

שתי שאלות בחשיבה הסתברותית

שאלה 1

חוקר בדק את הקשר בין הרגלי הצפיה (במונחים של צופה / לא צופה) בתכנית 'מי רוצה להיות מליונר' לבין גיל הצופה (מבוגר / צעיר).

התקבלו התוצאות הבאות:

$\frac{2}{3}$ צופים בתכנית.

$\frac{3}{4}$ מהצופים היו מבוגרים.

$\frac{4}{5}$ מבין המבוגרים צופים בתכנית.

א. מהו אחוז המבוגרים שהשתתפו במחקר?

ב. מבין הצעירים, מהו האחוז של אלו שאינם צופים בתוכנית?

ג. האם קיים קשר סטטיסטי בין צפיה בתכנית לבין גיל הצופה? מהי משמעות הקשר?

פתרון:

נסמן:

A - קבוצת הצופים בסדרה, \bar{A} - הקבוצה שאינה צופה בסדרה

B - קבוצת הצעירים, \bar{B} - קבוצת המבוגרים

נתון: $P(A) = \frac{2}{3}$, $P(\bar{B}/A) = \frac{3}{4}$, $P(A/\bar{B}) = \frac{4}{5}$

$$P(\bar{B}/A) = \frac{P(\bar{B} \cap A)}{P(A)} = \frac{P(\bar{B} \cap A)}{\frac{2}{3}} = \frac{3}{4} \Rightarrow P(\bar{B} \cap A) = \frac{2}{3} \cdot \frac{3}{4} = \frac{1}{2}$$

נסמן: $x = P(\bar{A} \cap \bar{B})$

ונרשום את הנתונים בטבלה דרמימדית:

	A	\bar{A}	Σ
B	$\frac{2}{3} - \frac{1}{2} = \frac{1}{6}$	$\frac{1}{3} - x$	$1 - (x + \frac{1}{2}) = \frac{1}{2} - x$
\bar{B}	$\frac{1}{2}$	x	$x + \frac{1}{2}$
Σ	$\frac{2}{3}$	$1 - \frac{2}{3} = \frac{1}{3}$	1

$$P(A/\bar{B}) = \frac{P(A \cap \bar{B})}{P(\bar{B})} = \frac{\frac{1}{2}}{x + \frac{1}{2}} = \frac{4}{5} \Rightarrow \frac{1}{2} = \frac{4}{5}(x + \frac{1}{2}) = \frac{4x}{5} + \frac{2}{5} \Rightarrow \frac{4x}{5} = \frac{1}{2} - \frac{2}{5} = \frac{1}{10} \quad / \cdot 10$$

$$8x = 1 \Rightarrow x = \frac{1}{8}$$

	A	\bar{A}	Σ
B	$\frac{1}{6}$	$\frac{1}{3} - \frac{1}{8} = \frac{5}{24}$	$\frac{1}{2} - \frac{1}{8} = \frac{3}{8}$
\bar{B}	$\frac{1}{2}$	$\frac{1}{8}$	$\frac{1}{8} + \frac{1}{2} = \frac{5}{8}$
Σ	$\frac{2}{3}$	$\frac{1}{3}$	1

א. ישירות מהטבלה: $P(\bar{B}) = \frac{5}{8} = 0.625 \equiv 62.5\%$

ב.

$$P(\bar{A}/B) = \frac{P(\bar{A} \cap B)}{P(B)} = \frac{\frac{5}{24}}{\frac{3}{8}} = \frac{5}{24} \cdot \frac{8}{3} = \frac{5 \times 8}{24 \times 3} = \frac{40}{72} \Rightarrow P(\bar{A}/B) = \frac{5}{9} \equiv 55.56\%$$

ג.

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{6}}{\frac{3}{8}} = \frac{1}{6} \cdot \frac{8}{3} = \frac{1 \times 8}{6 \times 3} = \frac{8}{24} = \frac{1}{3} = 0.3333$$

$$P(A/\bar{B}) = \frac{4}{5} = 0.8 \Rightarrow P(A/\bar{B}) > P(A/B)$$

מסקנה: יש קשר סטטיסטי בין צפיה בתכנית לבין גיל הצופה.

משמעות הקשר: שיעור הצופים בתכנית מבין המבוגרים,

גדול יותר משיעור הצופים בתכנית מבין הצעירים: 80% לעומת 33.33%.

שאלה 2

מחקר נערך על הקשר בין שתיית קפה למחלת לב, על מדגם בן 1000 גברים מעל גיל 50.

התקבלו הנתונים הבאים על קבוצת הנבדקים:

600 הינם שותי קפה, 420 הינם חולי לב ו-80 הינם חולי לב שאינם שותי קפה.

א. בנה טבלה דו־מימדית לתיאור הנתונים.

ב. האם קיים קשר בין מחלת לב לשתיית קפה? מהי משמעות הקשר?

יותר מאוחר הועלתה טענה שישנו גורם מתווך והוא העישון.

אלו התוצאות שהתקבלו לגבי חולי לב ומעשנים באותו מדגם:

	חולי לב	אינם חולי לב	
מעשנים	450	100	
לא מעשנים	50	400	

ואלה התוצאות שהתקבלו לגבי מעשנים ושותי קפה באותו מדגם:

	מעשנים	לא מעשנים
שותים קפה	500	100
לא שותים קפה	50	350

ג. האם קיים קשר סטטיסטי בין מחלת לב לעישון? מהי משמעות הקשר?

