תרגול 12 - ניתוח שונות דו-כיווני

ניתוח שונות דו-כיווני מצטלב

נשתמש במודל זה כאשר יש שני גורמים מסבירים, וכאשר מספר הרמות בכל אחד מהפקטורים הינו סופי.

המודל מאפשר לענות על 3 שאלות:

- האם יש השפעה לגורם המסביר הראשון על הגורם המוסבר?
- האם יש השפעה לגורם המסביר השני על הגורם המוסבר?
- האם שילוב של שני הגורמים המסבירים יוצר השפעה על הגורם המוסבר? (אינטראקציה)

סימונים:

a – מספר הרמות של הגורם הראשון, $i = 1, \dots, a$

b – מספר הרמות של הגורם השני, $j = 1, \dots., b$

n – מספר התצפיות שנדגמו בכל "תא" (שילוב רמות של הגורם הראשון והגורם השני), $k = 1, \dots, n$

$N = n \cdot a \cdot b$ – סך התצפיות בניסוי.

הנחת היסוד היא שמס' התצפיות בכל שילוב רמות הינו קבוע, n.

טבלת התצפיות:

		גורם II (β)			
	i \ j	1	2	…	b
גורם I (τ)	1	y_{111} y_{112} ⋮ y_{11n}	y_{121} y_{122} ⋮ y_{12n}	⋮	y_{1b1} y_{1b2} ⋮ y_{1bn}
	2	y_{211} y_{212} ⋮ y_{21n}	y_{221} y_{222} ⋮ y_{22n}	⋮	y_{2b1} y_{2b2} ⋮ y_{2bn}
	⋮	⋮	⋮	⋮	⋮
	a	y_{a11} y_{a12} ⋮ y_{a1n}	y_{a21} y_{a22} ⋮ y_{a2n}	⋮	y_{ab1} y_{ab2} ⋮ y_{abn}

טבלת הממוצעים:

i \ j	1	2	…	b	
1	$\bar{y}_{11\cdot}$	$\bar{y}_{12\cdot}$	⋮	$\bar{y}_{1b\cdot}$	$\bar{y}_{1\cdot\cdot}$
2	$\bar{y}_{21\cdot}$	$\bar{y}_{22\cdot}$	⋮	$\bar{y}_{2b\cdot}$	$\bar{y}_{2\cdot\cdot}$
⋮	⋮	⋮	⋮	⋮	
a	$\bar{y}_{a1\cdot}$	$\bar{y}_{a2\cdot}$	⋮	$\bar{y}_{ab\cdot}$	$\bar{y}_{a\cdot\cdot}$
	$\bar{y}_{\cdot 1\cdot}$	$\bar{y}_{\cdot 2\cdot}$		$\bar{y}_{\cdot b\cdot}$	$\bar{y}_{\cdot\cdot\cdot}$

ממוצע תא: $\overline{y_{ij\cdot}} = \frac{1}{n}\sum_{k=1}^{n} y_{ijk}$

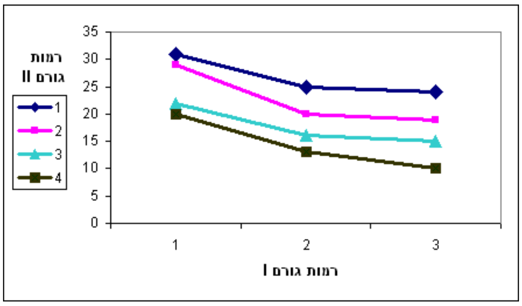
ממוצע רמה i: $\overline{y_{i\cdot\cdot}} = \frac{1}{b}\sum_{j=1}^{b} \overline{y_{ij\cdot}}$

ממוצע רמה j: $\overline{y_{\cdot j\cdot}} = \frac{1}{a}\sum_{i=1}^{a} \overline{y_{ij\cdot}}$

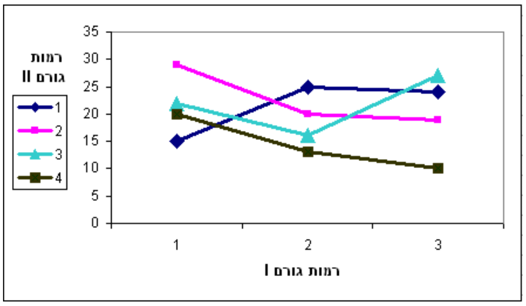
ממוצע כולל: $\bar{y} = \frac{1}{N}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk} = \frac{1}{a}\sum_{i=1}^{a}\overline{y_{i\cdot\cdot}} = \frac{1}{b}\sum_{j=1}^{b}\overline{y_{\cdot j\cdot}} = \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\overline{y_{ij\cdot}}$

