


פתרון תרגיל בנושא מבחנים אי פרמטריים ולוחות שכיחות

שאלה מספר 1


מחקר שנערך בשנות ה-80 בדק את ההשפעה של צריכת אלכוהול בזמן ההיריון על הופעת מומים בלידה. 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 הרביה של הילוד

Table 2

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

	Alcohol consumption (average # drinks/day)				
Malformation	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

שאלה 1 , סעיף א

בדקו את ההשערה כי אין קשר בין צריכת האלכוהול והופעת מומים בלידה.

Alcohol consumption (average # drinks/day)

Malformation	0	<1	1-2	3-5	6+	Total
Absent	17065.1	14460.6				32481
Present						93
Total	17114	14502	793	127	38	32574

$$\frac{17114}{32574}\cdot\frac{32481}{32574}\cdot 32574=\frac{17114\cdot 32481}{32574}\boxed{=17065.1}$$

$$\frac{14502\cdot 32481}{32574}\boxed{=14460.6}$$

שאלה 1 , סעיף א

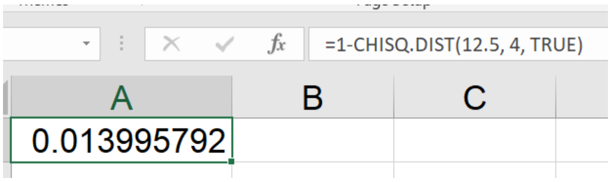
Observed	0	<1	1-2	3-5	6+
Absent	17066	14464	788	126	37
Present	48	38	5	1	1

Expected	0	<1	1-2	3-5	6+
Absent	17065.1	14460.6	790.7	126.6	37.9
Present	48.9	41.4	2.3	0.4	0.1

$(O-E)^2/E$	0	<1	1-2	3-5	6+
Absent	0.0000	0.0008	0.0092	0.0028	0.0214
Present	0.0166	0.2792	3.1696	0.9000	8.1000

$$\chi^2 = 12.50 \quad \mathrm{df} = 4$$

שאלה 1, סעיף א



p-value=0.0139 ולכן ברמת מובהקות של 5% נדחה את השערת האפס כי אין קשר בין צריכת האלכוהול בהריון ובין הופעת מומים באיברי הרביה של הילוד.

מסקנתנו היא כי קיים קשר בין שני המשתנים.

עם זאת, למרות שמתקבל על הדעת כי יש קשר סיבתי, אנו לא יכולים להסיק זאת מהניתוח שלנו.

שאלה 1, סעיף ב

חשבו את הערך של מתאם קרמר (Cramer's v)

$$v = \sqrt{\frac{\chi^2/n}{\min\,(R-1, C-1)}} = \sqrt{\frac{12.50/32574}{\min\,(2-1, 5-1)}} = \sqrt{\frac{12.50/32574}{1}} = 0.0195$$

למרות שדחינו את השערת האפס, קיבלנו ערך שלכאורה מעיד על חוסר קשר, וזאת בגלל שגודל המדגם גדול מאוד.
דוגמה זו מראה כי אין להסתמך רק על הערכים של מקדמי הקשר ולוותר על המבחן הסטטיסטי.

שאלה 1, סעיף ג

חשבו את יחס הסיכונים להופעת מום בלידה בין הנשים שצרכו פחות ממנת אלכוהול אחת ליום ובין אלה שצרכו מנה אחת או יותר ליום

Malformation	<1	1+
Absent	31530	951
Present	86	7
Total	31616	958

(average # drinks/day)

1-2	3-5	6+	Total
788	126	37	32481
5	1	1	93
793	127	38	32574

$$p_L = P(\text{malformation}|\text{low alcohol consumption}) = \frac{86}{31616} = 0.00272$$

$$p_L = P(\text{malformation}|\text{high alcohol consumption}) = \frac{7}{958} = 0.00731$$

$$RR = \frac{p_H}{p_L} = \frac{0.00731}{0.00272} = 2.686$$

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.

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.

בדקו את ההשערה כי התפלגות מספר הימים עד להדבקה משקפת אתת התפלגות הארצית

שאלה 2, סעיף א

חשבו את ההסתברויות למספרי הימים עד להדבקה על פי ההתפלגות הגיאומטרית עם $p = 0.3$.

