บทที่ 2

แผนงานทดลองแบบสุ่มสมบูรณ์

(Completely Randomized Design = CRD.)

ในบทต่าง ๆ ต่อไปนี้ จะเป็นการวิเคราะห์ข้อมูลโดยใช้เทคนิค "การวิเคราะห์ความแปรปรวน" ซึ่งเป็นผลงานของท่าน R.A. Fisher สำหรับการวางแผนแบบสุ่มสมบูรณ์ อยู่ในกลุ่ม "การจำแนกแบบทางเดียว" (One-Way Classification) เพราะใช้แต่เฉพาะ "กรรมวิธี" เป็นเกณฑ์ในการจำแนกข้อมูล

แผนงานทดลองนี้เหมาะสมมากกับงานทดลองที่หน่วยทดลองทั้งหลายมีลักษณะเป็นเอกภาพกัน หรือมีความผันแปรน้อยมาก การจัดกรรมวิธีให้กับหน่วยทดลองทำได้โดยไม่มีข้อจำกัด (no restriction) เหมือนกับการวางแผนแบบบล็อคสมบูรณ์ ซึ่งมีข้อจำกัดว่าทุกบล็อคต้องมีครบทุกกรรมวิธี ส่วนการกำหนดจำนวนซ้ำของกรรมวิธีไม่ถือว่าเป็น "ข้อจำกัด" หน่วยทดลองทุกหน่วยมีโอกาส หรือ ความน่าจะเป็นเท่า ๆ กันในการได้รับกรรมวิธีใดกรรมวิธีหนึ่ง การจัดกรรมวิธีให้กับหน่วยทดลองแบบสุ่ม มีขั้นตอน ดังนี้

สมมุติมีหน่วยทดลองทั้งหมด 15 หน่วย เพื่อจะรับกรรมวิธีทั้งหมด 3 กรรมวิธี ถ้าทุกกรรมวิธีมีจำนวนซ้ำเท่ากัน นั่นคือกรรมวิธีละ 5 ซ้ำ ให้เลขที่แก่หน่วยทดลองทั้งหมดคือหมายเลข 1, 2, ..., 14, 15 แล้วเลือกเลขจากตารางแบบสุ่ม อย่างต่ำต้องเป็นเลข 2 หลักขึ้นไป สมมุติได้เลข

11 87 01 78 99 65 88 63 89 84 13 96 80 26 79 ตัวที่ซ้ำต้องตัดทิ้งไป เลขเหล่านี้ เมื่อจัดตามอันดับ จะได้อันดับต่าง ๆ ดังนี้ (น้อยไปหามาก)

2 11 1 7 15 6 12 5 13 10 3 14 9 4 8 ซึ่งหมายความว่า หน่วยทดลองที่

2, 11, 1, 7, และ 15 ให้ รับกรรมวิธีที่ 1

6, 12, 5, 13 และ 10 รับกรรมวิธีที่ 2

และที่เหลือ คือหน่วยทดลองที่ 3, 14, 9, 4 และ 8 รับกรรมวิธีที่ 3

นอกจากวิธีจัดอันดับ ยังใช้วิธีอื่น ๆ ได้อีกหลายวิธี เช่นการเอาจำนวนหน่วยทดลองไปหาร

หมายความว่า

หน่วยทดลองที่ 11, 12, 01, 03, 09 รับกรรมวิธีที่ 1

หน่วยทดลองที่ 05, 13, 14, 06 04 รับกรรมวิธีที่ 2

หน่วยทดลองที่เหลือ คือ หน่วยที่ 15, 07, 02, 08 และ 10 รับกรรมวิธีที่ 3

(00 แทนหน่วยทดลองที่ 15)

และแผนผังของงานทดลอง (Experimental lay-out) จะเป็นดังนี้

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
ก	ค	ก	ข	ข	ข	ค	ค	ก	ค	ก	ก	ข	ข	ค

ก, ข, ค, คือกรรมวิธีที่ 1 2 และ 3 ตามลำดับ

ข้อควรปฏิบัติในระหว่างดำเนินการทดลอง คือต้องปฏิบัติต่อหน่วยทดลองทุกหน่วยในลักษณะเดียวกันเพื่อให้ความคลาดเคลื่อนจากการทดลองมีขนาดเล็กที่สุด

