

บทที่ 8

การวิเคราะห์ความแปรปรวน (Analysis of Variance)

การทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ยของประชากรสองกลุ่มสามารถใช้สถิติ Z หรือ T มาช่วยในการทดสอบ แต่ในกรณีที่ต้องการทดสอบสมมติฐานเกี่ยวกับความแตกต่างระหว่างค่าเฉลี่ยของประชากรตั้งแต่ 3 กลุ่มขึ้นไปเราไม่สามารถใช้สถิติทดสอบ Z ได้ แต่สถิติทดสอบ T สามารถใช้ได้โดยมีข้อจำกัด คือ สามารถใช้ทดสอบความแตกต่างระหว่างค่าเฉลี่ยเพียง 2 กลุ่มเท่านั้น สมมติว่าเราต้องการทดสอบค่าเฉลี่ยของประชากร k กลุ่ม โดยที่ $k > 2$ คือต้องการทดสอบ $H_0: \mu_1 = \dots = \mu_k$ ถ้าเราเลือกใช้สถิติทดสอบ T จะต้องทำการทดสอบครั้งละ 2 กลุ่ม นั่นคือจำนวนครั้งที่ต้องการทำการทดสอบคือ ${}^kC_2 = \frac{k!}{(k-2)!2!}$ ครั้ง ตัวอย่างเช่น จำนวนครั้งที่ต้องการทำการทดสอบคือ 3 ครั้ง โดยครั้งแรกทดสอบค่าเฉลี่ยของประชากรกลุ่มที่ 1 กับค่าเฉลี่ยของประชากรกลุ่มที่ 2 ($H_0: \mu_1 = \mu_2$) ครั้งที่สองทดสอบค่าเฉลี่ยของประชากรกลุ่มที่ 1 กับค่าเฉลี่ยของประชากรกลุ่มที่ 3 ($H_0: \mu_1 = \mu_3$) และครั้งสุดท้ายทดสอบค่าเฉลี่ยของประชากรกลุ่มที่ 2 กับค่าเฉลี่ยของประชากรกลุ่มที่ 3 ($H_0: \mu_2 = \mu_3$) นอกจากนี้ทุกครั้งที่ทำการทดสอบจะเกิดความคลาดเคลื่อนของการทดสอบขึ้น จึงทำให้เกิดความคลาดเคลื่อนมากเกินความจำเป็น อีกทั้งยังเสียเวลาในการทดสอบโดยไม่จำเป็น เพื่อแก้ปัญหาดังที่กล่าวมาในทางปฏิบัติจึงใช้เทคนิคการวิเคราะห์ความแปรปรวนมาใช้สำหรับทดสอบความแตกต่างระหว่างค่าเฉลี่ยของประชากรตั้งแต่ 3 กลุ่มขึ้นไปในครั้งเดียวกัน

การวิเคราะห์ความแปรปรวน (Analysis of Variance: ANOVA) เป็นการทดสอบความแตกต่างระหว่างค่าเฉลี่ยของประชากรตั้งแต่ 3 กลุ่มขึ้นไป ซึ่งนิยมเรียกย่อ ๆ ว่า ANOVA (อ่านว่า อะโนวา) โดยผู้ที่คิดค้นและเสนอวิธีการวิเคราะห์ความแปรปรวนก็คือ Ronald Fisher นักชีววิทยาและสถิติชาวอังกฤษ

8.1 ความเข้าใจเบื้องต้นที่เกี่ยวกับการวิเคราะห์ความแปรปรวน

1. การเปรียบเทียบค่าเฉลี่ยระหว่างกลุ่มหลาย ๆ กลุ่ม จะมีความแปรปรวนที่ต้องคำนึงถึงอยู่ 2 แบบ คือ ความแปรปรวนระหว่างกลุ่ม และความแปรปรวนภายในกลุ่ม