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

$$P(X = 0) = p \cdot (1 - p)^0 = 0.3 \cdot 0.7^0 = 0.3$$

$$P(X = 1) = p \cdot (1 - p)^1 = 0.3 \cdot 0.7^1 = 0.21$$

$$P(X = 2) = 0.3 \cdot 0.7^2 = 0.147$$

$$P(X = 3) = 0.3 \cdot 0.7^3 = 0.1029$$

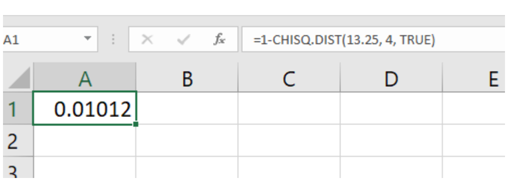
$$P(X \geq 4) = 1 - (0.3 + 0.21 + 0.147 + 0.1029) = 0.2401$$

שאלה 2 - המשך

K: מספר ימים עד

K	Observed	$P(X = k)$	Expected	$(O - E)^2/E$
0	22	0.3	30	2.133
1	15	0.21	21	1.714
2	14	0.147	14.7	0.033
3	14	0.1029	10.29	1.338
4	35	0.2401	24.01	5.030
סך הכל	100			13.25

$\chi^2 = 13.25$ df= 4



אנו דוחים את השערת האפס ולכן מסיקים כי התפלגות הזמן להדבקה בבית חולים זה אינה משקפת את ההתפלגות הארצית.

לא ניתן להסיק כי הצעדים שננקטו גרמו לירידה בסיכון להדבקה!

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.

שאלה 3 סעיף א

בעזרת מבחן הסימן בדקו את ההשערה כי החציון של השינוי המספר הנגעים (ההפרש בין מספר הנגעים בתחילת המחקר ובין מספרם בסופו) גדול מ-0. ניתן להניח כי גודל המדגם מספיק גדול.

SUM | =IF(C5<B5, 1, 0)

	A	B	C	D
1	subject_id	before	after	
2	1	1	0	1
3	2	4	1	1
4	3	1	1	0
5	4	3	1	=IF(C5<B5, 1, 0)
6	5	1	0	1
7	6	5	0	
8	7	1	1	

נחשב את הסימנים בעזרת אקסל:

במקום סימני פלוס ומינוס נשים אחד ואפס בהתאמה. כך יהיה לנו קל לספור כמה פלוסים יש

בסך הכל, יש 47 חולים עבורם הייתה ירידה במספר הנגעים מתוך 80

ההשערה שאנו רוצים לבדוק שקולה להשערות $H_0: p = 0.5 \quad H_1: p > 0.5$

שאלה 3 סעיף א

$H_0: p = 0.5 \quad H_1: p < 0.5$

נדחה את השערת האפס אם:

$$\hat{p} > p_0 + z_\alpha \cdot \sqrt{\frac{p_0 \cdot (1-p_0)}{n}} = 0.5 + 1.645 \cdot \sqrt{\frac{0.5 \cdot 0.5}{80}} = 0.592$$

אנו לא דוחים את השערת האפס כי

$$\hat{p} = \frac{47}{80} = 0.588$$

שאלה 3, סעיף ב

טיפול נחשב למוצלח אם הירידה במספר הנגעים היא לפחות 2. חשבו רווח סמך לפרופורציה של מספר החולים עבורם הטיפול מוצלח.

בעזרת האקסל נמצא כי הטיפול היה מוצלח עבור 41 חולים מתוך 80, ולכן הפרופורציה במדגם היא

$$\hat{p} = \frac{41}{80} = 0.5125$$

רווח הסמך הוא

$$\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}} = 0.5125 \pm 1.96 \cdot \sqrt{\frac{0.5125 \cdot 0.4875}{80}} = (0.403, 0.622)$$

שאלה 3, סעיף ג

בדקו את ההשערה כי הפרופורציה של החולים עבורם הטיפול יצליח גדולה מ-0.3. השתמשו ברמת מובהקות של $\alpha = 0.1$

נדחה את השערת האפס כאשר

$$\hat{p} > p_0 + z_\alpha \cdot \sqrt{\frac{p_0 \cdot (1 - p_0)}{n}} = 0.3 + 1.282 \cdot \sqrt{\frac{0.3 \cdot 0.7}{80}} = 0.366$$

הפרופורציה במדגם הייתה $\hat{p} = \frac{41}{80} = 0.5125$ ולכן נדחה את השערת האפס