ข้อดีและข้อเสียของการวางแผนแบบสุ่มสมบูรณ์

ข้อดี

1. เป็นแผนงานทดลองที่สามารถยืดหยุ่นได้ เพราะจำนวนกรรมวิธี และจำนวนซ้ำถูกจำกัดโดยจำนวนหน่วยทดลองที่มีอยู่ทั้งหมด ดังนั้นจำนวนซ้ำจึงขึ้นอยู่กับจำนวนกรรมวิธีและสามารถให้กรรมวิธีต่าง ๆ มีจำนวนซ้ำที่ต่างกันได้ แต่ถ้าไม่มีเหตุผลจำเป็น ควรให้มีจำนวนซ้ำเท่ากันทุกกรรมวิธี
2. การวิเคราะห์ง่ายมาก ไม่ยุ่งยากซับซ้อนแม้จะใช้จำนวนซ้ำต่างกัน
3. เมื่อเกิดข้อมูลสูญหาย จะมีผลกระทบกระเทือนน้อยมากเมื่อเทียบกับวิธีอื่น
4. มีจำนวน df. สำหรับใช้หาค่าประมาณของความคลาดเคลื่อนจากการทดลองสูงที่สุด

ข้อเสีย

เนื่องจากไม่มีข้อจำกัดในการจัดกรรมวิธีให้หน่วยทดลองแบบสุ่ม ทำให้ความคลาด-เคลื่อนจากการทดลองรวมกับความผันแปรของหน่วยทดลองทั้งหมด ยกเว้นความผันแปรจากกรรมวิธีซึ่งในกรณีที่หน่วยทดลองไม่มีความสม่ำเสมอกัน จะไม่สามารถแยกส่วนแตกต่างนี้ออกจากความคลาดเคลื่อนได้ จึงเป็นข้อเสียเปรียบ เมื่อเทียบกับวิธีอื่น ซึ่งจัดหน่วยทดลองเป็นกลุ่ม ๆ เรียกว่า บล็อค และสามารถตัดความผันแปรของบล็อคออกจากความคลาด-เคลื่อนได้

ตัวอย่าง

จำนวนน้ำมันที่ถูกดูดซึมในขนมโดนัท เมื่อทอดด้วยไขมันชนิดต่าง ๆ 4 ชนิด

ไขมัน	1	2	3	4	total
	64	78	75	55	
	72	91	93	66	
	68	97	78	49	
	77	82	71	64	
	56	85	63	70	
	95	77	76	68	
$\sum_j X_{ij}$	432	510	456	372	1770 = G
$\bar{X}_{i.}$	72	85	76	62	295
$\sum\sum X^2_{ij}$	31994	43652	35144	23402	134192
$(\sum X)^2/r$	31104	43350	34656	23064	132174
$\sum_j (X-\bar{X})^2$	890	302	488	338	2018
df	5	5	5	5	20

$$\text{pooled } s^2 = 2018/20 = 100.9$$

$$s_{\bar{d}} = \sqrt{2s^2/r} = \sqrt{2(100.9)/6} = 5.80$$

Analysis of Variance: One-Way Classification With Equal Replications

Source of Variation	df.	Sum of Squares: Definition	Sum of Squares: Working	Mean Square	F
Treatments (Among groups)	$t-1$	$r\sum_{i}^{t}(\bar{X}_{i.}-\bar{X}_{..})^2$	$\sum\frac{X_{i.}^2}{r}-\frac{X_{..}^2}{rt}$	SSTr./(t-1)	$\frac{MSTr.}{MSE}$
Experimental error	$t(r-1)$	$\sum_{ij}^{tr}\sum(X_{ij}-\bar{X}_{i})^2$	ลบเอา	SSE/t(r-1)	
Total	$tr-1$	$\sum_{ij}^{tr}\sum(X_{ij}-\bar{X}_{..})^2$	$\sum\sum X_{ij}^2-\frac{X_{..}^2}{tr}$		