אינטראקציה - השפעה נוספת מעבר להשפעה האדיטיבית של כל רמה בנפרד

אין אינטראקציה



יש אינטראקציה



<u>סכום ריבועי השגיאות:</u>

$$SS_{Tot}=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(y_{ijk}-\bar{y}_{...}\right)^2=$$

$$=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left[(\bar{y}_{i..}-\bar{y}_{...})+\left(\bar{y}_{.j.}-\bar{y}_{...}\right)+\left(\bar{y}_{ij.}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y}_{...}\right)+\left(y_{ijk}-\bar{y}_{ij.}\right)\right]^2=$$

$$=\underbrace{bn\sum_{i=1}^{a}\left(\bar{y}_{i..}-\bar{y}_{...}\right)^2}_{SS_A}+\underbrace{an\sum_{j=1}^{b}\left(\bar{y}_{.j.}-\bar{y}_{...}\right)^2}_{SS_B}+\underbrace{n\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\bar{y}_{ij.}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y}_{...}\right)^2}_{SS_E}+\underbrace{\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(y_{ijk}-\bar{y}_{ij.}\right)^2}_{SS_{AB}}$$

<u>דרגות החופש:</u>

$$df(SS_{Tot})=N-1$$

$$df(SS_E)=N-ab=(n-1)ab$$

$$df(SS_{Tr})=ab-1$$

$$df(SS_A)=a-1$$

$$df(SS_B)=b-1$$

$$df(SS_{AB})=(a-1)(b-1)$$

<u>נוסחאות מקוצרות:</u>

$$SS_A=nb\left[\sum_{i=1}^{a}\bar{y}_{i..}^2-a\bar{y}_{...}^{\ 2}\right]$$

$$SS_B=na\left[\sum_{j=1}^{b}\bar{y}_{.j.}^2-b\bar{y}_{...}^{\ 2}\right]$$

$$SS_e = \sum_i \sum_j \sum_k y_{ijk}^2 - n\sum_i \sum_j \bar{y}_{ij.}^2$$

$$SS_{Tot} = \sum_i \sum_j \sum_k y_{ijk}^2 - N\bar{y}_{...}^2$$

$$SS_{AB} = SS_{Tot} - SS_A - SS_B - SS_e$$

המודל הפרמטרי

$$\boxed{y_{ijk} = \mu + \tau_i + \beta_j + \tau\beta_{ij} + \varepsilon_{ijk}}$$

μ - הערך המרכזי של כלל התצפיות.

τ_i - הסטייה של רמה i של הגורם המסביר הראשון.

β_j - הסטייה של רמה j של הגורם המסביר השני.

$\tau\beta_{ij}$ - הסטייה שנגרמת מהאינטראקציה בין רמה i של הגורם הראשון לרמה j של הגורם השני.

$\varepsilon_{ijk} \sim N(0,\sigma^2)$ - התפלגות הרעש.

מס' הרמות של כל אחד מהגורמים המסבירים סופי ונכנס לתוך הדגימה. דוגמים את כל הרמות של כל הגורמים. מכיוון שיש מספר זהה של תצפיות בכל תא, ניתן לדרוש: $\sum_{i=1}^{a}\tau_i = 0, \sum_{j=1}^{b}\beta_j = 0$,

באותו אופן, סכום האינטראקציות על כל שורה או עמודה מתאפס:

$$\forall j \quad \sum_{i=1}^{a}\tau\beta_{ij} = 0 \qquad \text{ו-} \quad \forall i \quad \sum_{j=1}^{b}\tau\beta_{ij} = 0.$$

אמדים לשונויות:

$$E(MS_A) = \sigma^2 + \frac{bn\sum_{i=1}^{a}\tau_i^2}{a-1}$$

$$E(MS_B) = \sigma^2 + \frac{an\sum_{j=1}^{b}\beta_j^2}{b-1}$$

$$E(MS_{AB}) = \sigma^2 + \frac{n\sum_{i=1}^{a}\sum_{j=1}^{b}(\tau\beta)_{ij}^2}{(a-1)(b-1)}$$

$$E(MS_E) = \sigma^2$$

השערות האפס:

