

# תשובות למועד Y

## מבחן בסטטיסטיקה להנדסת תוכנה

סמסטר אביב, 2020  
הדס לפיד

**א.** בניסוי בדקו את השפעת החברות על יכולת הלמידה. בקבוצת ניסוי אחת, הושמו מספר חולדות בכלוב למשך שבועיים. אחרי אותם שבועיים, ערכו לחולדות משימת למידה קצרה ומיד אחריה בדקו את משך זמן ביצוע המשימה. בקבוצת ניסוי שניה, הושמו חולדות בכלובים נפרדים, בתנאי גידול שווים לקבוצה הראשונה (מבחינת צפיפות, מזון ומים), ולאחר שבועיים ערכו להן את אותן בדיקות. חלק מהחולדות לא הצליחו ללמוד את המשימה וזמן הביצוע שלהן היה חריג משמעותית מאלו שכן קלטו את המשימה.

בטבלה הבאה מסוכמים ממצאי הכשלונות לעומת ההצלחות בשתי הקבוצות:

סה"כ	בודדים	חברתיים	
12	10	2	כשלו
52	22	30	הצליחו
64	32	32	סה"כ

- נסחו השערה סטטיסטית למציאת הבדל בהתפלגות הכשלונות/הצלחות בין שתי הקבוצות.
- מה מספר דרגות החופש במבחן  $\chi^2$ ?
- איזה תנאי צריך להתקיים כדי שמבחן זה יחשב אמין?
- נתון כי  $\chi^2_{(df,0.975)} = 5.02$ ,  $\chi^2_{(df,0.95)} = 3.84$ ,  $\chi^2_{(df,0.90)} = 2.71$ .
- בהנחה שהתנאי למבחן מתקיים, חשבו את האומדן הסטטיסטי לבדיקת ההשערה שניסחתם בשאלה 1. הסיקו את המסקנה הסטטיסטית בנוגע להשערה ברמת מובהקות של 5%. השתמשו באומדן הקריטי המתאים לביסוס טענתכם.
- איזה קטגוריה/קבוצה מהמופעים נראים לכם חריגים? על איזה הבדל זה מרמז?
- חשבו את ה-Odd-Ratio בהתבסס על נתוני המדגם.
- מה ניתן להסיק ממדד זה לגבי ההבדלים בהתפלגות הצלחת הלמידה בין קבוצת החברתיים לקבוצת הבודדים?

## תשובות לשאלה א:

$$H_0: P_{\text{כשלו}}^{\text{חברתיים}} = P_{\text{כשלו}}^{\text{בודדים}} = P_{\text{כשלו}}^{\text{expected}}, P_{\text{הצליחו}}^{\text{חברתיים}} = P_{\text{הצליחו}}^{\text{בודדים}} = P_{\text{הצליחו}}^{\text{expected}} \quad 1.$$
$$H_A: \text{otherwise}$$

או ורסיה חלקית של ההשערה (רק עבור קטגוריה אחת)

2. מספר דרגות החופש במבחן:  $1 = (2-1) * (2-1)$  שמשמעו מספר הקטגוריות פחות 1 כפול מספר קבוצות ההשוואה פחות 1.

3. התנאי שצריך להתקיים בשימוש במבחן  $\chi^2$  להשוואה קטגורית בין קבוצות הוא שמספר התצפיות הצפוי יהיה גדול או שווה ל-5 בלפחות מ-80% מהתאים (תאים הכוונה – כל הקטגוריות בכל הקבוצות).

4. טבלת התדירויות הצפויות והאומדנים הסטטיסטים פר תא:

		חברתיים	בודדים
כשלונות	Expected counts	$32 * 12 / 64 = 6$	$32 * 12 / 64 = 6$
	cell $\chi^2$	$(2-6)^2 / 6 = 2.667$	$(10-6)^2 / 6 = 2.667$
הצלחות	Expected counts	$32 * 52 / 64 = 26$	$32 * 52 / 64 = 26$
	cell $\chi^2$	$(30-26)^2 / 26 = 0.615$	$(22-26)^2 / 26 = 0.615$

$$\chi^2 = 2 * 2.667 + 2 * 0.615 = 6.564$$

5.  $6.564 > 3.84$  ולכן נדחה את השערת האפס ונאמר כי התפלגות ההצלחות שונה בין קבוצת החברתיים לקבוצת הבודדים.

