การวิเคราะห์ความแปรปรวน

(Analysis of Variance)

การวิเคราะห์ความแปรปรวนใช้สำหรับทดสอบความแตกต่างระหว่างค่าเฉลี่ยของกลุ่มตัวอย่างตั้งแต่ 3 กลุ่มขึ้นไป (ใช้สำหรับกลุ่มตัวอย่าง 2 กลุ่มก็ได้ แต่กลุ่มตัวอย่าง 2 กลุ่ม ใช้ t-test สะดวกกว่า) เหตุที่ไม่ใช้ t-test กับกลุ่มตัวอย่างตั้งแต่ 3 กลุ่ม ขึ้นไป เนื่องจากใช้ t-test ทดสอบได้ทีละคู่เท่านั้น ถ้ามีกลุ่มตัวอย่างมากกว่า 2 กลุ่ม ต้องทำ t-test หลายครั้ง เช่น ถ้ามีกลุ่มตัวอย่าง 3 กลุ่ม แล้วทดสอบด้วย t-test จะต้องทำ t-test ถึง 3 ครั้ง คือ เปรียบเทียบกลุ่ม 1 กับ กลุ่ม 2 , กลุ่ม 1 กับ กลุ่ม 3 และ กลุ่ม 2 กับ กลุ่ม 3 ซึ่งเสียเวลามาก และทำให้มีความคลาดเคลื่อนแบบที่ 1 เพิ่มขึ้น กล่าวคือ ความน่าจะเป็นที่จะได้ผลการทดสอบแตกต่างกันอย่างมีนัยสำคัญโดยความบังเอิญมีมากขึ้น เพื่อขจัดปัญหาดังที่กล่าวมานี้ จึงจำเป็นต้องใช้การวิเคราะห์ความแปรปรวน ซึ่งสามารถทำการทดสอบความแตกต่างระหว่างค่าเฉลี่ยของกลุ่มตัวอย่างได้ทีละหลายกลุ่มในครั้งเดียวกัน

การทดสอบค่าเฉลี่ยมากกว่า 2 ค่าขึ้นไปจึงใช้เทคนิคการวิเคราะห์ความแปรปรวน ซึ่งเรียกย่อ ๆ ว่า แอนโนวา (ANOVA)

1. ความเข้าใจพื้นฐานเกี่ยวกับการวิเคราะห์ความแปรปรวน

1) ในการเปรียบเทียบค่าเฉลี่ยระหว่างกลุ่มหลาย ๆ กลุ่ม จะมีความแปรปรวนที่ต้องคำนวณอยู่ 2 ตัว คือ ความแปรปรวนระหว่างกลุ่ม และความแปรปรวนภายในกลุ่ม
   - ก. **ความแปรปรวนระหว่างกลุ่ม (Between-groups variance)** แสดงขนาดของความแตกต่างระหว่างค่าเฉลี่ยของกลุ่มต่าง ๆ ถ้าระหว่างกลุ่มมีค่าเฉลี่ยแตกต่างกันมาก ค่าความแปรปรวนระหว่างกลุ่มจะมีค่ามากด้วย
   - ข. **ความแปรปรวนภายในกลุ่ม (Within-groups variance)** แสดงการกระจายของคะแนนแต่ละตัวภายในแต่ละกลุ่มว่ามีการกระจายมากหรือน้อย ค่าที่คำนวณได้เรียกว่า**ความคลาดเคลื่อน**
2) สูตรที่ใช้ในการทดสอบคือ F-ratio โดยเอาความแปรปรวนระหว่างกลุ่มเป็นตัวตั้งหารด้วยความแปรปรวนภายในกลุ่ม แล้วเปรียบเทียบค่า F ที่คำนวณได้กับค่า F ในตารางค่าวิกฤต F (Critical values of F) เพื่อสรุปผลการวิเคราะห์ข้อมูล

2. ข้อตกลงเบื้องต้นในการวิเคราะห์ความแปรปรวน

1) ข้อมูลที่จะนำมาวิเคราะห์ต้องอยู่ในมาตราอันตรภาคหรืออัตราส่วน
2) กลุ่มตัวอย่างแต่ละกลุ่มต้องสุ่มมาจากประชากรที่มีการแจกแจงเป็นโค้งปกติ
3) กลุ่มตัวอย่างแต่ละกลุ่มต้องเป็นอิสระจากกัน
4) กลุ่มตัวอย่างแต่ละกลุ่มต้องได้มาจากประชากรที่มีความแปรปรวนเท่ากัน

