

ביוסטטיסטיקה מתקדמת



תוכן העניינים

1. אמידה נקודתית (אומד נראות מקסימלית, שיטת המומנטים' esm ועוד)..... 1
2. מבוא לבדיקת השערות על פרמטרים (תזכורת)..... 31
3. בדיקת השערות על שונויות..... 37
4. מבחני חי בריבוע לאי תלות..... 41
5. ניתוח שונות חד כיוונית..... 45
6. מבחנים אפרמטרים למדגמים מזווגים (ווילקוקוסון)..... 54
7. מבחנים אפרמטריים למדגמים בלתי תלויים (מאן וויטני, קרוסקל ואליס ופשיר)..... 65

ביוסטטיסטיקה מתקדמת

פרק 1 - אמידה נקודתית (אומד נראות מקסימלית, שיטת המומנטים 'mse' (ועוד)

תוכן העניינים

1	1. אומד חסר הטייה
8	2. אומד נראות מקסימלית
16	3. MSE
19	4. שיטת המומנטים
22	5. אומד עקיב
24	6. אומד חסר הטייה בעל שונות מינימלית
26	7. שאלות מסכמות

אומד חסר הטייה:

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	$1 - 60\theta$	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: $P(X = 3)$.

אומד 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 אומד חסר הטיה לשונות של X .

תזכורות חשובות:

אם: $Y = aX + b$, אזי: $E(Y) = aE(X) + b$, $V(Y) = a^2 \cdot V(X)$, $\sigma_Y = |a| \sigma_X$.

אם: $X_1, X_2, \dots, X_n$ משתנים מקריים, 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:

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

שאלות:

- (1) הציון במבחן מסוים של תלמידי כתה 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:

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

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

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

- איזה מן האומדים הוא חסר הטיה?
- הציעו תיקון לאומד המוטה כך שיהיה חסר הטיה.
- במדגם התקבלו הציונים הבאים: 100, 82, 58, 78, 65. חשבו את האומדנים המתקבלים עבור האומדים חסרי ההטיה.
- 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

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

- בדקו לגבי כל אומד אם הוא בלתי מוטה.
- 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 .

- פתחו אומד חסר הטיה ל- P .
- מהו אומד חסר הטיה לסיכוי לכישלון בניסיון בודד?
- מהו אומד חסר הטיה ל- $E(X)$?
- מצאו אומד חסר הטיה ל- $E(X^2)$.

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

פונקציית ההסתברות של X - מספר המניות שיעלו ביום מסוים:

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

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 מקרי 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. מצאו 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) קבעו 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 , ובמכונה 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, וב- Y את 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) יהיו T_1 ו- T_2 אומדים חסרי הטיה ובלתי תלויים לפרמטר θ .
 א. מצאו אומד חסר הטיה ל- θ^2 , המתבסס על T_1 ו- T_2 .
 ב. מצאו 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. הראו 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: $X_1 \cdot X_2$.

הראו 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. מצאו אומד חסר הטיה ל- μ^2 .

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

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

- א. הראו כי האומד $3\bar{X}$ הנו אומד בלתי מוטה ל- β .
 ב. מצאו את השונות של האומד מהסעיף הקודם.

(12) $X_1, X_2, \dots, X_n$ הינם משתנים מקריים רציפים בלתי תלויים בעלי פונקציית

$$f(x) = \begin{cases} X \cdot A & 0 \leq x \leq \theta \\ 0 & \text{אחרת} \end{cases} \quad \text{הצפיפות הבאה :}$$

- א. בטאו את ערכו של A באמצעות θ , כדי שפונקציית הצפיפות תהיה לגיטימית.
 ב. מצאו אומד חסר הטיה ל- θ , על סמך n התצפיות.

תשובות סופיות:

$$(1) \quad \text{א. } T_1 \text{ ו- } T_2. \quad \text{ב. } \frac{2}{3}T_3. \quad \text{ג. } T_1 = 76.6, T_2 = 110. \quad \text{ד. } T_1.$$

$$(2) \quad \text{א. ראו בוידאו.} \quad \text{ב. } T_2.$$

$$(3) \quad \text{א. } \frac{x}{n}. \quad \text{ב. } 1 - \frac{x}{n}. \quad \text{ג. } X. \quad \text{ד. } \theta.$$

$$(4) \quad \text{א. } \frac{3x}{2}. \quad \text{ב. } \frac{3\bar{x}}{2}.$$

$$(5) \quad \text{א. } 1 - \frac{x}{2}. \quad \text{ב. } 1 - \frac{1}{2}\bar{x}. \quad \text{ג. } 0.125. \quad \text{ד. } 3\left(1 - \frac{1}{2}\bar{x}\right).$$

$$\text{ה. לשונות } 0.917.$$

$$(6) \quad \text{א. נכון.} \quad \text{ב. לא נכון.}$$

$$(7) \quad \text{ב'.$$

$$(8) \quad \text{א. } T_1 \cdot T_2. \quad \text{ב. } T_1 - T_1 \cdot T_2.$$

$$(9) \quad \text{א. שאלת הוכחה.} \quad \text{ב. שאלת הוכחה.}$$

$$(10) \quad \bar{X}^2 - \frac{1}{n}.$$

$$(11) \quad \text{א. שאלת הוכחה.} \quad \text{ב. } V(3\bar{X}) = \frac{3 - \beta^2}{n}.$$

$$(12) \quad \text{א. } A = \frac{2}{\theta^2}. \quad \text{ב. } \theta = \frac{3 - \bar{X}}{2}.$$

אומד נראות מקסימלית:

רקע:

להלן נלמד את שיטת הנראות המקסימלית למציאת אומדים.

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

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

נבנה את פונקציית ההסתברות המשותפת (פונקציית הדגימה).

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

נגדיר את פונקציית הנראות:

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

פונקציית 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 פונקציית 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 שקיבלנו 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 $g(\hat{\theta})$, בהנחה והפונקציה היא חד-חד ערכית (אינווריאנטיות).

המשך דוגמה:

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

שאלות:

- (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 אומד נראות מקסימלית ל- λ , על סמך מספר הלקוחות שנכנסים ביום מסוים.
 ב. מצאו אומד 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, 5, 7 ו-3.
 א. פתחו אומד 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: $U(0, q)$.
 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.
 ד. מצאו אומדן נראות מקסימלית ל- θ על סמך המדגם הזה.
 ד. מצאו באופן כללי אומד נראות מקסימלית ל- θ , על סמך מדגם של n בני נוער - $X_1, \dots, X_n$.

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

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

ב. נדגמו 5 אנשים בלתי תלויים בעלי הגבהים: 170, 182, 165, 174, 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.

$$(7) \quad X \text{ הוא משתנה מקרי בעל פונקציית הצפיפות: } f(x) = \begin{cases} 2\theta x e^{-\theta x^2} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

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 ובכד ב' 5 כדורים שחורים ו-15 לבנים. דוגמים באקראי כדור, בלי לדעת 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 מקרי X , בעל התפלגות התלויה בפרמטר θ , באופן הבא:

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

בחמישה זמנים שונים שנבחרו באקראי נמצאו: 0, 0, 1, 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 לקבל בו תוצאה של 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 את 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 באיזו 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 המשחק.
 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 פונקציית 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: $2 - \frac{2}{X}$.

(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 ש-30 מהם עדיין פועלים.

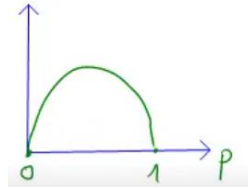
- מצא אומדן נראות מקסימלית עבור P .
- רשמו את האומד שבו השתמשתם בחלק א' באופן כללי, עבור מדגם של n מכשירים שמתוכם נמצאו Y מכשירים שעדיין פועלים לאחר חודש אחד.
- בהנחה שאורך החיים (בחודשים) של סוללה בודדת הוא מעריכי, עם פי צפיפות: $f(t) = \theta e^{-\theta t}$ עבור $t > 0$. מצא א.נ.מ. עבור θ , המבוסס 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.

- רשמו את פונקציית ההסתברות של המשתנה X – מספר הפעמים שהחייגן האוטומטי מחייג למספר הטלפון המבוקש, אם ההסתברות לקבלת צליל "פנוי" בשידור אחד של אות חיוג הוא P .
- 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: 1, 6, 2, 7, 3, 8, 2, 1, 5. מצאו אומדן נראות מקסימלית עבור P , על סמך התוצאות שהתקבלו.

תשובות סופיות:

להלן גרף:



$$(1) \quad L(p) = (1-p) \cdot p \quad \text{א.}$$

$$\text{ב. } 0.5 \quad \text{ג. } \frac{2}{9}$$

$$(2) \quad X \quad \text{א.} \quad \bar{X} \quad \text{ב.}$$

$$(3) \quad \frac{1}{\bar{X}} \quad \text{א.} \quad \frac{2}{9} \quad \text{ב.}$$

$$(4) \quad \text{א. } 1 \quad \text{ב. } 1 \quad \text{ג. } 3 \quad \text{ד. } \hat{\theta} = \max\{X_1, \dots, X_n\}$$

$$(5) \quad \frac{\sum_{i=1}^n (x_i - 170)^2}{n} \quad \text{א.} \quad \text{ב. } 40.2$$

$$(6) \quad \frac{x}{n}$$

$$(7) \quad \frac{n}{\sum X_i^2} \quad \text{א.} \quad \left(\frac{n}{\sum X_i^2} \right)^2 \quad \text{ב.}$$

$$(8) \quad \text{א. ראה סרטון.} \quad \text{ב. כד א'.$$

$$(9) \quad \text{א. } 32 \quad \text{ב. } 0.3916$$

$$(10) \quad \text{א. } 0.45 \quad \text{ב. } 0.81$$

$$(11) \quad \text{א. ראה סרטון.} \quad \text{ב. הוגן.$$

$$(12) \quad \text{א. } 0.08 \quad \text{ב. } 0.92 \quad \text{ג. } 11.5$$

$$(13) \quad \hat{\theta} = \begin{cases} 3 & X = 0,1 \\ 1 & X = 2 \end{cases} \quad \text{א.} \quad \hat{p} = \begin{cases} 2 \cdot 0.4 \cdot 0.6 & X = 0,1 \\ 2 \cdot 0.9 \cdot 0.1 & X = 2 \end{cases} \quad \text{ג.} \quad \text{ב. } 1$$

$$(14) \quad \max |X_i|$$

$$(15) \quad \text{שאלת הוכחה.}$$

$$(16) \quad \text{א. } 0.6124 \quad \text{ב. } \hat{p} = \sqrt{\frac{y}{n}} \quad \text{ג. } 0.49$$

$$(17) \quad \text{א.} \quad P(x) = \begin{cases} (1-p)^{x-1} p & 1 \leq x \leq 9 \\ (1-p)^9 & x = 10 \end{cases} \quad \text{ב. } 0.1818$$

נספח:

התפלגויות רציפות

ההתפלגות	פונקציית הצפיפות	פונקציית ההתפלגות המצטברת	תוחלת	שונות	הערות	אנ"מ
$X \sim U(a, b)$	$f(x) = \frac{1}{b-a}$ $a \leq x \leq b$	$\frac{t-a}{b-a}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$		$b = \max(X_i)$ $a = \min(X_i)$
$X \sim \exp(\lambda)$	$f(x) = \lambda e^{-\lambda x}$	$1 - e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	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		$\bar{\mu} = \bar{X}$ $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ $Z = \frac{X - \mu}{\sigma}$

התפלגויות בדידות

ההתפלגות	פונקציית ההסתברות $P(X = k)$	תוחלת	שונות	הערות	אנ"מ
בינומית $B(n, p)$ $0 \leq p \leq 1$	$\binom{n}{k} p^k (1-p)^{n-k}$ $k = 0, 1, \dots, n$	np	$np(1-p)$	(1)	$\hat{p} = \frac{Y}{n}$
גיאומטרית $G(p)$ $0 < p \leq 1$	$(1-p)^{k-1} p$ $k = 1, 2, \dots, \infty$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	(2)	$\hat{p} = \frac{1}{\bar{X}}$
אחידה $U(a, b)$	$\frac{1}{b-a+1}$ $K = a, \dots, b$	$\frac{a+b}{2}$	$\frac{(b-a+1)^2 - 1}{12}$	(3)	$b = \max(X_i)$ $a = \min(X_i)$
פואסונית $P(\lambda)$ $\lambda > 0$	$\frac{e^{-\lambda} \lambda^k}{k!}$ $k = 0, 1, \dots, \infty$	λ	λ	(4)	$\hat{\lambda} = \bar{X}$

(1) מספר ההצלחות ב- n ניסויי ברנולי ב"ת. p - ההסתברות להצלחה.(2) מספר הניסויים עד להצלחה הראשונה בסדרת ניסויי ברנולי ב"ת, p - ההסתברות להצלחה.(3) בחירה אקראית של מספר בין a ו- b .(4) מספר אירועים ביחידת זמן, λ - קצב האירועים.

