ללא חישוב האם הממוצע גדול/קטן/שווה לחציון.

ד. הסתבר שיש להוציא מספר תלמידים במחלקה בין 20-30 שקלים כיצד הדבר ישפיע על הממוצע, החציון והשכיח? הסבירו ללא חישוב.

פתרונות:**שאלה 1:**

החציון : 7

השכיח : 6

הממוצע : 6.9

שאלה 2:

א. 3

ב. שכיח : 3,4 חציון : 4

שאלה 3:

א. הממוצע : 1.7

החציון : 1.5

השכיח : 1

ב. הממוצע יגדל ויתר המדדים לא ישתנו.

שאלה 4:

א. 630

ב. 34.13%

ג. שכיח וחציון : 3

ממוצע : 2.952

שאלה 5:

תשובה : ב

שאלה 6 :

ב חציון : 2 שכיח : 2 אמצע טווח : 1.5

שאלה 7:

חציון וממוצע : 55

פרק 7 - סטטיסטיקה תיאורית - מדדי פיזור: הטווח, השונות וסטיית התקן

רקע:

המטרה : למדוד את הפיזור של הנתונים כלומר כמה הם רחוקים זה מזה ושונים זה מזה.

הטווח והמספר RANGE:

ההפרש בין התצפית הגבוהה ביותר לנמוכה ביותר : $R = X_{\max} - X_{\min}$

שונות וסטיית תקן:

השונות היא ממוצע ריבועי הסטיות מהממוצע וסטיית התקן היא שורש של השונות.

$$s_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2 \quad \text{עבור סדרת נתונים :}$$

דוגמה : נחשב את השונות של סדרת המספרים הבאה : 5,4,9

$$s_x^2 = \frac{\sum (x - \bar{x})^2 \cdot f}{n} = \frac{\sum x^2 \cdot f}{n} - \bar{x}^2 \quad \text{עבור טבלת שכיחויות :}$$

להלן התפלגות הציונים בכיתה מסוימת בה ממוצע הציונים הוא 7.44

הציון X	השכיחות f	$x^2 \cdot f$
5	2	50
6	4	144
7	8	392
8	5	320
9	4	324
10	2	200
סה"כ		1430

$$s_x^2 = \frac{\sum x^2 f(x)}{n} - \bar{x}^2 = \frac{1430}{25} - 7.44^2 = 1.8464$$

$$s = \sqrt{s_x^2} = \sqrt{1.8464} = 1.3588$$

כשיש מחלקות נעזר באמצע המחלקה כדי לחשב את השונות.

תרגילים:

1. להלן רשימת הציונים של 20 תלמידים שנבחנו במבחן הבנת הנקרא :

7, 6, 8, 9, 10, 6, 4, 5, 8, 7, 6, 7, 6, 8, 9, 6, 7, 8, 5, 6
חשבו את השונות, סטיית התקן והטווח של הציונים.

2. להלן התפלגות מספר המכוניות למשפחה ב"הגורן"

מספר מכוניות למשפחה	1	2	3	4	5
שכיחות	65	150	220	140	55

א. חשבו סטיית התקן.

ב. חשבו את הטווח של הנתונים.

הקפידו להסביר לגבי כל סעיף מה משמעות התוצאה שקיבלתם!

3. בחברה העוסקת בטלמרקטינג בדקו עבור כל עובד את מספר שנות הוותק שלו. התקבל שממוצע שנות הוותק הוא 4 שנים וסטיית התקן היא שנתיים.

א. האם הממוצע יגדל/יקטן/לא ישתנה וסטיית התקן תגדל/תקטן/לא תשנה כאשר יתווספו שני עובדים עם וותק של 4 שנים להתפלגות?

ב. האם הממוצע יגדל/יקטן/לא ישתנה וסטיית התקן תגדל/תקטן/לא תשנה כאשר יתווספו שני

עובדים אשר אחד עם וותק של 0 שנים והשני עם וותק של 8 שנים להתפלגות?

4. נתונה רשימה של 5 תצפיות, אך רק עבור 4 מהן נרשמו הסטיות שלהן מהממוצע:

2, 3, 2, -1. חשב את השונות של חמש התצפיות.

5. בשכונה בדקו בכל דירה את מספר החדרים לדירה. בשכונה 200 דירות.

מספר חדרים	פרופורציה
1	0.1
2	0.2
3	0.4
4	0.15
5	

א. מה הממוצע של מספר החדרים לשכונה בדירה?

ב. חשבו את סטיית התקן של מספר החדרים לדירה.

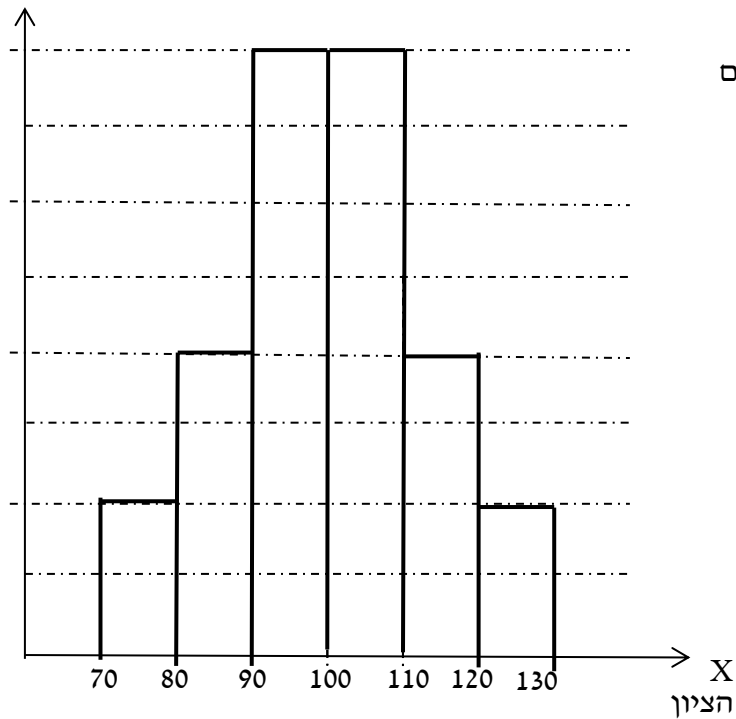
ג. חלק מבעלי הדירות בנו 2 החדרים הפכו את דירתם לדירת חדר. כיצד הדבר ישפיע (יקטין, יגדל, לא ישנה) כל מדד שחישבתם בסעיפים הקודמים.

6. להלן התפלגות המשקל של קבוצה מסוימת בק"ג:

משקל	מספר מקרים
40-45	10
45-50	20
50-60	30
60-65	20
65-70	10

מהי סטיית התקן של התפלגות המשקל?

7. להלן התפלגות הציונים במבחן אינטליגנציה:



12 = נבחנים



- א. מה הממוצע ומה החציון של ההתפלגות?
- ב. חשבו את סטיית התקן של הציונים.
- ג. מסתבר שיש להוסיף 20 תצפיות לכל אחת משתי המחלקות 90-100 ו-100-110. כיצד הדבר ישתנה את כל אחד מהמדדים של הסעיפים הקודמים?

פתרונות :**שאלה 1:**

השונות : 2.19

סטיית תקן : 1.48

טווח : 6

שאלה 2:

א. סטיית תקן: 1.106

ב. טווח 4

שאלה 3:

א. ממוצע לא ישתנה, סטיית התקן תקטן.

ב. ממוצע לא ישתנה, סטיית התקן תגדל.

שאלה 4:

10.8

שאלה 5:

א. 3.05

ב. 1.16

שאלה 6:

7.73

שאלה 7 :

א. 100

ב. 12.96

פרק 8 - סטטיסטיקה תיאורית - מדדי מיקום יחסי - ציון תקן

רקע:

המטרה למדוד איך תצפית ממוקמות יחסית לשאר התצפיות בהתפלגות.

ציון תקן:

הנוסחה לציון תקן של תצפית היא : $Z = \frac{X - \bar{X}}{S}$

ציון התקן נותן כמה סטיות תקן סוטה התצפית מהממוצע.

כלומר, ציון התקן מעיד על כמה סטיות תקן התצפית מעל או מתחת לממוצע.

ציון תקן חיובי אומר שהתצפית מעל הממוצע.

ציון תקן שלילי אומר שהתצפית מתחת לממוצע.

ציון תקן אפס אומר שהתצפית בדיוק בממוצע.

דוגמה : (פתרון בהקלטה)

במקום עבודה מסוים ממוצע המשכורות 8 אלפי ₪ עם סטית תקן של 2 אלפי ₪ באותו מקום עבודה ההשכלה הממוצעת של העובדים הנה 14 שנים עם סטית תקן של 1.5 שנים. ערן מרוויח במקום עבודה זה 11 אלף ₪ והשכלתו 16 שנים. מה ערן יותר באופן יחסי משכיל או משתכר ?

תרגילים

1. תלמידי כיתה ח' ניגשו למבחן בלשון ולמבחן במתמטיקה.
להלן התוצאות שהתקבלו :

המקצוע	ממוצע	סטיית תקן
לשון	74	12
מתמטיקה	80	16

- עודד קיבל: 68 בלשון ו70 במתמטיקה.
א. באיזה מקצוע עודד טוב יותר באופן יחסי לשכבה שלו?
ב. איזה ציון עודד צריך לקבל במתמטיקה כדי שיהיה שקול לציונו בלשון?

2. במפעל לייצור מצברים לרכב בדקו במשך 40 ימים את התפוקה היומית (מספר מצברים במאות) ואת מספר הפועלים שעבדו באותו היום.
להלן טבלה המסכמת את האינפורמציה שנאספה על שני המשתנים :

מספר פועלים	תפוקה	
15	48	ממוצע
2	10	סטיית תקן

- באחד הימים מתוך כלל הימים שנבדקו התפוקה הייתה 50 מאות מצברים ובאותו היום עבדו 13 פועלים.
מה יותר חריג באותו היום יחסית לשאר הימים שנבדקו נתוני התפוקה או כמות הפועלים?
בחר בתשובה הנכונה.
א. התפוקה.
ב. כמות הפועלים.
ג. חריגים באותה מידה.
ד. חסרים נתונים כדי לדעת זאת.

3. הגובה הממוצע של המתגייסים לצבא הוא 175 סנטימטר עם סטיית תקן 10 סנטימטר. המשקל הממוצע 66 ק"ג עם סטיית תקן 8 ק"ג. ערך התגייס, גובהו 180 ס"מ ומשקלו 59 ק"ג.
א. במה ערך חריג יותר ביחס לשאר המתגייסים- גובהו או משקלו?
ב. כמה ערך אמור לשקול כדי שמשקלו יהיה שקול לגובהו?

פתרונות:**שאלה 1:**

א. לשון

ב. 72

שאלה 2:

תשובה ב

שאלה 3:

א. משקל

ב. 70

פרק 9 - סטטיסטיקה תיאורית - מדדי מיקום יחסי - אחוזונים במחלקות

רקע:

האחוזון (המאון) ה- p הוא הערך בנתונים המחלק את הנתונים בצורה כזאת שעד אליו יש $p\%$

מהנתונים. מסמנים את האחוזון ה- p ב- X_p .

למשל, המאון ה- 25 הוא האחוזון ה- 25 או הרבעון התחתון : ערך ש- רבע מהתצפיות קטנות

ממנו והשאר גבוהות ממנו. מסומן : $X_{0.25}$

מציאת מאון במחלקות :

שלב א : נימצא את המחלקה הרלבנטית שמיקומה יהיה $\frac{np}{100}$.

שלב ב : נציב בנוסחה הבאה : $x_p = L_0 + \frac{\frac{n \cdot p}{100} - F(x_{m-1})}{f(x_m)} \cdot (L_1 - L_0)$

$F(x_{m-1})$ - שכיחות מצטברת של מחלקה אחת לפני המחלקה הרלבנטית.

$f(x_m)$ - השכיחות של המחלקה הרלבנטית.

L_0 - גבול התחתון של המחלקה.

L_1 - גבול העליון של המחלקה.

אם רוצים לחלץ את אחוז התצפיות שמתחת לערך מסוים נשתמש בנוסחה הבאה :

$$P_x = \left[\frac{(x - L_0)}{(L_1 - L_0)} \cdot f(x_m) + F(x_{m-1}) \right] \cdot \frac{100}{n}$$

דוגמה : (פתרון בהקלטה)

להלן התפלגות השכר של עובדים בחברה מסוימת :

שכר בש"ח	$f(x)$
4000-6000	140
6000-10000	128
10000-15000	60
15000-20000	54
20000-40000	18

א. מצאו את המאון ה-40.

ב. מהו אחוז העובדים שמשכרים מתחת ל-5,000 ₪?

תרגילים:

1. להלן התפלגות השכר (באלפי שקלים) בחברה:

שכר X	שכיחות מצטברת
6-10	48
10-15	100
15-20	120
20-30	132
30-60	136

א. חשבו את המאון ה-60.

ב. מהו העשירון העליון?

ג. 20% מהמשכורות הגבוהות ביותר הן משכורות של הבכירים, מהי המשכורת המינימאלית לבכיר?

ד. מה אחוז האנשים שמשתכרים מתחת ל-7000 ₪?

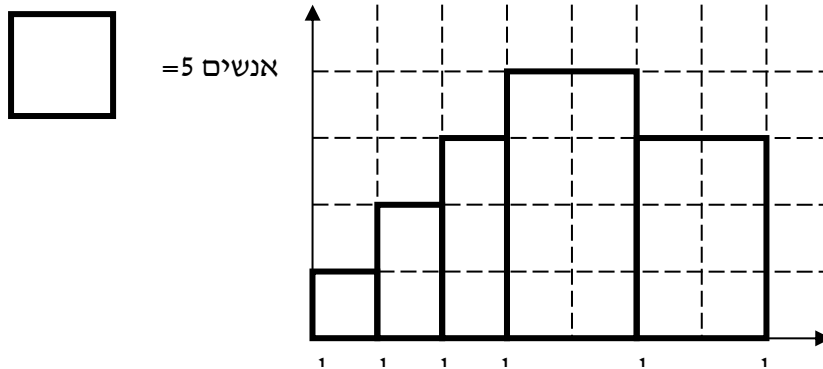
ה. איזה אחוז מהעובדים משתכרים מעל ל-25,000 ₪?

ו. איזה אחוז מהעובדים משתכרים בין 7000 ל-25,000 ₪?

2. למבחן ניגשו 400 נבחנים. נתון שהעשירון התחתון הוא הציון 60. הרבעון העליון הוא הציון 80. כמו כן ההתפלגות של הציונים היא סימטרית. מלאו את השכיחות החסרות.

ציון - X	$f(X)$
50-60	
60-70	
70-80	
80-90	
90-100	

3. להלן היסטוגרמה המתארת את התפלגות הגבהים בס"מ של קבוצה מסוימת:



א. העשירון התחתון.

ב. האחוזון ה-30.

ג. הגובה ש-20% מהתצפיות גדולות ממנו.

ד. את אחוז התצפיות מתחת לגובה 158 ס"מ.

ה. את אחוז התצפיות מעל לגובה 185 ס"מ.

ו. את אחוז התצפיות בין גובה 170 ס"מ ל-185 ס"מ.

פתרונות:

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

כתב ופתר - ברק קנדל ©

שאלה 1 :

א. 13.23

ב. 22

ג. 17.2

ד. 8.82%

ה. 7.36%

ו. 83.82%

שאלה 3 :

א. 162.5

ב. 170

ג. 183.33

ד. 3%

ה. 15%

ו. 55%ת

פרק 10 - סטטיסטיקה תיאורית - טרנספורמציה לינארית**רקע:**לפתרון מלא בסרטון וידאו היכנסו ל- www.GooL.co.il

כתב ופתר - ברק קנדל ©

מצב שבו מבצעים שינוי מסוג הוספה של קבוע (או החסרה) והכפלה של קבוע (או חילוק) לכל

$$y = a \cdot x + b \quad \text{התצפיות :}$$

וכך יושפעו המדדים השונים :

מדדי המרכז :

$$MR_y = a \cdot MR_x + b$$

$$Mo_y = a \cdot Mo_x + b$$

$$\bar{y} = a \cdot \bar{x} + b$$

$$Md_y = a \cdot Md_x + b$$

מדדי הפיזור :

$$R_y = |a| R_x$$

$$s_y = |a| s_x$$

$$s_y^2 = a^2 s_x^2$$

מדדי המיקום היחסי :

$$Y_p = a \cdot X_p + b$$

$$Z_y = \frac{a}{|a|} Z_x$$

שלבי העבודה :

1. נזהה שמדובר בטרנספורמציה ליניארית (שינוי קבוע לכל התצפיות).
2. נרשום את כלל הטרנספורמציה לפי נתוני השאלה.
3. נפשט את הכלל ונזהה את ערכי a ו b .
4. נציב בנוסחאות שלעיל בהתאם למדדים שנשאלים.

דוגמה : (פתרון בהקלטה)

השכר הממוצע של עובדים הנו 9000 ₪ וטווח 6000 ₪ חשבו את המדדים הללו לאחר שהעלו את כל המשכורות ב-10% ואחר כך קנסו אותם ב100 ₪.

תרגילים :

1. עבור סדרת נתונים התקבל :

$$\bar{X} = 80$$

$$S = 15$$

$$MO = 70$$

הוחלט להכפיל את כל התצפיות פי-4 ולהחסיר מהתוצאה 5. חשב את המדדים הללו לאחר השינוי.

2. בחברה מסוימת השכר הממוצע הוא 40 ₪ לשעה עם סטיית תקן של 5 ₪ לשעה. הוחלט להעלות את כל המשכורות ב-10%, אך זה לא סיפק את העובדים ולכן הם קיבלו לאחר מכן תוספת של 2 ₪ לשעה. מה הממוצע ומהי השונות של השכר לשעה לאחר כל השינויים.

3. במבחן הציון החציוני היה 73, טווח הציונים היה 40 נקודות. והעשירון העליון היה הציון 87. כיוון שהציונים בבחינה היו נמוכים, המורה החליט לתת פקטור של 4 נק' לכל התלמידים. חשבו את המדדים לאחר הפקטור.

4. דגמו מקו ייצור 50 קופסאות של גפרורים. בדקו בכל קופסא בה יש 40 גפרורים את כמות הגפרורים הפגומים. קבלו שבממוצע יש 3 גפרורים פגומים בקופסא. עם סטיית תקן של 1.5 גפרורים. מה יהיה הממוצע ומה תהיה סטיית התקן של מספר התקנים בקופסא?

5. חברת בזק הציעה את החבילה הבאה :
שלושים שקלים דמי מנוי חודשיים קבועים. ובנוסף 10 אגורות לכל דקה של שיחה יוצאת, אדם בדק במשך שנה את דקות השיחות היוצאות שלו, וקיבל שבממוצע בחודש יש לו 600 דקות שיחות יוצאות עם שונות 2500 דקות רבועות, כמו כן בחודש ינואר ציון התקן היה 2. חשבו את המדדים הללו עבור חשבון הטלפון החודשי של אותו אדם בשקלים אם היה משתמש בחבילה המוצעת לו על ידי בזק.

6. הוכח שאם כל התצפיות בהתפלגות עברו טרנספורמציה לינארית :

$$Y_i = a \cdot X_i + b$$

אזי הממוצע והשונות של כלל התצפיות לאחר הטרנספורמציה יהיו בהתאמה :

$$\bar{y} = a \cdot \bar{x} + b$$

$$s_y^2 = a^2 s_x^2$$

פתרונות :**שאלה 2 :**

הממוצע : 46
 השונות : 30.25

שאלה 1 :

הממוצע : 315
 סטיית התקן : 60
 השכיח : 275

שאלה 4 :

ממוצע : 37
 סטיית תקן : 1.5

שאלה 3 :

טווח : 40
 חציון : 77
 עשירון עליון : 91

שאלה 5 :

ממוצע : 90
 שונות : 25
 ציון תקן : 2

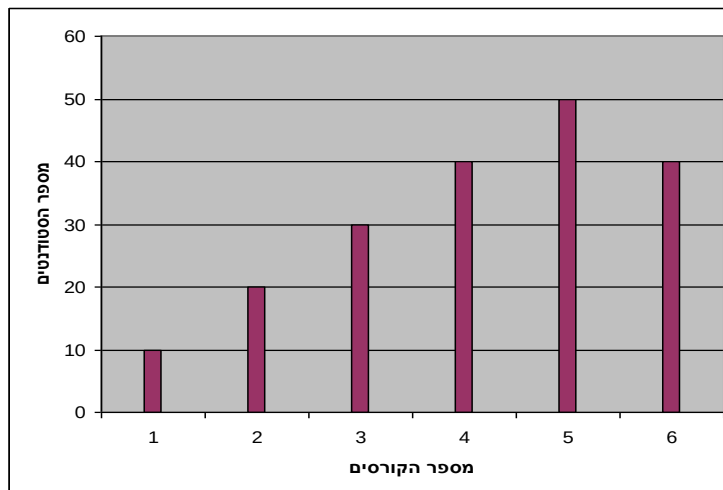
פרק 11 - סטטיסטיקה תיאורית - שאלות מסכמות

1. בדקו עבור 5 תלמידים את המשקל שלהם :

מספר תלמיד	משקל בק"ג
1	58
2	62
3	48
4	34
5	58

- מהו המשתנה הנחקר? האם הוא בדיד או רציף?
- מהו המשקל החציוני, הממוצע והשכיח?
- מה הטווח וסטיית התקן של המשקל?
- לאותם תלמידים חישבו גם את הגובה בס"מ וקיבלו גובה ממוצע של 168 וסטיית תקן 6. במה תלמיד מספר 3 שגובהו 162 יותר חריג במשקל או בגובה?
- הוסיפו עוד תלמיד השוקל 52 ק"ג בדיוק. הסבירו ללא חישוב כיצד הדבר ישפיע על הממוצע וסטיית התקן? (יגדיל יקטין או לא ישנה)

2. בפקולטה להנדסה אספה מזכירות הסטודנטים נתונים לגבי מס' הקורסים שכל סטודנט סיים בשנה הראשונה ללימודיו בשנת 2008.
להלן התוצאות שהתקבלו :



- מה המשתנה הנחקר? האם הוא בדיד או רציף?
 - מהי צורת ההתפלגות?
 - תאר את הנתונים בטבלת שכיחויות.
 - חשב את השכיח, החציון והטווח.
3. להלן התפלגות הציונים בבחינה בלשון שנעשתה עבור תלמידי כיתות ד'.

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

השתתפו במחקר 150 תלמידים.

ממוצע הציונים שהתקבל: $\bar{X} = 7\frac{1}{15}$

מספר התלמידים	ציון
12	4
16	5
	6
38	7
	8
14	9
10	10

א. השלם את השכיחויות החסרות בטבלה.

ב. חשב את הציון החציוני, השכיח.

ג. חשב שונות וסטיית תקן להתפלגות הציונים.

4. חברה סלולארית דגמה 200 אנשים. עבור כל אדם נבדקה מידת שביעות הרצון של הלקוח מהחברה (1 – שביעות רצון נמוכה ועד 5 שביעות רצון גבוהה) להלן ההתפלגות שהתקבלה:

שביעות רצון	מספר האנשים
1	40
2	60
3	50
4	30
5	20

- א. מה אחוז האנשים עם רמת שביעות רצון נמוכה?
- ב. מה המשתנה הנחקר ומאיזה סוג הוא?
- ג. מהי הדרך הגרפית המתאימה ביותר לתיאור הנתונים?
 - i. היסטוגרמה.
 - ii. דיאגרמת מקלות.
 - iii. דיאגרמת עוגה
- ד. חשבו את המדדים הבאים:
 1. טווח
 2. שכיח
 3. חציון

5. להלן התפלגות מספר שעות העבודה לשבוע של העובדים בחברת "סטאר".
בחברה 200 עובדים.

