# הסקה וניתוח נתונים

## פרק 14 - מבחני חי בריבוע

### תוכן העניינים

1. מבחן לאי תלות. ..... (ללא ספר)
2. מקדם המתאם של קרמר ..... 1
3. ניתוח פלטים במבחן אי תלות ..... 3

# 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.

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

בהקשר של קרמר הטבלה נקראת טבלת O (observed)

$X$ - מין (גבר/אישה) – סולם שמי.

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

**שלבים בחישוב $r_c$ :**

**שלב א':**

נבנה את טבלת E (Expected)

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

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

**שלב ב':**

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

**שלב ג':**

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

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


© כל הזכויות שמורות לברק קנדל

## שאלות

**1)** להלן תוצאות מחקר שבדק את הקשר בין מין להשכלה. לגבי כל נחקר נבדק המין שלו והשכלתו.
להלן התוצאות:

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

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.

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


© כל הזכויות שמורות לברק קנדל

## פלטים על מבחן לאי תלות – 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: $H_1$.

דרגות חופש: $d.f=(r-1)(c-1)$.

כאשר: $r$ - מספר הקטגוריות של המשתנה שבשורות.

$c$ - ספר הקטגוריות של המשתנה שבעמודות.

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

$O_i$ - השכיחות נצפית במדגם בתא $i$.

$E_i$ - שכיחות צפויה במדגם בתא $i$ בהנחת השערת האפס.

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

**הערה:**

תנאי כדי לבצע את המבחן הוא $E_i\geq 5$ לכל צ $i$. 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.


© כל הזכויות שמורות לברק קנדל

**דוגמה (פתרון בהקלטה) :**

במחקר רצו לבדוק את הקשר בין צבע שיער לבין צבע עיניים של אנשים.
הפלטים שהתקבלו מצורפים.

א. מהו המבחן הסטטיסטי המתאים?

ב. כמה קטגוריות יש לכל משתנה?

ג. רשמו את השערות המחקר.

ד. מה מספר דרגות החופש?

ה. כמה אנשים במדגם נמצאו עם שיער חום?

ו. כמה אנשים היית מצפה במדגם שיהיה להם שיער חום ועיניים ירוקות בהנחה ואין קשר בין צבע שיער לצבע עיניים?

ז. מתוך הבלונדינים מה אחוז בעלי עיניים כחולות במדגם?

ח. מתוך בעלי עיניים ירוקות מה אחוז הבלונדינים במדגם?

ט. מה ערכו של סטטיסטי המבחן ומהי מובהקות התוצאה?

י. מה מסקנת המחקר? $\alpha = 5\%$


© כל הזכויות שמורות לברק קנדל

להלן הפלטים שהתקבלו:

**Case Processing Summary**

| | Cases | | | | | |
|---|---|---|---|---|---|---|
| | Valid | | Missing | | Total | |
| | N | Percent | N | Percent | N | Percent |
| hair_color * eye_color | 78 | 100.0% | 0 | 0.0% | 78 | 100.0% |

**hair_color * eye_color Crosstabulation**

| | | | eye_color | | | Total |
|---|---|---|---|---|---|---|
| | | | brown | green | Blue | |
| hair_color | black | Count | 13 | 7 | 7 | 27 |
| | | Expected Count | 10.7 | 8.3 | 8.0 | 27.0 |
| | | % within hair_color | 48.1% | 25.9% | 25.9% | 100.0% |
| | | % within eye_color | 41.9% | 29.2% | 30.4% | 34.6% |
| | brown | Count | 12 | 12 | 6 | 30 |
| | | Expected Count | 11.9 | 9.2 | 8.8 | 30.0 |
| | | % within hair_color | 40.0% | 40.0% | 20.0% | 100.0% |
| | | % within eye_color | 38.7% | 50.0% | 26.1% | 38.5% |
| | blond | Count | 6 | 5 | 10 | 21 |
| | | Expected Count | 8.3 | 6.5 | 6.2 | 21.0 |
| | | % within hair_color | 28.6% | 23.8% | 47.6% | 100.0% |
| | | % within eye_color | 19.4% | 20.8% | 43.5% | 26.9% |
| | Total | Count | 31 | 24 | 23 | 78 |
| | | Expected Count | 31.0 | 24.0 | 23.0 | 78.0 |
| | | % within hair_color | 39.7% | 30.8% | 29.5% | 100.0% |
| | | % within eye_color | 100.0% | 100.0% | 100.0% | 100.0% |

**Chi-Square Tests**

| | Value | df | Asymp. Sig. (2-sided) |
|---|---|---|---|
| Pearson Chi-Square | 5.880[a] | 4 | .208 |
| Likelihood Ratio | 5.641 | 4 | .228 |
| Linear-by-Linear Association | 2.682 | 1 | .101 |
| N of Valid Cases | 78 | | |

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 6.19.