קריטריון MSE – תוחלת ריבוע הטעות:

רקע:

הקריטריון הנפוץ ביותר כדי לבדוק את טיב האומד הוא קריטריון MSE :
Mean Squared Error – תוחלת ריבוע טעות האמידה.

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

כאשר: $V(\hat{\theta})$ – 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: $MSE(T_1) > MSE(T_2)$, אז T_2 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 על סמך תצפית בודדת: $T_1 = 2X - 3$ ו- $T_2 = \frac{3X - 3}{2}$.

איזה אומד עדיף לאמידת הפרמטר θ ?

שאלות:

(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 בסעיפים א' ו-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 בין שני אומדים: $T_1 = \frac{X}{5}$ 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 בארץ בחודש אוקטובר מתפלג פואסונית עם תוחלת λ . נלקח מדגם של 10 חודשי אוקטובר. להלן שני אומדים אפשריים:

$$\hat{\lambda}_1 = \frac{\sum_{i=1}^{10} X_i}{10} \quad \text{ו-} \quad \hat{\lambda}_2 = \frac{2 \cdot \sum_{i=1}^5 X_i + 2 \cdot \sum_{i=6}^{10} X_i}{10}$$

כאשר: X_i = מספר השריפות בחודש אוקטובר ה- i .
 איזה מהאומדים עדיף, לצורך אמידת הפרמטר λ ?

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

תשובות סופיות:

- (1) שלוש תצפיות.
- (2) א. $2x$ ב. אומד מוטה. ג. סעיף ב.
- (3) א. T_1 ב. T_2 ג. T_1 .
- (4) $\hat{\lambda}_1$.
- (5) שאלת הוכחה.

שיטת המומנטים:

רקע:

מומנט מסדר ראשון של משתנה X מוגדר להיות: $E(X)$.

מומנט מסדר שני של משתנה X מוגדר להיות: $E(X^2)$.

באופן כללי, מומנט מסדר r מוגדר להיות: $E(X^r)$.

מומנט מסדר ראשון של n תצפיות בלתי תלויות מאותה התפלגות מוגדר

להיות: $\frac{\sum X_i}{n}$ - זהו מומנט מסדר ראשון של המדגם.

מומנט מסדר שני של n תצפיות בלתי תלויות מאותה התפלגות מוגדר

להיות: $\frac{\sum X_i^2}{n}$ - זהו המומנט מסדר שני של המדגם.

באופן כללי, מומנט מסדר r של n תצפיות בלתי תלויות מאותה התפלגות מוגדר

להיות: $\frac{\sum X_i^r}{n}$ - זהו מומנט ה- r של המדגם.

השיטה: משווים את המומנט המתאים של ההתפלגות לפי המומנט המתאים של המדגם.

דוגמה (פתרון בהקלטה):

נגיד שמספר הפעמים שאדם מתעטש ביום מתפלג פואסונית על ידי פרמטר λ (קצב ההתעטשויות ביום). רוצים לאמוד את λ בשיטת המומנטים.

שאלות:

- (1) X מתפלג אחיד רציף מהערך המינימלי a לערך המכסימלי 20. מצא אומד לערך מינימלי a לפי שיטת המומנטים על סמך n תצפיות מההתפלגות.
- (2) דוגמים n תצפיות בלתי תלויות מתוך התפלגות נורמאלית אשר תוחלתה היא μ והשונות שלה היא σ^2 . מצא אומדים לפרמטרים אלה לפי שיטת המומנטים.
- (3) אדם מטיל מטבע רגיל n פעמים. יש לאמוד את מספר הפעמים שהוא מטיל את המטבע וזאת על סמך X – מספר העצים שהוא קיבל.
א. מצא אומד בשיטת המומנטים ל- n על סמך X בודד.
ב. מצא אומד בשיטת המומנטים ל- n על סמך חזרה של m פעמים על אותו תהליך בו מטילים את המטבע ההוגן n פעמים.
ג. מהו האומדן אם האדם חזר על התהליך שלוש פעמים:
פעם אחת קיבל 5 עצים, בפעם השנייה הוא קיבל 4 עצים ובפעם השלישית הוא קיבל 7 עצים.
- (4) נתון ש- $X_i \sim \exp(\lambda)$. מצא אומד בשיטת המומנטים לפרמטר λ על סמך מדגם של n תצפיות.
- (5) נתונה פונקציית הצפיפות הבאה: $f(x) = \begin{cases} \theta \cdot x^{\theta-1} & 0 \leq x \leq 1 \\ 0 & \text{else} \end{cases}$.
א. בטא את $E(X)$ כפונקציה של הפרמטר θ .
ב. מצא אומד ל- θ על פי שיטת המומנטים.
- (6) הזמן בדקות להכנת לחם במאפייה מתפלג באופן הבא: $X_i \sim N(10, \sigma^2)$.
במדגם של הכנת ארבעה לחמים התקבלו התוצאות הבאות: 4, 6, 10, 5.
א. אמוד את σ^2 בשיטת המומנטים על סמך מדגם בגודל n .
ב. מצא את האומדן ל- σ^2 . מה הבעייתיות בתשובה?

תשובות סופיות:

$$\hat{a} = 2\bar{X} - 20 \quad (1)$$

$$\hat{\sigma}^2 = \frac{\sum X_i^2}{n} - \bar{X}^2, \quad \bar{X} = \hat{\mu} \quad (2)$$

$$\hat{n} = 2X \quad \text{א.} \quad \hat{n} = 2\bar{X}_m \quad \text{ב.} \quad \hat{n} = 10\frac{2}{3} \quad \text{ג.} \quad (3)$$

$$\hat{\lambda} = \frac{1}{\bar{X}} \quad (4)$$

$$\hat{\theta} = \frac{\bar{X}}{1 - \bar{X}} \quad \text{ב.} \quad \frac{\theta}{\theta + 1} \quad \text{א.} \quad (5)$$

$$\hat{\sigma}^2 = \frac{\sum X_i^2}{n} - 100 \quad \text{א.} \quad \hat{\sigma}^2 = -55.75 \quad \text{ב.} \quad (6)$$

אומד עקיב:

רקע:

יהי $\hat{\theta}_n$ אומד לפרמטר θ , המתבסס על n תצפיות.

אומד זה יקרא אומד עקיב, אם יתקיים ש: $\hat{\theta}_n \xrightarrow{n \rightarrow \infty} \theta$.

אם $\hat{\theta}_n$ אומד חסר הטיה לפרמטר θ ומתקיים ש: $\lim_{n \rightarrow \infty} V(\hat{\theta}_n) = 0$,

אזי $\hat{\theta}_n$ אומד עקיב ל- θ .

דוגמה (פתרון בהקלטה):

הסבר מדוע $\bar{X}$ אומד עקיב ל- μ .

אם $\hat{\theta}_n$ אומד נראות מקסימלית לפרמטר θ , מתקיים ש: $\hat{\theta}_n \xrightarrow{n \rightarrow \infty} \theta$.

כלומר, $\hat{\theta}_n$ אומד עקיב ל- θ .

דוגמה (פתרון בהקלטה):

הסבר מדוע בהתפלגות גיאומטרית $\frac{1}{X}$, אומד עקיב לפרמטר P .

שאלות:

(1) נתון כי: $X_i \sim U(0, \theta)$, כאשר: $i = 1, 2, \dots, n$.

$2\bar{X}$ מוצע להיות האומד ל- θ .

א. הראה שאומד זה הוא חסר הטיה.

ב. הסבר מדוע האומד הינו עקיב.

(2) נתון ש- $X \sim B(n, p)$. כמו כן, נתון ש: $\hat{p} = \frac{X}{n}$ הינו אומד ל- p .

הוכח שאומד זה הינו אומד עקיב ל- p .

(3) אורך חיי נורה מתפלג מעריכית עם קצב λ לשנה.

נגדיר את: $W_1, W_2, \dots, W_n$ כסדרת זמנים (בשנים) של n נורות בלתי תלויות.

א. מהו אומד הנראות המקסימלי עבור λ ? האם האומד עקיב?

ב. מצא אומד עקיב לסיכוי שנורה כלשהי תישרף תוך פחות משנתיים.

(4) נפח החלב בקרטון חלב מתפלג נורמלית עם תוחלת μ ושונות σ^2 .

מצא אומד עקיב לפרמטר σ^2 , המתבסס על n תצפיות בלתי תלויות.

תשובות סופיות:

(1) א. שאלת הוכחה. ב. ראה סרטון.

(2) שאלת הוכחה.

(3) א. $\frac{1}{\bar{W}}$. ב. $1 - e^{-\frac{2}{\bar{W}}}$.

(4) $\frac{\sum_i (X_i - \bar{X})^2}{n}$.

אומד חסר הטיה בעל שונות מינימלית:

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

רקע:

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

דוגמה (פתרון בהקלטה):

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

X_i - מספר הלקוחות שנכנסו לסניף A ביום i .

Y_j - מספר הלקוחות שנכנסו לסניף B ביום j .