מספר שעות עבודה	שכיחות יחסית (פרופורציה)	שכיחות
10-20	15%	
20-30	20%	
30-40	30%	
40-50	20%	
50-60		

א. השלם את הטבלה.

ב. חשב את החציון, השכיח, והממוצע של התפלגות מס' שעות העבודה בחברה.

ג. מהי סטיית התקן של מס' שעות העבודה?

ד. מה העשירון העליון של ההתפלגות?

ה. איזה אחוז מהעובדים עובדים מעל 45 שעות בשבוע?

ו. מה ציון התקן של רינה שעובדת 30 שעות בשבוע?

ז. כיצד ישתנה החציון, הממוצע וסטיית התקן אם מספר שעות העבודה המינימאלי אינו 10 אלא 15? הסבר.

6. חברה סלולארית דגמה 200 אנשים. עבור כל אדם נבדק מס' המסרונים ששלח במשך חודש. להלן ההתפלגות שהתקבלה:

מספר המסרונים	מספר האנשים
0-50	40
50-100	60
100-150	50
150-250	30
250-ומעלה	20

- א. מה אחוז האנשים ששלחו פחות מ-80 מסרונים בחודש?
 ב. מה אחוז האנשים ששלחו בין 50 ל-120 מסרונים?
 ג. הוחלט להעניק מתנה עבור $\frac{1}{4}$ מהלקוחות שמשלמים במספר הרב ביותר של מסרונים בחודש. החל מאיזה כמות של מסרונים תחולק המתנה?
 ד. ציינו איזה מדד ניתן לחשב ואיזה לא ניתן. אם ניתן חשב:
1. ממוצע
 2. שכיח
 3. חציון
 4. שונות

7. נתנו לקבוצת ילדים לבצע משימה מסוימת ובדקו את התפלגות זמן ביצוע המשימה בדקות. להלן ההתפלגות שהתקבלה:

זמן בדקות	מספר הילדים
0.5-3.5	20
3.5-9.5	18
9.5-19.5	14
19.5-29.5	8

- א. שרטט היסטוגרמה לתיאור התפלגות זמן ביצוע המשימה.
 ב. מתוך ההיסטוגרמה שבנית בסעיף א מהי צורת ההתפלגות?
 ג. חשב את השכיח והחציון של ההתפלגות.
 ד. הסבר, ללא חישובים, האם הזמן הממוצע לביצוע המשימה, קטן או גדול או שווה ביחס לשכיח ולחציון.

8. התפלגות ציוני מבחן אינטיליגנציה היא סימטרית.

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

הציון	מספר הנבחנים
50-70	
70-90	
90-100	
100-110	
110-130	
130-150	

נתון שהעשירון העליון הוא 130 והרבעון התחתון הוא 90.

נתון שלמבחן נגשו 500 מועמדים.

א. השלימו את הטבלה.

ב. מהו הממוצע והחציון של ההתפלגות?

ג. מהו הציון ש 40% מהתלמידים קיבלו מעליו? באיזה אחוזון מדובר?

ד. אם יוחלט להעלות את כל הציונים ב-10 נקודות. כיצד הדבר ישפיע על הממוצע וסטיית התקן של הציונים?

9. להלן מספר טענות, עבור כל טענה ציין אם היא נכונה או לא נכונה ונמקו.

א. בסדרה שבה כל התצפיות שוות זו לזו השונות הינה 0.

ב. ציון התקן של החציון תמיד יהיה 0.

ג. ציון התקן של האחוזון ה-70 בהתפלגות אסימטרית ימנית (חיובית) תמיד יהיה חיובי.

ד. אם נוסיף תצפיות לסדרה של תצפיות, הדבר בהכרח יגדיל את הממוצע של הסדרה.

ה. בסדרה החציון הינו 80. הוספו שתי תצפיות אחת 79 ואחת 100 לכן החציון יגדל.

ו. אם נוסיף את הערך 4 לכל התצפיות אז סטיית התקן לא תשתנה.

ז. אם נחלק את כל התצפיות בהתפלגות ב-2 אז השונות תקטן פי 2.

ח. אם נגדיל את ממוצע המשכורות של עובדים בחברה אז גם השונות תגדל.

פתרונות:

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

כתב ופתר - ברק קנדל ©

שאלה 1:

א. המשתנה הנחקר כאן הוא משקל תלמיד בק"ג והוא משתנה כמותי רציף.

ב.

$$\bar{X} = 52$$

$$Md = X_{\frac{n+1}{2}} = X_3 = 58$$

השכיח הוא 58

$$R = 28 \quad \text{ג.}$$

$$s = 10.12$$

ד. הוא חריג יותר בגובה כי שם ציון התקן בערך מוחלט יותר גבוה.

ה. הממוצע לא ישתנה אך סטיית התקן תקטן .

שאלה 2:

א. מספר הקורסים. בדיד.

ב. התפלגות אסימטרית שמאלית

ד. השכיח : 5

הטווח : 5

שאלה 3:

א. 20 תלמידים קיבלו ציון 6 ו-40

תלמידים קיבלו ציון 8.

החציון : 7

השכיח : 8

ג. השונות : 2.533

סטיית התקן : 1.592

שאלה 4:

א. 20%

ב. שביעות רצון (סדר)

ג. 2

ד. טווח : 4 שכיח : 2 חציון : 2.5

ה. חציון : 4

שאלה 5:

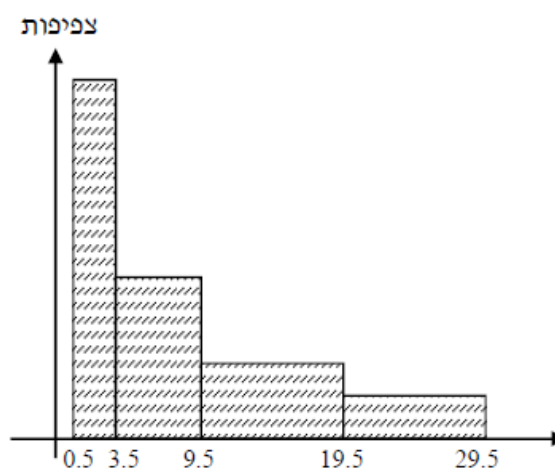
- ב. החציון : 35
 השכיח : 35
 הממוצע : 35
 ג. סטיית תקן : 12.65
 ד. 53.333
 ה. 25%
 ו. -0.395
 ז. חציון לא ישתנה, ממוצע יגדל סטיית התקן תקטן.

שאלה 6:

- א. 38%
 ב. 40%
 ג. 150
 ד. החציון : 100

שאלה 7:

א.



- ב. ההתפלגות היא א-סימטרית ימנית .
 ג. שכיח : 2 חציון : 6.83
 ד. בהתפלגות א-סימטרית ימנית מתקיים $Mo < Md < \bar{X} < MR$

שאלה 8:

א.

מספר הנבחנים	ציון
50	50-70
75	70-90
125	90-100
125	100-110
75	110-130
50	130-150

ב. 100

ג. 104

ד. הממוצע יעלה ב-10 נקודות אך סטיית התקן לא תשתנה .

שאלה 9:

א. נכון

ב. לא נכון

ג. לא נכון

ד. לא נכון

ה. לא נכון

ו. נכון

ז. לא נכון

ח. לא נכון

פרק 12 - רווח סמך לתוחלת (ממוצע האוכלוסייה)

רווח סמך כששונות האוכלוסייה ידועה

רקע:

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

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

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

A - גבול התחתון של רווח הסמך

B - הגבול העליון של רווח הסמך

$L = B - A$ - אורך רווח הסמך

דוגמה : (פתרון בהקלטה)

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

מהי אוכלוסיית המחקר?

מה המשתנה באוכלוסייה?

מה הפרמטר שהחוקר רצה לאמוד?

מהו רווח הסמך?

מה אורך רווח הסמך?

מהי רמת הביטחון של רווח הסמך?

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

הפרמטר שנרצה לאמוד : μ

האומד נקודתי : $\bar{x}$

התנאים לבניית רווח הסמך :

$$1 \quad X \sim N \text{ או } n \geq 30$$

2 σ^2 (שונות האוכלוסייה) ידועה

הנוסחה לרווח הסמך :

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

דוגמה : (פתרון בהקלטה)

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

מעוניינים לאמוד את תוחלת חיי סוללה.

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

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

שגיאת האמידה המקסימלית:

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

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

דוגמה: (פתרון בהקלטה)

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

קשרים מתמטיים ברווח הסמך:

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

תרגילים :

1. חוקר התעניין לאמוד את השכר הממוצע במשק. על סמך מדגם הוא קבע שבביטחון של 95% כי השכר הממוצע במשק נע בין 9200 ל-9800.
 - א. מי האוכלוסייה במחקר?
 - ב. מה המשתנה הנחקר?
 - ג. מה הפרמטר שאותו רוצים לאמוד?
 - ד. מה רווח הסמך לפרמטר?
 - ה. מהי רמת הסמך לפרמטר?
 - ו. מה אורך רווח הסמך?
 - ז. מה הסיכוי שטעות הדגימה תעלה על 300 ₪?
2. מעוניינים לאמוד את התפוקה היומית הממוצעת של מפעל מסוים ברמת סמך של 95%. במדגם אקראי של 100 ימים התקבלה תפוקה ממוצעת 4950 מוצרים ביום. לצורך פתרון הנח שסטיית התקן האמתית ידועה ושווה 150 מוצרים ביום. בנה את רווח הסמך.
 - א. בנו רווח סמך ברמת סמך של 90% לאורך החיים הממוצע של מכשיר.
 - ב. בנו רווח סמך ברמת סמך של 95% לאורך החיים הממוצע של מכשיר.
 - ג. הסבר כיצד ומדוע השתנה רווח הסמך.
3. מעוניינים לאמוד את ממוצע אורך החיים של מכשיר. מנתוני היצרן ידוע שאורך החיים מתפלג נורמאלי עם סטיית תקן של 20 שעות. נדגמו 25 מכשירים ונמצא כי ממוצע אורך החיים שלהם היה 230 שעות.
 - א. בנו רווח סמך ברמת סמך של 90% לאורך החיים הממוצע של מכשיר.
 - ב. בנו רווח סמך ברמת סמך של 95% לאורך החיים הממוצע של מכשיר.
 - ג. הסבר כיצד ומדוע השתנה רווח הסמך.
4. דגמו 200 עובדים מהמשק הישראלי. השכר הממוצע שלהם היה 9700 ₪. נניח שסטיית התקן של השכר במשק היא 3000 ₪.
 - א. בנו רווח סמך ברמת סמך של 95% לתוחלת השכר במשק.
 - ב. מה ניתן לומר בביטחון של 95% על הסטייה המרבית בין ממוצע המדגם לתוחלת השכר?
 - ג. מה היה צריך להיות גודל המדגם אם הינו רוצים להקטין את רווח הסמך ב-50%?
 - ד. אם היינו מגדילים את גודל המדגם ובונים רווח סמך באותה רמת סמך האם היה ניתן לטעון בביטחון רב יותר שרווח הסמך מכיל את הפרמטר?
5. בנו רווח סמך לממוצע הציונים של מבחן אינטליגנציה. ידוע שסטיית התקן היא 15 והמדגם מתבסס על 100 תצפיות. רווח הסמך שהתקבל הוא (99,105). שחזרו את :
 - א. ממוצע המדגם.
 - ב. שגיאת האמידה המקסימאלית.
 - ג. רמת הסמך.

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

11. חוקר בנה רווח סמך לממוצע וקיבל את רווח הסמך הבא: $63 < \mu < 83$.
 נתון שסטיית התקן בהתפלגות הייתה ידועה לו ושהמדגם התבסס על 40 תצפיות.
 א. אם החוקר היה רוצה לבנות רווח סמך באורך 10. כמה תצפיות עליו היה לדגום?
 ב. רווח הסמך שנבנה על ידי החוקר היה ברמת סמך של 95%. בנה את רווח הסמך שהיה מתקבל ברמת סמך של 98%.

12. נתון משתנה מקרי רציף מתפלג אחיד: $X_i \sim U(\mu - 0.5, \mu + 0.5)$. נרצה לאמוד את μ . מצאו רווח סמך ל- μ ברמת-בטחון של 0.95 אם במדגם של 45 תצפיות התקבל:
 $\bar{x} = 74$.

$$\text{Var}(X_i) = \frac{(b-a)^2}{12} \quad \text{(תזכורת על השונות בהתפלגות אחידה רציפה)}$$

פתרונות :**שאלה 2**

$$4920.6 < \mu < 4979.4$$

שאלה 3

$$\text{א. } 223.42 < \mu < 236.58$$

$$\text{ב. } 222.16 < \mu < 237.84$$

שאלה 5

$$\text{א. } 102$$

$$\text{ב. } 3$$

$$\text{ג. } 0.9544$$

שאלה 6

$$\text{א. } 3.58 < \mu < 4.42$$

$$\text{ב. יקטן פי } 2$$

$$\text{ג. גדל}$$

שאלה 7

$$\text{א. } 87$$

$$\text{ב. } 5$$

$$\text{ג. } 0.9544$$

שאלה 8

$$\text{א. } 139$$

$$\text{ב. } 21 < \mu < 25$$

שאלה 9

התשובה היא : ב

שאלה 10

התשובה היא : ג

שאלה 11

התשובה היא : ד

קביעת גודל מדגם באמידת תוחלת עם שונות אוכלוסייה ידועה

רקע:

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

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

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

דוגמה: (פתרון בהקלטה)

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

תרגילים:

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

פתרונות :**שאלה 1**

780

שאלה 2

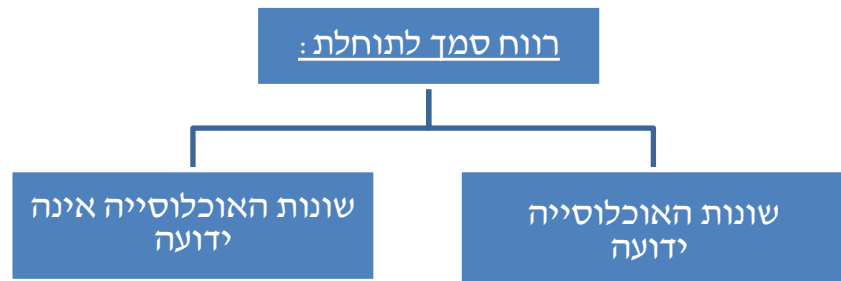
א. 139

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

רווח סמך לתוחלת (ממוצע האוכלוסייה) כששונות האוכלוסייה אינה ידועה

רקע:

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



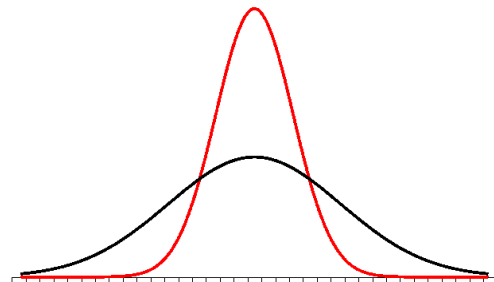
בפרק זה נעסוק במקרה ששונות האוכלוסייה (σ^2) אינה ידועה לנו. מקרה יותר פרקטי.

התנאי: $X \sim N$ או שהמדגם גדול

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

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 : 4.7, 5.2, 4.6, 5.3.
 בנו רווח סמך ברמת סמך של 95% לממוצע זמן הפתרון לשאלה בקרב תלמידי כיתה ח'.

פתרון :

$$4.39 < \mu < 5.51$$

תרגילים:

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

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

סטטיסטיקאי א : הניח $\sigma = 20$

סטטיסטיקאי ב : חישב לפי המדגם וקיבל $S = 20$

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

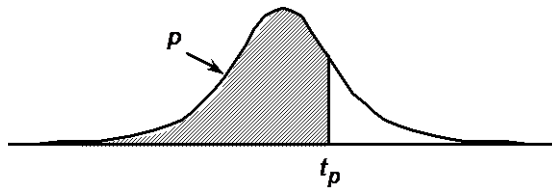
א. סטטיסטיקאי א

ב. סטטיסטיקאי ב

ג. אותו אורך רווח סמך לשני הסטטיסטיקאים.

ד. תלוי בתוצאות המדגם של כל סטטיסטיקאי.

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

נספח : טבלת התפלגות T 

P

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

פתרונות:**שאלה 1**

$$\text{א. } 79.88 < \mu < 89.72$$

שאלה 4

$$\text{א. } 96.63 < \mu < 107.37$$

$$\text{ב. } 96.90 < \mu < 107.10$$

שאלה 5

$$3.149 < \mu < 3.351$$

שאלה 8

90%

פרק 13 - רווח סמך לפרופורציה

רקע:

מטרה: לאמוד את P – פרופורציה באוכלוסייה.

האומד הנקודתי: $\hat{p} = \frac{y}{n}$ (Y – מספר ההצלחות שבמדגם)

רווח הסמך ל p: $\hat{p} \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

התנאי לבנות את רווח הסמך הינו מדגם של לפחות 30 תצפיות (לעיתים נותנים תנאי של מספר הצלחות ומספר כשלונות לפחות 5 או לפחות 10)

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

מתקיים ש: $\hat{P} = \frac{A+B}{2}$ $L = 2\varepsilon$

דוגמה: (פתרון בהקלטה)

1. במטרה לאמוד את אחוז המובטלים במשק נדגמו 200 אזרחים. מתוכם התקבל ש 24 היו מובטלים.

א. בנו רווח סמך לאחוז המובטלים באוכלוסייה ברמת סמך של 95%.

ב. מהו האומד לטעות התקן?

פתרון:

א. $7.5\% < p < 16.5\%$

ב. 2.29%

תרגילים:

1. נדגמו 200 דירות בעיר חיפה. 48 מתוכן נמצאו כבעלות ממ"ד.
 - א. בנו רווח סמך ברמת סמך של 95% לאחוז הדירות בחיפה עם ממ"ד.
 - ב. על סמך סעיף א' מה ניתן לומר על שגיאת האמידה המקסימאלית?
 - ג. בהנחה ובחיפה 80 אלף דירות, בנו רווח סמך ברמת סמך של 95% למספר הדירות בחיפה עם ממ"ד בפועל.
2. במדגם של 300 אנשי היי-טק התקבל ש-180 מהם אקדמאים.
 - א. בנו רווח סמך לפרופורציית אקדמאים ברמת סמך של 95% (בקרב אנשי היי-טק).
 - ב. כיצד רווח הסמך של סעיף א' היה משתנה אם היינו מקטינים את רמת הסמך?
 - ג. כיצד רווח הסמך היה משתנה אם הינו מגדילים את גודל המדגם?
3. במדגם של 400 נהגים התקבל רווח סמך לפרופורציית הנהגים החדשים: $0.08 < p < 0.18$
 - א. כמה נהגים במדגם היו נהגים חדשים?
 - ב. מהי רמת הסמך של רווח הסמך שנבנה?
4. במסגרת מערכת הבחירות בארה"ב נשאלו 840 אנשים עבור איזה מועמד יצביעו.

510 אנשים ענו כי יצביעו בעד ברק אובמה. בסקר פורסם שתתכן סטייה של $\pm 3\%$ מתוצאות האמת. באיזו רמת ביטחון הסקר השתמש?
5. במדגם של 300 נשים בגילאי 35-40 נמצא ש-140 היו נשואות, 80 היו גרושות, 60 רווקות והיתר אלמנות.
 - א. מצאו רווח סמך ברמה של 90% לאחוז הגרושות באוכלוסייה הנחקרת.
 - ב. מצאו רווח סמך ברמה של 99% לסיכוי שבאוכלוסייה הנחקרת תמצא אישה לא נשואה?
6. ביצעו מדגם באוכלוסייה. שיעור ההצלחות במדגם היה 10% ורווח הסמך ניבנה ברמת סמך של 95%. אורכו הינו 8.3156.

מהו גודל המדגם שנלקח?

פתרונות:**שאלה 3**

א. 52

ב. 0.997

שאלה 5א. $22.5\% < p < 30.9\%$ ב. $45.91\% < p < 60.72\%$ **שאלה 6**

200

קביעת גודל מדגם באמידת פרופורציה

רקע:

בפרק זה נדון איך קובעים גודל מדגם שבאים לאמוד פרופורציה באוכלוסייה מסוימת:

החוקר קובע מראש את רמת הסמך הרצויה: $1 - \alpha$.

החוקר קובע מראש את הטעות הסטטיסטית המרבית שבה הוא מעוניין: ε (או את אורך רווח הסמך).

$L = 2\varepsilon$ - אורך רווח הסמך.

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

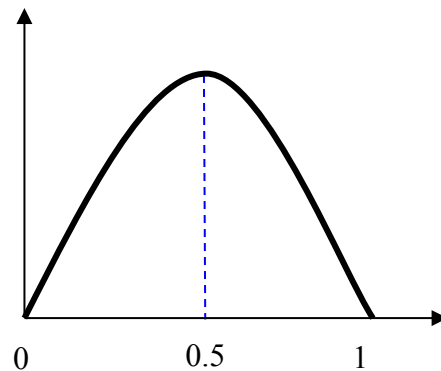
$$\varepsilon = z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

ויתעניין לדעת מהו גודל המדגם הרצוי לשם כך.

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

הבעיה שאין אנו יודעים את $\hat{p}$.

נתבונן בביטוי $\hat{p}(1-\hat{p})$:



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

$$\hat{p} = 0.5$$

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

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

דוגמה: (פתרון בהקלטה)

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

פתרון:

א. 423

ב. 271

תרגילים:

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

כתב ופתר - ברק קנדל ©

1. הממשלה אומדת מדי חודש את אחוז התמיכה בה. מהו גודל המדגם אשר יש לקחת אם דורשים שהאומדן לא יסטה מהאחוז האמיתי באוכלוסייה ביותר מ-3%, וזאת בביטחון של 95%?
2. משרד התקשורת מעוניין לדעת מה שיעור בתי האב עם אינטרנט.
 - א. כמה בתי אב יש לדגום אם מעוניינים שבביטחון של 90% אורך רווח הסמך לא יעלה על 8%?
 - ב. חזרו על סעיף א. אם ידעו שלפני חמש שנים ל-80% מבתי האב היה אינטרנט וכיום יש להניח שיש ליותר אינטרנט.
3. ערוץ טלוויזיה מעוניין לאמוד את הרייטינג של הערוץ בפריים טיים. המטרה שבביטחון של 95% הסטייה המרבית בין האומדן לרייטינג האמיתי לא תעלה על 4%.
 - א. כמה מכשירי PEOPLE METER יש להתקין לצורך האמידה?
 - ב. לפי הערכה מוקדמת הרייטינג של הערוץ לא יכול לעלות על 20%. בהנחה ומכשיר כזה עולה 500 ₪ ליחידה מה החיסכון הכספי מאינפורמציה זאת?
4. השאלות הבאות מתייחסות לסעיף 4 :
 - א. כמה אזרחים יש לדגום כדי לאמוד את אחוז התמיכה בממשלה עם אורך רווח הסמך שלא עולה על 9% ברמת סמך של 90%?
 - ב. בהנחה ובוצע מדגם שאת גודלו חישבתם בסעיף א והתקבל שאחוז התמיכה בממשלה במדגם הנו 42%. בנו רווח סמך לאחוז התמיכה בממשלה ברמת סמך של 95%.
 - ג. על סמך סעיף ב'. האם תקבל את הטענה שמיעוט האוכלוסייה תומך הממשלה?
5. משרד הבריאות מתכנן לבצע מדגם שמטרתו לבדוק את הסיכוי לחלות בשפעת עם לקיחת חיסון נגד שפעת. הוא מעוניין שבסיכוי של 98% טעות האמידה לא תעלה על 3%.
 - א. כמה מחוסנים יש לדגום ?
 - ב. משרד הבריאות ביצע את המדגם שאת גודלו חישבת בסעיף הקודם וקיבל ש 15% מבין אלה שקיבלו חיסון נגד שפעת בכל זאת חלו במשך החורף בשפעת. בנו ברמת סמך של 98% את הסיכוי לחלות בחורף בשפעת עם לקיחת חיסון נגד שפעת.
 - ג. בהמשך לסעיף הקודם. מהי טעות האמידה המרבית בביטחון של 98% ? מדוע הוא קטן מ-3% ?