© כל הזכויות שמורות לברק קנדל

## שאלות

**1)** בסקר שנעשה על ידי משרד ראש הממשלה נדגמו 60 אזרחים.
כל אזרח נשאל על מגדרו והאם הוא בעד הקמת מדינה פלסטינית.

א. מה ההשערות הנבדקות ומהו סטטיסטי המבחן?

ב. אם סטטיסטי המבחן היה גדל כיצד הדבר היה משפיע על SIG שבפלט.

ג. האם קיים קשר בין מגדר ודעה ברמת מובהקות של 5%?

ד. מהו האומדן לאחוז התומכים במדינה פלסטינית מתוך הגברים?

ה. איזה אחוז מהנשאלים שהיו בעד מדינה פלשתינית הם גברים?


© כל הזכויות שמורות לברק קנדל

להלן הפלטים:

## Chi-Square Tests

| | Value | df | Asymp. Sig. (2-sided) |
|---|---|---|---|
| Pearson Chi-Square | 1.973[a] | 2 | .373 |
| Likelihood Ratio | 1.987 | 2 | .370 |
| Linear-by-Linear Association | 1.882 | 1 | .170 |
| N of Valid Cases | 60 | | |

a. 0 cells (.0%) have expected count less than 5.

b. The minimum expected count is 7.25.

## gender * opinion Crosstabulation

| | | | opinion | | | Total |
|---|---|---|---|---|---|---|
| | | | yes | now | no opinion | |
| gender | Male | Count | 10 | 10 | 9 | 29 |
| | | Expected Count | 12.6 | 9.2 | 7.3 | 29.0 |
| | | % within gender | 34.5% | 34.5% | 31.0% | 100.0% |
| | | % within opinion | 38.5% | 52.6% | 60.0% | 48.3% |
| | | % of Total | 16.7% | 16.7% | 15.0% | 48.3% |
| | female | Count | 16 | 9 | 6 | 31 |
| | | Expected Count | 13.4 | 9.8 | 7.8 | 31.0 |
| | | % within gender | 51.6% | 29.0% | 19.4% | 100.0% |
| | | % within opinion | 61.5% | 47.4% | 40.0% | 51.7% |
| | | % of Total | 26.7% | 15.0% | 10.0% | 51.7% |
| Total | | Count | 26 | 19 | 15 | 60 |
| | | Expected Count | 26.0 | 19.0 | 15.0 | 60.0 |
| | | % within gender | 43.3% | 31.7% | 25.0% | 100.0% |
| | | % within opinion | 100.0% | 100.0% | 100.0% | 100.0% |
| | | % of Total | 43.3% | 31.7% | 25.0% | 100.0% |


© כל הזכויות שמורות לברק קנדל

**2)** להלן פלט על סמך סקר שנעשה בקרב סטודנטים, בסקר נשאלו הסטודנטים על המוזיקה אותה הם מעדיפים וצורת הבילוי המועדפת עליהם.

**Crosstab**

Count

| | | בילוי | | | Total |
|---|---|---|---|---|---|
| | | קריאה | ספורט | מועדון | |
| מוזיקה | רוק | 0 | 0 | 11 | 11 |
| | פופ | 1 | 6 | 8 | 15 |
| | קלאסי | 5 | 6 | 9 | 20 |
| Total | | 6 | 12 | 28 | 46 |

**Chi-Square Tests**

| | Value | df | Asymp. Sig. (2-sided) |
|---|---|---|---|
| Pearson Chi-Square | 11.929[a] | ? | .018 |
| N of Valid Cases | 46 | | |

a. 5 cells (55.6%) have expected count less than 5.

b. The minimum expected count is 1.43.

א. בין אלו משתנים נבדק הקשר? כמה קטגוריות לכל משתנה?

ב. האם התנאים של המודל מתקיימים?

ג. מה מספר דרגות החופש במבחן הנ"ל?

ד. מה ההשערות של המבחן?


© כל הזכויות שמורות לברק קנדל

**3)** מחקר התעניין לבדוק את הקשר בין רמת הכנסה של משפחה לבין צריכת עגבניות אורגניות. הפלטים מצורפים.

א. השלימו את שלושת המספרים החסרים בטבלה (היכן שיש סימני שאלה).

ב. מה ערכו של חי בריבוע הסטטיסטי.

