

סטטיסטיקה – מטלה 4

אלעזר פיין

א.

H0: Dist(Tel – Aviv) ~ Dist(Haifa) 1.

H1: Otherwise

2. נמלא טבלת שכיחויות לפי $E_{ij} = N_j \cdot P(X_i)$

n = 60	Haifa Expected	Haifa Observed	Tel – Aviv Expected	Tel – Aviv Observed	Observed Sum	$\chi^2 = \sum \frac{(x_i - m_i)^2}{m_i}$
מרוצים	$24 \cdot \frac{35}{60}$ = 14	18	$36 \cdot \frac{35}{60}$ = 21	17	35	$\frac{(18 - 14)^2}{14} + \frac{(17 - 21)^2}{21}$ = $\frac{40}{21}$
לא מרוצים	10	6	15	19	25	$\frac{(6 - 10)^2}{10} + \frac{(19 - 15)^2}{15}$ = $\frac{8}{3}$
sum	24	24	36	36	60	$\frac{40}{21} + \frac{8}{3} = \frac{32}{7} \approx 4.57$

לרמת מובהקות של 5% הערך הקריטי הוא 3.84, קיבלנו $3.84 < 4.57$ ולכן דוחים את השערת האפס ומסיקים כי יש הבדל ברמת שביעות הרצון בין תל – אביב לבין חיפה (לא מתפלגים בצורה דומה).

3. $P\text{value} < \alpha = 0.05$

4. Observed Probabilities Table:

	חיפה	תל אביב
מרוצים	$18/24 = 0.75$	$17/36 = 0.47$
לא מרוצים	$6/24 = 0.25$	$19/36 = 0.53$

$$OR = \frac{0.25 \cdot 0.47}{0.75 \cdot 0.53} = 0.29 < 1$$

מכיוון שהיחס קטן משמעותית מאחד מסיקים כי בחיפה משמעותית יותר מרוצים.

ב.

$$H_0: \mu_A = \mu_B = \mu_C = \mu_D$$

$$H_1: \text{otherwise}$$

2.

Source	df	SS	MS	F
Between Groups	$K - 1 = 4 - 1 = 3$	54.18	$\frac{54.18}{4 - 1} = 18.06$	$\frac{18.06}{2.344} = 7.705$
Error	$n - k = 60 - 4 = 56$	$2.344 \cdot 56 = 131.264$	2.344	-

3. $MS_{between}$: ממוצע השוניות של כל הקבוצות מהממוצע הכולל של המרחק בין הקבוצות. כאשר בוחנים ביחס לממוצע השוניות בתוך הקבוצות - יחס גבוה מערך קריטי F_c נותן דחייה להשערת האפס. בנוסף אפשר לקבל אינדיקציה מוקדמת בעזרתו לקבוצות חריגות בשונותם.

$$F > F_c \Leftrightarrow 7.705 > 2.769 \Rightarrow \text{Reject } H_0$$

מסיקים שאין שוויון תוחלות בין הקבוצות – כלומר יש תלות בין גיל לכמות פעילות גופנית.

4. נוסף לטבלה את הדרוש לבחון HOLMES:

i	Groups	Pv	Holmes CV $\frac{\alpha}{m + 1 - i}$	Reject H0
1	D-A	0.000125	$0.05/6+1-1 = 0.00833$	true
2	C-A	0.005909	$0.05/6+1-2 = 0.01$	true
3	B-A	0.021698	$0.05/6+1-3 = 0.0125$	false
4	D-B	0.080907	$0.05/6+1-4 = 0.0167$	false
5	D-C	0.300442	$0.05/6+1-5 = 0.025$	false
6	C-B	0.441925	$0.05/6+1-6 = 0.05$	false

קיימים 2 זוגות של קבוצות שיש ביניהם שוני משמעותי בתוחלת מספר שעות פעילות הגופנית:
A-C, A-D

2 זוגות אלו הם אלו ש"הרסו" לנו את השוויון תוחלות שחיפשנו במבחן ANOVA.
בכל שאר הזוגות אין הבדל משמעותי בתוחלות.

ג.

1. המשתנים מתפלגים נורמלית.

2. $H_0: \rho = 0$ (אין קשר לינארי בין המשתנים)

$H_1: \rho \neq 0$ (יש קשר לינארי בין המשתנים)

$$df = n - 2 = 30 - 2 = 28$$

$$r_1 = z_{x1} \cdot z_{y1} \quad 3.$$

$$r_1 = \frac{66.13 - 68}{3.837} \cdot \frac{2.56 - 3}{0.596} = 0.359$$

$$r = \frac{1}{n-1} \sum_{i=1}^n r_i : (n-1) \text{ מחברים ומחלקים ב-}$$

$$t = r \cdot \frac{\sqrt{n-2}}{\sqrt{1-r^2}} = 0.879 \cdot \frac{\sqrt{28}}{\sqrt{1-0.879^2}} = 9.75 \quad 4.$$

5. $9.75 > 2.048$ מכאן ניתן לדחות את השערת האפס ולהסיק כי המקדם מתאם לא אפס ויש קשר לינארי בין גובה תינוק למשקלו.

ד.

$$H_0: \mu = 4.5 \quad 1.$$

$$H_1: \mu \neq 4.5$$

2. יש להניח כי המשתנה מתפלג נורמלית.

$$CI = \bar{X} \pm t_{1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} = 3.8 \pm 2.0518 \cdot \frac{0.82}{\sqrt{28}} \quad 3.$$

$$ci_low = 3.48 \quad ci_upper = 4.11$$

4.5 לא ברווח סמך. למרות שמטרת רווח הסמך היא לא לקבל או לדחות השערות כעת יודעים מה תהיה תוצאת בדיקת ההשערות הדו צדדית: דוחים את H_0 ברמת בטחון 95%.

$$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} = \frac{3.8 - 4.5}{\frac{0.82}{\sqrt{28}}} = -4.52 \rightarrow |t| > 2.0518 \quad 4.$$

מכאן דוחים את השערת האפס ברמת מובהקות 0.05 – תוחלת האוכלוסייה אינה 4.5.