

R PROGRAMMING

Part 10



ผู้ช่วยศาสตราจารย์ ดร. อัจฉมานท์ รัตนเลิศนุสรณ์
สาขาสถิติประยุกต์ คณะวิทยาศาสตร์และเทคโนโลยี
มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี

สารบัญ

Contents

- การวิเคราะห์ความแปรปรวน
- การวิเคราะห์ความแปรปรวนทางเดียว
- วิเคราะห์ความแปรปรวนทางเดียวด้วยโปรแกรมอาร์

การวิเคราะห์ความแปรปรวน

(Analysis of Variance: ANOVA)

การวิเคราะห์ความแปรปรวน เป็นการทดสอบความแตกต่างระหว่างค่าเฉลี่ยของประชากรตั้งแต่ 3 กลุ่มขึ้นไป ซึ่งนิยมเรียกย่อ ๆ ว่า ANOVA (อ่านว่า อะโนว่า)

ข้อตกลงของการวิเคราะห์ความแปรปรวน

1. ข้อมูลแต่ละกลุ่มมีการแจกแจงปกติ
2. ข้อมูลแต่ละกลุ่มมีความแปรปรวนเท่ากัน
3. ข้อมูลแต่ละกลุ่มเป็นอิสระกัน

การวิเคราะห์ความแปรปรวนทางเดียว (One-way Analysis of Variance: One-way ANOVA)

- การวิเคราะห์ความแปรปรวนทางเดียว (One-way Analysis of Variance)

การวิเคราะห์ความแปรปรวนแบบทางเดียวเป็นการทดสอบความแตกต่างระหว่างค่าเฉลี่ยของประชากรตั้งแต่ 3 กลุ่มขึ้นไป โดยมีปัจจัย 1 ปัจจัย (หรือตัวแปร 1 ตัวเดียว) ซึ่งสามารถแบ่งออกเป็น k กลุ่ม (หรือระดับ) โดยที่แต่ละกลุ่มอาจมีจำนวนค่าสังเกตเท่ากันหรือไม่เท่ากันก็ได้

การวิเคราะห์ความแปรปรวนทางเดียว

(One-way Analysis of Variance: One-way ANOVA)

• รูปแบบข้อมูลของการวิเคราะห์ความแปรปรวนทางเดียว

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

และ $n = n_1 + n_2 + \dots + n_k$

การวิเคราะห์ความแปรปรวนทางเดียว (One-way Analysis of Variance: One-way ANOVA)

- สมมติฐานของการวิเคราะห์ความแปรปรวนทางเดียว

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

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

- สถิติทดสอบการวิเคราะห์ความแปรปรวนทางเดียว

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

โดยที่ F แทน ค่าสถิติเอฟที่คำนวณได้

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

MSE แทน ความแปรปรวนภายในกลุ่ม (Mean square within-groups)

การวิเคราะห์ความแปรปรวนทางเดียว

(One-way Analysis of Variance: One-way ANOVA)

- ผลการวิเคราะห์ความแปรปรวนสามารถสร้างเป็นตารางได้ดังนี้

ตารางที่ 1 แสดงสูตรการวิเคราะห์ความแปรปรวนแบบทางเดียว (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	$SST = \sum_{i=1}^k \sum_{j=1}^{n_i} X_{ij}^2 - \frac{X_{..}^2}{n}$		

การวิเคราะห์ความแปรปรวนทางเดียว

(One-way Analysis of Variance: One-way ANOVA)

โดยที่	F	แทน	คำสถิติเอฟที่คำนวณได้
SST_{rt}		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

ตัวอย่างการวิเคราะห์ความแปรปรวนทางเดียว (One-way ANOVA example)

ตัวอย่างที่ 1. 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

จงทดสอบว่าคะแนนเฉลี่ยของคะแนนสอบของวิธีการ
สอนทั้ง 3 วิธีแตกต่างกันหรือไม่ ที่ระดับนัยสำคัญ 0.05