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

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

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

שאלות:

- (1) T_1 ו- T_2 הינם אומדים חסרי הטיה ובלתי תלויים לפרמטר θ .
 כמו כן, נגדיר: $T = aT_1 + bT_2$.
 א. מה צריך להיות התנאי על a ו- b , כדי ש- T יהיה אומד חסר הטיה?
 ב. σ_1^2 ו- σ_2^2 הם השונות של T_1 ו- T_2 , בהתאמה.
 מצאו a ו- b , כך ש- T יהיה אומד חסר הטיה ל- θ , ובעל שונות מינימלית.
- (2) במפעל 3 מכונות המייצרות את אותו חלק.
 תוחלת הקוטר של החלקים המיוצרים בכל מכונה זהה.
 השונות של כל מכונה שונות, ומקיימות: $\sigma_2^2 = 2\sigma_1^2$, $\sigma_3^2 = 3\sigma_1^2$.
 הוחלט לדגום n חלקים מכל מכונה, ולחשב את ממוצע הקוטר המתקבל.
 $\bar{X}_i$ - יהיה הממוצע המתקבל במכונה i .
 יהי: $W = \sum_{i=1}^3 a_i \bar{X}_i$ האומד לתוחלת קוטר החלקים המיוצרים על ידי מכונה כלשהי.
 א. מה התנאי שצריך להתקיים על המשקלים a_i , כדי שהאומד המוצע יהיה בלתי-מוטה?
 ב. נניח ש- $a_1 = a_2$.
 מה במקרה זה המשקלים המביאים את האומד להיות MVUE?

תשובות סופיות:

- (1) א. $a + b = 1$.
 ב. $a = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$, $b = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}$.
- (2) א. $\sum_{i=1}^3 a_i = 1$.
 ב. $a_1 = a_2 = 0.4$, $a_3 = 0.2$.

שאלות מסכמות:

שאלות:

- (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 הראשונה,
 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- $Y = 3$, $X = 6$, מהו אומדן נראות מקסימלית ל- P ?
- (2) מספר תאונות הדרכים בקטע כביש א' מתפלג פואסונית עם קצב של λ
 תאונות בחודש, ומספר תאונות הדרכים בקטע כביש ב' מתפלג פואסונית
 עם קצב של 2λ תאונות בחודש.
 הוחלט לספור את כמות התאונות בחודש בכל אחד מקטעי הכביש.
 נסמן ב- X את מספר התאונות בחודש בקטע א' וב- Y בקטע ב'.
 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, ובו 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 שיחקה 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 אני"מ ל- θ (כאשר λ קבוע ידוע).
ג. מצא אני"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 האומד: $T_2 = X + 0.5Y$. האם האומד הנ"ל הוא חסר הטיה?
ג. איזה אומד יותר עדיף זה של סעיף א או זה של סעיף ב?
ד. מצא אומד נראות מקסימלית ל- θ על סמך X ו- Y .

- (7) נדגמו 2 תצפיות, X_1, X_2 , בלתי תלויות מהתפלגויות אחידות רציפות התלויות בפרמטר θ .

ידוע כי: $X_1 \sim U(0, \theta)$, $X_2 \sim U(0, a\theta)$ (כאשר a קבוע ידוע וחיובי).

- א. מצא אני"מ ל- θ , על סמך 2 התצפיות הנ"ל.
ב. חשב את תוחלת ושונות האני"מ מסעיף א'. האם האני"מ מוטה?
ג. מצא אחי"ה ל- θ על סמך סכומן של 2 התצפיות הנ"ל. מהי שונותו?

תשובות סופיות:

$$(1) \quad \text{א. } 0 \leq P \leq 0.5 \quad \text{ב. } \hat{p} = \frac{x+z}{60} \quad \text{ג. } 0.345$$

$$(2) \quad \text{א. } \frac{x+y}{3} \quad \text{ב. } 1 - e^{-\frac{x+y}{3}} \quad \text{ג. כן.}$$

(3) א. מאחר ולא התבקשתם לפתח, הרי שהאומד זהה לנוסחה הכללית (ראו נספח).
ב. כנ"ל. ג. כנ"ל (ראו הפרק על אומד נראות מקסמילי). ד. לא.

$$(4) \quad \text{א. } 0 \leq P \leq \frac{1}{3} \quad \text{ב. } \frac{1}{3} \quad \text{ג. } 0.389 \quad \text{ד. } -0.167$$

$$(5) \quad \text{א. אח"ה יהיה: } \hat{\lambda} = \frac{\theta+1}{\theta} \bar{x} \quad \text{ב. } \hat{\theta} = \frac{n}{n \ln \lambda - \sum_{i=1}^n \ln x_i}$$

$$\text{ג. } \hat{\lambda} = X_{\max}$$

$$(6) \quad \text{א. } T_1 = (x+y) \frac{2}{3} \quad \text{ב. כן.} \quad \text{ג. ב'.} \quad \text{ד. } \hat{\theta} = \max \left\{ x, \frac{1}{2} y \right\}$$

$$(7) \quad \text{א. } \hat{\theta} = \max \left(X_1, \frac{X_2}{a} \right) \quad \text{ב. } E(\hat{\theta}) = \frac{2}{3} \theta, \quad V(\hat{\theta}) = \frac{1}{18} \theta^2$$

$$\text{ג. } \tilde{\theta} = \left(\frac{2}{1+a} \right) (X_1 + X_2)$$

נספח: אומדי נראות מקסימלית 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. א.נ.מ עבור התוחלת $\frac{1}{p}$, הוא $\bar{X}$ והינו 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-הטיה).

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 :

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

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

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.ח.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.ח.ה.

בכל התפלגות:

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

אומד חסר-הטיה עבור σ^2 :

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

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

ביוסטטיסטיקה מתקדמת

פרק 2 - מבוא לבדיקת השערות על פרמטרים (תזכורת)

תוכן העניינים

31	1. הקדמה
35	2. סוגי טעויות

הקדמה:

רקע:

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

שלב א: נוהה את הפרמטר הנחקר.

שלב ב: נרשום את השערות המחקר.

השערת האפס המסומנות ב- H_0 .

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

השערה אלטרנטיבית (השערת המחקר) המסומנת ב- H_1 .

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

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

שלב ד: נרשום את כלל ההכרעה. בתהליך של בדיקת השערות יוצרים כלל שנקרא כלל הכרעה. הכלל יוצר אזורי שנקראים:

1. **אזור דחייה:**

דחייה של השערת האפס כלומר קבלה של האלטרנטיבה.

2. **אזור קבלה:**

קבלה של השערת האפס ודחייה של האלטרנטיבה. כלל ההכרעה מתבסס על איזשהו סטטיסטי. אזור הדחייה מוכתב על ידי סיכון שלוקח החוקר מראש שנקרא רמת מובהקות ומסומן ב- α .

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

שלב ו: להסיק מסקנה בהתאם לתוצאות המדגם.

דוגמה (פתרון בהקלטה):

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

$$n = 20, \bar{X} = 3120, S = 280$$

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

שאלות:

בשאלות הבאות, ענו על הסעיפים הבאים:

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

- (1) ממוצע הציונים בבחינת הבגרות באנגלית הנו 72 עם סטיית תקן 15 נקודות. מורה טוען שפיתח שיטת לימוד חדשה שתעלה את ממוצע הציונים. משרד החינוך החליט לתת למורה 36 תלמידים אקראיים. ממוצע הציונים של אותם תלמידים לאחר שלמדו בשיטתו היה 75.5.
- (2) לפי הצהרת היצרן של חברת משקאות מסוימת נפח הנוזל בבקבוק מתפלג נורמלית עם תוחלת 500 סמ"ק וסטיית תקן 20 סמ"ק. אגודת הצרכנים מתלוננת על הפחתת נפח המשקה בבקבוק מהכמות המוצהרת. במדגם שעשתה אגודת הצרכנים התקבל נפח ממוצע של 492 סמ"ק במדגם בגודל 25.
- (3) במשך שנים אחוז המועמדים שהתקבל לפקולטה למשפטים היה 25%. השנה מתוך מדגם של 120 מועמדים התקבלו 22. מחקר מעוניין לבדוק האם השנה מקשים על הקבלה לפקולטה למשפטים.
- (4) בחודש ינואר השנה פורסם שאחוז האבטלה במשק הוא 8% במדגם עכשווי התקבל שמתוך 200 אנשים 6.5% מובטלים. רוצים לבדוק ברמת מובהקות של 5% האם כיום אחוז האבטלה הוא כמו בתחילת השנה.

תשובות סופיות:

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

סוגי טעויות:

רקע:

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

1. אזור דחייה – דחייה של השערת האפס כלומר קבלה של האלטרנטיבה.
2. אזור קבלה – קבלה של השערת האפס ודחייה של האלטרנטיבה.

כלל ההכרעה מתבסס על איזשהו סטטיסטי.
בתהליך יש ללכת לתוצאות המדגם ולבדוק האם התוצאות נופלות באזור הדחייה או הקבלה וכך להגיע למסקנה – המסקנה היא בעירבון מוגבל כיוון שהיא תלויה בכלל ההכרעה ובתוצאות המדגם. אם נשנה את כלל ההכרעה אז אנחנו יכולים לקבל מסקנה אחרת. אם נבצע מדגם חדש אז אנחנו עלולים לקבל תוצאה אחרת. לכן יתכנו טעויות במסקנות שלנו:

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

הגדרת הטעויות:

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

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

דוגמה (פתרון בהקלטה):

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

שאלות:

- (1) לפי הצהרת היצרן של חברת משקאות מסוימת נפח הנוזל בבקבוק מתפלג נורמלית עם תוחלת 500 סמ"ק וסטיית תקן 20 סמ"ק. אגודת הצרכנים מתלוננת על הפחתת נפח המשקה בבקבוק מהכמות המוצהרת. במדגם שעשתה אגודת הצרכנים התקבל נפח ממוצע של 492 סמ"ק במדגם בגודל 25. בסופו של דבר הוחלט להכריע לטובת חברת המשקאות.
- רשמו את השערות המחקר.
 - מה מסקנת המחקר?
 - איזו סוג טעות יתכן וביצעו במחקר?
- (2) במחקר על פרמטר מסוים הוחלט בסופו של דבר לדחות את השערת האפס.
- האם ניתן לדעת אם בוצע טעות במחקר?
 - מה סוג הטעות האפשרית?
- (3) לפי נתוני משרד הפנים בשנת 1980 למשפחה ממוצעת היה 2.3 ילדים למשפחה עם סטיית תקן 0.4. ישנה טענה שכיום ממוצע מספר הילדים במשפחה קטן יותר. לצורך כך הוחלט לדגום 121 משפחות. במדגם התקבל ממוצע 2.17 ילדים למשפחה. על סמך תוצאות המדגם נקבע שלא ניתן לקבוע שבאופן מובהק תוחלת מספר הילדים למשפחה קטנה כיום.
- מהי אוכלוסיית המחקר?
 - מה המשתנה הנחקר?
 - מה הפרמטר הנחקר?
 - מה השערות המחקר?
 - מה מסקנת המחקר?
 - מהי סוג הטעות האפשרית במחקר?

