

מבחן בסטטיסטיקה להנדסת תוכנה

שאלון למועד X

סמסטר חורף, 2021

עמית שטקל

שאלה א (40 נק')

בסקר הצבעה לבחירות הקרובות ביקשו הסוקרים לבדוק האם קיימת תלות בין רמת ההשכלה של הבוחרים להעדפות הפוליטיות שלהם. בפרט, רצו לבדוק האם אחוזי ההצבעה מתפלגים באופן שונה בקרב קבוצות ההשכלה.

הנתונים שנמדדו סוכמו בטבלת השכיחויות הבאה.

כאשר X מתאר את רמת השכלה (1: יסודי, 2: תיכון, 3: אקדמאי)

ו-Y מתאר את המפלגה הפוליטית (A: מפלגה א', B: מפלגה ב', C: מפלגה ג').

X	Y		
Observed	A	B	C
1	100	200	300
2	100	100	100
3	80	10	10

- איזה מבחן מתאים לבדיקת ההשערה?
- אילו תנאים מקדימים צריכים להתקיים על מנת שנוכל להשתמש במבחן זה?
- בהנחה שהתנאים מתקיימים, נסחו השערה סטטיסטית מתאימה לבדיקת ההשערה.
- מה מספר דרגות חופש המתאים למבחן?
- חשבו את האומדן הסטטיסטי המתאים למדגם
- נתונים הערכים הקריטיים הבאים:

$$\chi^2_{(df,0.975)} = 11.14, \quad \chi^2_{(df,0.95)} = 9.49, \quad \chi^2_{(df,0.90)} = 7.779$$

נסחו מסקנה סטטיסטית לגבי ההשערה הנבדקת ברמת מובהקות של 5% בהתבסס על האומדן שחישבתם בשאלה 5. (השתמשו באומדן הקריטי המתאים לביסוס טענתכם, לא לשכוח לכתוב מסקנה מילולית).

שאלה ב (40 נק)

במחקר בדקו את הקשר בין הרגלי עישון לבין ריכוז הרטינול בפלסמת הדם. לצורך העניין חילקו את הנבדקים לארבע קבוצות: כאלו שלא עישנו מעולם, גמולים מעישון, מעשנים קל ומעשנים כבד. רמות הרטינול בדם של ארבע הקבוצות מסוכמות בטבלה הבאה:

הרגלי עישון	n	ממוצע רטינול בפלסמת הדם (ng/ml)	סטיית התקן של רמת הפלסמה בדם (ng/ml)
A לא עישנו מעולם	25	811.4	274.5
B גמולים מעישון	30	619.5	244.5
C מעשנים קל	13	525.6	143.7
D מעשנים כבד	12	404.3	221.2
סה"כ	80	631.9	

1. באיזה מבחן נשתמש על מנת לבדוק האם קיים שוני בין רמות הרטינול בדם של הקבוצות השונות?

נסחו השערה סטטיסטית לבדיקת ההשערה.

2. מהן ההנחות שצריכות להתקיים על מנת שנוכל לבצע מבחן זה?

3. מה מספר דרגות החופש לאומדן הסטטיסטי במבחן?

4. הדגימו את חישוב ה-MSB. שימו לב כי ממוצע הרטינול הכללי בדם נתון בטבלה המופיעה למעלה.

5. מה מייצג ה-MSB? (במשפט אחד, כולל פרוש הקיצור)

6. נתון כי ה- **SSE** שווה ל- 4,328,269 וה- **SSB** שווה ל- 1,578,354.

חשבו את האומדן הסטטיסטי F (מספיקה ספרת דיוק אחת אחרי הנקודה)

7. אם היינו מוציאים את הקבוצה "לא עישנו מעולם", מי מהאומדנים היה משתנה יותר?

- ה-MSB או ה-MSE?

- הסבירו מבלי לחשב מחדש את טבלת ה-ANOVA.

8. באיזה כוון ישתנה המדד המושפע מהוצאת קבוצת "לא עישנו מעולם"?

9. נתונים F סטטיסטי הבאים

$$F(0.95, df1, df2) = 2.725$$

$$F(0.975, df1, df2) = 3.293$$

$$F(0.05, df1, df2) = 0.117$$

האם ניתן לדחות את השערת האפס? נסחו את המסקנה הסטטיסטית בהתבסס על האומדן הקריטי המתאים לרמת מובהקות 0.05.

שאלה ג (20 נק')

במדגם הרגלי העישון שצויין בשאלה הקודמת רצו להשוות בין זוגות של קבוצות הניסוי השונות באמצעות התפלגות studentized (התפלגות Tukey).

נתון כי $q(p, df1, df2) = 3.7148$. כמו כן, נתון $MSE = 56,951$.

1. מהן מספר דרגות החופש לאומדן הסטטיסטי הקריטי q^* במקרה זה?
2. מהו ערך p המתאים לבדיקה ברמת מובהקות $\alpha=0.05$?
3. פרטו את חישוב האומדן הסטטיסטי המתאים לבחינת השאלה.
4. האם קיים הבדל בין רמות הרטינול בדם של קבוצת ה"גמולים מעישון" לקבוצת ה"מעשנים קל"? הסיקו מסקנה סטטיסטית התואמת ל- q_{B-C} המחושב.