ตัวอย่างการวิเคราะห์ความแปรปรวนทางเดียว (One-way ANOVA example)

วิธีทำ

1) ตั้งสมมติฐาน

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

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

2) กำหนดระดับนัยสำคัญ $\alpha = 0.05$

3) คำนวณค่าสถิติ F

นำข้อมูลจากตารางข้างต้น มาคำนวณตามสูตรในตารางที่ 2 ดังแสดง
รายละเอียดดังนี้

ตัวอย่างการวิเคราะห์ความแปรปรวนทางเดียว (One-way ANOVA example)

ตารางที่ 2 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		
$X_1 = 93$	$\sum_{j=1}^{n_1} X_{1j}^2 = 1099$	$X_2 = 129$	$\sum_{j=1}^{n_2} X_{2j}^2 = 1643$	$X_3 = 121$	$\sum_{j=1}^{n_3} X_{3j}^2 = 1641$

ตัวอย่างการวิเคราะห์ความแปรปรวนทางเดียว (One-way ANOVA example)

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

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

$$SST = \sum_{i=1}^k \sum_{j=1}^{n_i} X_{ij}^2 - \frac{X_{..}^2}{n} = 4383 - 3921.633 = 461.367$$

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

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

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

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

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

ตัวอย่างการวิเคราะห์ความแปรปรวนทางเดียว (One-way ANOVA example)

สามารถเสนอผลการวิเคราะห์ความแปรปรวนทางเดียวดังนี้

ตารางที่ 3 แสดงผลการคำนวณค่าในการวิเคราะห์ความแปรปรวนทางเดียว

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

ตัวอย่างการวิเคราะห์ความแปรปรวนทางเดียว (One-way ANOVA example)

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 หรือกล่าวอีกนัยหนึ่ง
ว่า วิธีสอนทั้ง 3 แบบให้ผลไม่แตกต่างกัน

ตัวอย่างการวิเคราะห์ความแปรปรวนทางเดียว (One-way ANOVA example)

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 โดยใช้การเปรียบเทียบพหุคูณ (Multiple comparison test) ตามวิธีของเชฟเฟ้ (Scheffe's method) การทดสอบ HSD ของทูกีย์ (Tukey's HSD test) หรือ วิธีของนิวแมนคูลส์ (Newman Keuls method) ซึ่งเป็นวิธี “post hoc” หมายถึง วิธีการที่ตามมาหลังการวิเคราะห์ความแปรปรวน

ตัวอย่างการวิเคราะห์ความแปรปรวนทางเดียว (One-way ANOVA example)

ตัวอย่างที่ 2 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

ตัวอย่างการวิเคราะห์ความแปรปรวนทางเดียว (One-way ANOVA example)

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 $n=35$)

ตัวอย่างการวิเคราะห์ความแปรปรวนทางเดียว (One-way ANOVA example)

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)

การวิเคราะห์ความแปรปรวนทางเดียว ด้วยโปรแกรมอาร์

ฟังก์ชัน/การทำงานของฟังก์ชันที่เกี่ยวข้อง

ฟังก์ชันในโปรแกรมอาร์	การทำงานของฟังก์ชัน
<code>aov ()</code>	สร้างตารางวิเคราะห์ความแปรปรวน ภายใต้ข้อสมมติว่าความแปรปรวนของแต่ละกลุ่มเท่ากัน (homogeneity of variance) คือก่อนที่จะใช้ ฟังก์ชันนี้ ต้องทดสอบ Bartlett test ก่อนและได้ผล การทดสอบว่ายอมรับ H_0 : ความแปรปรวนของแต่ละ กลุ่มมีค่าเท่ากัน
<code>bartlett.test ()</code>	ทดสอบความแปรปรวนของประชากรว่าเท่ากันหรือไม่ ใช้กับข้อมูลตั้งแต่ 3 กลุ่มขึ้นไป H_0 : ความแปรปรวนของแต่ละกลุ่มมีค่าเท่ากัน H_1 : มีความแปรปรวนอย่างน้อย 1 คู่ที่ไม่เท่ากัน