ג. תנו הערכה למובהקות התוצאה לבדיקת הקשר בין רמת הכנסה של משפחה לבין צריכת עגבניות אורגניות.

**רמת_הכנסה * צרכן עגבניות** Crosstabulation

| | | | צרכן עגבניות | | Total |
|---|---|---|---|---|---|
| | | | אורגני | לא אורגני | |
| רמת_הכנסה | הרבה מתחת לממוצע | Count | 17 | 42 | 59 |
| | | % within רמת_הכנסה | 28.8% | ? | 100.0% |
| | | % within צרכן עגבניות | 13.6% | 33.6% | 23.6% |
| | מתחת לממוצע | Count | 27 | 22 | 49 |
| | | % within רמת_הכנסה | 55.1% | 44.9% | 100.0% |
| | | % within צרכן עגבניות | ? | 17.6% | 19.6% |
| | ממוצע | Count | 31 | 29 | 60 |
| | | % within רמת_הכנסה | 51.7% | 48.3% | 100.0% |
| | | % within צרכן עגבניות | 24.8% | 23.2% | 24.0% |
| | מעל הממוצע | Count | 44 | 26 | 70 |
| | | % within רמת_הכנסה | 62.9% | 37.1% | 100.0% |
| | | % within צרכן עגבניות | 35.2% | 20.8% | 28.0% |
| | הרבה מעל הממוצע | Count | ? | 6 | 12 |
| | | % within רמת_הכנסה | 50.0% | 50.0% | 100.0% |
| | | % within צרכן עגבניות | 4.8% | 4.8% | 4.8% |
| Total | | Count | 125 | 125 | 250 |


© כל הזכויות שמורות לברק קנדל

**4)** חוקר בדק את הקשר בין צבע השיער לבין צבע העיניים בעזרת מבחן חי בריבוע בקרב 52 נבדקים. תוצאות המבחן מוצגות בטבלה. בנוסף ידוע כי סטטיסטי המבחן שהתקבל מעיבוד הנתונים הוא 8.08.

א. מה תהיה מסקנת המחקר ברמת מובהקות של 1%?

ב. מה ערכו של E עבור עיניים כחולות וצבע שיער כהה.

ג. מה יהיה בקירוב ערכו של מקדם המתאם של קרמר?

ד. מהי פרופורציית בעלי צבע השיער הבהיר מקרב בעלי העיניים הירוקות?

להלן הפלט:

**צבע עיניים** * **צבע שיער** Crosstabulation

| | | | צבע שיער | | Total |
|---|---|---|---|---|---|
| | | | כהה | בהיר | |
| צבע עיניים | כחול | Count | | | |
| | | % within צבע עיניים | 50.0% | 50.0% | 100.0% |
| | | % within צבע שיער | 21.6% | 53.3% | 30.8% |
| | | % of Total | 15.4% | 15.4% | 30.8% |
| | חום | Count | | | |
| | | % within צבע עיניים | 83.3% | 16.7% | 100.0% |
| | | % within צבע שיער | 27.0% | 13.3% | 23.1% |
| | | % of Total | 19.2% | 3.8% | 23.1% |
| | ירוק | Count | | | |
| | | % within צבע עיניים | 79.2% | 20.8% | 100.0% |
| | | % within צבע שיער | 51.4% | 33.3% | 46.2% |
| | | % of Total | 36.5% | 9.6% | 46.2% |
| Total | | Count | | | |
| | | % within צבע עיניים | 71.2% | 28.8% | 100.0% |
| | | % within צבע שיער | 100.0% | 100.0% | 100.0% |
| | | % of Total | 71.2% | 28.8% | 100.0% |


© כל הזכויות שמורות לברק קנדל

**5)** במחקר מסוים רצו לבדוק האם יש קשר בין המגדר להוצאה על לבוש במשך שנה. דגמו באופן מקרי גברים ונשים ובדקו את רמת ההוצאה שלהם על לבוש בשנה האחרונה. חוקר א' בדק האם קיים הבדל בתוחלות ההוצאה בין גברים לנשים. חוקר ב' קיבץ את ההוצאה לקטגוריות ובאופן הזה בדק האם קיים הבדל בהתפלגות ההוצאה בין גברים לנשים.
הקטגוריות חולקו לשלוש קבוצות הוצאה.

א. איזה פלט מתאים לאיזה אחד מהחוקרים? נמקו.

ב. מה מסקנתו של חוקר א'? בדקו ברמת מובהקות $\alpha = 0.05$. (רשמו השערות, נסחו הנחות, ציינו כלל החלטה ותנו מסקנה במונחי המשתנים).