א. הגורם הראשון לא משפיע:

$$\begin{cases} H_o: \tau_i = 0 \\ H_1: else \end{cases} \Rightarrow \begin{cases} H_o: E(MS_A)/E(MS_e) = 1 \\ H_1: E(MS_A)/E(MS_e) > 1 \end{cases}$$

$$F_{emp} = \left.\frac{MS_A}{MS_e}\right|_{H_0} \sim F(a-1,(n-1)ab) \qquad F_{cr} = F_\alpha(a-1,(n-1)ab)$$

ב. הגורם השני לא משפיע:

$$\begin{cases} H_o: \beta_j = 0 \\ H_1: else \end{cases} \Rightarrow \begin{cases} H_o: E(MS_B)/E(MS_e) = 1 \\ H_1: E(MS_B)/E(MS_e) > 1 \end{cases}$$

$$F_{emp} = \left.\frac{MS_B}{MS_e}\right|_{H_0} \sim F(b-1,(n-1)ab) \qquad F_{cr} = F_\alpha(b-1,(n-1)ab)$$

ג. כל האינטראקציות שוות לאפס – אין אינטראקציה.

$$\begin{cases} H_o: \tau\beta_{ij} = 0 \\ H_1: else \end{cases} \Rightarrow \begin{cases} H_o: E(MS_{AB})/E(MS_e) = 1 \\ H_1: E(MS_{AB})/E(MS_e) > 1 \end{cases}$$

$$F_{emp} = \left.\frac{MS_{AB}}{MS_e}\right|_{H_0} \sim F((a-1)(b-1),(n-1)ab) \qquad F_{cr} = F_\alpha((a-1)(b-1),(n-1)ab)$$

אם $F_{emp} > F_{cr}$ נדחה את H_0

לוח ניתוח שונות

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_{emp}
A Treatment	SS_A	a-1	$MS_A = \frac{SS_A}{a-1}$	$\frac{MS_A}{MS_e}$
B Treatment	SS_B	b-1	$MS_B = \frac{SS_B}{b-1}$	$\frac{MS_B}{MS_e}$
Interaction	SS_{AB}	(a-1)(b-1)	$MS_{AB} = \frac{SS_{AB}}{(a-1)(b-1)}$	$\frac{MS_{AB}}{MS_e}$
Error	SS_e	ab(n-1)	$MS_e = \frac{SS_e}{ab(n-1)}$	
Total	SS_{Tot}	abn-1		

אמדים לפרמטרים:

תוחלת כוללת: $\hat{\mu} = \bar{y}_{...}$

סטיית רמה i של הגורם הראשון: $\hat{\tau}_i = \bar{y}_{i..} - \bar{y}_{...}$

סטיית רמה j של הגורם השני : $\hat{\beta}_j = \bar{y}_{.j.} - \bar{y}_{...}$

אינטראקציה בין i ל-j : $\tau\beta_{ij} = \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}$

אמד לשונות הרעש : $\hat{\sigma}^2 = MS_e$

רב״ס לתוחלת : $\mu \in \bar{y}_{...} \pm t_{\frac{\alpha}{2}}[(n-1)ab]\sqrt{\frac{MS_e}{N}}$

רב״ס לתוחלת של רמה i של בגורם הראשון : $\mu + \tau_i \in \bar{y}_{i..} \pm t_{\frac{\alpha}{2}}[(n-1)ab]\sqrt{\frac{MS_e}{nb}}$

רב״ס לתוחלת של רמה j של בגורם השני : $\mu + \beta_j \in \bar{y}_{.j.} \pm t_{\frac{\alpha}{2}}[(n-1)ab]\sqrt{\frac{MS_e}{na}}$

רב״ס לתוחלת של תא מסוים : $\mu + \tau_i + \beta_j + \tau\beta_{ij} \in \bar{y}_{ij.} \pm t_{\frac{\alpha}{2}}[(n-1)ab]\sqrt{\frac{MS_e}{n}}$

רב״ס לשונות הרעש :

$$\frac{ab(n-1)MS_e}{\chi^2_{\frac{\alpha}{2}}\left(ab(n-1)\right)} \le \sigma^2 \le \frac{ab(n-1)MS_e}{\chi^2_{1-\frac{\alpha}{2}}\left(ab(n-1)\right)}$$

הערה : נאמוד את הסטיות $\hat{\tau}_i, \widehat{\beta_j}, \tau\beta_{ij}$ רק לאחר שדחינו את השערת האפס הרלוונטית.

