

הסקה סטטיסטית 30204

פרק 22 - מבחנים אפרמטריים - מבחני חי בריבוע (יחידה -12 ממן 14)

תוכן העניינים

1. מבחן טיב התאמה..... (ללא ספר)
2. מבחן טיב התאמה והקשר שלו לבדיקת השערות על פרופורציה אחת..... 1
3. מבחן לאי תלות..... (ללא ספר)
4. קשר בין מבחן אי תלות לבדיקת השערות להפרש פרופורציות..... 3
5. מקדם המתאם של קרמר..... 5

הקשר בין מבחן טיב התאמה לבדיקת השערות על הפרופורציה – רקע

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

דוגמה (פתרון בהקלטה) :

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

- א. באמצעות מבחן טיב התאמה.
- ב. באמצעות מבחן Z לפרופורציה בודדת.

שאלות

(1) בסקר שנעשה על 320 נשאלים, 43.75% טענו שהחיה המועדפת עליהם היא כלב. עד היום היה נהוג לחשוב ש-40% מהאנשים מעדיפים כלבים. בדקו ברמת מובהקות של 5% האם הסקר יישנה את הסברה שהייתה נהוגה עד היום לגבי העדפת כלב.

- א. באמצעות מבחן טיב התאמה.
ב. באמצעות מבחן על פרופורציה.

(2) לסוכנות מכוניות שלושה סניפים ברחבי הארץ. המכוניות נמכרות בסניפים השונים. מתוך 100 מכוניות נמצא ש-65 נמכרו בסניף תל-אביב, 23 בסניף ירושלים והיתר בסניף חיפה.

- א. בדקו ברמת מובהקות של 5% האם שיעור המכוניות שנמכרות בסניף ת"א גדול פי 2 מכל סניף אחר.
ב. בדקו באמצעות מבחן טיב התאמה האם 60% מהמכוניות נהוגות להימכר בסניף תל אביב. האם יש דרך אחרת לבדוק את ההשערה?

(3) בתחרות ריצה בית ספרית שלושה מסלולי ריצה.

ב-50 תחרויות בדקו באיזה מסלול היה הניצחון. התוצאות שהתקבלו מסוכמות בטבלה הבאה:

המסלול	1	2	3
מספר הניצחונות	20	15	15

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

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

$$\text{גבוה מ-} \frac{1}{3}.$$

תשובות סופיות

(1) לא נדחה H_0 .

(2) א. נדחה H_0 . ב. לא נדחה H_0 .

(3) א. לא נדחה H_0 . ב. לא נדחה H_0 .

הקשר בין מבחן לאי תלות ובדיקת השערות להפרש פרופורציות – רקע

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

דוגמה (פתרון בהקלטה) :

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

שאלות

- 1) בקרב מדגם של 200 נשים 120 טענו שהן תצבענה למועמד R לראשות העיר. בקרב מדגם של 200 גברים 80 טענו שהם יצביעו למועמד R האם קיים הבדל בין דפוס ההצבעה של הנשים ושל הגברים? בדוק ברמת מובהקות של 5%.
 - א. על ידי מבחן לאי תלות.
 - ב. על ידי בדיקת השערות לשתי פרופורציות.

- 2) נלקחו 200 אנשים שמתוכם 60 הצהירו שהם עוסקים בפעילות גופנית סדירה. מתוך אלו שעוסקים בפעילות גופנית סדירה 50 נמצאו במצב בריאותי תקין. מתוך אלו שלא עוסקים בפעילות גופנית סדירה 90 נמצאו במצב בריאותי תקין.
 - א. בנו טבלת שכיחות משותפת לנתונים שהוצגו בשאלה.
 - ב. האם ניתן להגיד שהסיכוי להימצא במצב בריאותי תקין גבוה יותר כאשר עוסקים בפעילות גופנית סדירה לעומת המצב שלא עוסקים בפעילות גופנית סדירה? בדוק ברמת בטחון של 90%.

תשובות סופיות

- (1) נדחה את השערת האפס לפירוט הדרך ראה וידאו.
 (2) א. להלן טבלה: ב. נדחה את השערת האפס.

	לא תקין	תקין	פעילות/מצב בריאותי
60	10	50	סדירה
140	50	90	לא סדירה
200	60	140	

מדדי קשר-מדד הקשר של קרמר – רקע

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

דוגמה (פתרון בהקלטה) :

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

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

בהקשר של קרמר הטבלה נקראת טבלת O (observed).

X - מין (גבר/אישה) – סולם שמי.

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

שלב ב' בחישוב r_c :

שלב א' :

נבנה את טבלת E (Expected)

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

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

שלב ב' :

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

שלב ג' :

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

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

שאלות

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

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

להלן התוצאות:

האם קיים קשר בין מין להשכלה? נמקו!

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

א. בנו טבלת שכיחות משותפת לנתונים שהוצגו בשאלה.

ב. האם קיים קשר בין פעילות גופנית למצב בריאותי?

חשבו לפי מדד הקשר של קרמר.

תשובות סופיות

- (1) קיים קשר בעוצמה בינונית בין המין להשכלה מקדם המתאם של קרמר הוא 0.595.
- (2) א. להלן טבלה: ב. מדד קרמר 0.19 מעיד על קשר בעוצמה נמוכה.

$f(x)$	לא תקין	תקין	y/x
60	10	50	כן
140	50	90	לא
200	60	140	$f(y)$