תשובות סופיות:

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

ביוסטטיסטיקה מתקדמת

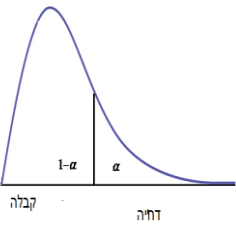
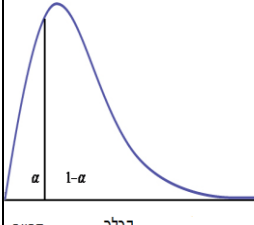
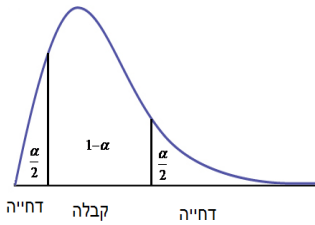
פרק 3 - בדיקת השערות על שונויות

תוכן העניינים

1. בדיקת השערות על שתי שונויות 37

בדיקת השערות על שתי שוניות:

רקע:

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

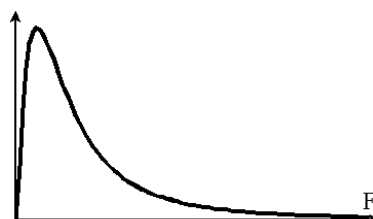
$$F = \frac{S_1^2}{S_2^2} : \text{סטטיסטי המבחן}$$

התפלגות F:

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

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

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



$$df_1 = n_1 - 1$$

$$df_2 = n_2 - 1$$

דוגמה (פתרון בהקלטה) :

מעוניינים להשוות בין נשים וגברים מבחינת השונות בזמנים שלהם לבצע משימה מסוימת. במדגם של 10 גברים התקבלו התוצאות הבאות לגבי זמני ביצוע המשימה : $\sum (y_i - \bar{y})^2 = 204$.

במדגם של 13 נשים התקבלו התוצאות הבאות : $\sum (x_i - \bar{x})^2 = 200$.
בדקו ברמת מובהקות של 2% האם קיים הבדל בין השונות? מה יש להניח?

טבלת ערכים קריטיים לפי התפלגות F																	
$\alpha = 0.05$																	
ד"ח מנהל"ח מנהל	1	2	3	4	5	6	7	8	9	10	12	16	20	24	60	120	∞
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	243.91	246.46	248.01	249.05	252.20	253.25	254.31
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	19.45	19.45	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.69	8.66	8.64	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.84	5.80	5.77	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.60	4.56	4.53	4.43	4.40	4.37
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.92	3.87	3.84	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.49	3.44	3.41	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.20	3.15	3.12	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	2.99	2.94	2.90	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.83	2.77	2.74	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.70	2.65	2.61	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.60	2.54	2.51	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.51	2.46	2.42	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.44	2.39	2.35	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.38	2.33	2.29	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.33	2.28	2.24	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.29	2.23	2.19	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.25	2.19	2.15	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.21	2.16	2.11	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.18	2.12	2.08	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.16	2.10	2.05	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.13	2.07	2.03	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.11	2.05	2.01	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.09	2.03	1.98	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.07	2.01	1.96	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.05	1.99	1.95	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.04	1.97	1.93	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.02	1.96	1.91	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.01	1.94	1.90	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	1.99	1.93	1.89	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.90	1.84	1.79	1.64	1.58	1.51
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.95	1.85	1.78	1.74	1.58	1.51	1.44
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.82	1.75	1.70	1.53	1.47	1.39
90	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99	1.94	1.86	1.76	1.69	1.64	1.46	1.39	1.30
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.73	1.66	1.61	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.64	1.57	1.52	1.32	1.22	1.00

טבלת ערכים קריטיים לפי התפלגות F $\alpha = 0.01$ ראה איור מסה.																	
ד"ח מונה/ד"ח מכנה	1	2	3	4	5	6	7	8	9	10	12	16	20	24	60	120	∞
1	4052.18	4999.50	5403.35	5624.58	5763.65	5858.99	5928.36	5981.07	6022.47	6055.85	6106.32	6170.10	6208.73	6234.63	6313.03	6339.39	6365.86
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.44	99.45	99.46	99.48	99.49	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.83	26.69	26.60	26.32	26.22	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.15	14.02	13.93	13.65	13.56	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.68	9.55	9.47	9.20	9.11	9.02
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.52	7.40	7.31	7.06	6.97	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.28	6.16	6.07	5.82	5.74	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.48	5.36	5.28	5.03	4.95	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.92	4.81	4.73	4.48	4.40	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.52	4.41	4.33	4.08	4.00	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.21	4.10	4.02	3.78	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	3.97	3.86	3.78	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.78	3.66	3.59	3.34	3.25	3.17
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.62	3.51	3.43	3.18	3.09	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.49	3.37	3.29	3.05	2.96	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.37	3.26	3.18	2.93	2.84	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.27	3.16	3.08	2.83	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.19	3.08	3.00	2.75	2.66	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.12	3.00	2.92	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.05	2.94	2.86	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	2.99	2.88	2.80	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.94	2.83	2.75	2.50	2.40	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.89	2.78	2.70	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.85	2.74	2.66	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.81	2.70	2.62	2.36	2.27	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.78	2.66	2.58	2.33	2.23	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.93	2.75	2.63	2.55	2.29	2.20	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.72	2.60	2.52	2.26	2.17	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.87	2.69	2.57	2.49	2.23	2.14	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.66	2.55	2.47	2.21	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.48	2.37	2.29	2.02	1.92	1.80
50	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.78	2.70	2.56	2.38	2.27	2.18	1.91	1.80	1.68
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.31	2.20	2.12	1.84	1.73	1.60
90	6.93	4.85	4.01	3.53	3.23	3.01	2.84	2.72	2.61	2.52	2.39	2.21	2.09	2.00	1.72	1.60	1.46
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.15	2.03	1.95	1.66	1.53	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.00	1.88	1.79	1.47	1.32	1.00

שאלות:

- (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. בדקו ברמת מובהקות של 10% את ההשערה ששונות 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

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

- יחידות מתכת מכל סוג והתקבלו התוצאות הבאות:
 יש להניח שרמת החוזק של המתכות מתפלגת נורמאלית.
 א. האם קיים הבדל בין שונות החוזק של מתכות?
 ב. האם קיים הבדל בין תוחלות החוזק של מתכות?
 בכל סעיף רמת מובהקות של 10%.

תשובות סופיות:

- (1) א. לא נדחה את H_0 . ב. מסקנה לא תשתנה.
 (2) א. לא נדחה את H_0 . ב. לא נדחה את H_0 .

ביוסטטיסטיקה מתקדמת

פרק 4 - מבחני חי בריבוע לאי תלות

תוכן העניינים

1. מבחן לאי תלות 41

מבחן חי בריבוע לאי תלות בין משתנים – רקע

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

מבנה המבחן:

השערות:

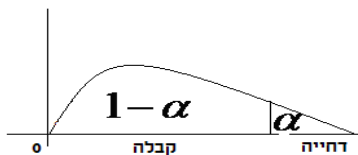
אין תלות בין המשתנים H_0 .

יש תלות בין המשתנים H_1 .

כלל הכרעה:

הערך הקריטי נקבע על סמך התפלגות חי בריבוע. התפלגות זו היא אסימטרית חיובית ותלויה בדרגות החופש $d.f = (r-1)(c-1)$. כאשר: r - מספר הקטגוריות של המשתנה שבשורות. c - מספר הקטגוריות של המשתנה שבעמודות.

הערך הקריטי הוא: $\chi^2_{1-\alpha, (r-1)(c-1)}$, כלומר האחוזון ה- $1-\alpha$ בהתפלגות חי בריבוע שדרגות החופש הן $(r-1)(c-1)$. אם $\chi^2 > \chi^2_{1-\alpha, (r-1)(c-1)}$ אז דוחים את השערת האפס.



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

כאשר:

O_i - השכיחות נצפית במדגם בתא i .

E_i - שכיחות צפויה במדגם בתא i בהנחת השערת האפס.

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

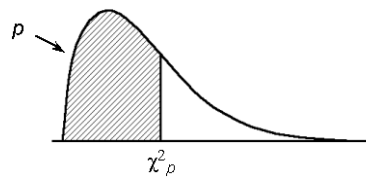
הערה:

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

דוגמה (הפתרון בהקלטה):

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

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

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

df	p												
	.005	.01	.025	.05	.10	.25	.50	.75	.90	.95	.975	.99	.995
1	0.004393	0.00157	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

שאלות

- (1) נבדקה 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? בדקו ברמת בטחון של 95%.

תשובות סופיות

- (1) נסיק שיש קשר בין גודל הארגון לשביעות הרצון של העובדים.
(2) נסיק שאין הבדל מובהק בין שיעור הפגומים במשמרות השונות.
(3) נסיק שאין קשר בין היות מוצר לייצוא למועד שבו הוא יוצר.

ביוסטטיסטיקה מתקדמת

פרק 5 - ניתוח שונות חד כיוונית

תוכן העניינים

1. כללי 45

ניתוח שונות חד כיוונית

רקע תיאורטי

ניתוח שונות (חד כיוונית) הוא מבחן להשוואת תוחלות $(\mu_1, \dots, \mu_k)$ של k אוכלוסיות שונות. לכן, בניתוח שונות, השערות המחקר הן:

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

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

ההנחות הדרושות לביצוע התהליך:

(2) בכל אוכלוסייה מתוך k האוכלוסיות ההתפלגות נורמלית.

(3) כל האוכלוסיות הן עם אותה שונות σ^2 .

(4) המדגמים בלתי תלויים זה בזה.

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

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

$\bar{X}_1$ - ממוצע המדגם הראשון, $\dots, \bar{X}_k$ - ממוצע המדגם ה- k .
 $\bar{X}$ - ממוצע כללי (של כל המדגמים).

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

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

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

$$SST = SSB + SSW$$

יש למלא את טבלת ניתוח השונות הבאה:

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

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

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

שאלות

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

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

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

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

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

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

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

(4) שלושה אופים נתבקשו להכין עוגת שוקולד. לכל אופה בדקו את משך זמן ההכנה בדקות. כל אופה נדרש לאפות בכל יום 4 עוגות.

האם קיים הבדל בין האופים מבחינת תוחלת זמני ההכנה של העוגות?

בדקו ברמת מובהקות של 5%.

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

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

ANOVA

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

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

ANOVA

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

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

ANOVA

pulse

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

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 5.000.

Post Hoc Tests

Multiple Comparisons

pulse
Tukey HSD

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

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

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

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

Oneway

ANOVA

grade

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

Post Hoc Tests

Multiple Comparisons

grade

Scheffe

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

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

Homogeneous Subsets

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.

תשובות סופיות

(1) א. משתנה בלתי תלוי : סוג התרופה. ב. ניתוח שונות חד כיוונית

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

$$H_1 : otherwise$$

משתנה תלוי : הזמן עד להשפעת התרופה בדקות.

ג. 1. מדגמים בלתי תלויים.

2. שוויון שונויות.

3. משתנים מתפלגים נורמלית.

(2) א. המבחן לניתוח שונות חד כיוונית.

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

$$H_1 : otherwise$$

הנחות :

1. מדגמים בלתי תלויים.

2. משתנים מתפלגים נורמלית.

3. שוויון שונויות.

ב. גודל המדגם : 14. משתנה ב"ת : בית הספר, בעל 3 רמות.

$$g. \bar{X} = 71.8, \hat{S} = 12.93, \bar{X} = 73.75, \hat{S} = 19.29, \bar{X} = 80.4, \hat{S} = 5.46$$

ד. להלן טבלה :

מקור השונות	SS	df	MS	F
B	200.6	2	100.3	
W	1904.75	11	173.2	
סה"כ	2105.35	13		0.58

$$h. F > 3.98$$

ו. נקבל את H_0 .

(3) א. $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$
 $H_1 : otherwise$
 ב. נדחה את H_0 .
 ג. כן.

הנחות :

1. מדגמים בלתי תלויים.

2. שוויון שונויות.

3. משתנים מתפלגים נורמלית.