אין משמעות לאמידת הפרמטרים הללו ללא דחיית השערת האפס!

<u>תרגיל 1</u>

ביחידה מובחרת יש שני צלפים ושני רובי צלפים. מפקד היחידה מתלבט האם להצמיד רובה מסוים לצלף מסוים, ולשם כך עורך ניסוי, בו כל צלף מבצע שלושה מחזורי ירי בכל רובה צלפים. הניקוד שצבר כל צלף בכל מחזור ירי נרשם, ולהלן התוצאות :

		רובה צלפים	
		1	**2**
צלף	**1**	355 395 323	415 403 346
	2	392 350 298	376 357 357

א. מהו המודל המתאים לניתוח הנתונים?
ב. מהן הנחות המודל?
ג. בדקו את השערות המודל :
 1. האם קיימת שונות בין הצלפים במובהקות של 0.01?
 2. האם קיימת שונות בין רובי הצלפים במובהקות של 0.01?
 3. האם קיימת שונות הנובעת מהאינטראקציה במובהקות של 0.01?
ד. מה המלצתך למפקד היחידה?

<u>פתרון:</u>

א. מהו המודל המתאים לניתוח הנתונים?

המודל המתאים לניתוח הנתונים הוא מודל ניתוח שונות דו כיווני מצטלב פרמטרי:

אנו מעוניינים לבחון את ההבדל בין הרמות של שני גורמים (צלף, רובה) ולכן מדובר במודל דו-כיווני.

המודל הינו פרמטרי, אין מדובר בבחירה של מספר רמות מתוך אוסף גדול יותר של רמות, אלא אוספים נתונים אודות כל הרמות הקיימות (לפי נתוני השאלה יש 2 רובים ו-2 צלפים בלבד).

משוואת המודל:

$$\boxed{y_{ijk} = \mu + \tau_i + \beta_j + \tau\beta_{ij} + \varepsilon_{ijk}}$$

τ_i - הסטייה של רמה i של הגורם המסביר הראשון (צלף).

β_j - הסטייה של רמה j של הגורם המסביר השני (רובה).

$\tau\beta_{ij}$ - הסטייה שנגרמת מהאינטראקציה בין רמה i של הגורם הראשון לרמה j של הגורם השני.

μ - הערך המרכזי של כלל התצפיות.

ב. מהן הנחות המודל?

- ניקוד הירי מתפלג נורמלית בכל רמה.
- הרעש מתפלג נורמלית, באופן אחיד על פני הרמות.
- אין תלות בין מחזורי הירי השונים (התצפיות ב״ת).

ג. בדיקת השערות המודל:

1. האם קיימת שונות בין הצלפים במובהקות של 0.01?

השערת האפס המתאימה תהיה כי כל רמות הגורם הראשון זהות.
הגורם הראשון (הצלף) אינו משפיע על ציון הירי:

$$\begin{cases} H_o : \tau_i = 0 \\ H_1 : else \end{cases} \Rightarrow \begin{cases} H_o : E(MS_A)/E(MS_e) = 1 \\ H_1 : E(MS_A)/E(MS_e) > 1 \end{cases}$$

2. האם קיימת שונות בין רובי הצלפים במובהקות של 0.01?

השערת האפס המתאימה תהיה כי כל רמות הגורם השני זהות, אם נצליח לדחות השערה זאת נוכל לטעון כי אין הבדל בין רובי הצלפים.

הגורם השני לא משפיע:

$$\begin{cases} H_o : \beta_j = 0 \\ H_1 : else \end{cases} \Rightarrow \begin{cases} H_o : E(MS_B)/E(MS_e) = 1 \\ H_1 : E(MS_B)/E(MS_e) > 1 \end{cases}$$

3. האם קיימת שונות הנובעת מהאינטראקציה במובהקות של 0.01?