3. ผลบวกของคะแนนเบี่ยงเบนยกกำลังสอง (The sum of squares)

ในการวิเคราะห์ความแปรปรวน ต้องหา sum of squares ทั้งหมด 3 ตัว คือ

1) **Total sum of squares (SS_T)** ซึ่งนำไปใช้หาความแปรปรวนรวม

(Mean square total: MS_T)

2) **Sum of squares between-groups (SS_B)** ซึ่งนำไปใช้หาความแปรปรวนระหว่างกลุ่ม

(Mean square between-groups: MS_B)

3) **Sum of squares within-groups (SS_W)** ซึ่งนำไปใช้หาความแปรปรวนภายในกลุ่ม

(Mean square within-groups: MS_W)

4. การวิเคราะห์ความแปรปรวนแบบทางเดียว (One-way analysis of variance) หรือ การวิเคราะห์ความแปรปรวนแบบมี 1 ตัวประกอบ

การวิเคราะห์ความแปรปรวนแบบทางเดียว เป็นการวิเคราะห์เพื่อตรวจสอบว่า ตัวแปรอิสระ 1 ตัว ซึ่งแบ่งออกเป็น k ประเภท จะให้ผลต่อตัวแปรตามแตกต่างกันหรือไม่ โดยมีกลุ่มตัวอย่าง k กลุ่มซึ่งจำนวนสมาชิกในแต่ละกลุ่มควรจะเท่ากันหรือใกล้เคียงกัน เช่น ต้องการเปรียบเทียบวิธีการจัดการเรียนรู้ 3 วิธี ว่าจะส่งผลให้ผลการเรียนรู้ของผู้เรียนในแต่ละวิธี แตกต่างกันหรือไม่

ก. ลักษณะข้อมูลที่ใช้ในการวิเคราะห์ความแปรปรวนแบบทางเดียว

ตาราง 1 แสดงข้อมูลของกลุ่มตัวอย่าง k กลุ่ม ที่เป็นผลจากตัวแปรอิสระ 1 ตัว แบ่งเป็น k ประเภท

กลุ่ม 1	กลุ่ม 2	กลุ่ม 3	…	กลุ่ม j	…	กลุ่ม k
X_{11}	X_{12}	X_{13}		X_{1j}		X_{1k}
X_{21}	X_{22}	X_{23}		X_{2j}		X_{2k}
⋮	⋮	⋮		⋮		⋮
$X_{n_1 1}$	$X_{n_2 2}$	$X_{n_3 3}$		$X_{n_j j}$		$X_{n_k k}$
T_1	T_2	T_3		T_j		T_k

X_{11} แทน ข้อมูลตัวที่ 1 ใน กลุ่มที่ 1

X_{21} แทน ข้อมูลตัวที่ 2 ใน กลุ่มที่ 1

X_{ij} แทน ข้อมูลตัวที่ i ใน กลุ่มที่ j

$X_{n_3 3}$ แทน ข้อมูลตัวสุดท้ายในกลุ่มที่ 3

T_1 แทน ผลรวมข้อมูลทั้ง n ตัว ในกลุ่มที่ 1

T_2 แทน ผลรวมข้อมูลทั้ง n ตัว ในกลุ่มที่ 2

T_3 แทน ผลรวมข้อมูลทั้ง n ตัว ในกลุ่มที่ 3

T_j แทน ผลรวมข้อมูลทั้ง n ตัว ในกลุ่มที่ j

ตาราง 2 ตัวอย่างคะแนน 3 กลุ่ม ที่เป็นผลจากตัวแปรอิสระ 1 ตัว คือวิธีสอน ซึ่งแบ่งเป็น 3 แบบ

กลุ่ม 1 ($n_1 = 9$)	กลุ่ม 2 ($n_2 = 11$)	กลุ่ม 3 ($n_3 = 10$)
7	11	17
6	12	6
15	9	18
8	7	14
15	17	16
6	12	8
12	19	7
8	11	15
16	13	9
	8	11
	10	
93	129	121

n_1, n_2 และ n_3 เป็นจำนวนข้อมูลในกลุ่มที่ 1, กลุ่มที่ 2 และ กลุ่มที่ 3 ตามลำดับ

ข. สูตรการทดสอบ

$$F = \frac{MS_B}{MS_W}$$

เมื่อ F แทน ค่าสถิติเอฟ

MS_B แทน ความแปรปรวนระหว่างกลุ่ม (Mean square between-groups)

MS_W แทน ความแปรปรวนภายในกลุ่ม (Mean square within-groups)