ก. **ความแปรปรวนระหว่างกลุ่ม (Between-groups variance)** จะแสดงขนาดของความแตกต่างระหว่างค่าเฉลี่ยของกลุ่มต่าง ๆ ถ้าระหว่างกลุ่มมีค่าเฉลี่ยแตกต่างกันมาก ค่าความแปรปรวนระหว่างกลุ่มจะมีค่ามากด้วย

ข. **ความแปรปรวนภายในกลุ่ม (Within-groups variance)** จะแสดงการกระจายของค่าสังเกตแต่ละค่าภายในแต่ละกลุ่มว่ามีการกระจายมากหรือน้อย บางครั้งเรียกว่า **ความคลาดเคลื่อน**

2) สถิติที่ใช้ทดสอบ คือ สถิติเอฟ (F-Statistics)

3) ข้อมูลที่จะนำมาวิเคราะห์ต้องอยู่ในมาตราอันดับหรืออัตราส่วน

4) กลุ่มตัวอย่างแต่ละกลุ่มต้องสุ่มมาจากระชากรที่มีการแจกแจงแบบปกติที่มีความแปรปรวนเท่ากัน และเป็นอิสระจากกัน

2. การคำนวณความแปรปรวน

ความแปรปรวนรวมทั้งหมด (Total sum of squares: SST) ในการวิเคราะห์ความแปรปรวนสามารถคำนวณจากความแปรปรวนระหว่างกลุ่ม (Sum of squares between-groups : SST_{tr}) รวมกับความแปรปรวนภายในกลุ่ม (Sum of squares within-groups : SSE) (หรือความแปรปรวนของความคลาดเคลื่อน) นั่นคือ

$$SST = SST_{tr} + SSE$$

สำหรับบทเรียนนี้จะกล่าวถึงรายละเอียดของการวิเคราะห์ความแปรปรวนทางเดียวดังนี้

3. การวิเคราะห์ความแปรปรวนแบบทางเดียว (One-way Analysis of Variance)

การวิเคราะห์ความแปรปรวนแบบทางเดียว เป็นการวิเคราะห์เพื่อตรวจสอบว่าตัวแปรอิสระ 1 ตัว ซึ่งแบ่งออกเป็น k กลุ่ม จะให้ผลต่อตัวแปรตามแตกต่างกันหรือไม่ โดยที่กลุ่มตัวอย่าง k กลุ่มอาจมีจำนวนค่าสังเกตในแต่ละกลุ่มเท่ากันหรือไม่เท่ากันก็ได้

ก. ลักษณะของข้อมูลที่ใช้ในการวิเคราะห์ความแปรปรวนแบบทางเดียว

ตารางที่ 8.1 ข้อมูลของตัวแปรอิสระ 1 ตัว แบ่งเป็น k กลุ่ม

กลุ่ม 1	กลุ่ม 2	...	กลุ่ม k
X_{11}	X_{21}	...	X_{k1}
X_{12}	X_{22}	...	X_{k2}
\vdots	\vdots	\vdots	\vdots
X_{1n_1}	X_{2n_2}	\vdots	X_{kn_k}
$X_{1.}$	$X_{2.}$...	$X_{k.}$

โดยที่

X_{ij} คือ ค่าสังเกตของกลุ่มที่ i ค่าที่ j

$X_{i.}$ คือ ผลรวมค่าสังเกตทั้งหมดในกลุ่มที่ i

n_1, n_2, \dots, n_k คือ จำนวนค่าสังเกตในกลุ่มที่ 1, 2, ..., k ตามลำดับ

$$\text{และ } n = n_1 + n_2 + \dots + n_k$$

ข. สมมติฐานที่ใช้ในการทดสอบ

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k$$

$$H_1 : \mu_i \neq \mu_j \text{ อย่างน้อย 1 คู่ โดยที่ } i \neq j$$

ค. สถิติทดสอบ

$$F = \frac{MSTrt}{MSE}$$

โดยที่ F แทน ค่าสถิติเอฟที่คำนวณได้

$MSTrt$ แทน ความแปรปรวนระหว่างกลุ่ม (Mean square between-groups)