ד. האם קיים קשר סטטיסטי בין עישון לשתיית קפה? מהי משמעות הקשר?

כאשר ניטרלו את גורם העישון התקבלו נתונים נוספים.

לגבי 550 המעשנים:

	חולי לב	אינם חולי לב
שותים קפה	300	200
אינם שותים קפה	40	10

לגבי 450 שאינם מעשנים:

	חולי לב	אינם חולי לב
שותים קפה	20	80
אינם שותים קפה	140	210

ה. האם קיים קשר סטטיסטי בין מחלת לב לשתיית קפה לגבי המעשנים? מהי משמעות הקשר?

ו. האם קיים קשר סטטיסטי בין מחלת לב לשתיית קפה לגבי הלא מעשנים? מהי משמעות הקשר?

ז. מה ניתן לומר על הקשר שבין מחלת לב לשתיית קפה,

על פי התוצאות של שתי הטבלאות האחרונות? נמק.

פתרון:

א. נסמן: A - קבוצת שותי הקפה, \bar{A} - קבוצת המתנזרים מקפה

B - קבוצת חולי הלב, \bar{B} - הקבוצה של אלו שאינם חולי לב

	A	\bar{A}	Σ
B	420	80	$420 + 80 = 500$
\bar{B}	$600 - 420 = 180$	$400 - 80 = 320$	$1000 - 500 = 500$
Σ	600	$1000 - 600 = 400$	1000

נרשום את טבלת השכיחויות במתאימה:

	A	\bar{A}	Σ
B	0.42	0.08	0.5
\bar{B}	0.18	0.32	0.5
Σ	0.6	0.4	1

ב.

$$P(B/A) = \frac{0.42}{0.6} = 0.7, \quad P(B/\bar{A}) = \frac{0.08}{0.4} = 0.2 \Rightarrow P(B/A) > P(B/\bar{A})$$

מסקנה: יש קשר סטטיסטי בין מחלת לב לבין שתיית קפה.

משמעות הקשר: שיעור חולי הלב באוכלוסית המעשנים גבוה יותר

מאשר שיעורם של חולי הלב בין אלו שאינם מעשנים: 70% לעומת 20%.

ג. נסמן: C - קבוצת המעשנים, \bar{C} - קבוצת המתנזרים מעישון

	B	\bar{B}	Σ
C	0.45	0.1	0.55
\bar{C}	0.05	0.4	0.45
Σ	0.5	0.5	1

לגבי חולי לב ומעשנים:

$$P(B/C) = \frac{0.45}{0.55} = \frac{9}{11} = 0.8182, \quad P(B/\bar{C}) = \frac{0.05}{0.45} = \frac{1}{9} = 0.1111 \Rightarrow P(B/C) > P(B/\bar{C})$$

מסקנה: יש קשר סטטיסטי בין עישון לבין מחלת לב.

משמעות הקשר: שיעור חולי הלב בקרב המעשנים גדול יותר מאשר

שיעור חולי הלב בקרב אלו שאינם מעשנים: 81.82% לעומת 11.11%.

ד.

	C	\bar{C}	Σ
A	0.5	0.1	0.6
\bar{A}	0.05	0.35	0.4
Σ	0.55	0.45	1

לגבי שותי קפה ומעשנים:

$$P(A/C) = \frac{0.5}{0.55} = 0.9191, \quad P(A/\bar{C}) = \frac{P(A \cap \bar{C})}{P(\bar{C})} = \frac{0.1}{0.45} = 0.2222 \Rightarrow P(A/C) > P(A/\bar{C})$$

מסקנה: יש קשר סטטיסטי בין שתיית קפה לבין עישון.

משמעות הקשר: שיעור שותי הקפה מבין המעשנים גדול יותר

מאשר שיעור שותי הקפה מבין אלו שאינם מעשנים: 91.91% לעומת 22.22%.

ה. לגבי מעשנים:

$$P(B/A) = \frac{P(B \cap A)}{P(A)} = \frac{\frac{300}{550}}{\frac{500}{550}} = \frac{3}{5} = 0.6, \quad P(B/\bar{A}) = \frac{\frac{40}{450}}{\frac{50}{450}} = \frac{4}{5} = 0.8 \Rightarrow P(B/A) < P(B/\bar{A})$$

מסקנה: יש קשר סטטיסטי בין שתיית קפה לבין מחלת לב בין המעשנים.

משמעות הקשר: שיעור חולי הלב בין שותי הקפה קטן יותר מאשר

שיעור חולי הלב מבין אלו המתנזרים משתיית קפה: 60% לעומת 80%.

ו. לגבי המתנזרים מעישון:

$$P(B/A) = \frac{P(B \cap A)}{P(A)} = \frac{\frac{20}{450}}{\frac{100}{450}} = \frac{20}{100} = 0.2, \quad P(B/\bar{A}) = \frac{\frac{140}{450}}{\frac{350}{450}} = \frac{2}{5} = 0.4 \Rightarrow P(B/A) < P(B/\bar{A})$$

מסקנה: יש קשר סטטיסטי בין שתיית קפה לבין מחלת לב בין המתנזרים מעישון.

משמעות הקשר: שיעור חולי הלב בין שותי הקפה קטן יותר מאשר

שיעור חולי הלב מבין אלו המתנזרים משתיית קפה: 20% לעומת 40%.

ז. העישון אכן הינו גורם מתווך והוא שגרם להיפוך הקשר' (אפקט סימפסון).

במקרה זה יש קשר סיבתי בין שתיית קפה למחלת לב.