หา MS_B และ MS_W ได้จากสูตรในตาราง 3

ตาราง 3 แสดงสูตรการวิเคราะห์ความแปรปรวนแบบทางเดียว (มีตัวแปรอิสระ 1 ตัว)

Source of variation	df	Sum of square (SS)	Mean square (MS)	F
Between-groups	$k-1$	$SS_B = \sum_{j=1}^{k}\left(\frac{T_j^2}{n_j}\right) - \frac{T^2}{N}$	$MS_B = \frac{SS_B}{k-1}$	$F = \frac{MS_B}{MS_W}$
Within-groups	$N-k$	$SS_W = SS_T - SS_B$	$MS_W = \frac{SS_W}{N-k}$	
Total	$N-1$	$SS_T = \sum_{j=1}^{k}\sum_{i=1}^{n_j} X_{ij}^2 - \frac{T^2}{N}$		

เมื่อ F แทน ค่าสถิติเอฟ

MS_B แทน ความแปรปรวนระหว่างกลุ่ม (Mean square between-groups)

MS_W แทน ความแปรปรวนภายในกลุ่ม (Mean square within-groups)

SS_B แทน ผลบวกของคะแนนเบี่ยงเบนยกกำลังสองระหว่างกลุ่ม (Sum of squares between groups)

SS_W แทน ผลบวกของคะแนนเบี่ยงเบนยกกำลังสองภายในกลุ่ม (Sum of squares within groups)

SS_T แทน ผลบวกของคะแนนเบี่ยงเบนยกกำลังสองทั้งหมด (Total Sum of squares)

n_j แทน จำนวนข้อมูลในแต่ละกลุ่มตัวอย่าง

k แทน จำนวนกลุ่มตัวอย่าง

N แทน จำนวนข้อมูลทั้งหมด

T_j แทน ผลรวมของคะแนนในแต่ละกลุ่มตัวอย่าง

T แทน ผลรวมของคะแนนทั้งหมด

$\sum_{j=1}^{k}\sum_{i=1}^{n_j} X_{ij}^2$ แทน ผลรวมทั้งหมดของคะแนนแต่ละตัวยกกำลังสอง

ค. ตัวอย่างขั้นตอนการทดสอบ (ใช้ตัวอย่างข้อมูลในตาราง 2)

1) ตั้งสมมติฐาน $H_0 : \mu_1 = \mu_2 = \mu_3$

$H_1 :$ อย่างน้อยที่สุดมีค่าเฉลี่ยของกลุ่มตัวอย่าง 2 กลุ่มไม่เท่ากัน

2) กำหนดระดับนัยสำคัญทางสถิติ : $\alpha = 0.05$

3) คำนวณค่าสถิติ F จากสูตร

ตัวอย่าง : นำข้อมูลจากตาราง 2 มาคำนวณตามสูตรในตาราง 3 ดังแสดงในตาราง 4

ตาราง 4 แสดงตัวอย่างการจัดกระทำข้อมูลตามสูตรในตาราง 3

กลุ่ม 1 (n_1= 9)		กลุ่ม 2 (n_2= 11)		กลุ่ม 3 (n_3 = 10)	
X_{i1}	X_{i1}^2	X_{i2}	X_{i2}^2	X_{i3}	X_{i3}^2
7	49	11	121	17	289
6	36	12	144	6	36
15	225	9	81	18	324
8	64	7	49	14	196
15	225	17	289	16	256
6	36	12	144	8	64
12	144	19	361	7	49
8	64	11	121	15	225
16	256	13	169	9	81
		8	64	11	121
		10	100		
T_1 = 93	$\sum_{i=1}^{n_1} X_{i1}^2$ = 1099	T_2 = 129	$\sum_{i=1}^{n_2} X_{i2}^2$ = 1643	T_3 = 121	$\sum_{i=1}^{n_3} X_{i3}^2$ = 1641
T_1^2 = 8649		T_2^2=16641		T_3^2=14641	