פתרונות:

שאלה 1

1068

שאלה 3

א. 601

ב. 108000 ₪.

פרק 14 - רווח סמך להפרש פרופורציות

רקע:

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

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

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

רווח סמך:

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

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

דוגמה: (פתרון בהקלטה)

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

פתרון:

(47%, 33%)

תרגילים:

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

פתרונות:**שאלה 2**

$$0.093 < P_A - P_B < 0.307 \quad \text{א.}$$

שאלה 3

$$0.625 < p < 0.7754 \quad \text{א.}$$

פרק 15 - רווח סמך להפרש תוחלות ממדגמים בלתי תלויים

כששונות האוכלוסייה ידועות

רקע:

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

האומד נקודתי: $\bar{x}_1 - \bar{x}_2$

התנאים לבניית רווח הסמך:

1. σ_1^2, σ_2^2 ידועות.

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

3. שני מדגמים בלתי תלויים.

רווח סמך:

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

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

דוגמה: (פתרון בהקלטה)

נדגמו 100 תושבים מאזור a והמשכורת הממוצעת הייתה ש"ח 9200.

כמו כן נדגמו 120 תושבים מאזור b וממוצע המשכורות שהתקבל ש"ח 8700.

לצורך פתרון נניח שסטיית התקן של המשכורות באוכלוסיית שני האזורים היא 1800 ש"ח.

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

תרגילים:

1. מעוניינים לבדוק האם קיים הבדל בין ממוצע ציוני הפסיכומטרי של חיילים לממוצע ציוני הפסיכומטרי של תלמידי תיכון. ידוע שציוני הפסיכומטרי מתפלגים נורמאלית עם סטיית תקן 100. במדגם של 16 נבחנים חיילים התקבל ממוצע 543. במדגם של 20 תלמידי תיכון התקבל ממוצע 508. בנו רווח סמך לפער תוחלות הציונים בין חיילים לתלמידי תיכון ברמת סמך של 90%. מה ניתן להסיק מרווח סמך זה?

2. ציוני I.Q. מתוכננים כך שיתפלגו נורמאלית עם סטיית תקן של 15. במדגם של 20 נבחנים ישראלים התקבל ממוצע ציונים 104. במדגם של 23 נבחנים אמריקאיים התקבל ממוצע ציונים 99.
 - א. בנו רווח סמך ברמת סמך של 95% לפער בין ישראל לארה"ב בממוצע הציונים במבחן ה-IQ.
 - ב. האם קיים הבדל בין ישראלים לאמריקאים מבחינת ממוצע הציונים?

3. חברה להנדסת בניין מעוניינת להשוות ברמת הקשיות של שני סוגי ברגים. ידוע שרמת הקשיות של ברגים מתפלגת נורמלית עם סטיית תקן של 4 יחידות. במדגם של 15 ברגים מסוג א' התקבל רמת קשיות ממוצעת של 28 יחידות ובמדגם של 12 ברגים מסוג ב' התקבל רמת קשיות ממוצעת של 25. עבור אילו רמות בטחון יקבע שאין הבדל בין שני סוגי הברגים מבחינת ממוצע רמת הקשיות שלהם?

פתרונות :**שאלה 1** $(-20, 90)$ **שאלה 3:**

רמות בטחון הגבוהות מ: 0.9476

בשונות האוכלוסייה אינן ידועות אך שוות והמדגמים בלתי תלויים

רקע:

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

האומד נקודתי: $\bar{x}_1 - \bar{x}_2$

התנאים לבניית רווח הסמך:

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

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

3. מדגמים בלתי תלויים.

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

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

דרגות החופש: $d.f = n_1 + n_2 - 2$

רווח סמך:

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

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

דוגמה: (פתרון בהקלטה)

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

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

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

תרגילים:

1. נדגמו 15 ישראלים ו-15 אמריקאים.

כל הנדגמים נגשו למבחן IQ. להלן תוצאות המדגם:

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

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

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

25	21	20	22	X
12	17	25	18	Y

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

פרק 16 - רווח סמך לתוחלת ההפרש במדגם מזווג

רקע:

מדגם מזווג: מדגם אחד שבו יש n צמדנים.

כל תצפית במדגם תנפק זוג ערכים: X ו- Y .

ניצור משתנה חדש:

$$D = x - y$$

הפרמטר שנרצה לאמוד: μ_D

התנאים לבניית רווח הסמך:

• $x, y \sim N$

• המדגם מזווג

נוסחת רווח הסמך:

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

כאשר דרגות החופש: $d.f = n - 1$

דוגמה: (פתרון בהקלטה)

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

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

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

תרגילים:

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

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

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. נגדמו באקראי 7 מדינות ועבור כל מדינה נבדקה עלות דקת שיחה. להלן התוצאות :

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

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

פרק 17 - רווח סמך לשונות וסטיית תקן

רקע:

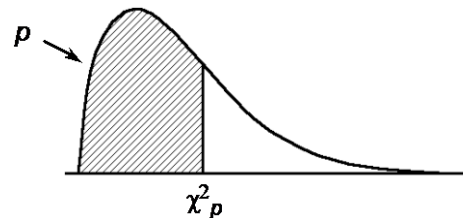
בפרק זה נדון על בניית רווח סמך לשונות האוכלוסייה.

התנאי לבניית רווח הסמך: המשתנה הנחקר מתפלג נורמלית, למרות שנהוג לא לדרוש את התנאי הזה אם המדגם מספיק גדול.

רווח הסמך יתבסס על התפלגות הנקראת חי בריבוע.

התפלגות זו היא התפלגות אסימטרית חיובית המתחילה מהערך אפס ותלויה בדרגות חופש.

דרגות החופש במקרה זה יהיו: $n-1$



$$\text{רווח הסמך לשונות: } \frac{(n-1)S^2}{\chi^2_{1-\alpha/2, n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{\alpha/2, n-1}}$$

$$\text{כאשר } S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} = \frac{\sum_{i=1}^n X_i^2 - n \cdot \bar{X}^2}{n-1} \text{ אומד לשונות הלא-ידועה.}$$

אם נרצה לבנות רווח סמך לסטיית תקן אז נוציא שורש לרווח סמך לשונות.

דוגמה: (פתרון בהקלטה)

זמן התגובה מתפלג נורמלית. במטרה לאמוד את שונות זמן התגובה נדגמו 4 תצפיות. להלן התוצאות בשניות: 4.7, 5.2, 4.6, 5.3. בנו רווח סמך, ברמת סמך של 95% לשונות זמן התגובה באוכלוסייה.

פתרון:

$$0.039 < \sigma^2 < 1.708$$

תרגילים :

1. חמישה מטופלים קבלו תרופה מסוימת. בדקו לכל מטופל את זמני התגובה שלו. להלן הזמנים שהתקבלו בדקות : 18,17,21,26,28.
בהנחה זמני התגובה מתפלגים נורמאלית, בנו רווח סמך ברמת סמך של 95 % לשונות זמן התגובה.
2. נדגמו 20 ימים אקראיים מחודשי יולי-אוגוסט ונמדדה בהם הטמפ' במעלות צלזיוס בת"א. במדגם התקבל טמפ' ממוצעת 30.8 וסטיית תקן מדגמית 1.1. בהנחה והטמפ' מתפלגת נורמאלית :
א. בנו רווח סמך לתוחלת הטמפ' בחודשים אלה בת"א ברמת סמך של 95%.
ב. בנו רווח סמך לסטיית התקן של הטמפ' בחודשים אלה בת"א ברמת סמך של 95%.
3. ציוני IQ בארה"ב מתפלגים נורמאלית עם ממוצע 100 וסטיית תקן 5. נבחנו 20 נבחנים ישראלים במבחן ה-IQ. להלן התוצאות שהתקבלו :

$$\sum_{i=1}^{20} X_i = 2080$$

$$\sum_{i=1}^{20} X_i^2 = 218,220$$

- נניח שגם בישראל הציונים מתפלגים נורמאלית.
- א. מצאו אומדנים לממוצע הציונים בישראל ולשונות הציונים בישראל באמצעות אומדנים חסרי הטיה.
 - ב. אמדו ברמת ביטחון של 95% את תוחלת הציונים של נבחנים בישראל.
 - ג. אמדו ברמת סמך של 90% את סטיית התקן של הציונים של נבחנים ישראלים.
 - ד. על סמך הסעיפים הקודמים, האם בישראל ממוצע הציונים וסטיית התקן של הציונים שונה מבארה"ב? הסבירו.

4. באוכלוסייה מסוימת נדגמו 10 תצפיות והתקבלו התוצאות הבאות :

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

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

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

נתון ש $X_i \sim N(\mu, \sigma^2)$

א. בנו רווח סמך ל- μ ברמת סמך של 95%.

ב. בנו רווח סמך ל- σ^2 ברמת סמך של 95%.

פתרונות :**שאלה 2**

א. $30.285 < \mu < 31.315$

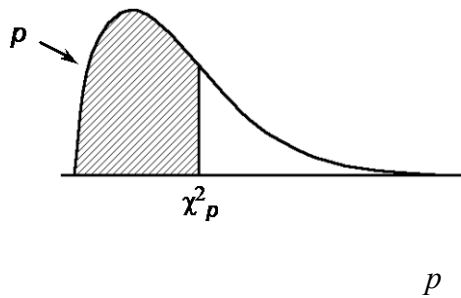
ב. $0.837 < \sigma < 1.607$

תשובה 3

א. לממוצע 104, לשונות 100.

ב. $99.32 < \mu < 108.68$

ג. $7.94 < \sigma < 13.7$

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

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

פרק 18 - רווח סמך ליחס שוניות

רקע:

נרצה לאמוד את ההבדל בין שתי שוניות משתי אוכלוסיות שונות.

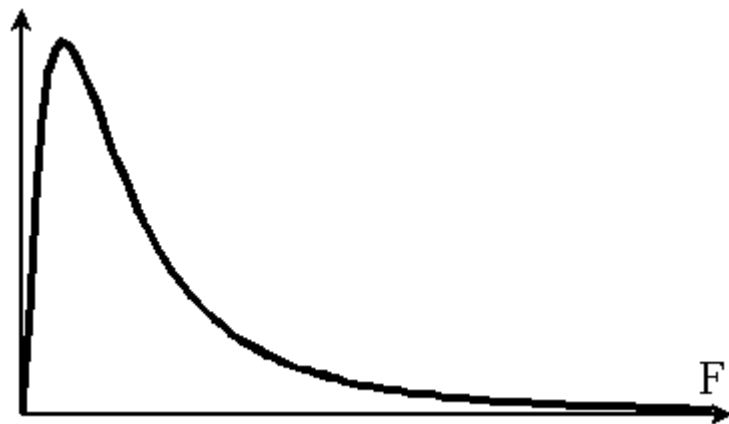
הפרמטר יהיה : $\frac{\sigma_1^2}{\sigma_2^2}$: כלומר היחס בין השוניות.

התנאים:

• $X_1, X_2 \sim N$ או מדגמים גדולים.

• מדגמים בלתי תלויים.

רווח הסמך יבנה על סמך התפלגות הנקראת התפלגות F. התפלגות זו היא אסימטרית חיובית ומושפעת משתי דרגות החופש זו של המונה וזו של המכנה.



$$df_1 = n_1 - 1$$

$$df_2 = n_2 - 1$$

רווח הסמך יהיה:

$$\frac{s_1^2}{s_2^2} \frac{1}{F_{1-\frac{\alpha}{2}}(n_1-1, n_2-1)} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{s_1^2}{s_2^2} F_{1-\frac{\alpha}{2}}(n_2-1, n_1-1)$$

דוגמה: (פתרון בהקלטה)

מחקר סוציולוגי מעוניין לחקור את הרגלי הבילויים בקבוצות גיל שונות :
 במדגם שנעשה על סטודנטים בגילאי 21-26 התקבל אומדן חוסר הטיה לשוונות ההוצאה החודשית
 על בילויים 10,000. כמות הסטודנטים שנדגמו 16.
 במדגם שנעשה על 11 מבוגרים בשנות השלושים התקבל אומדן חסר הטיה לשוונות ההוצאה
 החודשית על בילויים 490,000.
 נניח שההוצאה החודשית לבילוי בכל קבוצת גיל מתפלג נורמאלית.
 בנו רווח סמך ברמת סמך של 95% ליחס בין השונויות.

תרגילים:

1. בתחום הבינוי משתמשים בשני סוגי מתכות: מתכת A ומתכת B. מחקר מעוניין לבדוק האם קיים הבדל בין שני סוגי המתכות מבחינת שונות החוזק שלהן. דגמו מספר יחידות. מתכת מכל סוג והתקבלו התוצאות הבאות:

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

- יש להניח שרמת החוזק של המתכות מתפלגת נורמאלית.
- א. בנו רווח סמך ליחס השונות של רמות החוזק בין שני סוגי המתכות ברמת סמך של 90%.
- ב. בנו רווח סמך ליחס סטיות התקן של רמות החוזק בין שני סוגי המתכות ברמת סמך של 90%.
- ג. האם בבטחון של 90% ניתן לומר שסטיות התקן של שני סוגי המתכות שונות?

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

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

במדגם של 13 נשים התקבלו התוצאות הבאות:

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

- אמוד ברמת בטחון של 95% פי כמה גדולה השונות של הגברים באוכלוסייה מהשונות של הנשים. מה יש להניח לצורך פתרון?

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

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

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

פרק 19 - תרגול מסכם ברווחי סמך

1. מהירות הגלישה באינטרנט במקום מסוים מתפלגת נורמאלית. בדקו את מהירות הגלישה ב-30 זמנים אקראיים. מהירות הגלישה נמדדה ב-Mbps. מהירות מתחת ל-10 Mbps מוגדרת על ידי החברה כנמוכה.
- התוצאות שהתקבלו במדגם: ממוצע היה 87 עם סטיית תקן 17 ו-12 פעמים המהירות הייתה נמוכה. בנו רווחי סמך ברמת סמך של 95% לפרמטרים הבאים:
- א. תוחלת מהירות הגלישה.
- ב. סטיית תקן של מהירות הגלישה.
- ג. הסיכוי שמהירות הגלישה תהיה נמוכה.

2. 200 אנשים נשאלו כמה פעמים ביום הם שותים כוס קפה. להלן התפלגות התשובות:

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

- א. תנו רווח סמך לממוצע מספר כוסות הקפה שאנשים נוהגים לשתות ביום. $\alpha = 0.05$
- ב. אדם השותה לפחות 4 כוסות קפה ביום נקרא "מכור לקפה". בנו רווח סמך לאחוז "המכורים לקפה" $\alpha = 0.1$
3. חוקר בנה רווח סמך לאחוז האנשים שהתקררו לפחות פעם אחת בשנה. רווח הסמך שהתקבל הוא $81 < p < 91$ רווח הסמך הנ"ל התבסס על מדגם של 500 איש.
- א. כמה אנשים במדגם טענו שכלל לא התקררו השנה?
- ב. באיזו רמת סמך נבנה רווח הסמך?
- ג. בנו רווח סמך לאחוז האנשים שהתקררו לפחות פעם אחת השנה ברמת סמך של 95% על סמך תוצאות המדגם

4. ציוני IQ בארה"ב מתפלגים נורמאלית עם תוחלת 100. במדגם של 20 ישראלים שנבחנו במבחן

$$\sum_{i=1}^{20} x_i = 2040$$

ה-IQ התקבלו התוצאות הבאות :

$$\sum_{i=1}^{20} x_i^2 = 210740$$

א. אמדו ברמת ביטחון של 90% את ממוצע ציוני בחינת ה-IQ בישראל – מהי ההנחה

הדרושה לפתרון?

ב. על סמך רווח הסמך של סעיף א האם תקבלו את הטענה שבישראל ממוצע הציונים שונה

מארה"ב?

ג. מה היה קורה לרווח הסמך אם הינו מגדילים את רמת הסמך שלו?

5. להלן תוצאות מדגם שבדק עבור כל משפחה האם יש לה בבית מכשיר טאבלט.

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

א. בנו רווח סמך להבדל בין אחוז המשפחות עם טאבלט בגוש דן ואחוז המשפחות בעלי

טאבלט בשאר חלקי הארץ. ברמת סמך של 98%.

ב. בנו רווח סמך לפרופורצית משפחות בעלות טאבלט בכלל הארץ ברמת סמך של 95%.

6. הגובה של מתגייסים לצה"ל מתפלג נורמלית במדגם של 25 מתגייסים התקבלו התוצאות

הבאות :

$$\bar{x} = 176.2 \text{ cm}$$

$$\sum (x_i - \bar{x})^2 = 2832 \text{ cm}^2$$

א. אמדו את הגובה הממוצע של המתגייסים ברמת סמך של 98%.

ב. אמדו ברמת סמך של 90% את סטיית התקן של הגובה של מתגייסים של צה"ל.

7. בנק מתלבט האם לפתוח סניף באזור A או באזור B. לצורך פתרון נניח שסטית התקן של המשכורת באזור A היא 1200 ובאזור B 1500. הבנק דגם 50 אנשים מאזור A, המשכורת הממוצעת שהתקבלה במדגם היא 6,800 ₪. כמו כן נדגמו 40 אנשים מאזור B, המשכורת הממוצעת שהתקבלה במדגם היא 6,600 ₪.
- א. בנו רווח סמך ברמת סמך של 95% להפרש הממוצעים של המשכורות בשני האזורים. האם על סמך רווח הסמך ניתן להמליץ לבנק היכן לפתוח את הסניף. אם כן, היכן?
- ב. בנו רווח סמך לתוחלת המשכורת באזור A ברמת סמך של 95%.

8. להלן מדגם של שכר הדירה בש"ח של 5 דירות שלושה חדרים בשכונת בבלי בתל אביב :

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

- בנו רווח סמך ברמת סמך של 95% לתוחלת עליית שכר הדירה משנת 2012 לשנת 2013 בשכונת בבלי. ניתן להניח ששכר הדירה בשכונה מתפלג נורמלית.

פתרונות:**שאלה 1**

א. $80.65 \leq \mu \leq 93.35$

ב. $13.5 \leq \sigma \leq 22.9$

ג. $0.225 \leq p \leq 0.575$

שאלה 2

א. $1.21 \leq \mu \leq 1.65$

ב. $10.85\% \leq p \leq 19.15\%$

שאלה 3

א. 70

ב. 99.88%

ג. $83\% \leq p \leq 89\%$

שאלה 4

א. $97.4 \leq \mu \leq 106.6$

ב. לא

ג. יגדל

שאלה 5

א. $0.5\% < p_1 - p_2 < 19.5\%$

ב. $0.704 \leq p \leq 0.768$

שאלה 6

א. $170.8 \leq \mu \leq 181.6$

ב. $8.8 \leq \sigma \leq 14.3$

שאלה 7

$$\text{א. } -372 \leq \mu_A - \mu_B \leq 772$$

$$\text{ב. } 6467 \leq \mu \leq 7133$$

שאלה 8

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

פרק 20 - בדיקת השערות כללית

רקע:

תהליך של בדיקת השערות הוא תהליך מאד נפוץ בעולם הסטטיסטיקה. בתהליך זה ישנן שתי השערות שנבדקות :

השערת האפס המסומנות ב- H_0

והשערה אלטרנטיבית (השערת המחקר) המסומנת ב- H_1 .

בדרך כלל השערת האפס מסמנת את אשר היה מקובל עד עכשיו , את השגרה הנורמה ואילו ההשערה האלטרנטיבית את החדשנות בעצם ההשערה האלטרנטיבית מדברת על הסיבה שהמחקר נעשה .

למשל ,

ישנה תרופה קיימת למחלה A אשר גורמת ל – 10% מהמשתמשים בה לתופעות לוואי . חברת תרופות טוענת שפיתחה תרופה שיעילה באותה מידה , אך מקטינה את הסיכוי לתופעות הלוואי. לכן יש לבצע מחקר שעל סמך תוצאותיו ננסה להכריע איזה השערה נקבל :

H_0 : התרופה החדשה הנה קונבנציונאלית וגורמת ל-10% תופעות לוואי.

H_1 : התרופה החדשה מקטינה את אחוז הסובלים מתופעות לוואי מתחת ל -10%.

בתהליך של בדיקת השערות יוצרים כלל שניקרא כלל הכרעה :

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

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

לכן יתכנו טעויות במסקנות שלנו :

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

הגדרת הטעויות:

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

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

מה הן הטעויות האפשריות במחקר של התרופות? (בהקלטה)

נגדיר את ההסתברויות הבאות:

הסיכוי לבצע טעות מסוג 1 (רמת מובהקות)

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

הסיכוי לבצע טעות מסוג 2 :

$$\beta = P(H_0 \text{ לקבל} | H_1 \text{ נכונה}) = P_{H_1}(H_0 \text{ לקבל})$$

רמת בטחון :

$$(1-\alpha) = P(H_0 \text{ לקבל} | H_0 \text{ נכונה}) = P_{H_0}(H_0 \text{ לקבל})$$

עוצמה :

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

דוגמה: (פתרון בהקלטה)

בכד יש 10 כדורים. יתכן ש- 5 מהם לבנים והיתר שחורים (כד א- השערת האפס) או ש- 7 מהם לבנים והיתר שחורים (כד ב- השערה אלטרנטיבית).

כדי להחליט איזה מהכדים ברשותנו, הוחלט להוציא כדור ולהשתמש בכלל ההחלטה הבא: אם הכדור שהוצא הוא לבן שזהו כד ב' (H_1).

א. חשבו את רמת המובהקות ואת רמת הביטחון של המבחן המוצע.

ב. חשבו את הסיכוי לטעות מסוג שני והעוצמה של המבחן המוצע.

תרגילים:

1. אדם חשוד בביצוע פשע. מהן הטעויות האפשריות בהכרעת הדין?
2. ילד קנה שקית סוכריות אטומה שבה ציפה ל-10 סוכריות תות ו-5 לימון. ישנה שקית אחרת אותה הוא לא רצה בה 6 סוכריות תות ו-9 לימון. הוא החליט להוציא באקראי סוכרייה אם היא תהיה לימון הוא יחזיר את השקית לחנות. מה הסיכויים לכל סוג של טעות בהכרעתו?

3. יהי X מספר שלם הנבחר באקראי מבין המספרים השלמים. הסיכוי ש- X יקבל ערך

$$p(X = k) = \frac{1}{n} \quad \text{עבור } k = 1, 2, \dots, n$$

נתונות ההשערות הבאות לגבי התפלגות של X :

$$H_0 : n = 4$$

$$H_1 : n = 6$$

כמו כן נתון כלל ההכרעה הבא: נדחה את השערת האפס אם $X > 3$.
חשבו את הסיכוי לטעות מסוג ראשון וטעות מסוג שני ואת העוצמה?

4. איכות של מוצר מסווגת ל-4 רמות איכות: מצוין, טוב, בינוני וירוד. להלן התפלגות טיב המוצר בשני מפעלים:

מפעל	מצוין	טוב	בינוני	ירוד
"היוצר"	0.6	0.2	0.2	0
"שמשון"	0.1	0.2	0.3	0.4

בוחרים ממשלוח מוצר באקראי, אך לא יודעים מאיזה מפעל המשלוח הגיע. על סמך בדיקת האיכות מנסים להכריע האם מדובר במפעל "היוצר" (השערת האפס) או במפעל "שמשון" (השערה אלטרנטיבית).