การวิเคราะห์ความแปรปรวนทางเดียว ด้วยโปรแกรมอาร์

ฟังก์ชัน/การทำงานของฟังก์ชัน

ฟังก์ชันในโปรแกรมอาร์	การทำงานของฟังก์ชัน
<code>oneway.test ()</code>	สร้างตารางวิเคราะห์ความแปรปรวน โดยไม่ต้องคำนึงถึงข้อสมมติว่าความแปรปรวนของแต่ละกลุ่มเท่ากัน (homogeneity of variance) คือไม่ต้องทดสอบ Bartlett test
<code>TurkeyHSD ()</code>	ทดสอบ Multiple comparison โดยวิธีของ Turkey ก่อนที่จะใช้ฟังก์ชันนี้ต้องมีผลการทดสอบ one-way ANOVA ว่าปฏิเสธสมมติว่าง

การวิเคราะห์ความแปรปรวนทางเดียว ด้วยโปรแกรมอาร์

```
#####  
# one-way anova  
# date: 20/02/2021  
#####  
#example1:  
x=c(7,6,15,8,15,6,12,8,16,  
    11,12,9,7,17,12,19,11,13,8,10,  
    17,6,18,14,16,8,7,15,9,11)  
trt=as.factor(c(rep(1,9),rep(2,11),rep(3,10)))  
dataf=data.frame(x,trt)  
plot(x~trt) #plot trt[i] vs x  
bartlett.test(x~trt)  
#Bartlett test of homogeneity of variances
```

การวิเคราะห์ความแปรปรวนทางเดียว ด้วยโปรแกรมอาร์

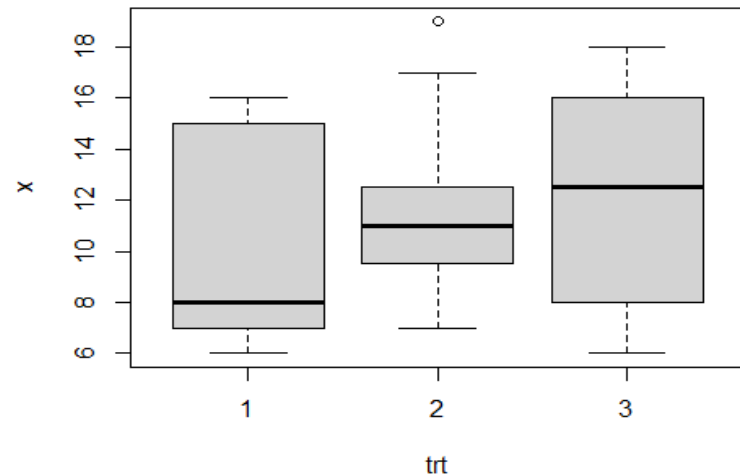
```
> bartlett.test(x~trt)
```

Bartlett test of homogeneity of
variances

data: x by trt

Bartlett's K-squared = 0.3945, df = 2,
p-value = 0.821

p-value (0.821) > $\alpha(0.05)$ แสดงว่า
ยอมรับ H_0
 $H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2$



การวิเคราะห์ความแปรปรวนทางเดียว ด้วยโปรแกรมอาร์

```
> summary(aov(x~trt,data=dataf))  
              Df Sum Sq Mean Sq F value Pr(>F)  
trt             2   16.3    8.142    0.494  0.616  
Residuals      27  445.1   16.485
```

สรุปค่าที่ได้จากโปรแกรม

$SSTrT = 16.2848$, $SSE = 445.0818$

df of treatment = 2 ($k-1$) , df of error = 27 ($n-k$)

