

תרגיל בנושא סטטיסטיקה אי פרמטרית ולוחות שכיחות

שאלה 1

מחקר שנערך בשנות ה-80 בדק את ההשפעה של צריכת אלכוהול בזמן ההיריון על הופעת מומים בלידה¹. כל אישה שהשתתפה במחקר ניהלה יומן בו דיווחה על מספר מנות האלכוהול ששתתה בכל יום, ולאחר הלידה נלקחו נתונים על הילוד.

בטבלה הבאה מובאים נתונים חלקיים מהמחקר. המשתנים הם:

- מספר מנות האלכוהול הממוצע ליום שצרכה כל אישה במהלך ההיריון (מחולק לקטגוריות)
- הופעת מום באיברי הרבייה של הילוד

Table 2

Presence or absence of congenital sex organ malformation categorized by alcohol consumption of the mother

Malformation	Alcohol consumption (average # drinks/day)				
	0	<1	1-2	3-5	≥6
Absent	17,066	14,464	788	126	37
Present	48	38	5	1	1
Total	17,114	14,502	793	127	38

- א. בדקו את ההשערה כי אין קשר בין צריכת האלכוהול והופעת מומים בלידה.
- חשבו את ערכו של סטטיסטי חי-בריבוע
 - מהו מספר דרגות החופש?
 - חשבו את ה-p-value בעזרת אקסל (הסבר בסוף התרגיל)
 - האם תדחו את השערת האפס? מהי מסקנתכם?
- ב. חשבו את הערך של מתאם קרמר (Cramer's v)
- ג. חשבו את ערכו של λ – מקדם הקשר בעזרת ניבוי
- ד. חשבו את יחס הסיכונים להופעת מום בלידה בין הנשים שצרכו פחות ממנת אלכוהול אחת ליום ובין אלה שצרכו מנה אחת או יותר ליום

שאלה 2

במגיפה עולמית של מחלה מדבקת במיוחד, הסיכון של אנשי צוות רפואי המטפלים בחולים להידבק במחלה הוא גבוה במיוחד. על פי נתונים ארציים, ההסתברות כי איש צוות ידבק ביום נתון היא $p = 0.3$. לכן למספר הימים שעוברים עד שאיש צוות רפואי נדבק במחלה יש התפלגות גיאומטרית. על פי התפלגות זו, ההסתברות כי מספר הימים k שיעברו, החל מנקודת זמן מסויימת, עד שאיש צוות יחלה שווה ל-

$$P(X = k) = p \cdot (1 - p)^k$$

כאשר הערכים ש- k יכול לקבל הם 0, 1, 2, 3 וכן הלאה. כאשר $k = 0$ המשמעות היא שאיש הצוות נדבק כבר ביום הראשון להצטרפותו לצוות.

¹ Graubard, Barry I., and Edward L. Korn. "Choice of column scores for testing independence in ordered 2 x K contingency tables." *Biometrics* (1987): 471-476.

בבית חולים מסויים החליטו לנקוט בצעדים שאמורים להקטין את הסיכון להדבקה. נאספו נתונים על משך הזמן עד להדבקה עבור 100 אנשי צוות העובדים במחלקות לטיפול בחולים במגיפה. הנתונים נמצאים בטבלה הבאה:

מספר ימים עד להדבקה	0	1	2	3	4+
מספר אנשי הצוות שנדבקו	22	15	14	14	35

הנהלת בית החולים מעוניינת לדעת האם לאחר שהצעדים האלה ננקטו הסיכון להדבקה ירד. אנו נסייע להנהלה בניתוח הסטטיסטי.

- א. בדקו את ההשערה כי התפלגות מספר הימים עד להדבקה משקפת את התפלגות הארצית, על פי השלבים הבאים:
- חשבו את ההסתברויות למספרי הימים עד להדבקה על פי התפלגות הגיאומטרית עם $p = 0.3$.
 - חשבו את מספרי הימים הצפויים עד להדבקה על פי ההסתברויות שחיבתם בסעיף הקודם
 - חשבו לכל תא בטבלה את $(O - E)^2 / E$
 - חשבו את ערכו של סטטיסטי χ^2
 - קבעו את מספר דרגות החופש של סטטיסטי χ^2
 - חשבו את ה-p-value בעזרת אקסל (הסבר בסוף התרגיל) וקבעו האם יש לדחות את השערת האפס ברמת מובהקות של $\alpha = 0.05$

ב. בסעיף א הייתם אמורים להסיק כי יש לדחות את השערת האפס. מה המשמעות של התוצאה?