- א. להלן כלל החלטה: אם מדובר במוצר שטיבו "טוב" נכריע שהמוצר בא ממפעל "שמשון", מהן ההסתברויות לסוגי הטעויות השונים?
- ב. להלן כלל החלטה: אם מדובר במוצר שטיבו "בינוני" או גרוע מכך נכריע שהמוצר בא ממפעל "שמשון", מה מהן ההסתברויות לסוגי הטעויות השונים?
- ג. איזה כלל החלטה עדיף? נמק!

5. במטרה לבדוק האם מטבע תקין הטילו אותו 8 פעמים. הוחלט שאם מספר העצים יהיה בין 1 ל 7 כולל יוחלט שהמטבע תקין, אחרת נחליט שהמטבע מזויף.
- א. רשמו את השערות המחקר.
- ב. מה ההסתברות לטעות מסוג ראשון?
- ג. מהי עצמת המבחן אם במציאות אכן המטבע אינו תקין כי הסיכוי לעץ בו הוא 20%.

6. להלן השערות:

$$H_0: X \sim t(5) \quad (\text{התפלגות } t \text{ עם 5 דרגות חופש})$$

$$H_1: X \sim Z \quad (\text{התפלגות נורמאלית סטנדרטית})$$

כלל החלטה: נדחה את השערת האפס אם X גדול מ-2.015.

- א. מהי רמת המובהקות של כלל ההחלטה?
- ב. מהי העוצמה של כלל ההחלטה?
7. במפעל מסוים נפלטים לאוויר חומרים רעילים. במצב שיגרה העוצמה הממוצעת של החומר הרעיל אמורה להיות 6,000 יחידות עם סטיית תקן 900. במצב חירום העוצמה הממוצעת היא 7,000 עם סטיית תקן 900. במפעל מערכת התראה נתמכת על ידי 9 חיישנים. אם ממוצע העוצמה של החומר הרעיל לפי תשעת החיישנים עולה על 6600 יחידות מופעלת מערכת ההתראה. נתון שעוצמת הזיהום מתפלגת נורמאלית.
- א. מה הסיכוי להתראת שווא? (באיזה סוג טעות מדובר)?
- ב. מה הסיכוי שבמצב חירום מערכת ההתראה לא תפעל? (באיזה סוג טעות מדובר)?
- ג. מה ההסתברות שאם המצב הוא מצב חירום מערכת ההתראה תפעל? (איך קוראים להסתברות זו)?
- ד. בסעיפים הבאים נשנה בכל סעיף נתון מסוים. כל סעיף עומד בפני עצמו, כיצד השינוי ישנה את הסיכוי לטעות מסוג ראשון ושני?
1. המפעל יקנה עוד 4 חיישנים.
 2. מצב חרום מוגדר כעת בתוחלת של 7500 יחידות.
 3. מערכת ההתראה תופעל אם ממוצע של תשעת החיישנים יהיה מעל 6700.

8. במטרה לבדוק האם במקום עבודה מסוים פרופורציית הבנים נמוכה מפרופורציית הבנות נדגמו באקראי 10 עובדים. הוחלט שאם מספר הבנים במדגם יהיה לכל היותר 2 תתקבל הטענה שפרופורציית הבנים נמוכה מפרופורציית הבנות.
- א. מה רמת המובהקות של כלל ההכרעה הנ"ל?
- ב. מהי העוצמה בהנחה ובחברה 30% בנים?

9. זמן ההשפעה של משכך הכאבים "אופטלנוס" מתפלג נורמאלית עם תוחלת של 40 דקות וסטיית תקן של 12 דקות.

חברת התרופות המייצרת את התרופה מנסה לשפר את התרופה כך שתוחלת הזמן עד להשפעה תתקצר. לצורך כך, דגמו 25 מטופלים שיקבלו את התרופה "אופטלנוס פורטה", ממוצע זמן התגובה של המטופלים היה 34.5 דקות. חברת התרופות החליטה מראש שאם ממוצע הזמן עד להשפעה יהיה נמוך מ-35 דקות, היא תמשיך בתהליך שיווק "אופטלנוס פורטה".

א. מהי רמת המובהקות של המבחן המוצע?

ב. על סמך תוצאות המדגם. מהי המסקנה ומהי הטעות האפשרית במסקנה?

ג. מהי עצמת המבחן המוצע אם במציאות התרופה "אופטלנוס פורטה" מפחיתה את התוחלת לכדי 32 דקות?

ד. כיצד תשתנה התשובה לסעיף ג' אם החברה הייתה מחליטה שהיא תמשיך בתהליך שיווק התרופה החדשה כאשר ממוצע המדגם יהיה נמוך מ-36 דקות?

פתרונות:**שאלה 2**

$$\beta = \frac{2}{5} \quad \alpha = \frac{1}{3}$$

שאלה 3

$$\beta = 0.5 \quad \alpha = 0.25$$

שאלה 4

$$\beta = 0.8 \quad \alpha = 0.2 \quad \text{א.}$$

$$\beta = 0.3 \quad \alpha = 0.2 \quad \text{ב.}$$

ג. כלל ב'

שאלה 5

$$\text{ב. } 0.00781$$

$$\text{ג. } 0.1678$$

שאלה 6

$$\text{א. } 0.05$$

$$\text{ב. } 0.022$$

שאלה 7

$$\text{א. } 0.0228$$

$$\text{ב. } 0.0918$$

$$\text{ג. } 0.9082$$

שאלה 8

$$\text{א. } 0.055$$

$$\text{ב. } 0.383$$

פרק 21 - בדיקת השערות על פרמטרים

הקדמה

רקע:

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

שלב א: נזהה את הפרמטר הנחקר.

שלב ב: נרשום את השערות המחקר.

השערת האפס המסומנות ב- H_0

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

השערה אלטרנטיבית (השערת המחקר) המסומנת ב- H_1 .

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

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

שלב ד: נרשום את כלל ההכרעה.

בתהליך של בדיקת השערות יוצרים כלל שניקרא **כלל הכרעה**:

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

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

שלב ה:

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

שלב ו:

להסיק מסקנה בהתאם לתוצאות המדגם.

דוגמה: (פתרון בהקלטה)

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

$$n = 20$$

$$\bar{X} = 3120$$

$$S = 280$$

א. מהי אוכלוסיית המחקר?

ב. מה המשתנה הנחקר?

ג. מה הפרמטר הנחקר?

ד. מהן השערות המחקר?

תרגילים:

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

א. מהי אוכלוסיית המחקר?

ב. מה המשתנה הנחקר?

ג. מה הפרמטר הנחקר?

ד. מהן השערות המחקר?

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

א. מהי אוכלוסיית המחקר?

ב. מה המשתנה הנחקר?

ג. מה הפרמטר הנחקר?

ד. מהן השערות המחקר?

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

א. מהי אוכלוסיית המחקר?

ב. מה המשתנה הנחקר?

ג. מה הפרמטר הנחקר?

ד. מהן השערות המחקר?

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

א. מהי אוכלוסיית המחקר?

ב. מה המשתנה הנחקר?

ג. מה הפרמטר הנחקר?

ד. מהן השערות המחקר?

טעויות בבדיקת השערות

רקע:

בתהליך של בדיקת השערות יוצרים כלל שניקרא כלל הכרעה :

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

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

לכן יתכנו טעויות במסקנות שלנו :

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

הגדרת הטעויות:

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

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

דוגמה : (פתרון בהקלטה)

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

תרגילים:

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

א. רשמו את השערות המחקר.

ב. מה מסקנת המחקר?

ג. איזו סוג טעות יתכן וביצעו במחקר?

2. במחקר על פרמטר מסוים הוחלט בסופו של דבר לדחות את השערת האפס.

א. האם ניתן לדעת אם בוצע טעות במחקר?

ב. מה סוג הטעות האפשרית?

3. לפי נתוני משרד הפנים בשנת 1980 למשפחה ממוצעת היה 2.3 ילדים למשפחה עם סטיית תקן

0.4. ישנה טענה שכיום ממוצע מספר הילדים במשפחה קטן יותר. לצורך כך הוחלט לדגום 121

משפחות. במדגם התקבל ממוצע 2.17 ילדים למשפחה. על סמך תוצאות המדגם נקבע שלא

ניתן לקבוע שבאופן מובהק תוחלת מספר הילדים למשפחה קטנה כיום.

א. מהי אוכלוסיית המחקר?

ב. מה המשתנה הנחקר?

ג. מה הפרמטר הנחקר?

ד. מה השערות המחקר?

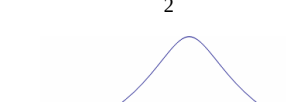
ה. מה מסקנת המחקר?

ו. מהי סוג הטעות האפשרית במחקר?

פרק 22 - בדיקת השערות על תוחלת (ממוצע)

כאשר שונות האוכלוסיה ידועה

רקע:

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

סטטיסטי המבחן :

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

חלופה אחרת לכלל הכרעה :

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

דוגמה : (פתרון בהקלטה)

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

תרגילים:

1. ממוצע הציונים בבחינת הבגרות באנגלית הנו 72 עם סטיית תקן 15 נקודות. מורה טוען שפיתח שיטת לימוד חדשה שתעלה את ממוצע הציונים. משרד החינוך החליט לתת למורה 36 תלמידים אקראיים. ממוצע הציונים של אותם תלמידים לאחר שלמדו בשיטתו היה 75.5. בהנחה שגם בשיטתו סטיית התקן תהייה 15 מה מסקנתכם ברמת מובהקות של 5%?
2. לפי הצהרת היצרן של חברת משקאות מסוימת נפח הנוזל בבקבוק מתפלג נורמלית עם תוחלת 500 סמ"ק וסטיית תקן 20 סמ"ק. אגודת הצרכנים מתלוננת על הפחתת נפח המשקה בבקבוק מהכמות המוצהרת. במדגם שעשתה אגודת הצרכנים התקבל נפח ממוצע של 492 סמ"ק במדגם בגודל 25.
 - א. מה מסקנתכם ברמת מובהקות של 2.5%?
 - ב. האם ניתן לדעת מה תהיה המסקנה עבור רמת מובהקות הגבוהה מ-5%?
3. מהנדס האיכות מעוניין לבדוק אם מכונה מכוילת (מאופסת). המכונה כוונה לחתוך מוטות באורך 50 ס"מ. לפי נתוני היצרן סטיית התקן בחיתוך המוטות היא 0.5 ס"מ. במדגם של 50 מוטות התקבל ממוצע אורך המוט 50.93 ס"מ. מה מסקנתכם ברמת מובהקות של 5%?
4. המשקל הממוצע של הספורטאים בתחום ספורט מסוים הוא 90 ק"ג, עם סטיית תקן 8 ק"ג. לפי דעת מומחים בתחום יש צורך בהורדת המשקל ובשימוש בדיאטה מסוימת שצריכה להביא להורדת המשקל. לשם בדיקת יעילות הדיאטה נלקח מדגם מקרי של 50 ספורטאים ובתום שנה של שימוש בדיאטה התברר שהמשקל הממוצע במדגם זה היה 84 ק"ג. יש לבדוק בר"מ של 10%, האם הדיאטה גורמת להורדת המשקל.
5. לפי מפרט נתון, על עובי בורג להיות 4 מ"מ עם סטיית תקן של 0.2 מ"מ. במדגם של 25 ברגים העובי הממוצע היה 4.07 מ"מ.

קבעו ברמת מובהקות 0.05, האם עובי הברגים מתאים למפרט. הניחו כי עובי של בורג מתפלג נורמלית וסטיית התקן של עובי בורג היא אכן 0.2 מ"מ.

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

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

ב

שאלה 7:

ג

שאלה 8:

א. אותה מסקנה

ב. לא ניתן לדעת.

סיכוי לטעויות ועוצמה כאשר שונות האוכלוסייה ידועה

רקע:

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

נגדיר את ההסתברויות הבאות:

הסיכוי לבצע טעות מסוג 1 (רמת מובהקות)

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

הסיכוי לבצע טעות מסוג 2:

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

רמת בטחון:

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

עוצמה:

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

התהליך לחישוב סיכוי לטעות מסוג שני:

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

כתב ופתר - ברק קנדל ©

$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$	השערת האפס : השערה אלטרנטיבית:
3. σ ידועה 4. $X \sim N$ או מדגם מספיק גדול			תנאים:
$\bar{X} > \mu_0 + Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$	$\bar{X} < \mu_0 - Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$	$\bar{X} > \mu_0 + Z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$ או $\bar{X} < \mu_0 - Z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$	כלל ההכרעה: אזור הדחייה של H_0 :
$P_{H_1}(\bar{X} < \mu_0 + Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}})$	$P_{H_1}(\bar{X} > \mu_0 - Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}})$	$P_{H_1}(\mu_0 - Z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \bar{X} < \mu_0 + Z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}})$	חישוב β :

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

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} : \text{התקנון}$$

דוגמה : (פתרון בהקלטה)

בתחילת השנה חשבון הטלפון הסלולארי הממוצע לאדם היה 200 ₪ עם סטיית תקן של 80 ₪ לחודש. בעקבות כניסתן של חברות טלפון סלולארית חדשות מעוניינים לבדוק האם כיום ממוצע חשבון הטלפון הסלולארי פחת. לצורך בדיקה דגמו באקראי 36 אנשים וחשבון הטלפון הסלולארי שלהם היה 150 ₪ בממוצע לחודש.

א. רשמו את השערות המחקר ובנו כלל הכרעה במונחי חשבון ממוצע מדגמי ברמת מובהקות של 5%.

ב. מה מסקנתכם? איזה סוג טעות אפשרית במסקנה?

ג. נניח שבמציאות כיום החשבון הממוצע הוא 160 ₪. מה הסיכוי לבצע טעות מסוג שני?

ד. אם נקטין את רמת המובהקות מסעיף א', כיצד הדבר ישפיע על התשובה מסעיף ג'?

תרגילים:

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

כתב ופתר - ברק קנדל ©

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. רשום כלל הכרעה במונחי ממוצע מדגם קריטי וברמת מובהקות של 10%.
ב. בהמשך לסעיף א מהי העצמה אם התוחלת שווה ל-195?
ג. הסבר ללא חישוב איך העצמה תשתנה אם רמת המובהקות תהייה 5%?
4. מפעל לייצור צינורות מייצר צינור שקוטרו מתפלג נורמלית עם תוחלת של 50 מ"מ וסטית תקן של 6 מ"מ. במחלקת ביקורת האיכות דוגמים בכל יום 81 צינורות ומודדים את קוטרים,

בכדי לבדוק, בעזרת מבחן סטטיסטי, האם מכונת הייצור מכוילת כנדרש או שקוטר הצינורות קטן מהדרוש.

- א. רשום את ההשערות ואת כלל ההכרעה ברמת מובהקות של 5%.
- ב. אם ביום כלשהו מכונת הייצור התקלקלה והיא מייצרת את הצינורות בקוטר שתוחלתו 48 מ"מ בלבד (סטית התקן לא השתנתה), מה ההסתברות שהתקלה לא תתגלה בביקורת האיכות? כיצד נקראת הסתברות זו?
- ג. הסבר ללא חישוב כיצד התשובה לסעיף ב תשתנה אם רמת המובהקות תגדל.
- ד. הסבר ללא חישוב כיצד התשובה לסעיף ב תשתנה אם התוחלת האמיתית היא 47 ולא 48 מ"מ.

5. להלן השערות של מחקר

$$H_0 : \mu = 50$$

$$H_1 : \mu = 58$$

- מעוניינים לדגום 100 תצפיות. ידוע שסטית התקן של ההתפלגות הינה 20.
- א. בנו כלל הכרעה שהסיכוי לטעות מסוג שני בו הוא 10% . מהי רמת המובהקות?
 - ב. כיצד הייתה משתנה רמת המובהקות אם (כל סעיף בפני עצמו) ?
 1. סטית התקן הייתה יותר גדולה .
 2. הסיכוי לטעות מסוג שני גדול יותר.

השאלות שלהלן הן שאלות רב בררתיות. בחר בכל שאלה את התשובה הנכונה ביותר :

6. אם חוקר החליט להגדיל את רמת המובהקות במחקר שלו אזי :

- א. הסיכוי לטעות מסוג ראשון גדל.
 - ב. העוצמה של המבחן גדלה.
 - ג. הסיכוי לטעות מסוג שני גדל.
 - ד. תשובות א ו-ב נכונות.
7. חוקר ביצע מחקר ובו עשה טעות מסוג שני לכן :
- א. השערת האפס נכונה.
 - ב. השערת האפס נדחתה.
 - ג. השערת האפס לא נדחתה.
 - ד. אף אחת מהתשובות לא נכונה בהכרח.

8. מה המצב הרצוי לחוקר המבצע בדיקת השערה :

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

9. נערך שינוי בכלל ההחלטה של בדיקת השערה מסוימת ובעקבותיו אזור דחיית

H_0 קטן. כל שאר הגורמים נשארו ללא שינוי. כתוצאה מכך :

א. 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 ?

א. טעות מסוג ראשון.

ב. טעות מסוג שני.

ג. טעות מסוג שלישי.

ד. אין טעות במסקנתו.

פתרונות :**שאלה 1:**

א. מעל 6.645

ב. 0.3632

שאלה 2:א. נדחה H_0 אם $\bar{X} < 2.24$ ב. נדחה H_0

ג. 1

שאלה 3:א. נדחה H_0 אם $\bar{X} > 203.29$ או $\bar{X} < 196.71$

ב. 0.8051

ג. תקטן.

שאלה 4:א. נדחה H_0 אם $\bar{X} < 48.9$

ב. 0.0885

ג. תקטן.

ד. תקטן.

שאלה 6:

ד

שאלה 7:

ג

שאלה 8:

ג

שאלה 9:

א

שאלה 10:

ב

קביעת גודל מדגם כששונות האוכלוסייה ידועה

רקע:

$$\begin{aligned} H_0 : \mu &= \mu_0 \\ H_1 : \mu &= \mu_1 \end{aligned}$$

השערות המחקר הן :

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

הנוסחה הבאה נותנת את גודל המדגם הרצוי :

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

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?

תרגילים:

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?)

$$H_0: \mu = \mu_0$$

$$H_1: \mu = \mu_1$$

3. השערות המחקר הן :

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 .

הוכח שגודל המדגם הרצוי לכך יהיה :

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

פתרונות :

שאלה 1:

41

שאלה 2 :

78

שאלה 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:

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

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

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

אם ההשערה היא דו צדדית:

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

מובהקות התוצאה היא גם האלפא המינימלית לדחיית השערת האפס.

השערת האפס: השערה אלטרנטיבה:	$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$
תנאים:	5. σ ידועה 6. $X \sim N$ 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 $2 \cdot P_{H_0}(\bar{X} \leq \bar{x}) \Leftarrow \bar{x} < \mu_0$

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

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

דוגמה: (פתרון בהקלטה)

המשקל הממוצע של מתגייסים לצבא לפני 20 שנה היה 65 ק"ג. מחקר מעוניין לבדוק האם כיום המשקל הממוצע של מתגייסים גבוה יותר. נניח שמשקל המתגייסים מתפלג נורמאלית עם סטיית תקן של 12 ק"ג. במדגם של 16 מתגייסים התקבל משקל ממוצע של 71 ק"ג.

א. מהי מובהקות התוצאה?

ב. מה המסקנה אם רמת המובהקות היא 5% ואם רמת המובהקות היא 1%?

תרגילים:

1. לפניך השערות של מחקר :

$$H_0 : \mu = 70$$

$$H_1 : \mu > 70$$

המשתנה הנחקר מתפלג נורמלית עם סטיית תקן 20. במדגם מאותה אוכלוסייה התקבלו התוצאות הבאות :

$$n = 100$$

$$\bar{x} = 74$$

מהי מובהקות התוצאה?

2. השכר הממוצע במשק בשנת 2012 היה 8800 ₪ עם סטיית תקן 2000. במדגם שנעשה אתמול על 100 עובדים התקבל שכר ממוצע 9500 ₪. מטרת המחקר היא לבדוק האם כיום חלה עליה בשכר. עבור אילו רמות מובהקות שיבחר החוקר יוחלט שחלה עליה בשכר הממוצע במשק?

3. אדם חושד שחברת ממתקים לא עומדת בהתחייבויותיה, ומשקלו של חטיף מסוים אותו הוא קונה מדי בוקר נמוך מ – 100 גרם. חברת הממתקים טוענת מצידה שהיא אכן עומדת בהתחייבויותיה. ידוע כי סטית התקן של משקל החטיף היא 12 גרם. האדם מתכוון לשקול 100 חפיסות חטיפים ולאחר מכן להגיע להחלטה. לאחר הבדיקה הוא קיבל משקל הממוצע של 98.5 גרם.

א. רשמו את השערות המחקר.

ב. מהי רמת המובהקות המינימלית עבורה דוחים את השערת האפס?

ג. מהי רמת המובהקות המקסימלית עבורה נקבל את השערת האפס?

ד. מה המסקנה ברמת מובהקות של 5%?

4. מכונה לחיתוך מוטות במפעל חותכת מוטות באורך שמתפלג 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.

5. אם מקבלים בחישובים אלפא מינימלית (P value) קטנה מאוד, סביר להניח כי החוקר ידחה את השערת האפס בקלות. נכון? לא נכון? נמק.

6. בבדיקת השערות התקבל שה- $p\text{-value}=0.02$.
מה תהיה מסקנת חוקר המשתמש ברמת מובהקות 1%? בחר בתשובה הנכונה.
א. יקבל את השערת האפס בכל מקרה.

ב. ידחה את השערת האפס מקרה.

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

ד. לא ניתן לדעת כי אין מספיק נתונים.

7. מובהקות התוצאה (PV) היא גם : (בחר בתשובה הנכונה)

א. רמת המובהקות המינימלית לדחות השערת האפס.

ב. רמת המובהקות המקסימלית לדחיית השערת האפס.

ג. רמת המובהקות שנקבעת מראש על ידי החוקר טרם קיבל את תוצאות המחקר.

ד. רמת המובהקות המינימלית לאי דחיית השערת האפס.

8. בבדיקת השערות מסוימת התקבל $p\text{ value}=0.0254$ לכן (בחר בתשובה הנכונה):

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

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

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

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

פתרונות :**שאלה 1:**

0.0228

שאלה 2:

עבור כל רמת מובהקות סבירה.

שאלה 3:

ב. 0.1056

ג. 0.1056

ד. נכריע שיש עמידה בהתחייבות של החברה.

שאלה 4:

א. 0.0006

ב. יקטן.

ג. נכריע שאין כיוול.

שאלה 5:

נכון

שאלה 6:

תשובה :א

שאלה 7:

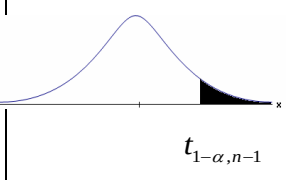
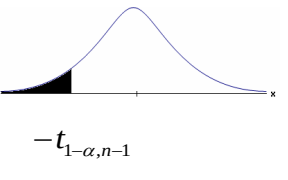
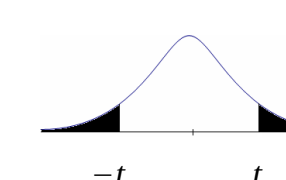
תשובה : א

שאלה 8:

תשובה : ג

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

רקע:

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

סטטיסטי המבחן :

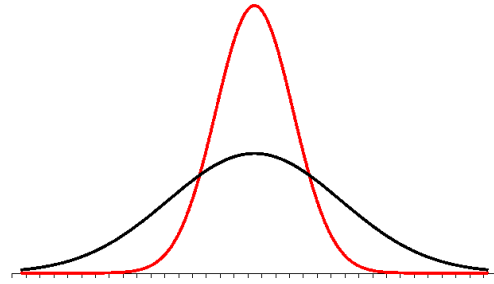
$$t_{\bar{x}} = \frac{\bar{x} - \mu_0}{S / \sqrt{n}}$$

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

התפלגות T:

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

כתב ופתר - ברוך קנדל ©



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. בדוק ברמת מובהקות של 5%.