(4) נקבל את H_0 : נכריע שאין הבדל מובהק בין האופים מבחינת תוחלת זמן הכנה.

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 \quad (5)$$

$$H_1 : otherwise$$

הנחות :

1. מדגמים בלתי תלויים.

2. שוויון שונות.

3. משתנים מתפלגים נורמלית.

נקבל את H_0 : לסוג סוללה אין השפעה של תוחלת החיים ברמת ביטחון של 10%.

(6) להלן טבלה :

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	357.450	2 ב	178.725 ג	18.36 ה	.000
Within Groups	165.5 א	17	9.735 ד		
Total	522.950	19			

$$H_0 : \mu_1 = \mu_2 = \mu_3 \quad (7) \quad \text{א.}$$

$$H_1 : otherwise$$

הנחות :

1. מדגמים בלתי תלויים.

2. משתנים מתפלגים נורמלית.

3. שוויון שונות.

נדחה את H_0 : ברמת ביטחון של 5% קיים הבדל במינונים השונים מבחינת תוחלת הדופק.

$$\text{ב. ראה וידאו. ג. כן. ד. } \mu_{20} = \mu_{10} > \mu_{30} .$$

$$(8) \quad \text{א. 5 ב. 25}$$

ג. נדחה את H_0 : יש לפחות שני בתי ספר בעיר עם תוחלת רמת ציונים שונה.

$$\text{ד. } (\mu_3 = \mu_4) < (\mu_1 = \mu_2 = \mu_3) .$$

ביוסטטיסטיקה מתקדמת

פרק 6 - מבחנים אפרמטרים למדגמים מזווגים (ווילקוקוסון)

תוכן העניינים

1. מבחן ווילקוקוסון..... 54
2. מבחן ווילקוקוסון - על ידי שימוש בטבלה לערכים קריטיים..... 57
3. מבחן ווילקוקוסון- פלטים..... 61

מבחן ויילקוקסון למדגמים מזווגים לפי שיטת המנייה – רקע

מתי נשתמש במבחן זה?

מבחן זה לא דורש הנחה של התפלגות נורמלית, אולם דורש ערכים מספריים המאפשרים חישוב הפרש בין ערכי X לערכי Y . מבחן זה הוא הגרסה הלא פרמטרית למבחן T למדגמים מזווג. נשתמש במבחן זה שיש משתנה כמותי שאינו מתפלג נורמלית או שיש משתנה מסולם סדר והמדגם הוא מזווג.

דוגמה (פתרון בהקלטה) :

שני קונדיטורים מתחרים על מקום עבודה. נתנו לשניהם להכין 8 מאפים שונים כאשר כל אחד מהמאפים נאפה על ידי שניהם. בסופו של דבר בעל הקונדיטוריה נתן ציון לכל אחד מהאופים בעבור כל אחד מהמאפים. להלן הציונים שהתקבלו, ורוצים לבדוק האם אופה א טוב יותר מאופה ב.

אופה א	אופה ב
10	9
9	8
7	7
8	9
9	6
10	6
7	5
8	4

א. מהו המבחן הסטטיסטי המתאים?

ב. מהן השערות המחקר?

חישוב סטטיסטי המבחן:

- נחשב את ההפרשים D_i לכל תצפית.
- נוציא מהמדגם את כל התצפיות עם ההפרשים ששווים ל-0.
- נדרג את ההפרשים הנותרים מהקטן אל הגדול בלי להתייחס לסימן ההפרש, כלומר מדרגים את הערכים המוחלטים של ההפרשים. הפרשים זהים מקבלים דרגה זהה שהיא הדרגה הממוצעת של המקומות שהם תופסים.
- מסכמים את הדרגות של ההפרשים החיוביים ($W+$) ואת הדרגות של ההפרשים השליליים ($W-$).

דוגמה (פתרון בהקלטה) :

חשבו את W על סמך תוצאות המדגם.

אופה א	אופה ב
10	9
9	8
7	7
8	9
9	6
10	6
7	5
8	4

חישוב מובהקות התוצאה :

כדי לחשב את מובהקות התוצאה נצטרך לדון בהתפלגות של W בהנחת השערת האפס. אם השערת האפס נכונה, לכל D_i יש את אותו סיכוי לקבל ערך חיובי או ערך שלילי כלומר 2 אפשרויות. ל n^* דרגות יש $2n^*$ אפשרויות שונות, שוות הסתברות. לפי רעיון זה ניתן יהיה לחשב את מובהקות התוצאה.

דוגמה (פתרון בהקלטה) :

א. מהי מובהקות התוצאה?

ב. מה המסקנה ברמת מובהקות של 5%?

שאלות

- (1) נדגמו 8 לקוחות שקיבלו שירות ממוקד טלפוני. לקוחות אלה נתבקשו לתת הערכה על יעילות השירות ועל האדיבות שבשירות. הציונים ניתנו בסקאלה מ-1 (הערכה הנמוכה) עד 10 (הערכה הגבוהה ביותר). להלן התוצאות שהתקבלו:

5	7	5	2	3	4	8	7	הערכה על יעילות השירות	X
4	7	10	8	6	7	7	8	הערכה על אדיבות השירות	Y

בדקו ברמת מובהקות של 5% האם קיים הבדל בין הערכה על יעילות השירות להערכה על אדיבות השירות?

- (2) רוצים לבדוק האם תרופה חדשה להקלת כאבי ראש יעילה יותר מתרופה מוכרת. לצורך כך נלקח מדגם בן 9 אנשים, שנתבקשו להשתמש בתרופה החדשה ובתרופה המוכרת, ולהשוות את יעילותה של התרופה החדשה ליעילות התרופה המוכרת. האנשים במחקר היו צריכים לתת הערכה של יעילות בסקלה של מ-1 עד 100. התוצאות שקיבל היו:

9	8	7	6	5	4	3	2	1	הנבדק
86	100	69	81	75	80	100	90	95	תרופה חדשה
60	60	50	70	75	49	65	76	80	תרופה מוכרת

האם התרופה החדשה משפרת את היעילות ביותר מ 10 נקודות? בדקו ברמת מובהקות של 1%.

תשובות סופיות

- (1) לא נדחה H_0 .
 (2) לא נדחה H_0 .

מבחן ווילקוקסון למדגמים מזווגים (על ידי שימוש בטבלה של ערכים

קריטיים) – רקע

מתי נשתמש במבחן זה?

מבחן זה לא דורש הנחה של התפלגות נורמלית, אולם דורש ערכים מספריים המאפשרים חישוב הפרש בין ערכי X לערכי Y . מבחן זה הוא הגרסה הלא פרמטרית למבחן T למדגמים מזווג. נשתמש במבחן זה שיש משתנה כמותי שאינו מתפלג נורמלית או שיש משתנה מסולם סדר על מדגם מזווג.

דוגמה (פתרון בהקלטה) :

שני קונדיטורים מתחרים על מקום עבודה. נתנו לשניהם להכין 8 מאפים שונים כאשר כל אחד מהמאפים נאפה על ידי שניהם. בסופו של דבר בעל הקונדיטוריה נתן ציון לכל אחד מהאופים בעבור כל אחד מהמאפים. להלן הציונים שהתקבלו, ורוצים לבדוק שאופה א טוב יותר מאופה ב.

אופה א	אופה ב
10	9
9	8
7	7
8	9
9	6
10	6
7	5
8	4

א. מהו המבחן הסטטיסטי המתאים?

ב. מהן השערות המחקר?

חישוב סטטיסטי המבחן:

- (1) נחשב את ההפרשים D_i לכל תצפית.
- (2) נוציא מהמדגם את כל התצפיות עם ההפרשים ששווים ל-0.
- (3) נדרג את ההפרשים הנותרים מהקטן אל הגדול בלי להתייחס לסימן ההפרש, כלומר מדרגים את הערכים המוחלטים של ההפרשים. הפרשים זהים מקבלים דרגה זהה שהיא הדרגה הממוצעת של המקומות שהם תופסים.
- (4) מסכמים את הדרגות של ההפרשים החיוביים (W^+) ואת הדרגות של ההפרשים השליליים (W^-).
- (5) W יהיה W^+ או W^- , זה שאמור להיות יותר קטן לפי השערת המחקר או הקטן מבניהם אם ההשערה היא דו צדדית.

דוגמה (פתרון בהקלטה) :

חשבו את W על סמך תוצאות המדגם.

אופה א	אופה ב
10	9
9	8
7	7
8	9
9	6
10	6
7	5
8	4

כלל הכרעה :

במבחן ווילקוקסון זה כלל ההכרעה הוא : נדחה את H_0 אם $W \leq W_c$.
 כאשר, W_c - הערך הקריטי ; W - הסטטיסטי.
 את הערכים הקריטיים נחלץ מתוך טבלה מתאימה :

n_1	חד-צדדי $\alpha = 0.01$ דו-צדדי $\alpha = 0.02$	חד-צדדי $\alpha = 0.025$ דו-צדדי $\alpha = 0.05$	חד-צדדי $\alpha = 0.05$ דו-צדדי $\alpha = 0.10$
5			1
6		1	2
7	0	2	4
8	2	4	6
9	3	6	8
10	5	8	11
11	7	11	14
12	10	14	17
13	13	17	21
14	16	21	26
15	20	25	30
16	24	30	36
17	28	35	41
18	33	40	47
19	38	46	54
20	43	52	60
21	49	59	68
22	56	66	75
23	62	73	83
24	69	81	92
25	77	90	101
26	85	98	110
27	93	107	120
28	102	117	130
29	111	127	141
30	120	137	152

דוגמה (פתרון בהקלטה) :

- א. רשמו את כלל ההכרעה המתאים ברמת מובהקות של 5%.
- ב. מה המסקנה ברמת מובהקות של 5%?

שאלות

- (1) נדגמו 8 לקוחות שקיבלו שירות ממוקד טלפוני. לקוחות אלה נתבקשו לתת הערכה על יעילות השירות ועל האדיבות שבשירות. הציונים ניתנו בסקאלה מ-1 (הערכה הנמוכה) עד 10 (הערכה הגבוהה ביותר). להלן התוצאות שהתקבלו:

5	7	5	2	3	4	8	7	הערכה על יעילות השירות	X
4	7	10	8	6	7	7	8	הערכה על אדיבות השירות	Y

בדקו ברמת מובהקות של 5% האם קיים הבדל בין הערכה על יעילות השירות להערכה על אדיבות השירות?

- (2) סטודנטים נתבקשו לתת חוות דעתם על רמת הקושי של הקורס (סקאלה של 1-5 כאשר 5=קשה ביותר) ועל רמת הקושי של הבחינות באותה סקאלה. הסטודנטים טוענים שהבחינה הייתה ברמה גבוהה יותר מהרמה של הקורס. להלן תוצאות המדגם:

4	5	1	2	3	4	2	3	4	1-קושי קורס
2	3	5	5	5	3	4	4	4	2-קושי בחינה

בדקו ברמת מובהקות של 5% את טענת הסטודנטים.