השערת האפס המתאימה תהיה כי אין השפעה לשילוב הגורם הראשון והשני יחד.
כלומר אין אינטראקציה:

$$\begin{cases} H_o : \tau\beta_{ij} = 0 \\ H_1 : else \end{cases} \Rightarrow \begin{cases} H_o : E(MS_{AB})/E(MS_e) = 1 \\ H_1 : E(MS_{AB})/E(MS_e) > 1 \end{cases}$$

נבצע מבחני F לבחינת שונות:

חישובי עזר:

טבלת הממוצעים :

i \ j	1	2	
1	357.66	388.00	372.83
2	346.66	363.33	355.00
	352.16	375.66	$\boxed{363.916}$

טבלת ריבועי הממוצעים :

i \ j	1	2	
1	127925.4	150544.0	139004.7
2	120177.8	132011.1	126025.0
	124021.4	141125.4	$\boxed{132435.3}$

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_{emp}
A Treatment	954.0833	1	954.0833	0.7635
B Treatment	1656.75	1	1656.75	1.3259
Interaction	140.0833	1	140.0833	0.1121
Error	9996	8	1249.5	
Total	12746.92	11		

$$F_{cr} = F_{0.01}(1,8) = 11.26$$

לא ניתן לדחות את השערות האפס. לכן לא ניתן לומר כי קיימת שונות בין הצלפים, בין רובי הצלפים או שונות הנובעת מהאינטראקציה.

ד. מה המלצתך למפקד היחידה?

ההמלצה למפקד היחידה תהיה לצוות לצלפים רובה בצורה אקראית מכיוון שהשונויות אינן מובהקות. לחילופין, ניתן להמליץ למפקד היחידה לבצע מחזורי ירי נוספים, ייתכן שאם נוסיף תצפיות, השונויות כן תהיינה מובהקות.

<u>תרגיל 2</u>

בקבוצת כדורגל אירופאית כלשהי מעוניינים לבחון האם למעמד המשחק יש השפעה על המרחק ששחקן כדורגל רץ במהלך המשחק. לשם כך נאספו נתוני מרחקי ריצה (בק"מ) של 4 שחקני שדה ממספר עונות. (כל הנתונים שנאספו מתייחסים למשחקים בהם השחקנים שיחקו משחק מלא).

מעמד \ שחקן	1	2	3	4
שמינית גמר	9.76	7.85	9.01	9.20
	10.52	7.05	9.20	9.41
	9.54	8.15	8.40	9.85

רבע גמר	10.07 9.67 10.21	8.60 7.95 8.11	9.50 9.41 10.67	8.98 9.47 10.46
חצי גמר	9.58 8.56 9.01	7.64 7.93 8.26	8.50 8.90 9.15	9.70 10.41 10.68

א. האם קיימת שונות בין השחקנים ברמת מובהקות של 0.01 ?

ב. האם קיימת שונות בין שלבי ההכרעה השונים ברמת מובהקות של 0.01?

ג. האם קיימת אינטראקציה בין הגורמים המסבירים?

ד. האם ניתן לקבוע כי שחקן 4 משקיע יותר מאמץ במשחקי חצי גמר מאשר במשחקי שמינית גמר?

<u>פתרון</u>

טבלת הממוצעים :

שחקן / **מעמד**	**1**	**2**	**3**	**4**	
שמינית גמר	9.94	7.68	8.87	9.49	9.00
רבע גמר	9.98	8.22	9.86	9.64	9.43
חצי גמר	9.05	7.94	8.85	10.26	9.03
	9.66	7.95	9.19	9.80	$\boxed{9.15}$

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_{emp}
A Treatment	1.378	2	0.689	2.90
B Treatment	19.072	3	6.357	26.77
Interaction	3.737	6	0.622	2.62
Error	5.698	24	0.237	
Total	29.886	35		

$F_{cr} = F_{0.01}(6,24) = 3.67$ $F_{cr} = F_{0.01}(2,24) = 5.61$ $F_{cr} = F_{0.01}(3,24) = 4.72$

ניתן לדחות את השערת האפס לגבי השונות בין השחקנים, כלומר ניתן לקבוע כי ישנה שונות במרחק הריצה בין השחקנים.

לא ניתן לדחות את השערת האפס לגבי השונות בין שלבי ההכרעה וכן את השערת האפס לגביי האינטראקציות.