ג. איזו טעות יכולה להיות במסקנתו של חוקר א'?
נסחו את הטעות במונחי השאלה.

ד. מהי מסקנתו של חוקר ב'? בדקו ברמת מובהקות $\alpha = 0.05$. (רשמו השערות, נסחו הנחות, ציינו כלל החלטה ורשמו מסקנה במונחי המשתנים).

ה. איזו טעות יכולה להיות במסקנתו של חוקר ב'? נסחו זאת במונחי השאלה.

ו. כיצד ניתן ליישב את מסקנות שני החוקרים?

להלן פלט ראשון:

## T-Test

### Group Statistics

| | gender | N | Mean | Std. Deviation | Std. Error Mean |
|---|---|---|---|---|---|
| dimension1 | female | 40 | 2.9000 | 1.15025 | .18187 |
| | male | 40 | 2.6000 | 2.52982 | .40000 |

**Independent Samples Test**

| | | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | 95% Confidence Interval of the Difference | |
| | | F | Sig. | t | df | Sig. (2-tailed) | Mean Difference | Std. Error Difference | Lower | Upper |
| expese | Equal variances assumed | 16.805 | .000 | .683 | 78 | .497 | .30000 | .43941 | -.57479 | 1.17479 |
| | Equal variances not assumed | | | .683 | 54.464 | .498 | .30000 | .43941 | -.58078 | 1.18078 |


© כל הזכויות שמורות לברק קנדל

להלן פלט שני:

## Crosstabs

### Case Processing Summary

| | Cases | | | | | |
|---|---|---|---|---|---|---|
| | Valid | | Missing | | Total | |
| | N | Percent | N | Percent | N | Percent |
| gender * category | 80 | 100.0% | 0 | .0% | 80 | 100.0% |

### gender * category Crosstabulation

| | | | category | | | Total |
|---|---|---|---|---|---|---|
| | | | a | b | c | |
| gender | Female | Count | 2 | 30 | 8 | 40 |
| | | Expected Count | 11.0 | 21.0 | 8.0 | 40.0 |
| | | % within gender | 5.0% | 75.0% | 20.0% | 100.0% |
| | | % within category | 9.1% | 71.4% | 50.0% | 50.0% |
| | Male | Count | 20 | 12 | 8 | 40 |
| | | Expected Count | 11.0 | 21.0 | 8.0 | 40.0 |
| | | % within gender | 50.0% | 30.0% | 20.0% | 100.0% |
| | | % within category | 90.9% | 28.6% | 50.0% | 50.0% |
| Total | | Count | 22 | 42 | 16 | 80 |
| | | Expected Count | 22.0 | 42.0 | 16.0 | 80.0 |
| | | % within gender | 27.5% | 52.5% | 20.0% | 100.0% |
| | | % within category | 100.0% | 100.0% | 100.0% | 100.0% |

### Chi-Square Tests

| | Value | df | Asymp. Sig. (2-sided) |
|---|---|---|---|
| Pearson Chi-Square | 22.442[a] | 2 | .000 |
| Likelihood Ratio | 25.064 | 2 | .000 |
| N of Valid Cases | 80 | | |

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 8.00.


© כל הזכויות שמורות לברק קנדל

## תשובות סופיות

**1)** א. 1.973 ב. קטן . ג. לא נדחה $H_0$ .
ד. 34.5% ה. 38.5%.

**2)** א. בילוי מועדף ומוזיקה מועדפת עם 3 קטגוריות לכל משתנה.
ב. לא ג. 4.
ד. $H_0$ אין תלות בין בילוי מועדף למוזיקה מועדפת.
$H_1$ יש תלות בין בילוי מועדף למוזיקה מועדפת.

**3)** א. 71.2%, 21.6%, 6 ב. 15.8 ג. קטן מ-0.005.

**4)** א. לא נדחה $H_0$ . ב. 11.4 ג. 0.394
ד. 20.8%.

**5)** א. פלא א'- חוקר א', פלא ב'- חוקר ב'. ב. נקבל את $H_0$ .
ג. טעות מסוג שני- הכרענו שאין הבדל בין גברים לנשים למרות שיש במציאות הבדל.
ד. נקבל את $H_1$ .
ה. טעות מסוג ראשון- הכרענו שיש קשר בין מין להוצאה למרות שבמציאות אין קשר.
ו. כל חוקר פעל בשיטה סטטיסטית שונה ובמצב כזה יתכן מסקנות סותרות.


© כל הזכויות שמורות לברק קנדל