MSE แทน ความแปรปรวนภายในกลุ่ม (Mean square within-groups)

ผลการวิเคราะห์ความแปรปรวนสามารถสร้างเป็นตารางวิเคราะห์ความแปรปรวนทางเดียวได้ดัง
ตารางที่ 8.2

ตารางที่ 8.2 สูตรการวิเคราะห์ความแปรปรวนแบบทางเดียว (One way ANOVA)

Source of variation	df	Sum of square (SS)	Mean square (MS)	F
Between-groups	k-1	$SSTrt = \sum_{i=1}^k \left(\frac{X_{i.}^2}{n_i} \right) - \frac{X_{..}^2}{n}$	$MSTrt = \frac{SSTrt}{k-1}$	$F = \frac{MSTrt}{MSE}$
Within-groups	n-k	$SSE = SST - SSTrt$	$MSE = \frac{SSE}{n-k}$	
Total	n-1	$SST = \sum_{i=1}^k \sum_{j=1}^{n_i} X_{ij}^2 - \frac{X_{..}^2}{n}$		

โดยที่ F แทน ค่าสถิติเอฟที่คำนวณได้

$SSTrt$ แทน ผลบวกของกำลังสองระหว่างกลุ่ม (Sum of squares between groups)

SSE แทน ผลบวกของกำลังสองภายในกลุ่ม (Sum of squares within groups)

SST แทน ผลบวกของยกกำลังสองทั้งหมด (Total Sum of squares)

n_i แทน จำนวนข้อมูลในกลุ่มที่ i

k แทน จำนวนกลุ่มตัวอย่างทั้งหมด

$n = \sum_{i=1}^k n_i$ แทน จำนวนค่าสังเกตทั้งหมด

$X_{i.}$ แทน ผลรวมของค่าสังเกตในกลุ่มที่ i

$X_{..}$ แทน ผลรวมของค่าสังเกตทั้งหมด

$\sum_{i=1}^k \sum_{j=1}^{n_i} X_{ij}^2$ แทน ผลรวมค่าสังเกตแต่ละค่ายกกำลังสอง

ตัวอย่างที่ 8.3 กำหนดให้ข้อมูลคะแนนสอบ (คะแนนเต็ม 30 คะแนน) ซึ่งได้จากวิธีสอน 3 วิธี ดังนี้

วิธีสอนที่ 1	วิธีสอนที่ 2	วิธีสอนที่ 3
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

จงทดสอบว่าคะแนนเฉลี่ยของคะแนนสอบของวิธีการสอนทั้ง 3 วิธีแตกต่างกันหรือไม่ ที่ระดับนัยสำคัญ 0.05

วิธีทำ

- 1) ตั้งสมมติฐาน

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

$$H_1 : \mu_i \neq \mu_j \text{ อย่างน้อย 1 คู่}$$

- 2) กำหนดระดับนัยสำคัญ $\alpha = 0.05$
- 3) คำนวณค่าสถิติ F

นำข้อมูลจากตารางข้างต้น มาคำนวณตามสูตรในตารางที่ 2 ดังแสดงรายละเอียดดังนี้

ตารางที่ 8.4 แสดงวิธีการคำนวณค่าต่าง ๆ ในการวิเคราะห์ความแปรปรวนทางเดียว

วิธีสอนที่ 1		วิธีสอนที่ 2		วิธีสอนที่ 3	
x_{1j}	x_{1j}^2	x_{2j}	x_{2j}^2	x_{3j}	x_{3j}^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		
$X_{1.} = 93$	$\sum_{j=1}^{n_1} X_{1j}^2 = 1099$	$X_{2.} = 129$	$\sum_{j=1}^{n_2} X_{2j}^2 = 1643$	$X_{3.} = 121$	$\sum_{j=1}^{n_3} X_{3j}^2 = 1641$