תרגילים:

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

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

$$n = 20$$

$$\bar{x} = 3120$$

$$S = 280$$

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

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

4. באוכלוסייה מסוימת נדגמו 10 תצפיות והתקבלו התוצאות הבאות:

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

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

נתון שההתפלגות היא נורמלית.

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

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

ליאור השתמש בטבלה של התפלגות Z .

רוני השתמשה בטבלה של התפלגות t .

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

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

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

ג. שני החוקרים בהכרח יגיעו לאותה מסקנה.

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

6. נתון $X \sim N(\mu, \sigma^2)$ 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.

בחר בתשובה הנכונה :

א. כעת בברור הוא ידחה את השערת האפס.

ב. כעת הוא דווקא יקבל את השערת האפס.

ג. כעת לא ניתן לדעת מה תהיה מסקנתו.

פתרונות:

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

כתב ופתר - ברק קנדל ©

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

התשובה היא : ב

שאלה 6:

התשובה היא : ג

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

רקע:

נזכיר שהמסקנה של המחקר תקבע לפי העיקרון הבא :

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

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

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

אם ההשערה היא דו צדדית :

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

מובהקות התוצאה היא גם האלפא המינימלית לדחיית השערת האפס.

השערת האפס : השערה אלטרנטיבה :	$H_0 : \mu = \mu_0$ $H_1 : \mu > \mu_0$	$H_0 : \mu = \mu_0$ $H_1 : \mu < \mu_0$	$H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$
תנאים :	9. σ אינה ידועה 10. $X \sim N$ 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 : 27,34,32,40,30 . הנח שזמן הנסיעה מתפלג נורמלית.

א. רשום את השערות המחקר.

ב. מצא חסמים למובהקות התוצאה.

ג. מה המסקנה ברמת מובהקות של 5% ?

תרגילים :

1. קו ייצור אריזות סוכר נארזות כך שהמשקל הממוצע של אריזות הסוכר צריך להיות אחד קילוגרם. בכל יום דוגמים מקו הייצור 5 אריזות במטרה לבדוק האם קו הייצור תקין. בבדיקה דגמו 5 אריזות סוכר ולהלן משקלן בגרמים :

$$1008, 1024, 996, 1005, 997$$
 - א. רשמו את השערות המחקר.
 - ב. מהי מובהקות התוצאה? הצג חסמים.
 - ג. מה המסקנה ברמת מובהקות של 5%?

2. חוקר בדק את הטענה כי פועלים העובדים במשמרת לילה איטיים יותר מפועלים העובדים ביום. ידוע כי משך הזמן הממוצע הדרוש לייצר מוצר מסוים ביום הוא 6 שעות. במדגם מיקרי של 25 פועלים שעבדו במשמרת לילה נמצא כי הזמן הממוצע לייצר אותו מוצר הוא 7 שעות עם סטית תקן של 3 שעות.
 מהי ה- α המינימלית שלפיה ניתן להחליט שאכן העובדים במשמרת לילה איטיים יותר ?

3. הגובה של מתגייסים לצה"ל מתפלג נורמלית . במדגם של 25 מתגייסים מדדו את הגבהים שלהם בס"מ והתקבלו התוצאות הבאות :

$$\bar{x} = 176.2$$

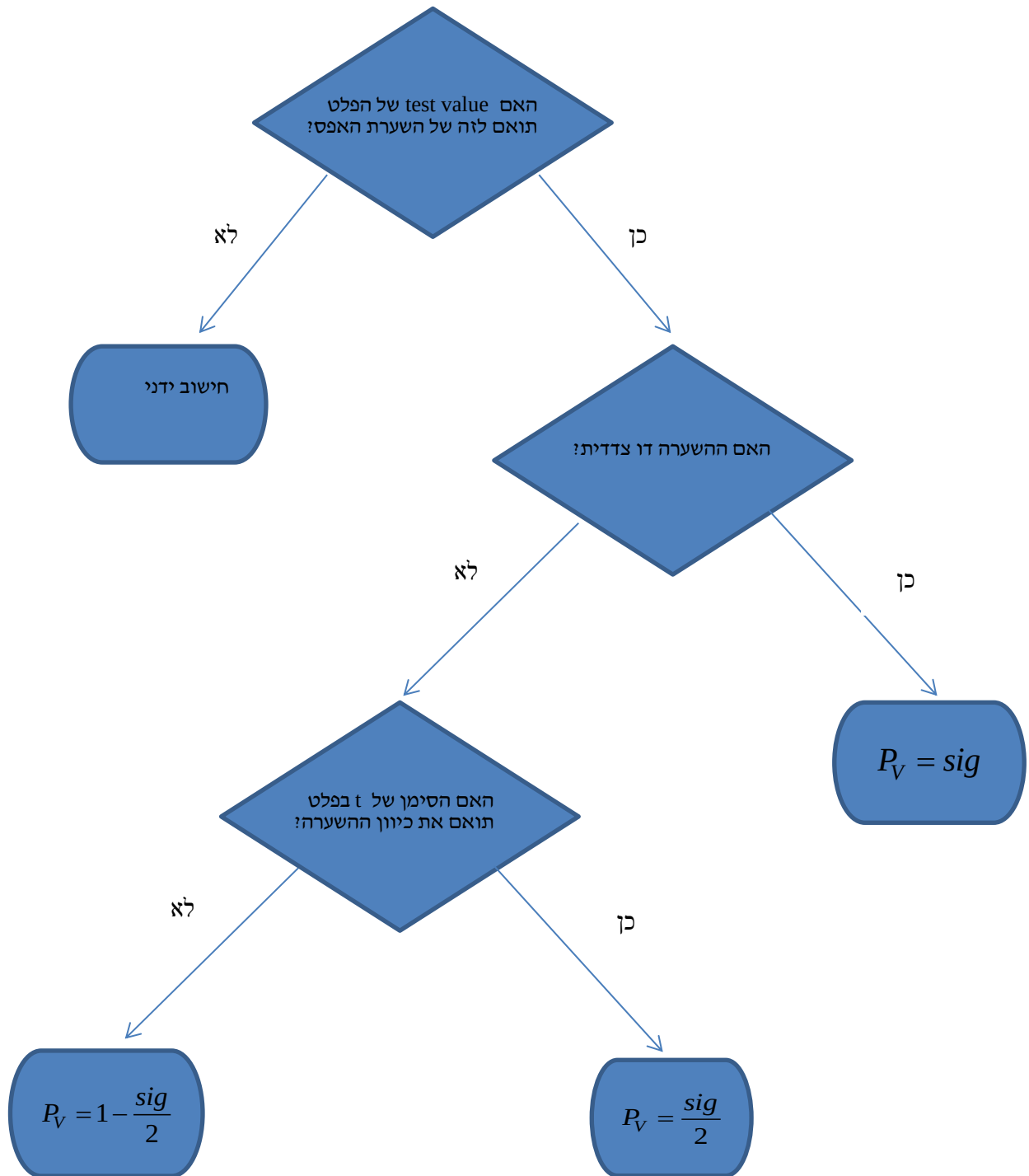
$$\sum (x_i - \bar{x})^2 = 2832$$
 מטרת המחקר היא לבדוק האם תוחלת הגבהים של המתגייסים גבוה מ-174 ס"מ באופן מובהק. מהי בקרוב מובהקות התוצאה ועל פיה מה תהיה המסקנה ברמת מובהקות של 6% ?

פתרונות :**שאלה 3:**נקבל H_0

ניתוח פלטי SPSS

רקע:

חישוב מובהקות התוצאה באמצעות פלט תוכנת SPSS



דוגמה: (פתרון בהקלטה)

One-Sample Statistics

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

One-Sample Test

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

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

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

ב. מהו P_V לבדיקת יעילותה של שיטת הלימוד החדשה?

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

תרגילים :

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

One-Sample Statistics

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

One-Sample Test

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

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

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

One-Sample Test

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

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

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

One-Sample Statistics

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

One-Sample Test

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

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

4. יצרן סיגריות מפרסם כי תוחלת הניקוטין בסיגריות שהוא מיצר קטנה מ- 27 מ"ג. בבדיקה מקרית של 5 סיגריות מתוצרתו נמצאו כמויות הניקוטין הבאות:

21, 20, 24, 22 מ"ג. הנח כי כמות הניקוטין בסיגריות מפולג נורמאלי.

One-Sample Statistics

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

One-Sample Test

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

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

פתרונות :**שאלה 1:**

ב. 23

ג. (22.104, 24.771)

ד. נקבל H_0 **שאלה 2:**

א. 25

ב. (4.3056, 4.5104)

ד. נדחה H_0 **שאלה 3**א. $n = 44$

$$\bar{X} = 67.66$$

$$t = -2.48$$

$$upper = -1.3735$$

ב. 0.017

ג. 0.0085

ד. 0.9915

שאלה 4:

א. נכריע שיש אמת בפרסום.

ב. מסקנה לא תשתנה.

ג. נכריע שהתוחלת שונה מ-26.

הקשר בין רווח סמך לבדיקת השערות על תוחלת

רקע:

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

$$H_0 : \mu = \mu_0$$

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

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

אם μ_0 נופל ברווח $\leftarrow$ נקבל את H_0

אם μ_0 לא נופל ברווח $\leftarrow$ נדחה את H_0

דוגמה: (פתרון בהקלטה)

חוקר ביצע בדיקת השערות לתוחלת. להלן השערותיו:

$$H_0 : \mu = 80$$

$$H_1 : \mu \neq 80$$

$$\alpha = 5\%$$

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

האם אפשר לדעת מה מסקנתו, ואם כן מהי?

תרגילים :

1. חוקר רצה לבדוק את ההשערות הבאות :

$$H_0 : \mu = 90$$

$$H_1 : \mu \neq 90$$

החוקר בנה רווח סמך לתוחלת ברמת סמך של 95% וקיבל את רווח הסמך הבא : (87,97).
 אם החוקר מעוניין לבצע בדיקת השערות ברמת מובהקות של 1% האם ניתן להגיע למסקנה ע"ס רווח הסמך? נמקו.
2. חוקר מעוניין לבדוק השפעת דיאטה חדשה על רמת הסוכר בדם. ידוע כי מספר מיליגרם הסוכר בסמ"ק דם הוא משתנה מקרי שמתפלג נורמלית עם סטיית תקן 10.4 מ"ג. נלקח מדגם של 60 נבדקים שניזונו מדיאטה זו. נמצא כי ממוצע מספר המיליגרם סוכר היה 115.5 מ"ג לסמ"ק.

א. בנה רווח סמך ברמת סמך 95% לתוחלת רמת הסוכר בדם אצל הניזונים מדיאטה זו.
 ב. ידוע שתוחלת רמת הסוכר בדם באוכלוסיה היא 90 מ"ג לסמ"ק. האם לדעתך ניתן להסיק על סמך תוצאת סעיף א שהדיאטה משפיעה על רמת הסוכר בדם? הסבר.
3. יצרן אנטיביוטיקה רושם על גבי התרופות שכמות הפנצלין היא 200 מ"ג לקפסולה. משרד הבריאות ביצע מדגם של 8 קפסולות אקראיות מקו הייצור ומצא שבממוצע יש 196 מ"ג פנצלין לקפסולה עם סטיית תקן מדגמית של 5 מ"ג. בהנחה וכמות הפנצלין בקפסולה מתפלגת נורמלית.

א. בנה רווח סמך ברמת סמך של 95% לממוצע כמות הפנצלין לקפסולה המיוצרת על ידי יצרן האנטיביוטיקה.
 ב. בדוק ברמת מובהקות של 5% האם יש אמת באינפורמציה המסופקת על ידי היצרן.

פתרונות :**שאלה 1:**

1. נקבל השערת H_0

שאלה 2:

א. $112.87 \leq \mu \leq 118.13$

ב. נכריע שהדיאטה משפיעה על תוחלת רמת הסוכר בדם.

שאלה 3:

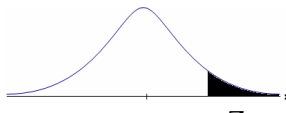
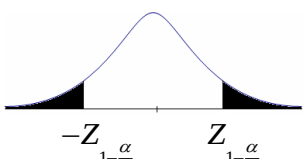
א. $191.8 \leq \mu \leq 200.2$

ב. נכריע שיש אמת בפרסום.

פרק 23 - בדיקת השערות על פרופורציה

התהליך

רקע:

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

סטטיסטי המבחן :

$$Z_{\hat{p}} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

חלופה אחרת לכלל הכרעה:

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

דוגמה: (פתרון בהקלטה)

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

תרגילים:

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

$$H_0 : p = p_0$$

$$H_1 : p > p_0$$

חוקר א השתמש ברמת מובהקות α_1 וחוקר ב ברמת מובהקות α_2 החוקר הראשון דחה את H_0 ואילו החוקר השני קיבל את H_0 . שניהם התבססו על אותם תוצאות של מדגם. בחר בתשובה הנכונה:

 - א. $\alpha_1 = \alpha_2$
 - ב. $\alpha_1 > \alpha_2$
 - ג. $\alpha_1 < \alpha_2$
 - ד. המצב המתואר לא אפשרי.

פתרונות :**שאלה 1:** H_0 נדחה**שאלה 2:** H_0 נדחה**שאלה 3:** H_0 נקבל**שאלה 4:**ב. נקבל H_0

ג. המסקנה לא תשתנה.

שאלה 5:א. נדחה H_0

ב. המסקנה לא תשתנה.

שאלה 6:

התשובה היא : ג.

סיכוי לטעויות ועוצמה

רקע:

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

נגדיר את ההסתברויות הבאות:

הסיכוי לבצע טעות מסוג 1 (רמת מובהקות):

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

הסיכוי לבצע טעות מסוג 2:

$$\beta = P(H_0 \text{ לקבל} | H_1 \text{ נכונה}) = P_{H_1}(H_0 \text{ לקבל})$$

רמת בטחון:

$$(1-\alpha) = P(H_0 \text{ לקבל} | H_0 \text{ נכונה}) = P_{H_0}(H_0 \text{ לקבל})$$

עוצמה:

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

התהליך לחישוב סיכוי לטעות מסוג שני:

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

כתב ופתר - ברק קנדל ©

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

כאשר : $\hat{P} \sim N(p, \frac{p(1-p)}{n})$

והתקנון : $Z_{\hat{p}} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$

דוגמה: (פתרון בהקלטה)

רופאי שיניים טוענים שיותר ממחצית האוכלוסייה הבוגרת בארץ אינם מבקרים אצל רופא שיניים באופן קבוע, כנדרש. כדי לבדוק טענה זו, נערך סקר בקרב 150 אנשים בוגרים.

א. רשמו את ההשערות וכלל הכרעה ברמת מובהקות של 10%.

ב. מהי עוצמת המבחן אם מסתבר ש 60% מהאוכלוסייה אינם מבקרים אצל רופא שיניים באופן קבוע.

תרגילים:

1. משרד הבריאות פרסם ש 10% מתושבי המדינה סובלים ממחלת האסטמה. מחקר דורש לבדוק האם בחיפה, בגלל זיהום האוויר, שיעור הסובלים מאסטמה גבוה יותר. לצורך המחקר נבדקו 260 מתושבי חיפה.
 - א. רשמו את השערות המחקר, וצרו מבחן ברמת מובהקות של 5% לבדיקתן.
 - ב. מהי עצמת המבחן של סעיף א' בהנחה ובחיפה 16% מהתושבים סובלים מאסטמה?
 - ג. כיצד תשנה התשובה לסעיף ב' אם מסתבר שבחיפה 18% סובלים מאסטמה?
 - ד. בהמשך לסעיף א' האם נכון לומר שבהסתברות של 5% ההשערה שבחיפה 10% מהתושבים סובלים מאסטמה אינה נכונה?
2. אחוז הסובלים מתופעות הלוואי מתרופה מסוימת הוא 15%. חברת תרופות טוענת שפיתחה תרופה שאמורה לצמצם את אחוז הסובלים מתופעות לוואי. לצורך בדיקת הטענה הוחלט לבצע מחקר שיכלול 120 חולים שיקבלו את התרופה הנבדקת.
 - א. נניח שהתרופה נבדקת אכן מורידה את פרופורציות הסובלים מתופעות הלוואי ל-10% מהי עצמת המבחן עבור רמת מובהקות של 5%?
3. בעיר מסוימת היו 20% אקדמאים. בעקבות פתיחת מכללה בעיר לפני כמה שנים מעוניינים לבדוק האם אחוז האקדמאים גדל. מעוניינים שהמחקר יכלול 200 אנשים והוא יהיה ברמת מובהקות של 5%.
 - א. חשבו את הסיכוי לבצע טעות מסוג שני בהנחה והיום יש 28% אקדמאים.
 - ב. כיצד התשובה לסעיף הקודם תשתנה אם נגדיל את רמת המובהקות?
4. מעוניינים לבדוק האם בפקולטה מסוימת ישנה העדפה לגברים. הוחלט לדגום 200 מתקבלים ועל סמך מספר הבנים לקבוע אם טענת המחקר מתקבלת.
 - א. חוקר א' קבע רמת מובהקות של 5% וחוקר ב' החליט לקבל את טענת המחקר אם במדגם יהיו לפחות 120 בנים. למי מבין החוקרים רמת מובהקות גדולה יותר?
5. חוקר ביצע מחקר ובו עשה טעות מסוג שני לכן (בחר בתשובה הנכונה)
 - א. השערת האפס נכונה.
 - ב. השערת האפס נדחתה.
 - ג. השערת האפס לא נדחתה.
 - ד. אף אחת מהתשובות לא נכונה בהכרח.
6. קבע אם הטענה הבאה נכונה :

"בבדיקת השערות לא ניתן לבצע בו זמנית טעות מסוג ראשון וטעות מסוג שני"

פתרונות:

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

כתב ופתר - ברק קנדל ©

שאלה 1:

ב. 0.9015

ג. תגדל

ד. טענה לא נכונה.

שאלה 2:

0.4404

שאלה 3:

א. 0.1446

ב. תקטן.

שאלה 4:

חוקר א.

שאלה 5:

התשובה הנכונה היא ג.

שאלה 6:

נכונה.

קביעת גודל מדגם**רקע:**

$$\begin{aligned} H_0 : p &= p_0 \\ H_1 : p &= p_1 \end{aligned} \quad \text{השערות המחקר הן :}$$

מעוניינים לבצע מחקר שרמת המובהקות לא תעלה על α והסיכוי לטעות מסוג שני לא יעלה על β .

הנוסחה הבאה נותנת את גודל המדגם הרצוי :

$$n \geq \left(\frac{Z_{1-\alpha} \sqrt{p_0 q_0} + Z_{1-\beta} \sqrt{p_1 q_1}}{p_0 - p_1} \right)^2$$

דוגמה: (פתרון בהקלטה)

רוצים לבדוק האם אחוז האנשים השוהים בשמש ללא הגנה ירד בעקבות הפרסומת על נזקי השמש. בעבר 60% מהאוכלוסייה שהתה בשמש ללא הגנה. מה גודל המדגם המינימלי שיש לקחת כדי לבדוק שהאחוז הנ"ל ירד ל-48% אם מעוניינים שהסיכוי לטעות מסוג ראשון יהיה 5% והסיכוי לטעות מסוג שני יהיה 1%?

תרגילים:

1. משרד התמ"ת פרסם שאחוז האבטלה במשק היום עומד על 8%. לעומתו, משרד הפנים טוען שחלה עלייה בשיעור האבטלה עד לכדי 11%.
כדי לבדוק מי מבניהם צודק, מה צריך להיות גודל המדגם שיענה על שני התנאים הבאים:
 - אם משרד התמ"ת צודק, נדחה את טענתו בסיכוי של 10%.
 - אם משרד הפנים צודק, נדחה את טענתו בסיכוי של 4%.

2. מפעיל קזינו מפרסם שהסיכוי לזכות במכונת מזל הינו 0.42.
אדם טוען שהסיכויים לזכות במשחק נמוכים יותר. כמה פעמים יש לשחק את המשחק כדי שאם טענת מפעיל הקזינו נכונה נקבל את טענת האדם בסיכוי של 1% ואם במציאות הסיכוי לזכות במכונה הוא 0.3 נקבל את מפעיל הקזינו בסיכוי של 8%.

פתרונות:

שאלה 1:

891

שאלה 2:

224

מובהקות התוצאה

רקע:

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

אם $p_v \leq \alpha$ דוחים את H_0

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

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

אם ההשערה היא דו צדדית:

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

מובהקות התוצאה היא גם האלפא המינימלית לדחיית השערת האפס.

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

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

והתקנון:

$$Z_{\hat{p}} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

דוגמה: (פתרון בהקלטה)

ישנה טענה שיש הבדל בין אחוז הבנים ואחוז הבנות הפונים ללמוד להנדסאי מחשבים. לשם כך נלקח מדגם מקרי של 200 תלמידים הלומדים מחשבים והתברר כי 112 מהם בנים.

א. מהי מובהקות התוצאה?

ב. מה המסקנה ברמת מובהקות של 5%?

תרגילים:

1. במשך שנים אחוז המועמדים שהתקבל לפקולטה מסוימת היה 25%. השנה מתוך מדגם של 120 מועמדים התקבלו 22. רוצים לבדוק האם השנה הקשו על תנאי הקבלה.
 - א. מהי מובהקות התוצאה?
 - ב. מה תהיה המסקנה ברמת מובהקות של 1% וברמת מובהקות של 5%?
2. נהוג לחשוב ש 60% מהילדים בגיל שלוש קמים מהמיטה במהלך הלילה לפחות פעם אחת. ישנה טענה שללא שנת צהריים פחות מ-60% מהילדים בגיל זה יקומו לפחות פעם אחת במהלך הלילה. נדגמו 80 ילדים בגיל 3 אשר אינם ישנים בצהריים מתוכם התקבל ש 41 קמו במהלך הלילה.
 - א. מהי רמת המובהקות המינימלית עבורה תתקבל הטענה במחקר?
 - ב. מהי רמת המובהקות המקסימלית עבורה לא תתקבל טענת המחקר?
 - ג. עבור אילו רמות מובהקות נקבל את טענת המחקר?
 - ד. מה תהיה מסקנת המחקר ברמת מובהקות של 6%?
3. במטרה לבדוק האם מטבע הוא הוגן מטילים אותו 80 פעמים. התקבל ש 60 מההטלות הראו עץ. רשמו את השערות המחקר, חשבו את מובהקות התוצאה והסיקו מסקנה ברמת מובהקות של 5%.
4. בבדיקת השערות על פרופורציה התקבל שה- $p\text{-value}=0.02$.
 - מה תהיה מסקנת חוקר המשתמש ברמת מובהקות 5% : (בחר בתשובה הנכונה)
 - א. יקבל את השערת האפס
 - ב. ידחה את השערת האפס.
 - ג. לא ניתן לדעת כי אין מספיק נתונים.
5. קבע אם הטענה הבאה נכונה :

"במבחן לבדיקת השערות חד-צדדי התקבל ערך $p\text{-value}$ של 3% לכן אם היינו מבצעים מבחן דו-צדדי (כאשר יתר הנתונים ללא שינוי) היינו מקבלים ערך $p\text{-value}$ של 6%"
6. במפעל 10% מהעובדים נפגעים לפחות פעם אחת בשנה מתאונות עבודה. לאור זאת, המפעל החליט לצאת בתוכנית לצמצום שיעור הנפגעים. תכנית זו נוסתה על 100 עובדים. מתוכם 12 נפגעו בתאונות עבודה במשך השנה. מהי רמת המובהקות הקטנה ביותר עבורה יוחלט שהתכנית יעילה?

פתרונות :

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

כתב ופתר - ברק קנדל ©

שאלה 1:

א. 0.0455

שאלה 2:

א. 0.0548

ב. 0.0548

ג. מעל 0.0548

ד. נכריע לטובת טענת המחקר.