N = 9 + 11 + 10 = 30 T = 93 + 129 + 121 = 343 T^2 = 117649

$$SS_B = \sum_{j=1}^{k}\left(\frac{T_j^2}{n_j}\right) - \frac{T^2}{N} = \left(\frac{T_1^2}{n_1} + \frac{T_2^2}{n_2} + \frac{T_3^2}{n_3}\right) - \frac{T^2}{N}$$

$$= \frac{8649}{9} + \frac{16641}{11} + \frac{14641}{10} - \frac{117649}{30} = 961 + 1512.818 + 1464.1 - 3921.633$$

$$= 16.285$$

$$SS_T = \sum_{j=1}^{k}\sum_{i=1}^{n_j} X_{ij}^2 - \frac{T^2}{N} = 1099 + 1643 + 1641 - \frac{117649}{30}$$

$$= 4383 - 3921.633 = 461.367$$

$$SS_W = SS_T - SS_B = 461.367 - 16.285 = 445.082$$

$$MS_B = \frac{SS_B}{k-1} = \frac{16.285}{2} = 8.1425 \qquad MS_W = \frac{SS_W}{N-k} = \frac{445.082}{27} = 16.4845$$

$$F = \frac{MS_B}{MS_W} = \frac{8.1425}{16.4845} = 0.4939$$

4) เปิดตารางค่าวิกฤต F (Critical values of F)

df ของตัวเศษ คือ df ของ MS_B และ df ของตัวส่วน คือ df ของ MS_W

ตามสูตรในตาราง 3 df ของ MS_B คือ $k - 1 = 3 - 1 = 2 = df_1$

df ของ MS_W คือ $N - k = 30 - 3 = 27 = df_2$

df ของตัวเศษ คือค่า df_1 ตามแนวนอนในตารางค่าวิกฤต F (ตารางค่าวิกฤต F มีแนบท้ายหนังสือสถิติฯ)

df ของตัวส่วน คือค่า df_2 ตามแนวตั้งในตารางค่าวิกฤต F

df_1 (degrees of freedom ของตัวเศษ)

df_2 (degrees of freedom ของตัวส่วน)	DEGREES OF FREEDOM FOR DENOMINATOR	DEGREES OF FREEDOM FOR NUMERATOR						
		1	2	…	30	…	500	∞
	1	161	200		250		254	254
		4052	4999		6258		6361	6366
	. . .							
	27	4.21	3.35		1.94		1.74	1.73
		7.68	5.49		2.58		2.23	2.21
	. . .							
	∞	3.84	2.99		1.46		1.11	1.00
		6.64	4.60		1.69		1.15	1.00

ในแถวคู่แต่ละคู่ แถวบนเป็นค่าวิกฤต F ที่ $\alpha = 0.05$ แถวล่างเป็นค่าวิกฤต F ที่ $\alpha = 0.01$

ที่ $df_1=2$ พบ $df_2 = 27$ นั้น แสดงค่าวิกฤต $F = 3.35$ ที่ $\alpha = 0.05$ และค่าวิกฤต $F= 5.49$ ที่ $\alpha = 0.01$

นำค่า F ที่คำนวณได้ในข้อ 3 คือ $F = 0.49$ มาเปรียบเทียบกับค่าวิกฤต $F = 3.35$ ที่ $\alpha = 0.05$ พบว่า ค่า F ที่คำนวณได้ < ค่าวิกฤต F ดังนั้น จึงยอมรับ $H_0 : \mu_1 = \mu_2 = \mu_3$

เสนอผลการวิเคราะห์ด้วยตารางได้ดังนี้

แหล่งความแปรปรวน	df	SS	MS	F
ระหว่างกลุ่ม	2	16.285	8.1425	0.49
ภายในกลุ่ม	27	445.082	16.4845	
รวมทั้งหมด	29	461.367		

$F_{.05\,(2,\,27)} = 3.35$

สรุปได้ว่า วิธีสอนทั้ง 3 แบบให้ผลแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ หรือกล่าวอีกอย่างหนึ่งว่า วิธีสอนทั้ง 3 แบบให้ผลไม่แตกต่างกัน

ในกรณีที่ ค่า F จากการคำนวณ $\geq$ ค่าวิกฤต F จะเป็นการปฏิเสธ H_0 ยอมรับ H_1 (Reject H_0 Accept H_1) แสดงว่า มีค่าเฉลี่ยอย่างน้อย 1 คู่ แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ณ ระดับ α ที่กำหนด ต้องทำการทดสอบต่อไปว่าค่าเฉลี่ยคู่ใดบ้างที่แตกต่างกัน และแตกต่างกันอย่างไร โดยใช้การเปรียบเทียบพหุคูณ (Multiple comparison test) ตามวิธีของเชฟเฟ่ (Scheffe's method) การทดสอบ HSD ของทูกีย์ (Tukey's HSD test) หรือ วิธีของนิวแมนคูลส์ (Newman Keuls method) ซึ่งเป็น "post hoc" หมายถึง วิธีการที่ตามมาหลังการวิเคราะห์ความแปรปรวน

ที่มา : http://www.krupai.net/stat/stat.htm