

שאלון למועד א

מבחן בסטטיסטיקה להנדסת תוכנה

סמסטר אביב, 2020
הדס לפיד

א. בניסוי בדקו את רמות הגופרית בצמחים מגידול אורגני לעומת צמחים מגידול לא אורגני.

נתגלו רמות הגופרית הבאות לקילוגרם יבול (נתונים פיקטיביים)

גודל המדגם	שונות מדגמית	ממוצע המדגם (mg/kg)	
32	65000	2456	אורגני
48	68000	2870	לא אורגני

1. באיזה מבחן תשתמשו בכדי לבדוק האם רמות הגופרית בגידולים הלא אורגניים גבוהות מרמות הגופרית בגידולים האורגניים?

2. מהם התנאים שצריכים להתקיים בכדי שתוכלו להשתמש במבחן שהצעתם?

3. כיצד הייתם בודקים את קיום התנאים המקדימים למבחן?

4. בהנחה שהתנאים הדרושים מתקיימים, נסחו את ההשערה הסטטיסטית לבדיקת השאלה.

5. מה מספר דרגות החופש המתאים למבחן?

6. חשבו את האומדן הסטטיסטי של המבחן.

7. נתון כי $t_{0.975,df} = 1.991$, $t_{0.95,df} = 1.664$

נסחו את המסקנה, האם ניתן לקבל את הטענה כי רמות הגופרית בגידולים הלא אורגניים גבוהות מרמות הגופרית בגידולים האורגניים ברמת מובהקות של 5%? (השתמשו באומדן הקריטי המתאים לביסוס טענתכם)

ב. נתון כי סף רמת הגופרית המותר בצמח הוא 2800 (mg/kg).

באותו ניסוי, התפלגות הגידולים ביחס לסף רמות גופרית במדגם האורגני לעומת המדגם הלא אורגני היתה כדלקמן:

סה"כ	לא אורגני	אורגני	
14	12	2	מעל לסף <2800 (mg/kg)
66	36	30	מתחת לסף >2800 (mg/kg)

80	48	32	סה"כ
----	----	----	------

1. נסחו השערה סטטיסטית למציאת הבדל בין התפלגויות הגידולים האורגניים והלא אורגניים.

2. מה מספר דרגות החופש במבחן χ^2 ?

3. איזה תנאי צריך להתקיים בשביל לבצע מבחן זה?

4. בהנחה שהתנאי מתקיים, חשבו את האומדן הסטטיסטי לבדיקת ההשערות שניסחתם ב-1.

5. נתון: $\chi^2_{(df,0.975)} = 5.02$, $\chi^2_{(df,0.95)} = 3.84$, $\chi^2_{(df,0.90)} = 2.71$. הסיקו מסקנה סטטיסטית בנוגע להשערה שניסחתם ב-1 ברמת מובהקות של 5%. השתמשו באומדן הקריטי המתאים לביסוס ההשערה.

6. חשבו את ה-Odd-Ratio בהתבסס על נתוני המדגם.

7. מה ניתן להסיק ממדד זה לגבי ההבדלים בהתפלגות הרעילות בין הגידולים האורגניים והגידולים הלא אורגניים?

ג. בבדיקת הקשר בין הזמן שסטודנטים מבלים ברשתות החברתיות לבין ממוצע הציונים שלהם, סוכמו הנתונים באופן הבא:

זמן ברשתות החברתיות (שעות ביום)	גודל המדגם (n_i)	ממוצע ציונים	
		ממוצע (\bar{Y}_i)	סטיית התקן (s_i)
A: 1 or less	12	92.04	7.32
B: 2-3	68	85.14	9.48
C: 4 and more	23	78.43	9.33
All	103	84.45	

דווח גם כי הסכום המשוקלל של ריבועי ההפרשים בין ממוצע המדגמים לממוצע הכללי (ה- MS_{between} groups) היה 779.4, וכי השונות התוך קבוצתית (ה- SSE) היתה 8528.

1. נסחו השערת אפס והשערת אלטרנטיבית לבדיקת שוויון תוחלות הציונים בין קבוצות זמן החשיפה.

2. סכמו את הנתונים בטבלת ANOVA. לצורך כך, השלימו את המדדים החסרים (ה- SS_{between} וה- MS_{Error}) וחשבו את האומדן הסטטיסטי, F .

Source	df	SS	MS	F
Between groups	2		779.4	
Error	100	8528		

3. מה משמעות מדד ה- MS_{Error} ?

4. נתון כי $F(0.95, df1=2, df2=100) = 3.087$. נסחו את מסקנתכם בנוגע להשערה ברמת מובהקות של 5%, בהתבסס על האומדן הסטטיסטי שחישבתם בשאלה 2.

5. בהשוואות contrast בין תוחלות כל זוגות הקבוצות האפשריים, נמצאו רמות המובהקות הבאות.

Groups	Pv
C-A	0.000218
B-A	0.018919
C-B	0.006511

נתחו את ההבדלים לפי שיטת Binyamini-Hochberg (BH) עבור רמת מובהקות של 5%.

6. סכמו את הממצאים במילים.

7. איזה קריטריון אנו מגבילים לפי שיטת BH?

8. מה משמעות הקריטריון הזה?

9. לאיזה גודל הוא מוגבל כאשר מבצעים תיקון BH?