שאלה 3:

$$p_v = 0$$

שאלה 4:

התשובה הנכונה : ב

שאלה 5:

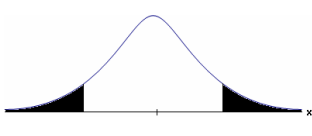
הטענה נכונה

שאלה 6:

0.7486

פרק 24 - בדיקת השערות על הפרש פרופורציות

רקע:

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

סטטיסטי המבחן :

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

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

חלופה אחרת לכלל הכרעה:

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

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

התפלגות של $\hat{p}_1 - \hat{p}_2$:

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

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

תקנון:

דוגמה: (פתרון בהקלטה)

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

כתב ופתר - ברק קנדל ©

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

תרגילים:

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

פתרונות:

שאלה 1:

לא נדחה את H_0

שאלה 2:

א. נדחה H_0

ב. נדחה H_0

שאלה 3:

א. 0.0139

ב. נדחה H_0

שאלה 4:

ב. נדחה H_0

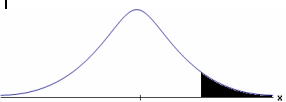
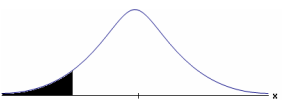
ג. 0.8238

ד. תגדל

פרק 25 - בדיקת השערות על הפרש תוחלות במדגמים בלתי תלויים

כשהשונות של האוכלוסייה ידועות

רקע:

$H_0 \quad \mu_1 - \mu_2 = c$ $H_1 \quad \mu_1 - \mu_2 > c$	$H_0 \quad \mu_1 - \mu_2 = c$ $H_1 \quad \mu_1 - \mu_2 < c$	$H_0 \quad \mu_1 - \mu_2 = c$ $H_1 \quad \mu_1 - \mu_2 \neq c$	השערות האפס: השערה אלטרנטיבית:
1. מדגמים בלתי תלויים 2. σ_1, σ_2 ידועות 3. $X_1, X_2 \sim N$ 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	$Z_{\bar{x}_1 - \bar{x}_2} > Z_{1-\frac{\alpha}{2}}$ או $Z_{\bar{x}_1 - \bar{x}_2} < -Z_{1-\frac{\alpha}{2}}$  $-Z_{1-\frac{\alpha}{2}} \quad Z_{1-\frac{\alpha}{2}}$ דוחים את H_0	כלל ההכרעה: אזור הדחייה של H_0:

סטטיסטי המבחן:

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

חלופה אחרת לכלל הכרעה:

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

התפלגות הפרש הממוצעים:

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

התקנון:

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

דוגמה: (פתרון בהקלטה)

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

כתב ופתר - ברק קנדל ©

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

תרגילים :

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

פתרונות:**שאלה 1:**נדחה H_0 **שאלה 2:**לא נדחה את H_0 **שאלה 3:**

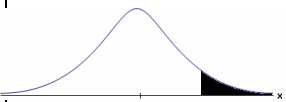
א. 0

ב. נדחה H_0 **שאלה 4:**א. נדחה H_0 אם במדגם הגברים יהיו גבוהים בממוצע מהנשים ביותר מ-10.72 ס"מ.

ב. 0.6331

בששונות האוכלוסיה לא ידועות ומניחים שהן שוות

רקע:

השערת האפס : השערה אלטרנטיבית:			תנאים:
$H_0 \quad \mu_1 - \mu_2 = c$ $H_1 \quad \mu_1 - \mu_2 > c$	$H_0 \quad \mu_1 - \mu_2 = c$ $H_1 \quad \mu_1 - \mu_2 < c$	$H_0 \quad \mu_1 - \mu_2 = c$ $H_1 \quad \mu_1 - \mu_2 \neq c$	
4. מדגמים בלתי תלויים 5. σ_1, σ_2 לא ידועות אך שוות 6. המשתנים בכל אוכלוסיה מתפלגים נורמלית			אזור הדחייה של 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	$t_{\bar{x}_1 - \bar{x}_2} < -t_{1-\alpha}^{(n_1+n_2-2)}$  $-t_{1-\alpha, n_1+n_2-2}$ דוחים את H_0	$t_{\bar{x}_1 - \bar{x}_2} < -t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)}$ 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	

סטטיסטי המבחן :

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

חלופה אחרת לכלל הכרעה:

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

דוגמה : (פתרון בהקלטה)

חברה המייצרת מוצרי בנייה טוענת שפיתחה סגסוגת (תערובת מתכות) שטמפרטורת ההתכה שלה גבוהה משמעותית מטמפרטורת ההתכה של הסגסוגת לבנייה שמשמשים בה כיום לבניית בניינים.

לצורך בדיקת טענת המחקר נדגמו 10 יחידות של מתכות מהסוג הישן ו-12 יחידות של מתכות מהסוג החדש.

להלן תוצאות המדגם:

טמפרטורת ההתכה הממוצעת במתכת הישנה 1170 מעלות עם אומד חסר הטיה לשונות $S^2 = 200$.

טמפרטורת ההתכה הממוצעת במתכת החדשה 1317 מעלות עם אומד חסר הטיה לשונות $S^2 = 260$.

נניח לצורך פתרון שטמפרטורת ההתכה מתפלגת נורמאלית עם אותה שונות במתכות השונות. בדקו ברמת מובהקות של 5%.

תרגילים:

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

כתב ופתר - ברק קנדל ©

1. להלן נתונים של שטחי דירות מתוך דירות שנבנו בשנת 2012 ובשנת 2013 (מטרים רבועים):

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

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

2. נדגמו 15 ישראלים ו-15 אמריקאים. הנדגמים נגשו למבחן IQ. להלן תוצאות המדגם:

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

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

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

הקבוצה	2-60W	1-100W
$\bar{x}$	1007	956
S	80	72
n	13	15

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 שנוורות מסוג W60 דולקות בממוצע יותר מאשר נורות מסוג W100?

ג. בדקו ברמת מובהקות של 5% האם נורות מסוג W60 דולקות יותר מ-1000 שעות. רשמו את כל ההנחות הדרושות.

פתרונות:

שאלה 1:

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

כתב ופתר - ברק קנדל ©

לא נדחה H_0

שאלה 2:

שאלה 3:

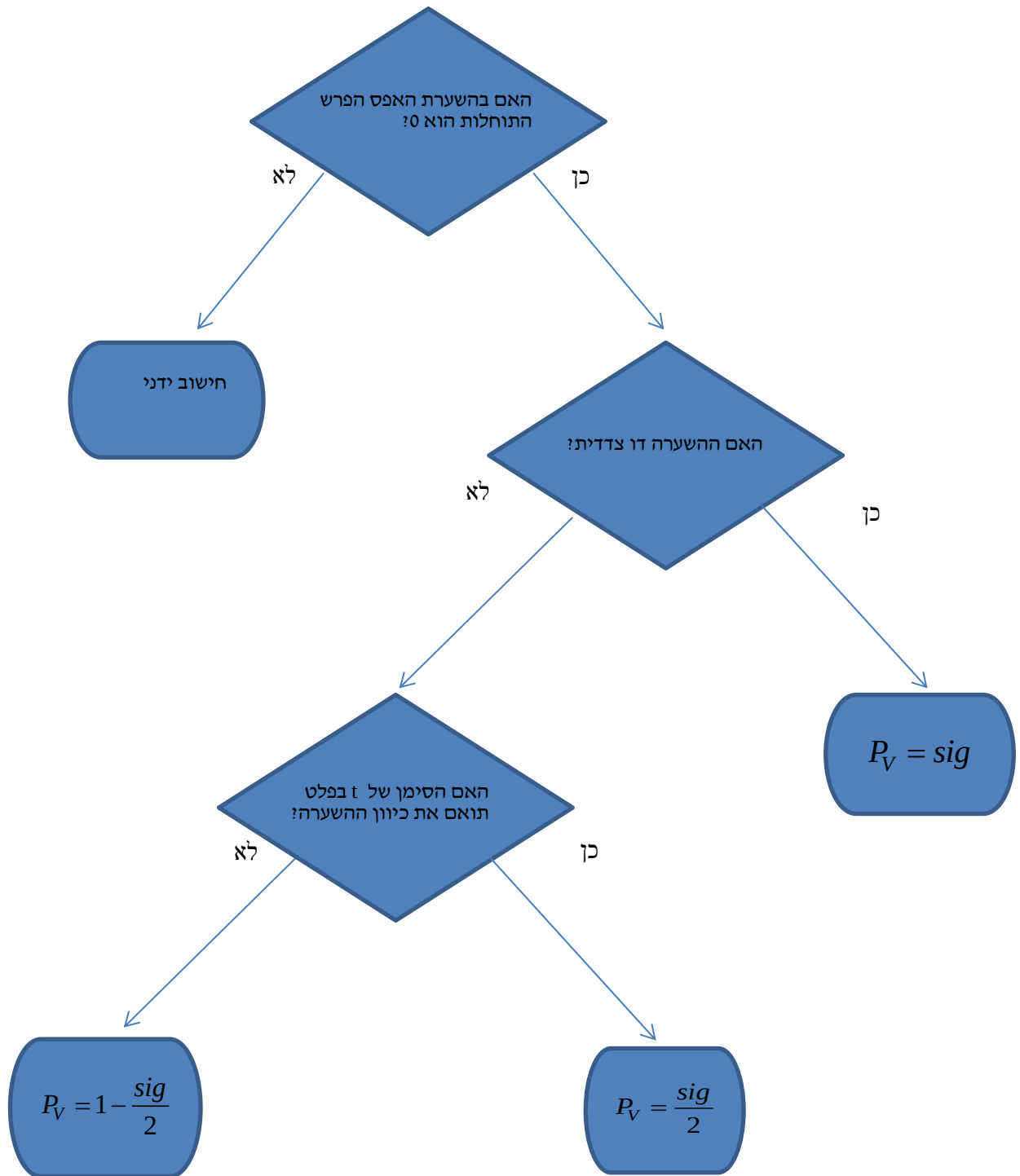
א. נדחה H_0

ב. רמות מובהקות של לפחות 5%

ג. לא נדחה H_0

ניתוח פלטיםרקע:

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 שההפרש הממוצע בין-"South East"

לבין-"North East" חיובי ?

תרגילים :

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

Group Statistics

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

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
optimizm	Equal variances assumed	.383	.536	1.935	1199	.053	.05500	.02842	-.00076	?
	Equal variances not assumed			1.938	1190.977	.053	.05500	.02838	-.00068	.11067

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

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

T-Test

Group Statistic:

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

Independent Samples Test

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

מקרא:

A = נגשו פעם אחת.

B = נגשו יותר מפעם אחת.

- א. רשמו את השערות המחקר והסבירו מהו המבחן המתאים כאן.
 - ב. כיצד הייתה משתנה התשובה לסעיף הקודם אם היה מדובר על אותם אנשים שציונם נבדק פעם אחרי המבחן הראשון שעשו ופעם אחרי המבחן השני?
 - ג. האם ניתן לומר כי מידת הפיזור של ציוני אנשים הנבחנים בפעם הראשונה שונה ממידת הפיזור של ציוני האנשים אשר נבחנים בפעם השנייה. בדוק ברמת מובהקות של $\alpha=0.05$.
 - ד. האם נכונה טענת הפסיכולוגים ברמת מובהקות של $\alpha=0.01$.
3. כחלק ממחקר בנושא הנישואין בישראל, אחד החוקרים העלה השערה שיש הבדל בממוצע גיל הנישואין (הראשונים), בין נשים הגרות בערים מרכזיות לבין נשים הגרות בערים מרוחקות מהמרכז.
- לשם כך נדגמו 50 כלות מכל אחת משתי ערים א'-מרכזית ועיר ב'-מרוחקת ונרשם גילן. תוצאות עיבוד הנתונים מופיעות בטבלאות שלהלן:

T-Test**Group Statistics**

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

Independent Samples Test

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

א. מהו המבחן הסטטיסטי שנעשה כאן?

ב. מצא רווח סמך ברמת סמך של 95% להפרש בין עיר א לעיר ב מבחינת גיל הנשים הממוצע בנישואין הראשונים.

ג. האם ניתן לומר ברמת מובהקות של 1% שנשים בערים מרכזיות מתחתנות בגיל מאוחר יותר מאשר נשים הגרות בערים מרוחקות?

4. להלן פלט של תוכנת SPSS:

T-Test

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

Independent Samples Test

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

א. השלם את סימני השאלה בטבלה.

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

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

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

פתרונות:**שאלה 1:**

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

שאלה 2:

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

שאלה 3:

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

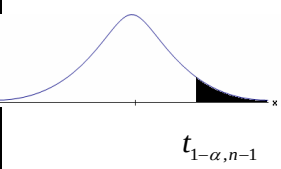
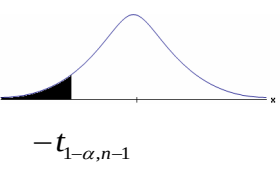
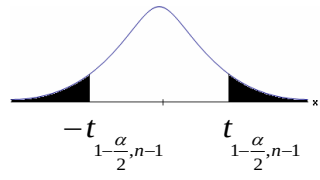
שאלה 4:

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

**פרק 26 - בדיקת השערות על תוחלת הפרשים במדגמים מזווגים
(תלויים)**

בדיקת השערות למדגמים מזווגים

רקע:

$H_0: \mu_D = C$ $H_1: \mu_D > C$	$H_0: \mu_D = C$ $H_1: \mu_D < C$	$H_0: \mu_D = C$ $H_1: \mu_D \neq C$	השערת האפס : השערה אלטרנטיבית:
7. σ_D אינה ידועה 8. $D \square N$ או מדגם מספיק גדול			תנאים:
$t_{\bar{D}} > t_{1-\alpha}^{(n-1)}$  $t_{1-\alpha, n-1}$ ■ - דוחים את H_0	$t_{\bar{D}} < -t_{1-\alpha}^{(n-1)}$  $-t_{1-\alpha, n-1}$ ■ - דוחים את H_0	$t_{\bar{D}} < -t_{1-\frac{\alpha}{2}}^{(n-1)}$ או $t_{\bar{D}} > t_{1-\frac{\alpha}{2}}^{(n-1)}$  $-t_{1-\frac{\alpha}{2}, n-1}$ $t_{1-\frac{\alpha}{2}, n-1}$ ■ - דוחים את H_0	כלל ההכרעה: אזור הדחייה של H_0 :
$\bar{D} > C + t_{1-\alpha}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$	$\bar{D} < C - t_{1-\alpha}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$	$\bar{D} > C + t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$ ו $\bar{D} < C - t_{1-\frac{\alpha}{2}}^{n-1} \cdot \frac{S_D}{\sqrt{n}}$	חלופה לכלל הכרעה: נדחה H_0 אם מתקיים:

סטטיסטי המבחן :

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

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

דוגמה : (פתרון בהקלטה)

חברה שיווקית מעוניינת לבדוק את טענת רשת השיווק "מגה בעיר" הטוענת שמחיריה נמוכים מהמחירים מרשת השיווק "שופרסל".

לצורך בדיקה נבחרו באקראי 4 מוצרים שונים. המחירים נבדקו בשתי הרשתות. להלן המחירים :

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

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

תרגילים:

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

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

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

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

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

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

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

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

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

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

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

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

נניח שהמשקלים מתפלגים נורמלית.

המבחן שיש לבצע כאן הוא :

א. מבחן Z למדגם יחיד.

ב. מבחן T למדגם יחיד.

ג. מבחן T למדגמים בלתי תלויים.

ד. מבחן T למדגמים מזווגים.

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

המבחן שיש לבצע כאן הוא :

א. מבחן Z למדגם יחיד.

ב. מבחן T למדגם יחיד.

ג. מבחן T למדגמים בלתי תלויים.

ד. מבחן T למדגמים מזווגים.

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

ב. 0.25

ג. לא נדחה H_0 **שאלה 4:**

התשובה היא ד.

שאלה 5:

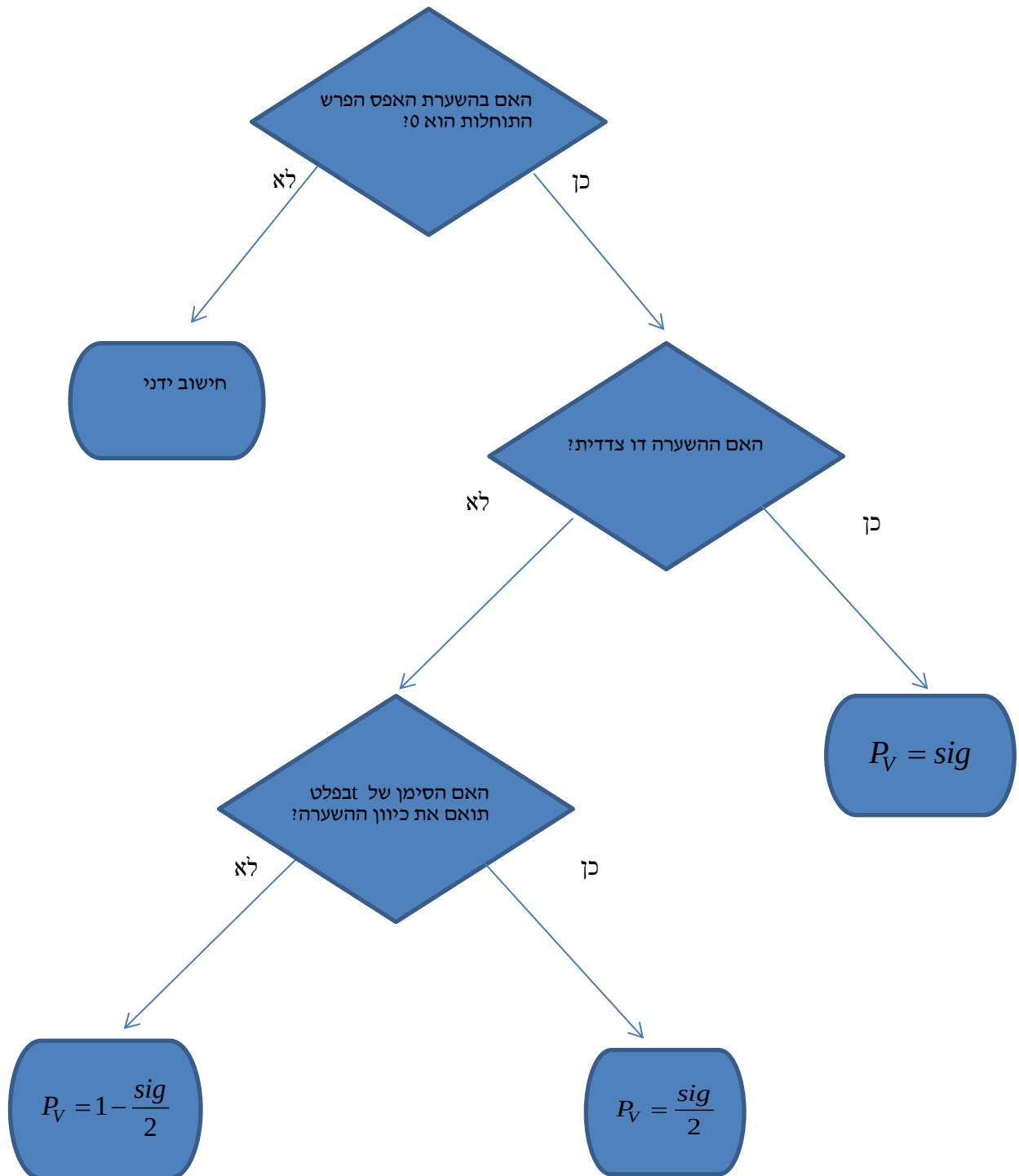
התשובה היא ד.

שאלה 6:

התשובה היא ג.

מדגמים מזווגים - ניתוח פלטים

רקע:



דוגמה : (פתרון בהקלטה)

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

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

Paired Samples Statistics

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

Paired Samples Test

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

בדוק ברמת מובהקות של 5% האם הקורס יעיל .

תרגילים :

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

Paired 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	Highest Year School Completed, Father - Highest Year School Completed, Mother	-.007	3.115	.100	-.203	.189	-.072	973	.943

- א. תן אומדן להפרש הממוצעים.
 ב. תן אומדן לטעות התקן של הפרש הממוצעים.
 ג. האם קיים הבדל מובהק בין השכלת האבות להשכלת האימהות ברמת מובהקות של 5%?

2. בתחרות קפיצה למים שופטים באופן קבוע שופט איטלקי ושופט דרום קוריאני. להלן פלט המנתח את הציונים ששופטים אלה נתנו בתחרויות השונות :

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

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

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

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

Paired Samples Test

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

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

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
	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% .

פתרונות:

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

כתב ופתר - ברק קנדל ©

שאלה 1:א. -0.007 ב. 0.1

ג. אין הבדל מובהק

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

Sig=0

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

ג. יגדל

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

פרק 27 - הקשר בין רווח סמך לבדיקת השערות על הפרש תוחלות

רקע:

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

$$H_0 : \mu_1 - \mu_2 = C$$

$$H_1 : \mu_1 - \mu_2 \neq C$$

על ידי בניית רווח סמך ברמת סמך של $1 - \alpha$ ל $\mu_1 - \mu_2$:

אם C נופל ברווח $\leftarrow$ נקבל את H_0

אם C לא נופל ברווח $\leftarrow$ נדחה את H_0

דוגמה: (פתרון בהקלטה)

חוקר ביצע בדיקת השערות לתוחלת ההפרש במדגם מזווג. להלן השערותיו :

$$H_0 : \mu_D = 80$$

$$H_1 : \mu_D \neq 80$$

$$\alpha = 5\%$$

החוקר בנה רווח סמך ברמה של 90%

$$78 < \mu_D < 83$$

האם אפשר לדעת מה מסקנתו, ואם כן מהי?

תרגילים:

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

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

- א. בנו רווח סמך ברמת סמך של 95% לתוחלת פער הציונים בין סמסטר א לבין סמסטר ב.
 ב. פורסם שתלמידים שמסיימים את סמסטר ב משפרים בממוצע את הציונים ב-5 נק' לעומת סמסטר א' האם יש אמת בפרסום?

2. הוחלט להשוות הציונים אצל מרצה X ואצל מרצה Y. נבחרו באקראי 6 סטודנטים, 3 סטודנטים של מרצה X ו- 3 סטודנטים של מרצה Y, עבורם התקבלו הציונים הבאים :

68	90	82	מרצה X
64	81	68	מרצה Y

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

שאלות אמריקאיות:

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

4. במטרה לבדוק האם קיים הבדל בין קווי זהב לבזק מבחינת ממוצע המחירים לשיחות בינ"ל. נדגמו באקראי 7 מדינות ועבור כל מדינה נבדקה עלות דקת שיחה. בהנחה

והמחירים מתפלים נורמלית בנו רווח סמך לממוצע ההפרשים וקיבלו :

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

לכן מסקנת המחקר היא :

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

פתרונות:**שאלה 1:**

א. $-3.8 \leq \mu_D \leq 19$

ב. נכריע שיש אמת בפרסום.

שאלה 2:

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

ב. נכריע שאין הבדל.

שאלה 3:

התשובה היא ג.

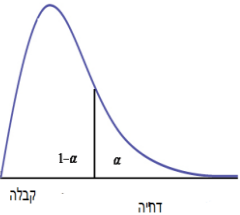
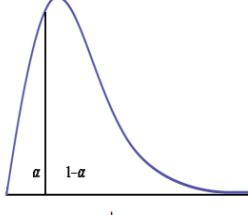
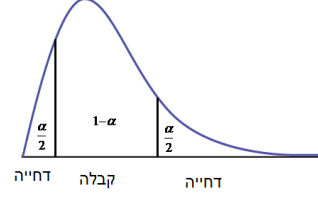
שאלה 4:

התשובה היא א.

