מחסן מקריאת דאוטה - תרגיל 5

[redacted] [redacted]

השותפים שלי הם: [redacted]

שאלה 1

(a) precent agreement לפי הגדרה יחושב כך:

$$\frac{\sum a}{N} = \frac{\sum \text{agreements}}{\#\text{tags}} = \frac{100+130+160}{250+200+450} = \frac{490}{900} \approx 0.54$$

(b)

(1) $\sum \text{agreements} = 490 \quad (\sum a), \qquad N = 900$

(2) $\left\{ \begin{array}{l} \text{expected liberal agreement: } \frac{250 \times 300}{900} \approx 83.3 \\ \text{expected cons. agreement: } \frac{200 \times 200}{900} \approx 44.4 \\ \text{expected neutral agreement: } \frac{450 \times 400}{900} \approx 200 \end{array} \right.$

$\sum e \approx 83.3 + 44.4 + 200 = 327.7$

$$\Rightarrow \text{Kappa} = \frac{\sum a - \sum e}{N - \sum e} \approx \frac{490 - 327.7}{900 - 327.7} = \frac{162.3}{572.3} \approx 0.28$$

fair agreement!

החלפתי ביניהם

(c) מנתוני השאלה נמלא את הטבלת שכיחות: (נשים לב שהוגניות כפי שניתן מתרגמת לסכומים שווים של השורות והעמודות.)

טבלת שכיחות		A: l	A: c	A: n	
B	l	a_{11}	a_{12}	a_{13}	300
	c	a_{21}	a_{22}	a_{23}	300
	n	a_{31}	a_{32}	a_{33}	300
		300	300	300	900

1) $\sum a = 490, \quad N = 900$ (כמו בסעיף א' נמלא:)

2) $el = ec = en = \frac{300 \times 300}{900} = 100 \Rightarrow \sum e = 300$

$\Rightarrow \text{Kappa} = (490-300)/(900-300) \approx 0.317$

שאלה 2#

(a) p_value הוא הסיכוי להתקיימות הנתונים, בהינתן שהשערת האפס נכונה.

⊛ מכך ש- $p < 0.05$, החוקרים דחו את השערת האפס. מכאן שהנתונים הראו שנשים שלקחו X הראו עליה בצפיפות העצם, וזה גרם לדחית ההשערה, סימן שההשערה, היא "תרופה X לא מועלה בשיפור צפיפות העצם אצל נשים".

⊛ באותו האופן, $p > 0.05$ והחוקרים לא יכלו לדחות. הנתונים הראו שנשים שלקחו Y לא הראו עליה בצפיפות העצם, ולכן השערת האפס כנראה היא "תרופה Y לא מועלה בשיפור צפיפות העצם אצל נשים".

⇐ לגבי תרופה Y לא יודעים, ותרופה X לא מועלה. לכן פסלנו את (2),(3). ונותרנו עם (1) שמתאימה למה שהצהרתי בשורה מעל.

(b) $p < 0.05$ ולכן השערת האפס נדחו בהתאם לנתונים:

⊛ רמות ויטמין A נמוכות יותר נפוצות בקרב אנשים הסובלים משברי-עצמות.

⊛ דיאטה עשירה בשומן נפוצה יותר באנשים הסובלים משברי-עצמות.

לכן השערת האפס כנראה היא: שאין קשר בין רמת ויטמין A ובין רמות שומן בדיאטה, לבין הופעת שברי עצמות. לכן נסיק שהשערותיהן לא נכונות. אולם, מכיוון שהבדיקה נעשתה על נתונים שונים, ייתכן וכמו בדוגמא מהכיתה קיבלנו תוצאות מקריות. לכן לא ניתן לדעת בוודאות. לפיכך התשובה: (4).

כי בהנחה
שהשערות
האפס
נכונות ו
הצפיות
לא סבירות

שאלה 3#

(a) ① • השערת האפס: התלמידים בקורס הקריאה המהירה מצליחים לקרוא ולהבין פי 2 יותר מילים בדקה.

② נגדיר את התפלגות האפס להיות נורמלית, כאשר: mean = 2,

$\text{standard error} = \frac{2.2}{\sqrt{100}} = 0.22$ ⇐ standard deviation = 2.2

③ בניסוי על קבוצה רנדומית של N=100 תלמידים, קיבלנו שיעור הקריאה השתפרה פי 4.25

④ $$z = \frac{|(\text{sample - mean}) - (\text{hyp. mean})|}{\text{standard error}} = \frac{|4.25-5|}{0.22} \cong 3.41$$

ונמיר לערכי p, עבור $\alpha = 0.05$: $p = 0.000325$

⑤ ומתקיים: $p < 0.05$, ולכן נדחה את טענת האפס.

(b) תחום ה-Confidence יחושב ע"י: (כאשר: $\text{conf.level} = 95\% = 1.96$)

$$(\text{point estimate}) \pm (\text{conf. level} \times \text{standard error})$$

$$\Rightarrow [4.25 - \underbrace{(1.96 \times 0.22)}_{0.4312}, 4.25 + (1.96 \times 0.22)] = [3.8188, 4.6812]$$

(c) בהסתברות של 95% נימצא מתחום ה-Confidence, ונבחין ש-5 לא

נמצא בו. לכן גם 4.25 לא נמצא ב תחום ה-conf. של 5!

(כי שאם השערת האפס הייתה נכונה, אז קיבלנו בניסוי תוצאה לא צפויה ($< 5\%$)

ולכן אכן היינו דוחים אותה.) לכן ניתן לומר שתוצאת הניסוי ומגלות

התחום - מובהקות.