$$n = 9 + 11 + 10 = 30$$

$$X_{..} = 93 + 129 + 121 = 343$$

$$SST = \sum_{i=1}^k \sum_{j=1}^{n_i} X_{ij}^2 - \frac{X_{..}^2}{n} = 4383 - 3921.633 = 461.637$$

$$SSTrt = \sum_{i=1}^k \left(\frac{X_{i.}^2}{n_i} \right) - \frac{X_{..}^2}{n} = 961 + 1512.818 + 1464.1 - 3921.633 = 16.285$$

$$SSE = SST - SSTrt = 461.367 - 16.285 = 445.082$$

$$MSTrt = \frac{SSTrt}{k-1} = \frac{16.285}{2} = 8.1425$$

$$MSE = \frac{SSE}{n-k} = \frac{445.082}{27} = 16.4845$$

$$\text{จะได้ว่าค่าสถิติทดสอบ } F = \frac{MSTrt}{MSE} = \frac{8.1425}{16.4845} = 0.49$$

สามารถเสนอผลการวิเคราะห์ความแปรปรวนทางเดียวดังนี้

ตารางที่ 8.5 ผลการวิเคราะห์ความแปรปรวนทางเดียว

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

4) เปิดตารางค่าวิกฤต F (Critical values of F)

df_1 คือ df ของ MSTrt และ df_2 คือ df ของ MSE

ดังนั้นให้เปิดตารางเอฟ $df_1 = 2$ และ $df_2 = 27$ ที่ $\alpha = 0.05$

จะได้ค่าวิกฤต $F = 3.35$

จากนั้นนำค่า F ที่คำนวณได้ในข้อ 3) คือ $F = 0.49$ มาเปรียบเทียบกับค่าวิกฤต $F = 3.35$ ที่

$\alpha = 0.05$ พบว่า ค่า F ที่คำนวณได้ < ค่าวิกฤต F

ดังนั้น จึงยอมรับ $H_0 : \mu_1 = \mu_2 = \mu_3$

สรุปได้ว่า วิธีสอนทั้ง 3 แบบให้ผลไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ หรือกล่าวอีกนัยหนึ่งว่า วิธีสอนทั้ง 3 แบบให้ผลไม่แตกต่างกัน

หมายเหตุ: ในกรณีที่ ค่า F จากการคำนวณ \geq ค่าวิกฤต F จะเป็นการปฏิเสธ H_0 หรืออีกนัยหนึ่งเป็น ยอมรับ H_1 ซึ่งแสดงว่า มีค่าเฉลี่ยอย่างน้อย 1 คู่ แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ณ ระดับ α ที่กำหนด ต้องทำการทดสอบต่อไปว่าค่าเฉลี่ยคู่ใดบ้างที่แตกต่างกัน และแตกต่างกันอย่างไร โดยการใช้การเปรียบเทียบพหุคูณ (Multiple comparison test) ตามวิธีของเชฟเฟ (Scheffe's method) การทดสอบ HSD ของทูกีย์ (Tukey's HSD test) หรือ วิธีของนิวแมนคูลส์ (Newman Keuls method) ซึ่งเป็นวิธี "post hoc" หมายถึง วิธีการที่ตามมาหลังการวิเคราะห์ความแปรปรวน

ตัวอย่างที่ 8.4 ให้ตารางการวิเคราะห์ความแปรปรวนทางเดียวมีลักษณะดังนี้

แหล่งความแปรปรวน	df	SS	MS	F
ระหว่างกลุ่ม	3			
ภายในกลุ่ม		25.0		
รวมทั้งหมด	34	150.0		

จงตอบคำถามดังนี้

1. เติมตัวเลขในตารางวิเคราะห์ความแปรปรวนทางเดียวให้สมบูรณ์
2. จำนวนกลุ่มมีกี่กลุ่ม(k)
3. จำนวนข้อมูลทั้งหมด (n)
4. จงเขียนสมมติฐานว่าง และสมมติฐานทางเลือก
5. ถ้ากำหนดให้ $\alpha = 0.05$ จงหาค่าวิกฤต F
6. เขียนสรุปผลการทดสอบสมมติฐาน