$MSTrt = SSTrT/2 = 16.2848/2 = 8.1424$, $MSE = SSE/27 = 445.0818/27 = 16.4845$

$F = MSTrt/MSE = 8.1424/16.4845 = 0.4939$

การวิเคราะห์ความแปรปรวนทางเดียว ด้วยโปรแกรมอาร์

$$F = \text{MSTrt}/\text{MSE} = 8.1424/16.4845 = 0.4939$$

เนื่องจาก $F_{0.05}(2, 27) = 3.3541$ (ค่าวิกฤติ)

`> qf(0.05,df1=2,df2=27,lower.tail=FALSE)`

`[1] 3.354131`

เปรียบเทียบ F ที่คำนวณได้คือ 0.4939 กับ F ค่าวิกฤติ (3.3541)

พบว่า $0.4939 < 3.3541$ แสดงว่า ยอมรับ $H_0 : \mu_1 = \mu_2 = \mu_3$

นั่นคือ ค่าเฉลี่ยของคะแนนสอบของวิธีการสอนทั้ง 3 วิธีไม่แตกต่างกัน

การวิเคราะห์ความแปรปรวนทางเดียว ด้วยโปรแกรมอาร์

```
#####  
#example1:  
x=c(7,6,15,8,15,6,12,8,16,  
    11,12,9,7,17,12,19,11,13,8,10,  
    17,6,18,14,16,8,7,15,9,11)  
trt=as.factor(c(rep(1,9),rep(2,11),rep(3,10)))  
dataf=data.frame(x,trt)  
oneway.test(x~trt,data=dataf, var.equal = TRUE)
```

```
> oneway.test(x~trt,data=dataf, var.equal = TRUE)
```

One-way analysis of means

data: x and trt

F = 0.49394, num df = 2, denom df = 27,

p-value = 0.6156

การวิเคราะห์ความแปรปรวนทางเดียว ด้วยโปรแกรมอาร์

เปรียบเทียบ p-value (0.6156) กับ alpha (0.05)

พบว่า $0.6156 > 0.05$ แสดงว่า ยอมรับ $H_0 : \mu_1 = \mu_2 = \mu_3$

นั่นคือ ค่าเฉลี่ยของคะแนนสอบของวิธีการสอนทั้ง 3 วิธีไม่แตกต่างกัน

กรณีที่เปรียบเทียบ p-value กับ alpha (0.05)

แล้วพบว่า p-value < alpha ก็แสดงว่า ปฏิเสธ H_0 : มี $\mu_i \neq \mu_j$ อย่างน้อย 1 คู่ที่ไม่เท่ากัน

นั่นคือ ค่าเฉลี่ยของคะแนนสอบของวิธีการสอนทั้ง 3 วิธีให้ผลแตกต่างกันต้องทำการ
เปรียบเทียบค่าเฉลี่ยรายคู่ต่อไป

ฝึกปฏิบัติการวิเคราะห์ความแปรปรวนทางเดียว ด้วยโปรแกรมอาร์

นักวิจัยตลาดคนหนึ่งต้องการทราบว่า รถตู้ที่กำลังได้รับความนิยมในตลาด 3 ยี่ห้อ คือ A , B และ C จะกินน้ำมันต่างกันหรือไม่ จึงสุ่มตัวอย่างรถแต่ละยี่ห้อมาอย่างละ 4 คัน แล้วให้น้ำมันคันละ 3 ลิตร บันทึกระยะทางที่รถแต่ละคันวิ่งได้จนน้ำมันหมด ดังข้อมูลดังตาราง

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

จงสรุปผลการทดสอบดังกล่าว ที่ระดับนัยสำคัญ 0.05

ฝึกปฏิบัติการวิเคราะห์ความแปรปรวนทางเดียว ด้วยโปรแกรมอาร์

ข้อมูลต่อไปนี้เป็นจำนวนผลผลิตที่ขำรดจากการตรวจพบในแต่ละวันของเครื่องจักร 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

Q&A