6. ה- $\chi^2$  של החברתיים ושל הבודדים בקטגוריית הכשלונות גדול מ-2, ולכן ניתן לחשוד כי מספר הכשלונות שונה סטטיסטית בין הקבוצות.

7. נחשב את טבלת ההסתברויות הנצפית:

	חברתיים	בודדים
כשלונות	$2/32 = 0.0625$	$10/32 = 0.3125$
הצלחות	$30/32 = 0.9375$	$22/32 = 0.6875$

נעת נחשב את ה-Odd-Ratio:

$$OR = \frac{0.0625 * 0.6875}{0.3125 * 0.9375} = 0.146$$

8. מכיוון שה- $OR = 0.146 < 0.5$  ניתן להסיק כי אחוז הכשלונות גבוה משמעותית בקרב הבודדים בהשוואה לחברתיים.

**ב.** בהמשך לאותו הניסוי, סוכמו זמני התגובה של החולדות שהצליחו לבצע את המשימה בשתי קבוצות המחקר באופן הבא:

גודל המדגם	שונות מדגמית	ממוצע המדגם (שניות)	חברתיים
30	16.8	15.6	חברתיים
22	14.4	19.2	בודדים

1. באיזה מבחן תשתמשו בכדי לבדוק האם החולדות החברתיות ביצעו מהר יותר את משימת הלימוד?

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  $t_{0.975,df} = 2.008$ ,  $t_{0.95,df} = 1.6759$

האם ניתן לקבל את הטענה כי החולדות החברתיות היו מהירות יותר מהחולדות הבודדות ברמת מובהקות של 5%? נסחו מסקנתכם, השתמשו באומדן הקריטי המתאים לביסוס טענתכם.

## תשובות לב':

1. מבחן t לא מצומד חד צדדי.
2. בכדי להשתמש במבחן זה יש להניח שויון שוניות באוכלוסיות הנדגמות והתפלגות נורמליות משותפת של שני המדגמים.
3. בדיקת שויון שוניות תתבצע על ידי מבחן ליון.
- בדיקת נורמליות משותפת בודקים על ידי מבחן שפירו-ווילק על אוסף הנתונים של שני המדגמים יחדיו, המוסטים להתפלגות סביב ה-0 ע"י חיסור הממוצע המדגמי מכל אחת מהדגימות.

$$H_0: \mu_{\text{חברתיים}} = \mu_{\text{בודדים}}$$

$$H_A: \mu_{\text{בודדים}} > \mu_{\text{חברתיים}}$$

$$5. \text{ מספר דרגות החופש למבחן הוא } 30+22-2 = 50$$

$$6. \text{ נחשב תחילה את הסטיית התקן המשותפת של המדגמים:}$$

$$S^2 = \frac{\sum_{i=1}^2 (n_i - 1) \cdot S_i^2}{\sum_{i=1}^2 (n_i - 1)} = \frac{29 \cdot 16.8 + 21 \cdot 14.4}{29 + 21} = 15.79$$

$$S = \sqrt{15.79} = 3.974$$

ומכאן נחשב את האומדן הסטטיסטי:

$$t = \frac{|\bar{Y}_1 - \bar{Y}_2|}{s \cdot \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{|15.6 - 19.2|}{3.974 \cdot \sqrt{\left(\frac{1}{30} + \frac{1}{22}\right)}} = 3.227$$

$$3.227 > 1.6759 \quad t > t_{(0.95,df)} \quad 7.$$

ולכן ניתן להסיק כי החולדות החברתיות היו מהירות יותר בביצוע משימת הלמידה מהחולדות הבודדות ברמת בטחון של 95%.