- (3) רוצים לבדוק את הטענה שהציונים במבחן בסטטיסטיקה ב גבוהים מאשר בסטטיסטיקה א. נלקחו 10 סטודנטים שסיימו את סטטיסטיקה ב. עבור כל סטודנט נבדק מה הציון בסטטיסטיקה א ומה הציון בסטטיסטיקה ב. להלן התוצאות שהתקבלו:

80	78	84	65	67	82	94	68	74	62	א
82	79	86	80	67	77	90	80	80	70	ב

- א. בדקו ברמת מובהקות של 5% באמצעות מבחן ווילקוקסון.
- ב. כיצד התשובה לסעיף הקודם הייתה משתנה אם יוחלט לתת פקטור של 2 נקודות לכל הסטודנטים בשני המועדים?
- ג. כיצד הייתה משתנה התשובה אם מסתבר שנפלה טעות ועבור הסטודנט הראשון ברשימה יש להחליף בנתונים את הציון של סטטיסטיקה ב עם סטטיסטיקה א?

- (4) רוצים לבדוק האם תרופה חדשה להקלת כאבי ראש יעילה יותר מתרופה מוכרת. לצורך כך נלקח מדגם בן 9 אנשים, שנתבקשו להשתמש בתרופה החדשה ובתרופה המוכרת, ולהשוות את יעילותה של התרופה החדשה ליעילות התרופה המוכרת.
- האנשים במחקר היו צריכים לתת הערכה של יעילות בסקלה של מ-1 עד 100. התוצאות שקיבל היו:

הנבדק	1	2	3	4	5	6	7	8	9
תרופה חדשה	95	90	100	80	75	81	69	100	86
תרופה מוכרת	80	76	65	49	75	70	50	60	60

האם התרופה החדשה משפרת את היעילות ביותר מ 10 נקודות? בדקות ברמת מובהקות של 1%.

תשובות סופיות

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

ניתוח פלטי SPSS במבחן ווילקוקסון למדגמים מזווגים – רקע

מבחן זה לא דורש הנחה של התפלגות נורמלית, אולם דורש ערכים מספריים המאפשרים חישוב הפרש בין ערכי X לערכי Y . מבחן זה הוא הגרסה הלא פרמטרית למבחן t למדגם מזווג. נשתמש במבחן זה שיש משתנה כמותי שאינו מתפלג נורמלית או שיש משתנה מסולם סדר. נראה איך מנתחים פלט של תכנת SPSS במבחן זה על ידי דוגמה.

דוגמה (פתרון בהקלטה) :

קופות החולים טוענות כי רכישת תרופות שאינן דורשות מרשם רופא, הינן זולות יותר אצלן מאשר ברשתות הפארם. דגמו 11 תרופות ובדקו את מחירן בבית המרקחת של קופות החולים וברשת הפארם. המחיר המוצג הינו עבור קפסולה בודדת :

שם התרופה	קופת חולים	פארם
אדוויל	1.2	1.5
אקמול	2.6	2.6
אופטלגין	0.9	1.4
פוסטינור	3.5	3.2
סטרפסיל	1.1	1.4
נורפן	1.7	1.8
לורסטין	0.8	1.1
קולדקס	1.5	2
אלרגיז	2	2.8
נוסידקס	2	2.5
קורמיר	3	3.3

להלן פלט שמתקבל מהרצת מבחן ווילקוקסון על הנתונים הללו:

Wilcoxon Signed Ranks Test

		Ranks		
		N	Mean Rank	Sum of Ranks
HMO-PHARM	Negative Ranks	9 ^a	???	51.00
	Positive Ranks	1 ^b	4.00	4.00
	Ties	1 ^c		
	Total	11		

a. HMO < PHARM

b. HMO > PHARM

c. HMO = PHARM

Test Statistics^a

	HMO-PHARM
Z	-2.434 ^b
Asymp. Sig. (2-tailed)	.015
Exact Sig. (2-tailed)	.014
Exact Sig. (1-tailed)	.007

a. Wilcoxon Signed Ranks Test

b. Based on positive ranks.

א. הסבירו מדוע מבחן ווילקוקסון למדגמים מזווגים מתאים למקרה זה?

ב. השלם את הערך החסר שמסומן בסימני שאלה בטבלה.

ג. מה תהיה מסקנת המחקר ברמת מובהקות של 5%?

שאלות

1) רוצים לבדוק את הטענה שהציונים במבחן בסטטיסטיקה ב גבוהים מאשר בסטטיסטיקה א. נלקחו 10 סטודנטים שסיימו את סטטיסטיקה ב. עבור כל סטודנט נבדק מה הציון בסטטיסטיקה א ומה הציון בסטטיסטיקה ב. להלן תוצאות הפלט שהתקבל:

Wilcoxon Signed Ranks Test

Ranks				
		N	Mean Rank	Sum of Ranks
STAS2- STAS1	Negative Ranks	2 ^a	6.50	13.00
	Positive Ranks	8 ^b	5.25	42.00
	Ties	0 ^c		
	Total	10		

a. STAS2 < STAS1

b. STAS2 > STAS1

c. STAS2 = STAS1

Test Statistics^a

	STAS2- STAS1
Z	-1.483 ^b
Asymp. Sig. (2-tailed)	.138

a. Wilcoxon Signed Ranks Test

b. Based on negative ranks.

א. בדקו ברמת מובהקות של 5% את הטענה.
 ב. כיצד התשובה לסעיף הקודם הייתה משתנה אם יוחלט לתת פקטור של 2 נקודות לכל הסטודנטים בשני המועדים?

(2) מחקר בדק את רמת שביעות הרצון משירות לקוחות אחרי רפורמה שבוצעה בחברה. להלן תוצאות שהתקבלו כאשר שביעות הרצון הייתה בסקלה מ-1 כלל לא מרוצה ועד 5 מרוצה מאד.

4	1	1	4	5	4	2	5	4	5	4	2	Before
5	5	5	4	3	5	5	5	4	4	4	4	After

Wilcoxon Signed Ranks Test

		Ranks		
		N	Mean Rank	Sum of Ranks
after-before	Negative Ranks	?????a	3.25	?????
	Positive Ranks	?????b	4.92	?????
	Ties	?????c		
	Total	12		

a. after < before

b. after > before

c. after = before

Test Statistics^a

	after-before
Z	-1.622
Asymp. Sig. (2-tailed)	.105

- א. השלימו את הערכים עם סימני השאלה אשר בטבלה.
 ב. בדקו ברמת מובהקות של 5% האם הרפורמה הייתה יעילה.
 ג. כיצד הייתה משתנה מובהקות התוצאה אם היו מוסיפים עוד 2 תצפיות שעבורן הפרש שביעות הרצון היה אפס?
 ד. כיצד הייתה משתנה מובהקות התוצאה של הפלט אם הינו מוסיפים תצפית כזו:

before	after
5	4

תשובות סופיות

(1) א. לא נדחה H_0 . ב. לא תשתנה המסקנה.

(2) א. 2, 6, 4, 6.5, 29.5 ב. לא נדחה H_0 .

ג. לא משתנה. ד. גדלה.

ביוסטטיסטיקה מתקדמת

פרק 7 - מבחנים אפרמטריים למדגמים בלתי תלויים (מאן וויטני, קרוסקל ואליס ופסיר)

תוכן העניינים

1. מבחן ווילקוקסון למדגמים בלתי תלויים..... 65
2. מבחן מאן וויטני - שימוש בפלטים..... 72
3. מבחן קרוסקל ווליס - ניתוח פלטים..... 76
4. מבחן פיסר..... 80

מבחן ווילקוקסון למדגמים בלתי תלויים – רקע

מבחן ווילקוקסון למדגמים בלתי תלויים נכנס לקטגוריות המבחנים האפרמטריים. מבחן זה רלבנטי כאשר רוצים להשוות בין שתי אוכלוסיות על סמך שני מדגמים בלתי תלויים. המשתנה התלוי הוא משתנה כמותי שאינו מתפלג נורמלית או משתנה מסולם סדר. מבחן זה הוא החלופה האפרמטרית למבחן הפרמטרי להשוואת תוחלות על סמך שני מדגמים בלתי תלויים.

דוגמה:

מחקר חינוכי מעוניין להשוות בין 2 שיטות חינוך. המחקר רוצה לבדוק האם קיים הבדל ברמת ביטחון העצמי של הילדים בשיטות החינוך השונות. נבחרו באקראי 5 ילדים שחונכו בשיטת A. כמו כן נדגמו באקראי 5 ילדים שחונכו בשיטת B. פסיכולוגים בחנו את 10 הילדים ונתנו ציון לביטחון העצמי בסקאלה של 1-20. מהן ההשערות ומהו המבחן הסטטיסטי המתאים?

שיטה A	שיטה B
16	14
17	14
20	19
10	9
18	8

כדי לבצע את המבחן יש לחשב על סמך תוצאות המדגם את סטטיסטי המבחן שנסמן באות U.

השלבים לחישוב סטטיסטי המבחן:

- נסדר את כלל התצפיות של המחקר בסדר עולה מהנמוך ביותר לגבוה ביותר אך יש לדעת כל תצפית מאיזה מדגם היא באה.
- נדרג את כלל התצפיות של המחקר (אם יש תצפיות עם ערכים זהים הדירוג שלהן יהיה ממוצע המקומות שהם תופסים)
- נחשב את W_1 - סכום הדירוגים של התצפיות השייכות למדגם 1,
- W_2 - סכום הדירוגים של התצפיות השייכות למדגם 2.
- נחשב את הגדלים הבאים:

$$U_1 = W_1 - \frac{n_1(n_1+1)}{2} \quad U_2 = W_2 - \frac{n_2(n_2+1)}{2}$$

- הסטטיסטי U: במבחן דו צדדי $U = \min(U_1, U_2)$ במבחן חד צדדי U יהיה ה- U_i , שאמור להיות יותר קטן לפי השערת המחקר.

כדי להגיע למסקנה יש שני סוגים של טבלאות סטטיסטיות.
 טבלה מהסוג הראשון: עוזרת לנו לחשב את מובהקות התוצאה לאחר שחישבנו את
 ה- U הסטטיסטי.

טבלה מהסוג השני שקובעת מראש את הערך הקריטי של U שנסמן ב U_c .

טבלה מהסוג הראשון	טבלה מהסוג השני
$n_1 - n_2 \leq 8$	n_1 או $n_2 \geq 9$
עוזרת לחשב על סמך תוצאות המדגם את P_v אם $P_v \leq \alpha$ דוחים את H_0 .	נותנת את הערך הקריטי U_c כל הכרעה: נדחה את H_0 אם $U \leq U_c$

דוגמה:

מה המסקנה ברמת מובהקות של 5%?

