מבחן סטטיסטי מס׳ 13: ניתוח שונות חד כיווני – ANOVA

מטרה – השוואת תוחלות של מספר אוכלוסיות $(k \geq 3)$

$$H_0 \quad \mu_1 = \mu_2 = \ldots = \mu_k$$
$$H_1 \quad otherwise$$

הנחות

1. האוכלוסיות מפולגות נורמלית.
2. שוויון שונויות של k האוכלוסיות.
3. המדגמים בלתי תלויים.

מבנה הנתונים

מדגם	1	2	...	k
	$X_{1,1}$	$X_{2,1}$	$\cdots$	$X_{k,1}$
המספר הסידורי של התצפית	$\vdots$	$\vdots$		$\vdots$
	X_{1,n_1}	X_{2,n_2}	$\cdots$	X_{k,n_k}

$x_{i,j}$

$i = 1,2,\ldots,k$ - המספר הסידורי של המדגם.

$j = 1,2,\ldots n_i$ - המספר הסידורי של התצפית בתוך המדגם.

חישובי עזר

1. עבור כל אחד מבין המדגמים מחשבים את $S_i, \overline{x_i}, n_i$ $(i = 1,2,\ldots,k)$
2. מחשבים ממוצע כללי $\overline{x} = \dfrac{\sum_{i=1}^{k} n_i \overline{x_i}}{n}$ ($\overline{x}$ - ממוצע כללי, n - סה"כ מספר הנתונים)
3. מחשבים את סכום הריבועים בין המדגמים $SS_b = \sum_{i=1}^{k} n_i \left(\overline{x_i} - \overline{x}\right)^2$ (between)
4. מחשבים את סכום הריבועים בתוך המדגמים $SS_w = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(x_{i,j} - \overline{x}_i\right)^2 = \sum_{i=1}^{k}(n_i - 1)S_i^2$ (within)

בניית לוח ניתוח שונות (ANOVA)

מקור ההשתנות	סכומי הריבועים (SS)	דרגות החופש (DF)	ממוצעי סכומי הריבועים (MS)	יחס F
בין (B)	SS_b	k-1 ד מס' הקבוצות -1	$MS_b = \dfrac{SS_b}{k-1}$	$F_{calculated} = \dfrac{MS_b}{MS_w}$
בתוך (W)	SS_w	n-k	$MS_w = \dfrac{SS_w}{n-k}$	
סה"כ (T)	$SS_t = SS_b + SS_w$	n-1	$MS_t = S^2 = \dfrac{SS_t}{n-1}$ $(MS_t \neq MS_b + MS_w)$	$R = \left\{F_{calculated} > F_{1-\alpha,k-1,n-k}\right\}$

השוואות מרובות - ניתוח POST-HOC (LSD):

קיים שוני בין תוחלת של אוכלוסיה i לבין תוחלת של אוכלוסיה j אם:

$$\frac{|\bar{x}_i - \bar{x}_j|}{\sqrt{MSw\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}} > t_{1-\frac{\alpha}{2},n-k}$$

נוסחה לחישוב מספר השוואות מרובות שיש לבצע ל-k אוכלוסיות:

$$\binom{k}{2} = \frac{k(k-1)}{2}$$

הערה: נשתמש במבחן המשך (POST- HOC) רק אם דוחים את השערת האפס.