פרק 28 - בדיקת השערות על שונות

בדיקת השערות על שונות האוכלוסייה כאשר התוחלת לא ידועה

רקע:

השערת האפס : השערה אלטרנטיבית :	$H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 > \sigma_0^2$	$H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 < \sigma_0^2$	$H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 \neq \sigma_0^2$
תנאים :	$X \sim N$		
נדחה את השערת האפס אם :	 $\chi^2 > \chi_{1-\alpha}^{2(n-1)}$	 $\chi^2 < \chi_{\alpha}^{2(n-1)}$	 $\chi^2 < \chi_{\frac{\alpha}{2}}^{2(n-1)} \text{ או } \chi^2 > \chi_{1-\frac{\alpha}{2}}^{2(n-1)}$

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} : \text{סטטיסטי המבחן}$$

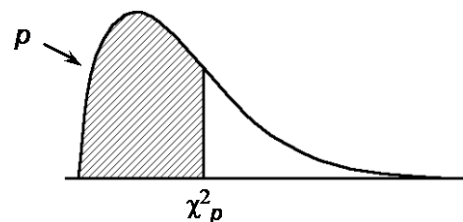
התפלגות חי בריבוע :

$$\frac{(n-1)S^2}{\sigma_0^2} \sim \chi_{2(n-1)}^2 : \text{אם } X_i \sim N(\mu, \sigma^2) \text{ והפרמטר } \mu \text{ אינו ידוע מתקיים ש:}$$

התפלגות זו היא התפלגות אסימטרית חיובית המתחילה מהערך אפס וערכיה שואפים לאינסוף.

התפלגות זו תלויה בדרגות החופש. אם μ אינו ידוע אז :

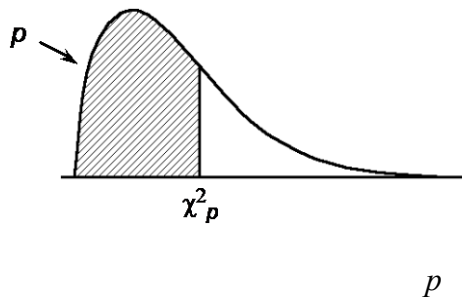
$$d.f = n - 1$$



דוגמה : (פתרון בהקלטה)

ציוני IQ לפי סטנדרטים אמריקאים מתפלגים נורמאלית עם $\sigma = 15$. מעוניינים לבדוק האם שונות הציונים של נבחנים ישראלים שונה מאמריקה. במדגם של 20 ישראלים התקבל :

$$\sum_{i=1}^{20} (x_i - \bar{x})^2 = 3420 \quad \text{מה המסקנה ברמת מובהקות של 5\%?}$$

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

df	.005	.01	.025	.05	.10	.25	.50	.75	.90	.95	.975	.99	.995
1	0.004393	0.00157	0.00982	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

תרגילים:

1. חברה אורזת סוכר במשקל עם סטיית תקן 20 גרם. משקל הסוכר באריזה מתפלג נורמאלית. החברה החליפה את מכוונת האריזה במטרה לדייק יותר במשקל הנארו. (רוצים שסטיית התקן תהיה קטנה יותר).
לצורך בדיקה דגמו 5 אריזות סוכר ולהלן משקלן בגרמים: 1008, 1024, 996, 1005, 997.
מה המסקנה ברמת מובהקות של 5%?
2. זמן ההחלמה ממחלה מסוימת כאשר משתמשים בטיפול מסוים מתפלג נורמלית עם סטיית תקן של 80 שעות. תרופה חדשה נוסתה על 5 חולים. זמני ההחלמה שלהם בשעות היו: 38, 72, 90, 110, 50.
א. ברמת מובהקות של 5% בדקו האם סטיית התקן של זמן החלמה של התרופה החדשה נמוכה מהתרופה המקורית.
ב. האם ניתן לדעת מה תהיה התשובה לסעיף א אם נגדיל את רמת המובהקות?
ג. האם ניתן לדעת מה תהיה התשובה לסעיף א אם נקטין את רמת המובהקות?
ד. האם ניתן לדעת מה תהיה התשובה לסעיף א אם נוסף תצפית שערכה 70?
3. הגובה של אוכלוסייה מסוימת נחשב כמתפלג נורמלית עם ממוצע של 174 ס"מ וסטיית תקן של 12. במדגם של 20 אנשים מהאוכלוסייה התקבל ממוצע 171 וסטיית תקן מדגמית 23.
א. בדקו ברמת מובהקות של 5% האם חל שינוי בשונות הגבהים באוכלוסייה.
ב. בדקו ברמת מובהקות של 5% האם חל שינוי בתוחלת הגבהים באוכלוסייה, בבחירת המבחן המתאים הסתמך על המסקנה מסעיף א'.
4. השערות המחקר הן: $H_0: \sigma^2 = 100$
 $H_1: \sigma^2 > 100$
מתכננים לבצע מדגם בגודל 10 תצפיות. רמת המובהקות היא 5%.
א. מה תהיה עוצמת המבחן אם $\sigma_1^2 = 150$?
ב. איזו השערה אלטרנטיבית תיתן עוצמה של 90%?
5. השערות המחקר הן: $H_0: \sigma = 2$
 $H_1: \sigma < 2$
במדגם של 21 תצפיות התקבל סטיית תקן 1.43. תן הערכה למובהקות התוצאה.

פתרונות:**שאלה 1**

לא נדחה H_0

שאלה 2

א. נדחה H_0

ב. לא תשתנה

ג. לא ניתן לדעת

ד. לא תשתנה

שאלה 3

א. נדחה H_0

ב. לא נדחה H_0

שאלה 4

א. בין 25% ל- 50%

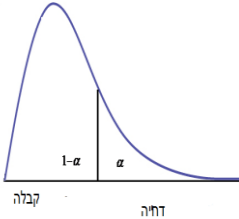
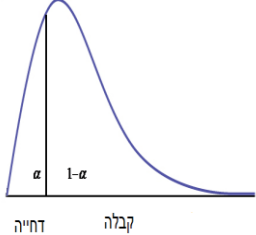
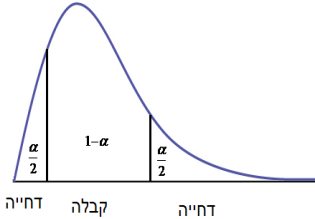
ב. 405.3

שאלה 5

$$0 < P_v < 0.005$$

בדיקת השערות על שתי שונות

רקע:

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

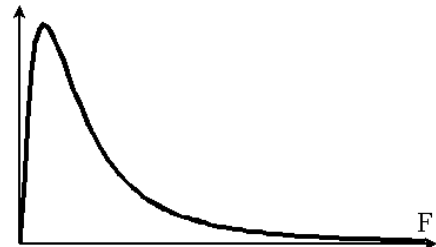
$$F = \frac{S_1^2}{S_2^2} : \text{סטטיסטי המבחן}$$

התפלגות F:

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

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

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



$$df_1 = n_1 - 1$$

$$df_2 = n_2 - 1$$

דוגמה: (פתרון בהקלטה)

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

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

במדגם של 13 נשים התקבלו התוצאות הבאות:

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

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

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

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

m

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

תרגילים :

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

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

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

ובשנת 2013 שוות. מה הן ההנחות הדרושות לביצוע הבדיקה?

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

להפחית 10 מ"ר מכל הדירות שמופיעות במדגם?

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

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

התוצאות הבאות :

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

יש להניח שרמת החוזק של המתכות מתפלגת נורמאלית.

א. האם קיים הבדל בין שונויות החוזק של מתכות?

ב. האם קיים הבדל בין תוחלות החוזק של מתכות ?

בכל סעיף רמת מובהקות של 10% .

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

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

500 ₪ וסטיית תקן של 300 ₪.

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

על בילויים 10,000. כמות הסטודנטים שנדגמה 16.

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

החודשית על בילויים 490,000.

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

גילאי 21-26 נמוכה מהשונויות אצל כלל המבוגרים.

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

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

4. נתון $X_i \sim N(\mu_x, \sigma^2)$ וכמו כן $Y_i \sim N(\mu_y, \sigma^2)$

מאוכלוסייה X נדגמו 7 תצפיות ומאוכלוסייה Y נדגמו 13 תצפיות.

א. כיצד $\frac{S_x^2}{S_y^2}$ מתפלג?

ב. מה ההסתברות ש S_x^2 גדולה ביותר מפי 3 מאשר S_y^2 ?

פתרונות:**שאלה 1**

- א. לא נדחה H_0
ב. מסקנה לא תשתנה

שאלה 2

- א. לא נדחה H_0
ב. לא נדחה H_0

שאלה 3

- א. נדחה H_0
ב. נדחה H_0

שאלה 4

- א. $F(6,12)$
ב. 5%

פרק 29- שאלות מסכמות בבדיקת השערות על פרמטרים

שאלות מסכמות בבדיקת השערות על פרמטרים

1. שני חוקרים נתבקשו לבדוק את ההשערות הבאות :

$$H_0 : \mu = 520$$

$$H_1 : \mu > 520$$

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

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

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

א. למי מהחוקרים הסתברות לטעות מסוג ראשון קטנה יותר?

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

ג. הסבר ללא חישוב נוסף , האם ההסתברות לטעות מסוג שני עבור $\mu_1 = 525$, של חוקר א' ,

שווה/קטנה/גדולה לזו של חוקר ב' .

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

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

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

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

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

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

מוצר	A	B
1	5	5
2	4	5
3	5	3
4	7	4

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

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

$$n = 10$$

$$\sum X_i = 1020$$

$$\sum X_i^2 = 105120$$

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

$$n = 14$$

$$\sum X_i = 1386$$

$$\sum X_i^2 = 138644$$

נתון שציוני הבחינה מתפלגים נורמלית בכל מדינה.

- א. בדוק 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:

						באוניברסיטאות
						במכללות

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

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

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

- א. בדוק את הטענה שממוצע משכורותיהם של אזרחי שומקום גבוה מאשר ממוצע משכורותיהם של אזרחי טרפפו ברמת מובהקות של 5%. בהנחה שסטיית התקן זהה בשתי המדינות.
- ב. חזור על הסעיף הקודם ללא ההנחה הנ"ל.
- ג. ישנה טענה שסטיית התקן במדינת שומקום גבוהה מזו של טרפפו. בדוק ברמת מובהקות של 5%.

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

- 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. בדוק 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 :

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

בדוק ברמת מובהקות של 5% :

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

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

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. חזור על סעיף א אם ההשערה האלטרנטיבית הייתה :

$$H_1 : \mu < 40$$

ג. חזור על סעיף א אם ההשערה האלטרנטיבית הייתה :

$$H_1 : \mu \neq 40$$

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.

פתרונות:**שאלה 1:**

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

שאלה 2:

- א. לפחות 0.0808
- ב. 0.7995

שאלה 3:

- א. לא נדחה H_0
- ב. לא נדחה H_0

שאלה 4:

- א. לא נדחה H_0
- ב. לא נדחה H_0

שאלה 5:

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

שאלה 6:

- א. נדחה H_0
- ב. נדחה H_0
- ג. נדחה H_0

שאלה 7:

- א. נדחה H_0
- ב. נדחה H_0

שאלה 8:ב. נדחה H_0

ג. טעות מסוג ראשון בסיכוי של 0.05

שאלה 9:א. לא נדחה H_0

ב. 0.8643

ג. 0.8749

שאלה 10:א. נדחה H_0 ב. נדחה H_0 **שאלה 11:**

א. 0.0062

ב. 0.9938

ג. 0.0124

שאלה 12:א. לא נדחה H_0 ב. לא נדחה H_0

שאלות מסכמות בסגנון רב ברירה (אמריקאיות) על בדיקת השערות

1. בבדיקת השערה חד-צדדית ימנית ברמת מובהקות $\alpha=0.01$, נדחתה השערת האפס. מה הייתה המסקנה לו נבדקה אותה ההשערה באמצעות אותם נתונים ברמת מובהקות $\alpha=0.05$?
 - א. השערת האפס הייתה נדחית.
 - ב. השערת האפס לא הייתה נדחית.
 - ג. ההשערה המחקרית הייתה נדחית.
 - ד. בהעדר נתונים נוספים, לא ניתן לדעת.

2. על מנת לבדוק האם ההסתברות ללידת בן הנה חצי, נבחר מדגם מקרי של 200 ילדים, ונמצא שישנם 120 בנים.

מהן ההשערות האלטרנטיביות להשערת האפס?

 - א. $H_1 : p = 0.5$
 - ב. $H_1 : p = 0.6$
 - ג. $H_1 : p > 0.5$
 - ד. $H_1 : p \neq 0.5$

3. לצורך בדיקת השפעת היפנוזה על לימוד אנגלית, נבחרו 10 זוגות תאומים זהים. אחד התאומים למד אנגלית בהשפעת היפנוזה, והשני ללא היפנוזה. לאחר מכן נערך לשניהם מבחן באנגלית. 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:

68	82	93	69	X-אין
71	84	88	80	Y-ואין

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

א. $H_1: \mu_d < 0, H_0: \mu_d = 0$

ב. $H_1: \mu_X - \mu_Y < 0, H_0: \mu_X - \mu_Y = 0$

ג. $H_1: \mu_X - \mu_Y < 0, H_0: \mu_X - \mu_Y = 0$

ד. $H_1: \mu_d > 0, H_0: \mu_d = 0$

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

א. השערת האפס נכונה.

ב. השערת האפס נדחתה.

ג. השערת האפס לא נדחתה.

ד. אף אחת מהתשובות לא נכונה בהכרח.

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

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

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

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

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

7. מובהקות התוצאה (PV) היא גם:

א. רמת המובהקות המינימאלית לדחות השערת האפס.

ב. רמת המובהקות המקסימאלית לדחיית השערת האפס.

ג. רמת המובהקות שנקבעת מראש על ידי החוקר טרם קיבל את תוצאות המחקר.

ד. רמת המובהקות המינימאלית לאי דחיית השערת האפס.

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

אין	69	93	82	68
ואין	80	88	84	71

באיזה התפלגות משתמשים לבדיקת ההשערות, ובכמה דרגות חופש:

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

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?

- א. טעות מסוג ראשון.
- ב. טעות מסוג שני.
- ג. טעות מסוג שלישי.
- ד. לא ניתן לדעת כיוון שאנו לא יודעים מה התוחלת האמיתית אצל הצמחוניים.

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

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

13. שני חוקרים העוסקים בתחום מחקרי משותף החליטו להסתמך על נתונים של מדגם שפורסם על ידי הלשכה המרכזית לסטטיסטיקה. חוקר א' ניסח השערה דו צדדית ואילו חוקר ב' ניסח השערה חד צדדית. מסקנתו של איזה מבין המשפטים הבאים הוא הנכון בנוגע למסקנות החוקרים?

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

14. ידוע מנתוני העבר כי תוחלת הציונים בבחינה בפסיכולוגיה היא 79. הועלתה השערה כי תוחלת הציונים בקרב העולים החדשים נמוכה יותר. לצורך בדיקת הטענה נלקח מדגם מקרי של 47 סטודנטים עולים ונמצא ממוצע של 75. מה משמעות הפרמטר בניסוח ההשערות?
- תוחלת ציוני העולים באוכלוסייה.
 - ממוצע ציוני העולים במדגם.
 - תוחלת ציוני האוכלוסייה מנתוני העבר.
 - ממוצע ציוני שאר האוכלוסייה במדגם.
15. חוקר ביצע מחקר וידוע כי עשה טעות מסוג 1. מה מהבאים נכון?
- החוקר דחה את השערת H_0 כאשר היא הייתה נכונה.
 - החוקר דחה את השערת H_1 כאשר היא הייתה נכונה.
 - החוקר לא דחה את השערת H_0 כאשר היא הייתה לא נכונה.
 - המדגם של החוקר שייך בפועל להתפלגות הדגימה של H_1 .
16. חוקר ביקש לבחון האם תאומים זהים אשר הופרדו בילדותם שונים מתאומים זהים אשר גדלו יחדיו מבחינת מידת הפער בין התאומים בלחץ הדם. הוא דגם 20 זוגות תאומים מכל אוכלוסייה ומדד את הפרש בין לחץ הדם בכל זוג תאומים. מהו המבחן הסטטיסטי המתאים?
- מבחן t למדגמים בלתי תלויים עם 38 דרגות חופש.
 - מבחן t למדגמים מזווגים, עם 39 דרגות חופש.
 - מבחן t למדגמים בלתי תלויים עם 39 דרגות חופש.
 - מבחן t למדגמים מזווגים עם 38 דרגות חופש.
17. בינואר השנה פורסם שהשכר הממוצע במשק הוא 8,900 ₪. במדגם שנעשה בחודש יוני על 60 עובדים נרשם עבור כל עובד במדגם האם השכר שלו נמוך או לא נמוך מהשכר הממוצע שפורסם בחודש ינואר. מהו המבחן המתאים כדי לבדוק שרוב העובדים בחודש יוני קיבלו שכר הנמוך מהשכר הממוצע שפורסם בחודש ינואר?
- מבחן Z על פרופורציה.
 - מבחן t על תוחלת אחת.
 - מבחן t על שתי תוחלות במדגמים בלתי תלויים.
 - מבחן t על שתי תוחלות במדגמים תלויים.

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

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

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

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

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

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

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

- א. ליאור השתמש בטבלה של התפלגות Z .
- ב. רוני השתמש בטבלה של התפלגות t .
- מה נוכל לומר בנוגע להחלטת המחקר שלהם?
- א. אם ליאור ידחה את השערת האפס אז גם בהכרח רוני.
- ב. אם רוני תדחה את השערת האפס אז גם בהכרח ליאור.
- ג. שני החוקרים בהכרח יגיעו לאותה מסקנה.
- ד. לא ניתן לדעת על היחס בין דחיית השערת האפס של שני החוקרים.

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

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu < \mu_0$$

חוקר בדק את ההשערות הללו על סמך מדגם שכלל 10 תצפיות. σ^2 לא הייתה ידועה לחוקר. החוקר החליט לדחות את השערת האפס ברמת מובהקות של 5% לאחר מכן כדי לחזק את קביעתו הוא דגם עוד 5 תצפיות ושקלל את תוצאות אלה גם למדגם כך שכלל עכשיו 15 תצפיות.

א. כעת בברור הוא ידחה את השערת האפס.

ב. כעת הוא דווקא יקבל את השערת האפס.

ג. כעת לא ניתן לדעת מה תהיה מסקנתו.

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

א. הסיכוי לטעות מסוג ראשון גדל.

ב. העוצמה של המבחן גדלה.

ג. הסיכוי לטעות מסוג שני גדל.

ד. תשובות א ו-ב נכונות.

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

א. השערת האפס נכונה.

ב. השערת האפס נדחתה.

ג. השערת האפס לא נדחתה.

ד. אף אחת מהתשובות לא נכונה בהכרח.

25. מה המצב הרצוי לחוקר המבצע בדיקת השערה :

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

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

א. הן 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

השוונות האמתית.

מספר דרגות החופש במבחן הוא :

א. 9

ב. 19

ג. 18

ד. 8

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

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

- נניח שהמשקלים מתפלגים נורמלית.
- המבחן שיש לבצע כאן הוא :
- א. מבחן Z למדגם יחיד.
- ב. מבחן T למדגם יחיד.
- ג. מבחן T למדגמים בלתי תלויים.
- ד. מבחן T למדגמים מזווגים.
35. כדי להשוות בין שני אצים נדגמו 5 תוצאות מריצת 100 מטר של כל אצן. זמני הריצה נרשמו ויש להניח שמתפלגים נורמלית. המטרה להשוות בין האצנים.
- המבחן שיש לבצע כאן הוא :
- א. מבחן Z למדגם יחיד.
- ב. מבחן T למדגם יחיד.
- ג. מבחן T למדגמים בלתי תלויים.
- ד. מבחן T למדגמים מזווגים.
36. סטטיסטיקאי ערך מבחן סטטיסטי. הוא חישב את עוצמת המבחן וקיבל 0. המשמעות של תוצאה זו היא :
- א. לעולם לא לדחות את השערת האפס כאשר היא לא נכונה.
- ב. תמיד לדחות את השערת האפס כאשר היא נכונה.
- ג. לעולם לא לדחות את השערת האפס כאשר היא נכונה.
- ד. תמיד לדחות את השערת האפס כאשר היא לא נכונה.

37. סטטיסטיקאי נתבקש לאמוד את הפרש הממוצעים של שני טיפולים לפי שני מדגמים מקריים בלתי תלויים.

הוא חישב רווח סמך להפרש ברמת סמך 0.98, וקיבל את הרווח $-2 < \mu_1 - \mu_2 < 4.5$. אילו יתבקש החוקר לבדוק לפי אותם נתונים את ההשערות:

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

ה. לדחות את השערת האפס.

ו. לא לדחות את השערת האפס.

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

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

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

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

בהנחה והמחירים מתפללים נורמלית בנו רווח סמך לממוצע ההפרשים וקיבלו:

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

לכן מסקנת המחקר היא:

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

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

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

של D.

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

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

ב. לא נדחה את H_0 אף פעם.

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

ד. כל התשובות לא נכונות.

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

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

תוצאות המדגם הוא חישב וקיבל : $t_{\bar{x}} = -2.63$. לכן המסקנה היא :

- א. הוא ידחה H_0 ברמת מובהקות 0.1 אך לא כן ברמת מובהקות 0.05.
- ב. הוא ידחה H_0 ברמת מובהקות 0.05 אך לא כן ברמת מובהקות 0.025.
- ג. הוא ידחה H_0 ברמת מובהקות 0.025 אך לא כן ברמת מובהקות 0.01.
- ד. 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?

- א. 0.005
- ב. 0.03
- ג. 0.0287
- ד. 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. $POWER + \beta + \alpha = 1$

ב. $\beta - POWER = .05$

ג. $POWER + \alpha = 1$

ד. $\beta + \alpha = 1$

ה. הכול לא נכון

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

א. היא אומרת שהשונות המדגמיות שוות.

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

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

ד. אף תשובה אינה נכונה.

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

ברמת מובהקות של 2α על סמך אותם נתונים האם ההשערה תדחה?

א. ההשערה תדחה.

ב. ההשערה לא תדחה.

ג. התשובה תלויה בעוצמת המבחן.

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

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

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

א. אפס

ב. חיובי

ג. שלילי

ד. לא ניתן לדעת

47. עוצמה שווה ל-1 פרושה:

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

ב. תמיד לדחות את השערת האפס כאשר היא נכונה.

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

48. מה מהבאים נכון לגבי מבחן t מדגמים מזווגים?

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

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

$$H_0 : \mu \geq 10$$

$$H_1 : \mu < 10$$

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

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

$$H_1 : \mu \neq 10$$

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

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

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

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

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

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

א. שלושת הטענות אינן נכונות.

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

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

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

פתרונות :

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

פרק 30 - מבחני חי בריבוע

מבחן טיב התאמה

רקע:

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

מבנה המבחן:

השערות:

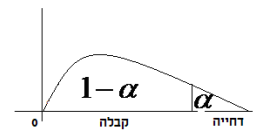
H_0 : המשתנה מתפלג לפי התפלגות מסוימת

H_1 : אחרת

כלל הכרעה:

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

$d.f = K - 1$, 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.

סטטיסטי המבחן:

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

O_i - השכיחות שנצפתה במדגם בקטגוריה i .

p_i - הסתברות לקטגוריה i לפי השערת האפס.

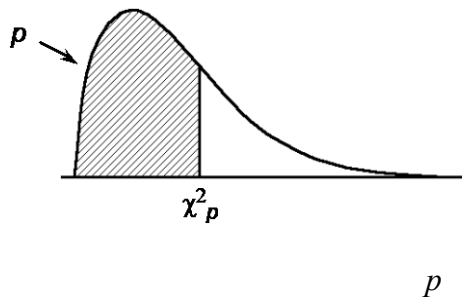
$E_i = np_i$ שכיחות צפויה במדגם לקטגוריה i בהנחת השערת האפס.