טבלאות למציאת מובהקות התוצאה במבחן ווילקוקסון למדגמים בלתי תלויים
 $n_2 = 3$

u	n_1		
	1	2	3
0	0.250	0.100	0.050
12	0.500	0.200	0.100
3	0.750	0.400	0.200
4		0.600	0.350
5			0.500
			0.650

 $n_2 = 4$

u	n_1			
	1	2	3	4
0	0.200	0.067	0.028	0.014
1	0.400	0.133	0.057	0.029
2	0.600	0.267	0.114	0.057
3		0.400	0.200	0.100
4		0.600	0.314	0.171
5			0.429	0.243
6			0.571	0.343
7				0.443
8				0.557

 $n_2 = 5$

u	n_1				
	1	2	3	4	5
0	0.167	0.047	0.018	0.008	0.004
1	0.333	0.095	0.036	0.016	0.008
2	0.500	0.190	0.071	0.032	0.016
3	0.667	0.286	0.125	0.056	0.028
4		0.429	0.196	0.095	0.048
5		0.571	0.286	0.143	0.075
6			0.393	0.206	0.111
7			0.500	0.278	0.155
8			0.607	0.365	0.210
9				0.452	0.274
10				0.548	0.345
11					0.421
12					0.500
13					0.579

$$n_2 = 6$$

u	n_1					
	1	2	3	4	5	6
0	0.143	0.036	0.012	0.005	0.002	0.001
1	0.286	0.071	0.024	0.010	0.004	0.002
2	0.428	0.143	0.048	0.019	0.009	0.004
3	0.571	0.214	0.083	0.033	0.015	0.008
4		0.321	0.131	0.057	0.026	0.013
5		0.429	0.190	0.086	0.041	0.021
6		0.571	0.274	0.129	0.063	0.032
7			0.357	0.176	0.089	0.047
8			0.452	0.238	0.123	0.066
9			0.548	0.305	0.165	0.090
10				0.381	0.214	0.120
11				0.457	0.268	0.155
12				0.545	0.331	0.197
13					0.396	0.242
14					0.465	0.294
15					0.535	0.350
16						0.409
17						0.469
18						0.531

$$n_2 = 7$$

u	n_1						
	1	2	3	4	5	6	7
0	0.125	0.028	0.008	0.003	0.001	0.001	0.000
1	0.250	0.056	0.017	0.006	0.003	0.001	0.001
2	0.375	0.111	0.033	0.012	0.005	0.002	0.001
3	0.500	0.167	0.058	0.021	0.009	0.004	0.002
4	0.625	0.250	0.092	0.036	0.015	0.007	0.003
5		0.333	0.133	0.055	0.024	0.011	0.006
6		0.444	0.192	0.082	0.037	0.017	0.009
7		0.556	0.258	0.115	0.053	0.026	0.013
8			0.333	0.158	0.074	0.037	0.019
9			0.417	0.206	0.101	0.051	0.027
10			0.500	0.264	0.134	0.069	0.036
11			0.583	0.324	0.172	0.090	0.049
12				0.394	0.216	0.117	0.064
13				0.464	0.265	0.147	0.082
14				0.538	0.319	0.183	0.104
15					0.378	0.223	0.130
16					0.438	0.267	0.159
17					0.500	0.314	0.191
18					0.562	0.365	0.228
19						0.418	0.267
20						0.473	0.310
21						0.527	0.355
22							0.402
23							0.451
24							0.500
25							0.549

$$n_2 = 8$$

u	n_1							
	1	2	3	4	5	6	7	8
0	0.111	0.022	0.006	0.002	0.001	0.000	0.000	0.000
1	0.222	0.044	0.012	0.004	0.002	0.001	0.000	0.000
2	0.333	0.089	0.024	0.008	0.003	0.001	0.001	0.000
3	0.444	0.133	0.042	0.014	0.005	0.002	0.001	0.001
4	0.556	0.200	0.067	0.024	0.009	0.004	0.002	0.001
5		0.267	0.097	0.036	0.015	0.006	0.003	0.001
6		0.356	0.139	0.055	0.023	0.010	0.005	0.002
7		0.444	0.188	0.077	0.033	0.015	0.007	0.003
8		0.556	0.248	0.107	0.047	0.021	0.010	0.005
9			0.315	0.141	0.064	0.030	0.014	0.007
10			0.387	0.184	0.085	0.041	0.020	0.010
11			0.461	0.230	0.111	0.054	0.027	0.014
12			0.539	0.285	0.142	0.071	0.036	0.019
13				0.341	0.177	0.091	0.047	0.025
14				0.404	0.217	0.114	0.060	0.032
15				0.467	0.262	0.141	0.076	0.041
16				0.533	0.311	0.172	0.095	0.052
17					0.362	0.207	0.116	0.065
18					0.416	0.245	0.140	0.080
19					0.472	0.286	0.168	0.097
20					0.528	0.331	0.198	0.117
21						0.377	0.232	0.139
22						0.426	0.268	0.164
23						0.475	0.306	0.101
24						0.525	0.347	0.221
25							0.389	0.253
26							0.433	0.287
27							0.478	0.323
28							0.522	0.360
29								0.399
30								0.439
31								0.480
32								0.520

טבלה למציאת U_c

ברמת מובהקות של 5% למבחן חד צדדי או ברמת מובהקות של 10% למבחן דו צדדי במבחן ווילקוקסון למדגמים בלתי תלויים.

n_1	n_2											
	9	10	11	12	13	14	15	16	17	18	19	20
1											0	0
2	1	1	1	2	2	2	3	3	3	4	4	4
3	3	4	5	5	6	7	7	8	9	9	10	11
4	6	7	8	9	10	11	12	14	15	16	17	18
5	9	11	12	13	15	16	18	19	20	22	23	25
6	12	14	16	17	19	21	23	25	26	28	30	32
7	15	17	19	21	24	26	28	30	33	35	37	39
8	18	20	23	26	28	31	33	36	39	41	44	47
9	21	24	27	30	33	36	39	42	45	48	51	54
10	24	27	31	34	37	41	44	48	51	55	58	62
11	27	31	34	38	42	46	50	54	57	61	65	69
12	30	34	38	42	47	51	55	60	64	68	72	77
13	33	37	42	47	51	56	61	65	70	75	80	84
14	36	41	46	51	56	61	66	71	77	82	87	92
15	39	44	50	55	61	66	72	77	83	88	94	100
16	42	48	54	60	65	71	77	83	89	95	101	107
17	45	51	57	64	70	77	83	89	96	102	109	115
18	48	55	61	68	75	82	88	95	102	109	116	123
19	51	58	65	72	80	87	94	101	109	116	123	130
20	54	62	69	77	84	92	100	107	115	123	130	138

שאלות

- (1) מעוניינים להשוות בין שתי קבוצות כדורסל. נלקחו 5 משחקים מקבוצה א' ושישה משחקים מקבוצה ב'. נבדק בכל משחק ועבור כל קבוצה מספר הנקודות שצברה במשחק. בדקו ברמת מובהקות של 5% האם קיים הבדל בין הקבוצות מבחינת הניקוד שצברה במשחק.

קבוצה א	קבוצה ב
68	82
82	74
78	82
94	64
87	67
	65

- (2) מעוניינים לבדוק האם קורס קיץ באנגלית משפר את יכולות האנגלית לתלמידי חטיבת ביניים. נלקחו 20 ילדים בגיל חטיבת הביניים ברמת אנגלית דומה. 12 מהם נשלחו לקורס קיץ והיתר לא. בסוף הקיץ כולם נבחנו במבחן באנגלית הציון הגבוה ביותר התקבל בקרב אחד שלא עשה את הקורס ושבעת הציונים הנמוכים ביותר היו גם בקרב תלמידים שלא עשו את הקורס. מה המסקנה ברמת מובהקות של 5%?
- (3) במחקר לבדיקת יעילות ויטמין C נבחרו 15 מתנדבים מבין עובדי המפעל. תשעה מהם נבחרו מקרית וקיבלו טיפול שוטף בוויטמין C, ואילו שאר המתנדבים (קבוצת הביקורת) קבלו גלולת סוכר. במשך שלוש שנות המחקר היו מספר ימי ההיעדרות בגלל ההצטננות:
קבוצת הטיפול: 1, 3, 9, 3, 4, 0, 8, 12, 16.
קבוצת הביקורת: 19, 7, 28, 13, 23, 12.
 בדקו ברמת מובהקות של 5% שמספר ימי המחלה במשך שלוש שנים מצטמצם ביותר מ-4 ימים עם לקיחת ויטמין C.

תשובות סופיות

- (1) לא נדחה את H_0 .
- (2) נדחה את H_0 .
- (3) לא נדחה את H_0 .

\\

מבחנים אפרמטריים למדגמים בלתי תלויים

ניתוח פלטים במבחן מן-וויטני – רקע

מבחן מן-וויטני מיועד לבדוק האם לשתי אוכלוסיות התפלגות שווה. המבחן בוחן באופן רוחבי את כל תחום הערכים ולא מתמקד בערך מרכזי אחד. נשתמש במבחן זה כאשר יש שני מדגמים בלתי תלויים והמשתנה הכמותי הנחקר אינו מתפלג נורמלית או שמדובר במשתנה מסולם סדר. המבחן מתבסס על דירוג כל התצפיות. בעצם, מבחן זה הוא המענה האפרמטרי למבחן הפרמטרי להפרש תוחלות במדגמים בלתי תלויים.

דוגמה (פתרון בהקלטה) :

להלן תוצאות הערכות שקיבלו שני מורים : ד"ר A ופרופסור B. סטודנטים נתנו משוב כללי על המורים בסקלה של 1 (גרוע) עד 5 (מצוין). הטענה היא שד"ר A הוא מרצה טוב יותר מאשר פרופסור B.

Mann-Whitney Test

Ranks

	teacher	N	Mean Rank	Sum of Ranks
grade	dr A	17	25.00	425.00
	prof B	20	13.90	278.00
	Total	37		

Test Statistics^a

	grade
Mann-Whitney U	68.000
Wilcoxon W	278.000
Z	-3.249
Asymp. Sig. (2-tailed)	.001
Exact Sig. [2*(1-tailed Sig.)]	.001 ^b

a. Grouping Variable: teacher

b. Not corrected for ties.

- הסבירו מדוע נעשה כאן מבחן מן-וויטני.
- מה המסקנה ברמת מובהקות של 1%?

שאלות

- 1) מחקרים טוענים שקיים הבדל בין שעות השינה של גברים לשעות השינה של נשים. כיוון שלא ניתן להוכיח ששעות שינה הינו משתנה המתפלג נורמלית ביצעו מבחן מן-וויטני בו לקחו נשים וגברים אקראיים ובדקו את שעות השינה שלהם.

Ranks				
	gender	N	Mean Rank	Sum of Ranks
sleeptime	male	10	???	135.00
	female	???	7.50	75.00
	Total	20		

Mann-Whitney Test

Test Statistics ^a	
	sleeptime
Mann-Whitney U	20.000
Wilcoxon W	75.000
Z	-2.319
Asymp. Sig. (2-tailed)	.020

- א. השלימו את סימני השאלה החסרים בפלט.
 ב. מהי המסקנה ברמת מובהקות של 5%?
 ג. מה הייתה מובהקות התוצאה אם טענת המחקר הייתה שגברים ישנים יותר מנשים?

2) שני אנשים נתבקשו לבדוק את מספר תאונות הדרכים בשבוע בשני קטעי כביש שונים. כל אחד בחר את השבועות באופן אקראי ובלתי תלוי באחר וספר כמה תאונות היו בכל כביש בשבוע. הפלטים שהתקבלו:

מספר תאונות

	Frequency
.00	2
1.00	1
2.00	2
3.00	1
Total	6

a. road = 1.00

מספר תאונות

	Frequency
.00	2
1.00	1
2.00	1
3.00	1
4.00	1
5.00	1
Total	7

a. road = 2.00

Ranks

	road	N	Mean Rank	Sum of Ranks
מספר תאונות	1.00	A	B	C
	2.00	D	E	F
	Total	G		

Test Statistics^a

	VAR00002
Asymp. Sig. (2-tailed)	.465

- א. השלימו בטבלה השלישית את המספרים החסרים במקום האותיות.
- ב. מהי מובהקות התוצאה לבדיקת ההשערה שאין הבדל בין הכבישים מבחינת התפלגות תאונות הדרכים?
- ג. כיצד הייתה משתנה התשובה של הסעיף הקודם אם כל חוקר היה מוסיף נתונים על שבוע נוסף לכל כביש?
- ד. מה המסקנה ברמת מובהקות של 5%?
- ה. מהי מובהקות התוצאה לבדיקת ההשערה שכביש מספר 1 עם התפלגות תאונות גבוהה יותר מאשר כביש מספר 2?