วิธีทำ

1. ตารางวิเคราะห์ความแปรปรวนทางเดียวที่สมบูรณ์คือ

แหล่งความแปรปรวน	df	SS	MS	F
ระหว่างกลุ่ม	3	114	38.00	32.76
ภายในกลุ่ม	31	36.0	1.61	
รวมทั้งหมด	34	150.0		

2. จำนวนกลุ่มมีกี่กลุ่ม(k)
k = 4 หรือ 4 กลุ่ม (หาจาก k-1 = 3 ดังนั้น k=4)
3. จำนวนข้อมูลทั้งหมด (n)
n = 35 (หาจาก n-1 = 34 ดังนั้น n=35)
4. เขียนสมมติฐานว่าง และสมมติฐานทางเลือก ได้ดังนี้

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$

$$H_1 : \mu_i \neq \mu_j \text{ อย่างน้อย 1 คู่}$$

5. ถ้า $\alpha = 0.05$ จะได้ค่าวิกฤต F คือ $F_{0.05,(3,31)} = 2.9113$
6. เนื่องจาก $F = 32.76$ ซึ่งมากกว่าค่าวิกฤต 2.9113 ดังนั้นจึงสรุปว่าปฏิเสธสมมติฐานว่าง นั่นคือมีค่าเฉลี่ยอย่างน้อย 1 คู่ที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ จึงต้องทำการทดสอบต่อไปว่าค่าเฉลี่ยคู่ใดบ้างที่แตกต่างกัน โดยใช้การเปรียบเทียบพหุคูณ (Multiple comparison test)

แบบฝึกหัดท้ายบท

1. การวิเคราะห์ความแปรปรวน (ANOVA) คืออะไร จงอธิบายมาพอสังเขป

2. กรณีการวิเคราะห์ความแปรปรวนทางเดียว (one way ANOVA) มีความแปรปรวนอะไรบ้าง

3. ถ้ามีกลุ่มข้อมูล 5 กลุ่ม จงเขียนสมมติฐานว่าง และสมมติฐานทางเลือกของการวิเคราะห์ความแปรปรวนทางเดียว