במקרה זה כדאי לשקול ביצוע ניתוח שונות חד כיווני כשהגורם המסביר היחידי בו הוא השחקן : נוכל לבטל את פרמטר 'מעמד המשחק', לכל שחקן יהיו 9 תצפיות ונוכל להגיע לתוצאות מדויקות יותר.

כדי לקבוע האם יש הבדל בין הרמות של שחקן 4 ניתן להשתמש במבחן של השוואות מרובות.

ניתוח שונות דו-כיווני היררכי

קיים גורם ראשי בעל a רמות (למשל מכונות), הגורם השני מקונן (nested) בתוך הגורם הראשון ולכן הוא משני (עובדים שונים לכל מכונה - הגורמים המשניים).

Machine								
1			2			3		
Operator			Operator			Operator		
1	2	3	1	2	3	1	2	3
y_{111}	y_{121}	y_{131}	y_{211}	y_{221}	y_{231}	y_{311}	y_{321}	y_{331}
…	…	…	…	…	…	…	…	…
y_{11n}	y_{12n}	y_{13n}	y_{21n}	y_{22n}	y_{23n}	y_{31n}	y_{32n}	y_{33n}
$\bar{y}_{11.}$	$\bar{y}_{12.}$	$\bar{y}_{13.}$	$\bar{y}_{21.}$	$\bar{y}_{22.}$	$\bar{y}_{23.}$	$\bar{y}_{31.}$	$\bar{y}_{32.}$	$\bar{y}_{33.}$
$\bar{y}_{1..}$			$\bar{y}_{2..}$			$\bar{y}_{3..}$		

מניחים כי לכל רמה יש אותו מספר גורמים משניים (b), ובנוסף מניחים כי בכל תא מספר דגימות זהה n.

המודל

$$y_{ijk} = \mu + \tau_i + \beta_{ij} + \varepsilon_{ijk}$$

μ- ערך מרכזי של כלל הרמות

τ_i- הסטייה של רמה i של הגורם הראשי

β_{ij}- הסטייה של רמה j של הגורם המשני בתוך רמה i של הגורם הראשי

$\varepsilon_{ijk} \sim N(0, \sigma^2)$ - התפלגות הרעש

דוגמים את כל הרמות הן של הראשי והן של המשני ולכן : $\sum_{i=1}^{a} \tau_i = 0 \quad \forall i = 1 \ldots a \quad \sum_{j=1}^{b} \beta_{ij} = 0$

פרמטרים

a – מספר הרמות של הגורם הראשי $i = 1 \ldots a$

b – מספר הרמות של הגורם המשני (עבור כל גורם ראשי) $j = 1 \ldots b$

k – מספר הדגימה בתא ה-ij, $k = 1 \ldots n$

N – סך הדגימות $N = a \cdot b \cdot n$

$\overline{y_{ij}} = \frac{1}{n} \sum_{k=1}^{n} y_{ijk}$ $\qquad$ $\overline{y_i} = \frac{1}{b} \sum_{j=1}^{b} \overline{y_{ij\cdot}}$ $\qquad$ $\bar{y} = \frac{1}{a} \sum_{i=1}^{a} \overline{y_{i\cdot\cdot}}$

$$SS_{Tot} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} \left(y_{ijk} - \overline{y_{\cdots}}\right)^2 =$$

$$\underbrace{nb \sum_{i=1}^{a} (y_{i\cdot\cdot} - \overline{y_{\cdots}})^2}_{SS_A} + \underbrace{n \sum_{i=1}^{a} \sum_{j=1}^{b} (y_{ij\cdot} - \overline{y_{i\cdot\cdot}})^2}_{SS_{B/A}} + \underbrace{\sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} (y_{ijk} - \overline{y_{ij\cdot}})^2}_{SS_e}$$

נוסחאות מקוצרות ודרגות חופש:

$SS_A = nb\sum_{i=1}^{a}\bar{y}_{i..}^2 - N\bar{y}_{...}^2$ $\qquad df = a-1$

$SS_{B/A} = n\sum_{i=1}^{a}\sum_{j=1}^{b}\bar{y}_{ij.}^2 - nb\sum_{i=1}^{a}\bar{y}_{i..}^2$ $\qquad df = ab-a = a(b-1)$