**ג.** בבדיקת הקשר בין אומדן ההערכה העצמית (core self-evaluation) לבין אומדן הבריאות הסובייקטיבי (מידת הבריאות שאנשים מדווחים לגבי עצמם, מ-1 עד 10) בקרב אנשים מקבוצת הגיל השלישי, סוכמו הנתונים באופן הבא:

אומדן ההערכה העצמית	גודל המדגם ( $n_i$ )	אומדן הבריאות	
		ממוצע ( $\bar{Y}_i$ )	סטיית התקן ( $s_i$ )
A (best)	20	8.51	1.72
B (average)	54	7.76	1.97
C (worst)	16	6.20	1.66
All	90	7.65	

דווח גם כי סכום ריבועי הפרשים בין ממוצע המדגמים לממוצע הכללי (ה- $SS_{\text{between groups}}$ ) היה 49.09, וכי ה-Sum of Squared Errors (SSE) היה 303.69.

1. נסחו השערת אפס והשערת אלטרנטיבית לבדיקת שוויון תוחלות הציונים בין קבוצות ההערכה העצמית השונות.

2. סיכמו את הנתונים בטבלת ה-ANOVA הבאה:

Source	df	SS	MS	F
Between groups	2	49.03		
Error	87	303.69		

השלימו את המדדים החסרים (ה- $MS_{\text{between}}$ , וה- $MS_{\text{error}}$ ) וחשבו את האומדן הסטטיסטי,  $F$ .

3. מה משמעות מדד ה- $MS_{\text{between groups}}$ ?

4. נתון כי  $F(0.95, df_1=2, df_2=87) = 3.1013$ . האם ניתן לדרוח את השערת האפס? נסחו את המסקנה הסטטיסטית בהתבסס על האומדן הקריטי.

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

Groups	Pv
C-A	0.001186
C-B	0.008782
B-A	0.126546

נתחו את ההבדלים לפי שיטת Bonferroni (עבור  $\alpha=0.05$ )

6. סכמו מילולית את מסקנותיכם לגבי שוויון התוחלות בכל אחת מההשוואות, בהתאם לתיקון בונפרוני שביצעתם ב-5.

7. איזה קריטריון אנו מגבילים לפי שיטת Bonferroni?

8. מה המשמעות של הקריטריון הזה?

9. האם הוא נחשב קריטריון מתירני או שמרני? למה?

## תשובות לשאלה ג:

1.  $H_0: \mu_A = \mu_B = \mu_C$   
 $H_A: otherwise$

2.

Source	df	SS	MS	F
Between groups	2	49.03	$49.03/2=24.516$	$24.516/3.491=7.023$
Error	87	303.69	$303.69/87=3.491$	

3.  $MS_{\text{between groups}}$  הוא אומדן לשונות הבין קבוצתית הממוצעת, הוא משמש אותנו במונה בחישוב האומדן הסטטיסטי ב-  $F$  test. הוא אומד את המרחק המנומל בין ממוצעי קבוצות המדגם לממוצע הכללי בניסוי, תחת הנחת שוויון שונות.

4. במבחן אנאליזת שונות זה נתקבל אומדן סטטיסטי  $7.023 > 3.1013$ , ז"א שהאומדן הסטטיסטי גדול מהאומדן הקריטי ברמת מובהקות של 5%. מכאן ניתן להסיק כי יש הבדלים מובהקים בין תוחלות אומדן הבריאות של קבוצות ההערכה העצמית השונות.

5.

Index	Groups	Pv	Bonferroni Correction $\frac{\alpha}{m}$	Hypothesis assertion
1	C-A	0.001186	0.0166	*
2	C-B	0.008782	0.0166	*
3	B-A	0.126546	0.0166	NA

כאשר  $m$  מייצג את מספר ההשוואות (3 במקרה זה).

6. לפי תיקון Bonferroni ל-FWER, נתגלו הבדלים משמעותיים סטטיסטית בין הקבוצות C ו-A ובין C ו-B.

7. Bonferroni בא להגביל את ה-FWER (Family-wise error rate)

8. משמעות ה-FWER הוא הסיכוי שנטעה בלפחות השוואה אחת מתוך כלל ההשוואות ונטען  $H_A$ , בעוד  $H_0$  היא הנכונה עבור אותה ההשוואה. ע"י Bonferroni אנו מגבילים את ה-FWER לרמת המובהקות  $\alpha$ .

9. בונפרוני נחשב לתיקון שמרני, כי הוא לא מאפשר הרבה דחיות של  $H_0$  בגלל שסיכוי הסף שהוא מחייב לדחיית  $H_0$  הוא הנמוך בהשוואה לרמת המובהקות  $\alpha$ .