4. สถิติทดสอบที่ใช้ในการวิเคราะห์ความแปรปรวนทางเดียวคืออะไร

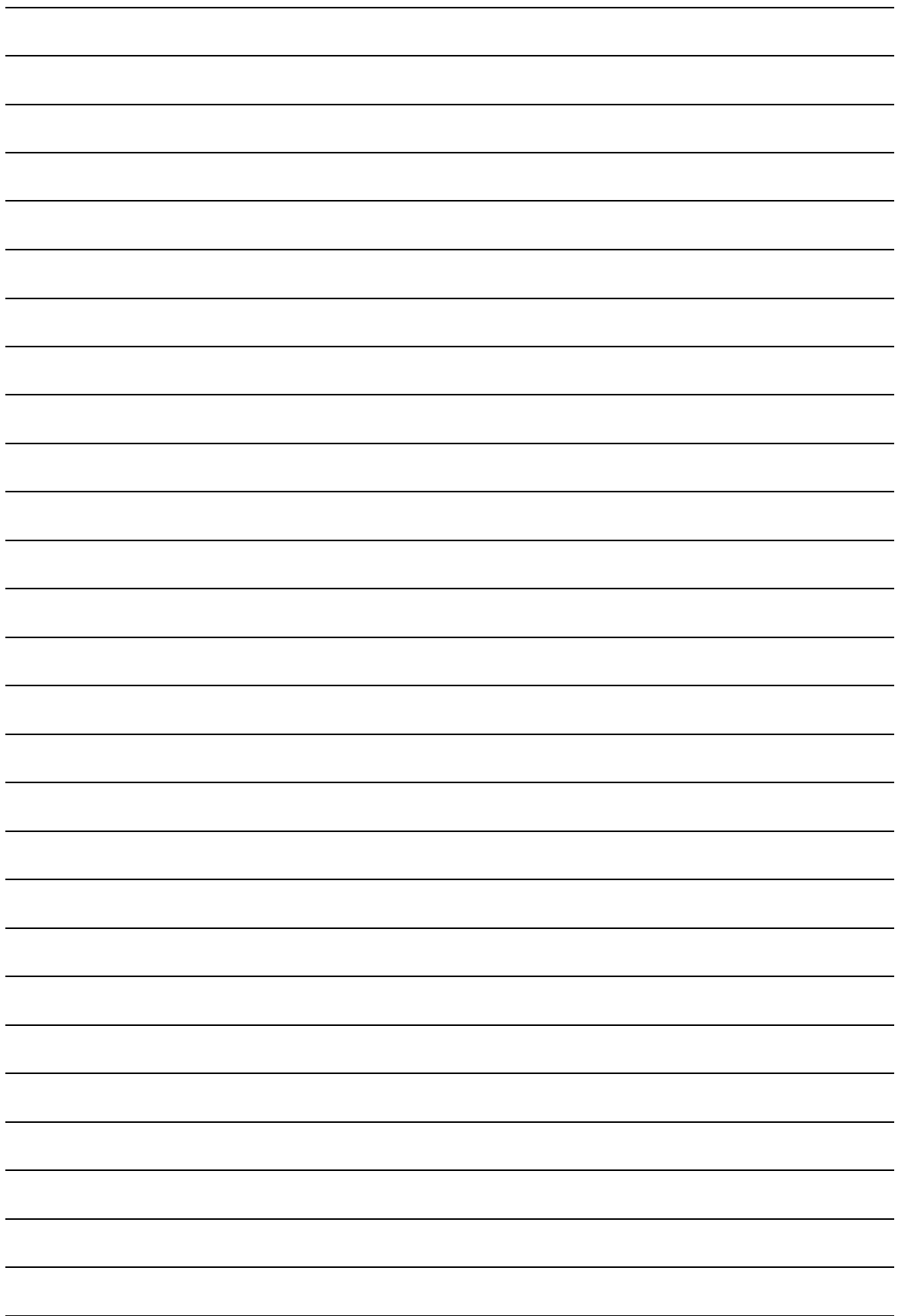
5. จงเขียนตารางการวิเคราะห์ความแปรปรวนทางเดียวพร้อมสูตรคำนวณ

This image shows a single sheet of white paper with horizontal ruling lines. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.

6. ในการวิเคราะห์ความแปรปรวนทางเดียว จะปฏิเสธสมมติฐานว่าง เมื่อใด จงอธิบายมาพอสังเขป

- | ยี่ห้อรถ | ระยะทางที่วิ่งได้ (กิโลเมตร) | | | | รวม |
|----------|------------------------------|----|----|----|-----|
| A | 21 | 26 | 25 | 20 | 92 |
| B | 23 | 26 | 25 | 18 | 92 |
| C | 35 | 38 | 35 | 32 | 140 |

[illegible]



8. ข้อมูลต่อไปนี้เป็นจำนวนผลผลิตที่ขำรุดจากการตรวจพบในแต่ละวันของเครื่องจักร 5 เครื่องในระยะเวลา 5 วัน จงทดสอบว่าประสิทธิภาพของเครื่องจักรทั้ง 5 เครื่องแตกต่างกันหรือไม่ ที่ระดับนัยสำคัญ 0.01

วัน	เครื่องจักร				
	A	B	C	D	E
1	7	12	14	19	7
2	7	17	18	25	10
3	15	12	18	22	11
4	11	18	19	19	15
5	9	18	19	23	11
รวม	49	77	88	108	54

[illegible]