$SS_e = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}y_{ijk}^2 - n\sum_{i=1}^{a}\sum_{j=1}^{b}\bar{y}_{ij.}^2$ $\qquad df = (n-1)ab$

אמדים לשונויות:

$E(MS_A) = \sigma^2 + nb\frac{\sum_{i=1}^{a}\tau_i^2}{a-1}$ $\qquad E\left(MS_{B/A}\right) = \sigma^2 + n\frac{\sum_{i=1}^{a}\sum_{j=1}^{b}\beta_{ij}^2}{a(b-1)}$ $\qquad E(MS_e) = \sigma^2$

לוח ניתוח שונות

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_{emp}
A Treatment	SS_A	a-1	$MS_A = \frac{SS_A}{a-1}$	$\frac{MS_A}{MS_e}$
B Treatment	$SS_{B\|A}$	a(b-1)	$MS_{B\|A} = \frac{SS_{B\|A}}{a(b-1)}$	$\frac{MS_{B\|A}}{MS_e}$
Error	SS_e	ab(n-1)	$MS_e = \frac{SS_e}{ab(n-1)}$	
Total	SS_{Tot}	abn-1		

השערות האפס:

1. הגורם הראשי לא משפיע:

$$\begin{cases} H_o: \tau_i = 0 \\ H_1: else \end{cases} \Rightarrow \begin{cases} H_o: E(MS_A)/E(MS_e) = 1 \\ H_1: E(MS_A)/E(MS_e) > 1 \end{cases}$$

$$F_{emp} = \left.\frac{MS_A}{MS_e}\right|_{H_0} \sim F(a-1,(n-1)ab) \qquad F_{cr} = F_\alpha(a-1,(n-1)ab)$$

2. הגורם המשני לא משפיע:

$$\begin{cases} H_o: \beta_{ij} = 0 & \forall i,j \\ H_1: else \end{cases} \Rightarrow \begin{cases} H_o: E(MS_{B|A})/E(MS_e) = 1 \\ H_1: E(MS_{B|A})/E(MS_e) > 1 \end{cases}$$

$$F_{emp} = \left.\frac{MS_{B|A}}{MS_e}\right|_{H_0} \sim F(a(b-1),(n-1)ab) \qquad F_{cr} = F_\alpha(a(b-1),(n-1)ab)$$

אם הצלחנו לדחות את השערת האפס השנייה ניתן לבדוק לכל רמה של הגורם הראשי בנפרד האם הגורם המשני משפיע.

3. השערה - הרמות של הגורם המשני זהות עבור רמה i של הגורם הראשי:

סך השונות הנובעת מהרמות של הגורם המשני בתוך רמה i של הגורם הראשי - $SS_{B/i}$

$SS_{B/i} = n\sum_{j=1}^{b}(\overline{y_{ij\cdot}} - \overline{y_{i\cdot\cdot}})^2$ $\qquad$ $E(MS_{B/i}) = \sigma^2 + n\frac{\sum_{j=1}^{b}\beta_{ij}^2}{b-1}$

$F_{emp} = \frac{MS_{B/i}}{MS_e}\Big|_{H_0} \sim F(b-1,(n-1)ab)$

<u>אמדים ורב"סים:</u>

תוחלת כוללת: $\hat{\mu} = \bar{y}_{\dots}$

סטיית רמה i של הגורם הראשי: $\widehat{\tau_i} = \bar{y}_{i\cdot\cdot} - \bar{y}_{\dots}$

סטיית רמה j של הגורם המשני: $\widehat{\beta_{ij}} = \bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot}$

אמד לשונות הרעש: $\hat{\sigma}^2 = MS_e$

רב"ס לתוחלת: $\mu \in \bar{y}_{\dots} \pm t_{\frac{\alpha}{2}}[(n-1)ab]\sqrt{\frac{MS_e}{N}}$

רב"ס לתוחלת של רמה i של הגורם הראשי: $\mu + \tau_i \in \bar{y}_{i\cdot\cdot} \pm t_{\frac{\alpha}{2}}[(n-1)ab]\sqrt{\frac{MS_e}{nb}}$

רב"ס לתוחלת של רמה j של הגורם המשני בתוך רמה i של הגורם הראשוני:

$$\mu + \tau_i + \beta_{ij} \in \bar{y}_{ij\cdot} \pm t_{\frac{\alpha}{2}}[(n-1)ab]\sqrt{\frac{MS_e}{n}}$$