תשובות סופיות

- (1) א. 10, 13.5 ב. נדחה H_0 ג. 0.01
- (2) א. 6, 6.17, 7, 7.71, 13, 54 ב. 0.465
- ג. לא ניתן לדעת. ד. לא נדחה H_0 ה. 0.7675

מבחן קרוסקל וואליס – ניתוח פלטים – רקע

כאשר לא ניתן לבצע מבחן ניתוח שונות חד כיווני (למשל, כאשר השונויות לא שוות, המשתנה התלוי אינו מתפלג נורמאלי או המשתנה התלוי נמדד בסולם סדר) נבחר במבחן לא-פרמטרי להשוואת k קבוצות בלתי תלויות.

מבחן Kruskal-Wallis הוא מבחן אפרמטרי להשוואת מספר אוכלוסיות בלתי תלויות. מבחן מן-ווייטני הינו מקרה פרטי של מבחן Kruskal-Wallis כאשר יש שתי קבוצות בלתי תלויות.

השערות המבחן:

- השערת האפס: אין הבדל בין הקבוצות הנבדקות.
 - השערת המחקר: קיים הבדל בין 2 קבוצות לפחות.
- מבחן זה מתבסס על דירוג כלל התצפיות במחקר מכל המדגמים.

דוגמה (פתרון בהקלטה):

השוו את שביעות הרצון מהשירות של 4 חברות טלפון סלולרי שונות. שביעות הרצון הוא ציון שהלקוח היה צריך לתת לרמת השירות של החברה מ-1 ועד 10.

- מהו המשתנה התלוי ומהו המשתנה הבלתי תלוי במחקר זה?
- מהן השערות המחקר?
- מהו המבחן הסטטיסטי במקרה זה?
- מה מסקנות המחקר ברמת מובהקות של 5%?

Count

		company				Total
		1	2	3	4	
שביעות רצון	4.00	0	2	0	0	2
	5.00	0	1	0	0	1
	6.00	2	1	0	1	4
	7.00	2	2	1	0	5
	8.00	1	2	2	2	7
	9.00	1	1	3	2	7
	10.00	1	1	0	0	2
Total		7	10	6	5	28

Kruskal-Wallis Test

Ranks

	company	N	Mean Rank
שביעות רצון	1	7	13.93
	2	10	11.40
	3	6	18.50
	4	5	16.70
	Total	28	

Test Statistics^{a,b}

	שביעות רצון
Chi-Square	3.363
df	3
Asymp. Sig.	.339

a. Kruskal Wallis Test

b. Grouping Variable: company

שאלות

- 1) מחקר פסיכולוגי רצה להשוות 3 טיפול שונים לטיפול בחרדה. נלקחו נבדקים הסובלים מחרדה. הם חולקו באקראי ל-3 קבוצות וכל קבוצה קיבלה שיטת טיפול אחרת לחרדה. להלן תוצאות המבחן הסטטיסטי שבוצע:

Kruskal-Wallis Test

Ranks

	group	N	Mean Rank
anxiety	1.00	8	15.88
	2.00	4	6.88
	3.00	8	6.94

Test Statistics^{a,b}

	anxiety
Chi-Square	11.149
df	2
Asymp. Sig.	.004

- א. כמה נבדקים השתתפו במחקר?
 ב. באיזו שיטת טיפול רמת החרדה הייתה הנמוכה ביותר במדגם?
 ג. האם, עבור ר"מ 0.05, יש הבדל ברמת החרדה המתקבלת בשיטות הטיפול השונות?

(2) חברת תרופות לקחה 15 אנשים ברמת בריאות דומה. החברה חילקה את האנשים לשלוש 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 נבדק קצב פעימות הלב של כל אדם (pulse). הנתונים שלהלן הוזנו לתוכנה סטטיסטית, והתקבלו התוצאות הבאות:

30	30	30	30	30	20	20	20	20	20	10	10	10	10	10	מינון
72	73	75	67	68	82	82	82	76	79	79	81	83	87	87	דופק

Kruskal-Wallis Test

Ranks

	dosage	N	Mean Rank
pulse	10.00	A	E
	20.00	B	F
	30.00	C	G
	Total	D	

Test Statistics^{a,b}

	pulse
Chi-Square	10.204
df	2
Asymp. Sig.	.006

- א. השלימו את המספרים החסרים במקום האותיות A עד G בטבלת הדירוגים.
- ב. בדוק ברמת מובהקות של 5% האם קיים הבדל בין המינונים השונים מבחינת הדופק של האנשים?
- ג. הסבירו ללא חישוב כיצד הייתה משתנה התשובה לסעיף הקודם אם הינו מעלים את הדופק של כל התצפיות במחקר ב-2.

תשובות סופיות

- (1) א. 20 ב. מדגם 2 ג. נדחה H_0 .
- (2) א. 3, 9.3, 11.7, 5, 5, 5, 15 ב. נדחה H_0 . ג. לא תשתנה.

מבחן פישר – רקע

מבחן זה הוא מבחן הנכנס לקטגוריית המבחנים האפרמטריים. משתמשים במבחן כאשר מתעניינים להשוות בין שתי אוכלוסיות והמשתנה התלוי הוא דיכוטומי, כלומר משתנה שיש לו שני ערכים אפשריים. במבחן זה יוצרים מדגם אחד שמקבל טיפול כלשהו ומדגם אחר המהווה קבוצת ביקורת ואינו מקבל את הטיפול. מבחן זה הוא החלופה האפרמטרית למבחן הפרמטרי להשוואת שתי פרופורציות על סמך שני מדגמים בלתי תלויים. במבחן הפרמטרי דורשים שבכל מדגם מספר ההצלחות וגם מספר הכישלונות יהיה לפחות 10 (הפחות מחמירים דורשים לפחות 5). מבחן פישר, אותו נעשה בפרק זה, נקרא בספרות המקצועית: "Fisher exact probability test".

דוגמה:

מעוניינים לבדוק האם שיעורי עזר יעילים בשיפור ההישגים בקורס סטטיסטיקה. נלקחו 2 כיתות הלומדות סטטיסטיקה בנות 15 תלמידים כל אחת. בכיתה אחת ניתנו שיעורי עזר ובכיתה השנייה לא ניתנו שיעורי עזר. בכיתה בה ניתנו שיעורי עזר נכשל בקורס ובכיתה שבה לא ניתנו שיעורי עזר 3 נכשלו בקורס. מהן השערות המחקר ומהו המבחן הסטטיסטי המתאים?

הטכניקה הנוחה ביותר למבחן פישר היא לחשב את מובהקות התוצאה ולדחות את השערת האפס אם $\alpha \geq PV$. מובהקות התוצאה היא הסיכוי לתוצאות של המדגם וקיצוני יותר בהנחת השערת האפס. כדי לחשב את מובהקות התוצאה נבנה טבלת שכיחות משותפת במבנה הבא:

סה"כ	"הצלחה"	"כישלון"	
A+B	B	A	קבוצת טיפול
C+D	D	C	קבוצת ביקורת
N	B+D	A+C	סה"כ

בנו את טבלת השכיחות המשותפת המתאימה לדוגמה:

סה"כ			
A+B	B	A	קבוצת טיפול
C+D	D	C	קבוצת ביקורת
N	B+D	A+C	סה"כ

דוגמה:

בהנחת השערת האפס התפלגות של A הינו משתנה מקרי היפר גיאומטרי שבו יש אוכלוסייה בגודל N , מתוכם $C + A$ "מיוחדים" ואנו דוגמים מתוכם מדגם בגודל $B + A$.

פונקציית ההסתברות של ההתפלגות ההיפר גיאומטרית במקרה זה,

$$\text{כאשר } X \text{ מייצג את השכיחות } A \text{ תהיה: } \frac{\binom{A+C}{x} \binom{B+D}{A+B-x}}{\binom{N}{A+B}}.$$

חשבו את מובהקות התוצאה בדוגמה ומה תהיה המסקנה ברמת מובהקות של 5%?

שאלות

(1) פסיכיאטרים נשאלו האם תרופה אנטי דיכאונית מסוימת אכן משפיעה על מצב הרוח. נלקחו 28 אנשים שהתלוננו על דיכאון ברמה דומה והם חולקו באקראי לשתי קבוצות: 16 נטלו את התרופה האנטי דיכאונית הנחקרת והיתר היוו קבוצת ביקורת ונטלו פלסיבו. כעבור 3 חודשים נבדק מצבם הנפשי של כל משתתפי המחקר. בקרב אלו שנטלו את התרופה רק 2 התלוננו על דיכאון ובקרב אלו שנטלו הפלסבו 6 התלוננו על דיכאון. מה המסקנה ברמת מובהקות של 10%?

(2) חנות פרחים מעוניינת לבדוק את הטענה שתאורה אולטרה סגולה מגדילה את אורך החיים של הפרחים. נדגמו 20 פרחים מאותו סוג. הם חולקו באקראי ל-2 קבוצות: 10 פרחים יהיו בקבוצת הניסוי, כלומר בתאורה אולטרה סגולה והפרחים הנותרים יהיו בקבוצת הביקורת – באותם התנאים בדיוק אך ללא תאורה אולטרה סגולה. כעבור 5 ימים נבדקו כלל הפרחים.

תקין	נבול	תאורה / מצב הפרח לאחר 5 ימים
9	1	אולטרה סגולה
3	7	רגילה

א. מה היא רמת המובהקות המינימלית עבורה יוסק שתאורה אולטרה סגולה מגדילה את אורך החיים של הפרחים?
 ב. כיצד התשובה לסעיף הקודם הייתה משתנה אם בקבוצת הביקורת היו נמצאים פחות פרחים נבולים?

(3) משרד החינוך הזמין מחקר שמטרתו היה לבדוק האם שנת צהריים קבועה בזמן לימודי התיכון משפיעה על הזכאות לבגרות. נדגמו בוגרי תיכון אקראיים שנשאלו שתי שאלות:
 Q1 - האם בזמן התיכון נהגת לישון צהריים באופן קבוע?
 Q2 - האם את/ה זכאית לתעודת בגרות?
 להלן התוצאות שהתקבלו.
 א. רשמו את השערות המבחן. מהו המבחן הסטטיסטי המתאים? נמקו.
 ב. מהי מובהקות התוצאה?
 ג. בדקו את השערות המחקר ברמת מובהקות של 6%.

שאלון/ מספר נשאל	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Q1	כן	לא	לא	לא	כן	כן	לא	לא	לא	לא	לא	כן	לא	לא
Q2	כן	כן	כן	לא	כן	כן	לא	לא	כן	כן	לא	כן	כן	לא

תשובות סופיות

- (1) נדחה את H_0 .
- (2) א. 0.0099 ב. תקטן.
- (3) א. מבחן פישר. ב. 0.8112 ג. לא נדחה H_0 .