שאלה 3

במהלך פיתוח תרופות עבור טרשת נפוצה, מקובל לבצע ניסוי קליני על פי התכנון הבא:

- בתחילת הניסוי מבצעים לכל חולה בדיקת MRI לספירת מספר הנגעים (lesions) במוחו של החולה.
- לאחר מספר חודשים של טיפול, נערכת לכל חולה בדיקת MRI נוספת, ונספרים כמה נגעי מוח יש לחולה לאחר תקופת הטיפול.
- בתרופה יעילה מצפים כי מספר הנגעים יקטן לאחר תקופת הטיפול

השיטה הסטטיסטית המקובלת לניתוח נתונים מסוג זה היא רגרסיה בינומית שלילית, אולם אנו נשתמש בכלים פשוטים יותר אשר למדנו בקורס זה.

בדרייב של הקורס, בתיקיית הנתונים, נמצא קובץ בשם ms.xlsx ובו נתונים פיקטיביים של ניסוי כזה שנערך בהשתתפות 80 חולים.

- בעזרת מבחן הסימן בדקו את ההשערה כי החציון של השינוי המספר הנגעים (ההפרש בין מספר הנגעים בתחילת המחקר ובין מספרם בסופו) קטן מ-0. ניתן להניח כי גודל המדגם מספיק גדול.
- טיפול נחשב למוצלח אם הירידה במספר הנגעים היא לפחות 2. חשבו רווח סמך לפרופורציה של מספר החולים עבורם הטיפול מוצלח.
- בדקו את ההשערה כי הפרופורציה של החולים עבורם הטיפול יצליח גדולה מ-0.5. השתמשו ברמת מובהקות של $\alpha = 0.1$
- הסבירו כיצד תחשבו את העצמה של המחקר עבור ההשערה שבדקתם אם הפרופורציה האמיתית של מספר החולים עבורם הטיפול מצליח היא 0.6. יש לערוך את החישוב עד לנקודה בה יש צורך לחשב את ההסתברות כי (משהו $P(Z > \dots)$)

ה. הסבירו כיצד תחשבו את ה-p-value עבור ההשערה שבדקתם בסעיף 5 על פי הנתונים שלנו. יש לערוך את החישוב עד לנקודה בה יש צורך לחשב את ההסתברות כי $P(Z > \text{משהו})$

חישוב ערך p-value של סטטיסטי חי-בריבוע בעזרת אקסל

החישוב נעשה בעזרת הפונקציה CHISQ.DIST. לפונקציה זו שלושה פרמטרים:

- x - ערכו של סטטיסטי חי בריבוע
- deg_freedom - מספר דרגות החופש
- cumulative – משתנה לוגי המגדיר את סוג החישוב. יש להשתמש בערך TRUE

הפונקציה מחזירה את ההסתברות כי ערכו של חי-בריבוע קטן מערך כלשהו. מאחר ואנחנו דוחים את השערת האפס אם ערכו של חי בריבוע גדול מאפס באופן מובהק, אנו רוצים את ההסתברות המשלימה.

דוגמא: חישבנו את ערכו של חי בריבוע ומצאנו כי הוא שווה ל-5.28 עם דרגת חופש אחת. נזין את הנתונים באופן הבא:



ונקבל כי התוצאה היא 0.02157

The image shows a portion of an Excel spreadsheet. The formula bar at the top contains `=1-CHISQ.DIST(5.28,1,TRUE)`. Below it, a table with three columns labeled A, B, and C is shown. Cell A1 contains the value 0.02157.

A	B	C
0.02157		

הערה: לאקסל יש גם פונקציה שמתיימרת לבצע מבחן חי-בריבוע לבדיקת אי-תלות בשם CHISQ.TEST, אך היא לא שימושית במיוחד מכיוון שעליכם לחשב את ערכי ה-expected בעצמכם.