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

קטגוריות סמוכות עד שהתנאי יתקיים.

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 χ^2_p 

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

תרגילים:

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

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

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

מה המסקנה ברמת מובהקות של 5%.

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

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

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

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

X						
זות על התהליך						

א. בהנחה והמטבע הוגן, מהי ההתפלגות של X ?

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

8. להלן השערות מחקר :

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

$$H_1 : \text{else}$$

תוצאות המדגם הן :

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

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

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

נכריע שהקובייה אינה הוגנת.

שאלה 6 :

0.005

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

14

הקשר בין מבחן טיב התאמה לבדיקת השערות על הפרופורציה

רקע:

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

דוגמה: (פתרון בהקלטה)

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

א. באמצעות מבחן טיב התאמה.

ב. באמצעות מבחן Z לפרופורציה בודדת.

תרגילים:

1. בסקר שנעשה על 320 נשאלים, 43.75% טענו שהחיה המועדפת עליהם היא כלב. עד היום היה נהוג לחשוב ש-40% מהאנשים מעדיפים כלבים.

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

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

2. לסוכנות מכוניות שלושה סניפים ברחבי הארץ. המכוניות נמכרות בסניפים השונים. מתוך 100 מכוניות נמצא ש-65 נמכרו בסניף תל-אביב, 23 בסניף ירושלים והיתר בסניף חיפה.
א. בדקו ברמת מובהקות של 5% האם שיעור המכוניות שנמכרות בסניף ת"א גדול פי 2 מכל סניף אחר.

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

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

המסלול	1	2	3
מספר הניצחונות	20	15	15

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

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

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

מבחן חי בריבוע לאי תלות בין משתנים

רקע:

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

מבנה המבחן:

השערות:

H_0 : אין תלות בין המשתנים

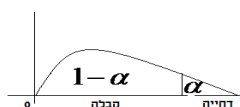
H_1 : יש תלות בין המשתנים

כלל הכרעה:

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

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

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



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

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

סטטיסטי המבחן:

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

O_i - השכיחות נצפית במדגם בתא i .

E_i - שכיחות צפויה במדגם בתא i בהנחת השערת האפס.

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

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

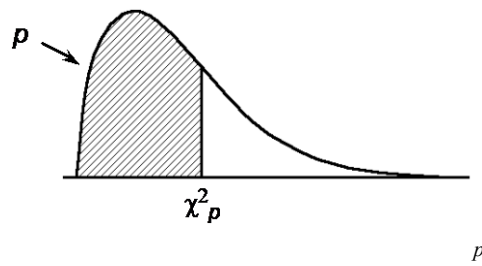
לאחד קטגוריות סמוכות עד שהתנאי יתקיים. תנאי חלופי: אין E קטן מ-1 וגם אין ביותר מ-20% מהתאים E קטן מ-5.

דוגמה: (הפתרון בהקלטה)

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

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

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



df	.005	.01	.025	.05	.10	.25	.50	.75	.90	.95	.975	.99	.995
1	0.004393	0.00157	0.00982	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

תרגילים :

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

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

מה המסקנה ברמת מובהקות של 2.5%?

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

לילה	ערב	יום	
70	60	50	פגומים
800	700	600	תקינים

האם יש הבדל בין שיעורי הפגומים

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

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

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

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

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

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

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

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

כתב ופתר - ברק קנדל ©

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

צבע	שחור	לבן	אדום	כחול
מספר הפריטים	15	20	15	50

כמו כן בדקו את התפלגות הצבעים בחנות שכנה B :

צבע	שחור	לבן	אדום	כחול
מספר הכדורים	60	20	10	20

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

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

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

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

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

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

פתרונות:**שאלה 1:**

נסיק שיש קשר בין גודל הארגון לשביעות הרצון של העובדים.

שאלה 2:

נסיק שאין הבדל מובהק בין שיעור הפגומים במשמרות השונות.

שאלה 3:

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

שאלה 4:

א. 10%

ב. קטן

ג. (0.223, 0.297)

שאלה 5:

א. נסיק שהתפלגות הצבעים בחנות היא כמו שמצוין.

ב. נסיק שיש הבדל בין החנויות מבחינת התפלגות הצבעים.

שאלה 6:

נכון

הקשר בין מבחן לאי תלות ובדיקת השערות להפרש פרופורציות

רקע:

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

דוגמה: (פתרון בהקלטה)

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

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

תרגילים:

1. בקרב מדגם של 200 נשים טענו שהן תצבענה למועמד R לראשות העיר. בקרב מדגם של 200 גברים 80 טענו שהם יצביעו למועמד R האם קיים הבדל בין דפוס ההצבעה של הנשים ושל הגברים? בדוק ברמת מובהקות של 5%.
 - א. על ידי מבחן לאי תלות.
 - ב. על ידי בדיקת השערות לשתי פרופורציות.

2. נלקחו 200 אנשים שמתוכם 60 הצהירו שהם עוסקים בפעילות גופנית סדירה. מתוך אלו שעוסקים בפעילות גופנית סדירה 50 נמצאו במצב בריאותי תקין. מתוך אלו שלא עוסקים בפעילות גופנית סדירה 90 נמצאו במצב בריאותי תקין.
 - א. בנה טבלת שכיחות משותפת לנתונים שהוצגו בשאלה.
 - ב. האם ניתן להגיד שהסיכוי להימצא במצב בריאותי תקין גבוה יותר כאשר עוסקים בפעילות גופנית סדירה לעומת המצב שלא עוסקים בפעילות גופנית סדירה? בדוק ברמת בטחון של 90%.

פתרונות:

שאלה 1:

נדחה את השערת האפס.

שאלה 2:

ב. נדחה את השערת האפס.

פרק 31 - ניתוח שונות חד כיוונית

רקע תיאורטי:

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

ולכן בניתוח שונות השערות המחקר הן:

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

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

התנחות הדרושות לביצוע התהליך הן:

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 .

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

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

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

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

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

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

$$SST = SSB + SSW$$

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

© כתב ופתר - ברק קנדל

יש למלא את טבלת ניתוח השונות הבאה :

טבלת ניתוח שונות

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

$$F = \frac{SS_B / (k - 1)}{SS_W / (n - k)} \sim F(k - 1, n - k)$$

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

תרגילים:

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

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

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

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

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

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

א. רשמו את השערות המחקר וההנחות הדרושות כדי לבצע את המבחן המתאים.

ב. מה מסקנת המחקר ברמת מובהקות של 5%?

ג. האם יש צורך לבצע השוואות מרובות?

4. שלושה אופים נתבקשו להכין עוגת שוקולד. לכל אופה בדקו את משך זמן ההכנה בדקות.

כל אופה נדרש לאפות בכל יום 4 עוגות.

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

האם קיים הבדל בין האופים מבחינת תוחלת זמני ההכנה של העוגות? בדקו ברמת מובהקות של 5%.

5. להלן טבלת ניתוח שונות חד כיוונית. במחקר בחנו 4 סוגי סוללות. רצו לבדוק האם לסוג הסוללה השפעה על תוחלת אורך החיים שלה. הפעילו את כל הסוללות על אותו מכשיר ובדקו את אורך החיים של כל סוללה בשעות.

ANOVA

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

מה המסקנה ברמת מובהקות של 10%? רשמו את ההשערות וההנחות הדרושות.

6. להלן טבלת ANOVA בטבלה הושמטו חלקים. השלם את החלקים בטבלה שהושמטו ומסומנים באותיות.

ANOVA

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

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

ANOVA

pulse

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

Post Hoc Tests

Multiple Comparisons

pulse
Tukey HSD

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

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

Homogeneous Subsets

pulse

Tukey HSD^a

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

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 5.000.

- א. בדוק ברמת מובהקות של 5% האם קיים הבדל בין המינונים השונים מבחינת תוחלת הדופק של האנשים? רשמו את ההשערות וההנחות הדרושות לצורך פתרון.
- ב. הסבירו ללא חישוב כיצד הייתה משתנה התשובה לסעיף הקודם אם הינו מעלים את הדופק של כל התצפיות במחקר ב- 2.
- ג. האם יש צורך במחקר בהשוואת מרובות. נמק!
- ד. לטבלת ה ANOVA צורפו טבלאות של השוואות מרובות בשיטה הנקראת "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:
- א. כמה בתי ספר יש בעיר?
- ב. כמה תלמידים השתתפו בסך הכול במחקר?
- ג. האם קיים הבדל בין בתי הספר בעיר מבחינה רמת הציונים? בדקו ברמת מובהקות של 1%.
- ד. 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

grade

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 - ניתוח שונות חד כיוונית

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

פרק 32 - מדדי קשר - מדד הקשר הלינארי (פירסון)

רקע:

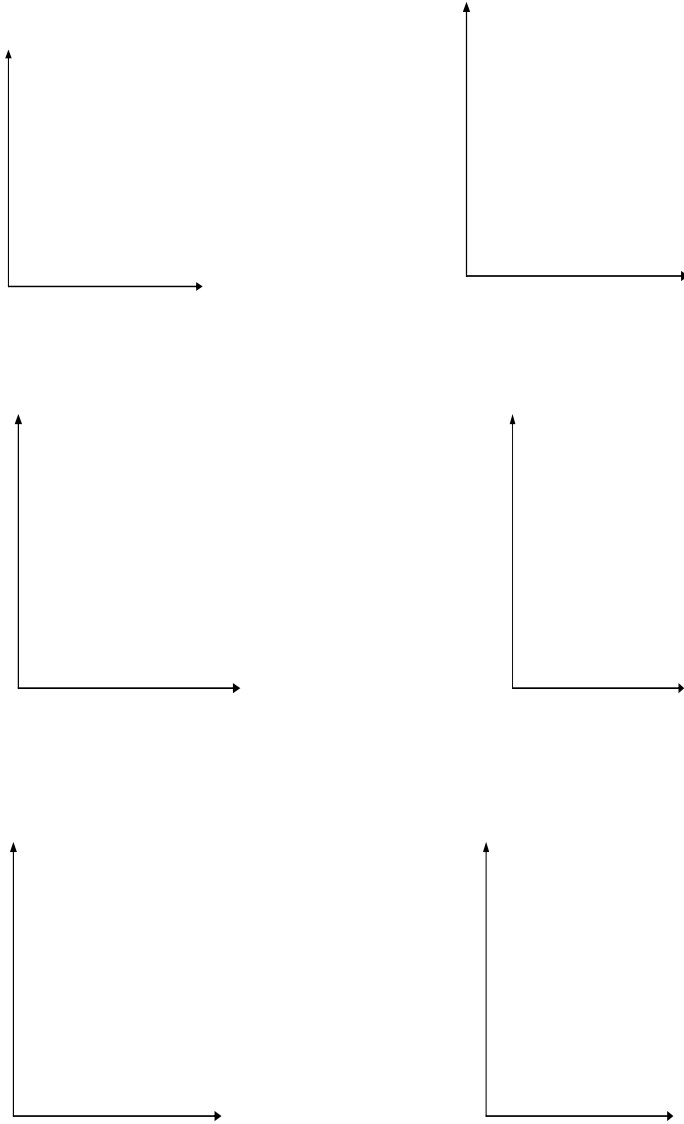
המטרה היא לבדוק האם קיים קשר (קורלציה, מתאם) של קו ישר בין שני משתנים כמותיים. מבחינת סולמות המדידה קשר בין סולמות רווחים ומנה. בדרך כלל, X הוא המשתנה המסביר (הבלתי תלוי) ו Y הוא המשתנה המוסבר (התלוי). למשל, נרצה להסביר כיצד השכלה של אדם הנמדדת בשנות לימוד X מסבירה את ההכנסה שלו Y . במקרה זה שנות ההשכלה זהו המשתנה המסביר (או הבלתי תלוי) ואנחנו מעוניינים לבדוק כיצד שינויים בשנות ההשכלה של אדם יכולים להסביר את השינויים שלו בהכנסה, ולכן רמת ההכנסה זהו המשתנה המוסבר התלוי במשתנה המסביר אותו. בשלב הראשון, נהוג לשרטט דיאגרמת פיזור. זו דיאגרמה שנותנת אינדיקציה ויזואלית על טיב הקשר בין שני המשתנים. למשל, בבניין של 5 דירות בדקו את הנתונים הבאים: X - מס' חדרים בדירה. Y - מס' נפשות הגרות בדירה. להלן התוצאות שהתקבלו:

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

נשרטט מנתונים הללו דיאגרמת פיזור:



נתבונן בכמה מקרים של דיאגרמות פיזור וננתח אותן :



בשלב השני, מחשבים את מקדם המתאם (מדד הקשר) שבודק עד כמה קיים קשר לינארי בין שני המשתנים. המדד (ניקרא גם מדד הקשר של פירסון) מכמת את מה שנראה בשלב הראשון רק בעין.

המדד בודק את כיוון הקשר (חיובי או שלילי).

ואת עוצמת הקשר (חלש עד חזק).

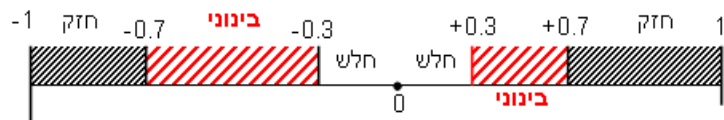
מקדם מתאם זה מקבל ערכים בין -1 ל 1.

מקדם מתאם -1 או 1 אומר שקיים קשר לינארי מוחלט ומלא בין המשתנים שניתן לבטאו על ידי הנוסחה: $y = bx + a$.

מתאם חיובי מלא (מקדם מתאם 1) אומר שקיים קשר לינארי מלא בו השיפוע b יהיה חיובי ואילו מתאם שלילי מלא אומר שקיים קשר לינארי מלא בו השיפוע b שלילי (מקדם מתאם -1).

מתאם חיובי חלקי אומר שככל שמשנתנה אחד עולה לשני יש נטייה לעלות בערכו אבל לא קיימת נוסחה לינארית שמקשרת את X ל- Y באופן מוחלט ואילו מתאם שלילי חלקי אומר שככל שמשנתנה אחד עולה לשני יש נטייה לרדת אבל לא קיימת נוסחה לינארית שמקשרת את X ל- Y באופן מוחלט.

ככל שערך מקדם המתאם קרוב לאפס נאמר שעוצמת הקשר חלשה יותר וככל שמקדם המתאם רחוק מהאפס נאמר שעוצמת הקשר חזקה יותר.



מקדם המתאם יסומן באות r .

כדי לחשב את מקדם המתאם, יש לחשב את סטיות התקן של כל משתנה ואת השונות המשותפת.

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

כתב ופתר - ברק קנדל ©

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

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

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

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

תרגילים:

1. להלן נתונים לגבי שישה תלמידים שנגשו למבחן. בדקו לגבי כל תלמיד את הציון שלו בסוף הקורס וכמו כן את מספר החיסורים שלו מהקורס.

מספר חיסורים	ציון
2	80
1	90
0	90
2	70
3	70
4	50

- א. שרטט דיאגרמת פיזור לנתונים. מה ניתן להסיק מהדיאגרמה על טיב הקשר בין מספר החיסורים של תלמיד לציונו? מיהו המשתנה הבלתי תלוי ומיהו המשתנה התלוי?
- ב. חשב את מדד הקשר של פירסון. האם התוצאה מתיישבת עם תשובתך לסעיף א'?
- ג. הסבר ללא חישוב כיצד מקדם המתאם היה משתנה אם היה מתווסף תלמיד שהחסיר 4 פעמים וקיבל ציון 80?

2. במחקר רפואי רצו לבדוק האם קיים קשר בין רמת ההורמון X בדם החולה לרמת ההורמון Y שלו. לצורך כך מדדו את רמת ההורמונים ההלו עבור חמישה חולים. להלן התוצאות שהתקבלו:

x	y
10	12
14	15
15	15
18	17
20	21

- א. מה הממוצע של כל רמת הורמון?
- ב. מהו מקדם המתאם בין ההורמונים? ומה משמעות התוצאה?

3. נסמן ב- X את ההכנסה של משפחה באלפי ₪. נסמן ב- Y את ההוצאות של משפחה באלפי ₪. נלקחו 20 משפחות והתקבלו התוצאות הבאות :

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

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

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

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

ב. מה המשמעות של התוצאה שקיבלת בסעיף א?

4. נסמן ב- X את ההכנסה של משפחה באלפי ₪. נסמן ב- Y את ההוצאות של משפחה באלפי ₪. נלקחו 20 משפחות והתקבלו התוצאות הבאות :

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

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

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

חשב את מדד הקשר הלינארי בין X ל- Y .

5. במוסד אקדמי ציון ההתאמה מחושב כך : מכפילים את הציון הממוצע בבגרות ב- 3 ומפחיתים 2 נקודות. ידוע שעבור 40 מועמדים סטיית התקן של ממוצע הציון בבגרות הייתה 2. מה מקדם המתאם בין ציון ההתאמה לציון הממוצע בבגרות שלהם ?

6. להלן רשימת טענות, לגבי כל טענה קבע נכון/לא נכון ונמק!

א. מתווך דירות המיר מחירי דירות מדולר לשקל. נניח שדולר אחד הוא 3.5 ₪. אם מתווך הדירות יחשב את מדד הקשר של פירסון בין מחיר הדירה בשקלים למחיר הדירה בדולרים הוא יקבל 1.

ב. לסדרה של נתונים התקבל $\bar{X} = \bar{Y} = 6$ $S_x = S_y = 1$ לכן מדד הקשר של פירסון יהיה 1.

ג. אם השונות המשותפת של X ושל Y הינה 0 אז בהכרח גם מקדם המתאם של פירסון יהיה 0.

שאלות אמריקאיות:

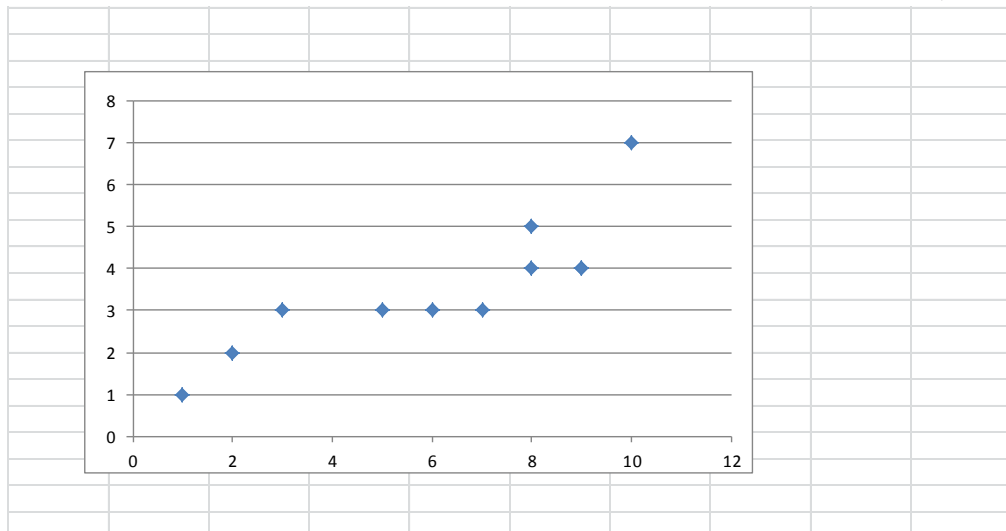
7. נמצא שקיים מקדם מתאם שלילי בין הציון בעברית לציון בחשבון בבחינה לכן :

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

8. נלקחו 20 מוצרים וניבדק ביום מסוים המחיר שלהם בדולרים והמחיר שלהם בש"ח (באותו היום ערך הדולר היה - 4.2 ₪) מהו מקדם המתאם בין המחיר בדולר למחיר בש"ח?

- א. 1
- ב. 0
- ג. 4.2
- ד. לא ניתן לדעת.

9. להלן דיאגרמת פיזור :



מה יהיה מקדם המתאם בין שני המשתנים?

- א. 1
- ב. 0.85
- ג. 0.15
- ד. 0

פתרונות:**שאלה 1:**

א. בהקלטה

ב. -0.9325

שאלה 2:א. $\bar{y} = 16$ $\bar{x} = 15.4$ ב. $r_{xy} = 0.96$ **שאלה 3:**

א : 0.8

שאלה 4:

0.8

שאלה 5:

1

שאלה 6:

א. נכון

ב. לא נכון

ג. נכון

שאלה 7:

התשובה : ג

שאלה 8:

התשובה : א

שאלה 9:

התשובה : ב

פרק 33 - מדדי קשר - רגרסיה ליניארית

רקע:

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

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

a - בעצם נותן את ערך Y כאשר X הנו אפס על גבי קו הניבויים. הוא ניקרא החותך של הקו.

b - הוא שיפוע הקו נותן בכמה בעצם Y משתנה כאשר X גדל ביחידה אחת על גבי קו הניבויים. להלן המשוואות למציאת הפרמטרים של קו הרגרסיה:

$$\tilde{Y} = bX + a$$

$$b = r \frac{S_y}{S_x}$$

$$a = \bar{Y} - b\bar{X}$$

אם נרצה לבנות קו ניבויים לניבוי X על סמך Y נצטרך לעדכן את הנוסחאות בהתאם.

תרגילים:

1. נסמן ב- X את ההכנסה של משפחה באלפי ₪. נסמן ב- Y את ההוצאות של משפחה באלפי ₪. נלקחו 20 משפחות והתקבלו התוצאות הבאות:

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

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

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

- א. חשב את מדד הקשר הלינארי בין X ל- Y . מיהו המשתנה התלוי?
 ב. מצא את קו הרגרסיה לניבוי ההוצאה של משפחה על סמך הכנסה שלה. הסבר את משמעות הפרמטרים של קו הרגרסיה.
 ג. משפחת כהן הכניסה 15,000 ₪, מה ההוצאה הצפויה שלה?

2. נסמן ב- X את ההשכלה של אדם בשנות למוד. נסמן ב- Y את הכנסתו באלפי ₪. במחקר התקבלו התוצאות הבאות:

$$S_y = 5 \quad S_x = 2$$

$$\bar{Y} = 8 \quad \bar{X} = 14$$

$$COV(X, Y) = 7.5$$

- א. חשב את מדד הקשר של פירסון בין ההשכלה להכנסה.
 ב. מה ההכנסה הצפויה לאדם שהשכלתו 12 שנים?
 ג. מה ההשכלה הצפויה לאדם שהכנסתו 10,000 ₪?

3. חוקר רצה לחקור את הקשר הקווי שבין הציון המבחן בסטטיסטיקה לבין מספר שעות ההכנה של הסטודנטים למבחן. במדגם של 100 סטודנטים שנבחנו בקורס נרשמו התוצאות הבאות: הציון הממוצע של הסטודנטים היה 65 עם סטיית תקן של 27. מספר שעות ההכנה הממוצע היה 30 עם סטיית תקן של 18. מקדם המתאם בין הציון לשעות ההכנה היה 0.8.
 א. על פי משוואת הרגרסיה שעת הכנה נוספת משפרת את ציון המבחן ב?
 ב. על פי משוואת הרגרסיה תלמיד שייגש למבחן ללא שעות הכנה כלל יקבל ציון ?
 ג. מהו קו הרגרסייה לניבוי הציון לפי שעות ההכנה?

4. נתונים 2 משתנים Y, X . כמו כן נתון: X ממוצע = 1.5, שונות X = שונות Y = 4, וכן שקו הרגרסיה של Y על בסיס X הינו $Y = -0.2X + 0.5$. חשב מהו מקדם המתאם בין X ל- Y ?

פתרונות:**שאלה 1:**

א. 0.8

ב. $\tilde{Y} = 0.8X + 0.4$

ג. 12.4

שאלה 2:

א. 0.75

ב. 4.25 אלפי ש"ח

ג. 14.6 שנים

שאלה 3:

א. 1.2

ב. 29

ג. $y = 1.2x + 29$

שאלה 4:

-0.2