רב"ס לשונות הרעש: $\frac{ab(n-1)MS_e}{\chi^2_{\frac{\alpha}{2}}(n-1)ab} \le \sigma^2 \le \frac{ab(n-1)MS_e}{\chi^2_{1-\frac{\alpha}{2}}(n-1)ab}$

<u>תרגיל 3</u>

במחקר שנערך בבית-חולים, נבדקת יעילותן של 3 תרופות שונות המיועדות להורדת רמת הכולסטרול בדם אצל נשים. התרופות נרכשו מ-6 חברות שונות. מכיוון שהתרופות מיוצרות בחברות שונות, מעוניינים בבית החולים לבדוק בנוסף האם יש חברות אשר מייצרות תרופות טובות יותר.

להלן טבלה המסכמת את רמות הכולסטרול שנמדדו אצל החולות לאחר נטילת התרופות:

תרופה 3		תרופה 2		תרופה 1	
חברה 6	**חברה 5**	**חברה 4**	**חברה 3**	**חברה 2**	**חברה 1**
112	131	141	150	161	170
124	128	135	132	156	165
117	126	161	147	144	178
128	107	138	167	118	151
134	96	110	122	180	174
102	114	123	115	147	162

א. הסבר מדוע לא ניתן לנתח את הנתונים באמצעות מודל ניתוח שונות דו-כיווני מצטלב. באיזו סיטואציה ניתן היה לנתח את הנתונים באמצעות מודל ניתוח שונות דו כיווני מצטלב?

ב. ברמת מובהקות 0.01, האם ניתן לומר כי יש הבדל בין התרופות?

ג. ברמת מובהקות 0.01, האם ניתן לומר כי יש הבדל בין החברות?

ד. איזו תרופה תמליץ לבית החולים לרכוש? מאיזו חברה ?

<u>פתרון:</u>

א. לא ניתן לנתח את הנתונים באמצעות מודל ניתוח שונות דו-כיווני מצטלב מכיוון שעבור כל רמה של הגורם הראשוני יש ברשותנו רמות שונות של הגורם המשני. במודל ניתוח שונות דו כיווני מצטלב נשתמש כאשר בכל רמה של הגורם הראשוני יהיו ברשותנו רמות זהות של הגורם המשני.

ב. טבלת הממוצעים:

תרופה 1		תרופה 2		תרופה 3	
חברה 1	חברה 2	חברה 3	חברה 4	חברה 5	חברה 6
166.67	151.00	138.83	134.67	117.00	119.50
158.83		136.75		118.25	
137.94					

טבלת ריבועי הממוצעים:

תרופה 1		תרופה 2		תרופה 3	
חברה 1	חברה 2	חברה 3	חברה 4	חברה 5	חברה 6
27777.78	22801.00	19274.69	18135.11	13689.00	14280.25
25228.03		18700.56		13983.06	
19028.67					

$$SS_A = nb\sum_{i=1}^{a}\bar{y}_i^2 - N\bar{y}^2 = 6\cdot 2\cdot 57911.65 - 36\cdot 19028.67 = 9907.72$$

$$SS_{B|A} = n\sum_{i=1}^{a}\sum_{j=1}^{b}\bar{y}_{ij}^2 - nb\sum_{i=1}^{a}\bar{y}_i^2 = 6\cdot 115957.83 - 12\cdot 57911.65 = 807.18$$

$$SS_e = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}^2 - n\sum_{i=1}^{a}\sum_{j=1}^{b}\bar{y}_{ij}^2 = 757$$

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_{emp}
A Treatment	9907.72	2	4953.86	19.61
B Treatment	807.18	3	269.06	1.06
Error	7575	30	252.5	
Total	18289.9	35		

ב. ברמת מובהקות 0.01, ניתן לומר כי לתרופות השפעה שונה על החולות.

ג. ברמת מובהקות 0.01, לא ניתן לומר כי למוצרי החברות השונות השפעה שונה.

$$F_{0.01}(3,30) = 4.51 \quad F_{0.01}(2,30) = 5.39$$

ד. נמליץ לבית החולים לרכוש את תרופה 3, אין העדפה לגבי החברה.