Correction Factor CF $= \frac{X_{..}^2}{tr} = G^2/N$

$$SS(Total) = (64^2+72^2+\ldots+70^2+68^2)-(1770)^2/24$$

$$= 134192 - 130538 = 3654 \;;\; df = 23$$

$$SS(Treatments) = (432^2+510^2+456^2+372^2)/6-(1770)^2/24$$

$$= 132174 - 130538$$

$$= 1636 \;;\; df. = 3$$

$$SS(error) = SS(total) - SS(Treatments)$$

$$= 3654 - 1636 = 2018 \;;\; df. = 23-3 = 20$$

หรือคิดโดยใช้สูตรนิยาม

$$SS(Treatments) = 6[(72-73.75)^2+(85-73.75)^2+(76-73.75)^2+(62-73.75)^2]$$

$$= 1636 \quad (\bar{X}_{..} = 73.75 = 1770/24)$$

หรืออีกวิธี คือ

$$SS(Treatments) = 6 [72^2 + 85^2 + 76^2 + 62^2 - (295)^2/4] = 1636$$

ทั้งนี้เพราะว่า

$$432^2/6 = (6 \times 72)^2/6 = 6(72)^2$$

$$1770^2/24 = (6 \times 295)^2/24 = 6(295)^2/4$$

$$SS(Error) = \Sigma\Sigma(X-\bar{X})^2 = 890+302+488+338 = 2018; \ df. = 20$$

$$s_p^2 = SSE/df. = 2018/20 = 100.9$$

ANOVA

SOV	df	SS	MS	F- ratio
Between treatments (ไขมัน)	3	1636	545.3	5.4
Experimental error	20	2018	100.9	$f_{3,20(.01)} = 4.94$
total	23	3654		

การสรุปผล

เนื่องจากค่าสถิติเอฟที่คำนวณได้ใหญ่กว่าค่าที่เปิดได้จากตารางที่ระดับนัยสำคัญ 0.01 จึงสรุปว่าปริมาณไขมันที่ถูกดูดซึมในโดนัทโดยเฉลี่ยแล้วไม่เท่ากันทั้งหมด

ในกรณีที่สามารถปฏิเสธ Ho ได้เช่นกรณีนี้ จะเห็นค่าสถิติเอฟเป็นเพียงการสรุปอย่างกว้างกว้างว่าค่าเฉลี่ยของประชากรเหล่านี้ไม่เท่ากันทั้งหมด แต่ไม่ได้บอกละเอียดว่าถ้าเปรียบเทียบความแตกต่างเป็นคู่ ๆ แล้ว จะมีคู่ใดที่ต่างกันอย่างมีนัยสำคัญบ้าง ซึ่งมีวิธีการทดสอบหลายวิธี ซึ่งจะได้กล่าวโดยละเอียดต่อไป

การวิเคราะห์ข้อมูลเมื่อแต่ละวิธีการมีจำนวนซ้ำไม่เท่ากัน

ต้องการทดสอบความทนทานของส่วนผสมของยาง 4 ชนิด ได้เตรียมส่วนผสมไว้ชนิดละ 4 ตัวอย่าง รวมเป็น 16 ตัวอย่าง เพื่อใช้ทดสอบ แต่ปรากฏว่าส่วนผสมอันหนึ่งของชนิดที่ 1 เสีย ใช้ทดสอบไม่ได้ จึงเหลือแค่ 3 ตัวอย่างสำหรับชนิดที่ 1 ผลการตรวจสอบ ได้ข้อมูล ดังนี้

A	B	C	D
3210	3225	3220	3545
3000	3320	3410	3600
3315	3165	3320	3580
	3145	3370	3485

แปลงค่าข้อมูลโดยนำ 3000 ไปหักออกทุกค่า จะไม่ทำให้ค่าสถิติเอฟจากตารางวิเคราะห์ความแปรปรวนเปลี่ยนแปลง ดังนี้ $X'_{ij} = (X_{ij} - 3000)/5$.

	A	B	C	D	
	42	45	44	109	
	0	64	82	120	
	63	33	64	116	
		29	74	97	
total	105	171	264	442	952

1. CF. $982^2/15 = 64288.27$ (1 df.)

2. $\Sigma\Sigma X^2_{ij} = 42^2+0^2+\ldots\ldots+116^2+97^2 = 81162$ (15 df.)

3. $\Sigma X^2_{i.}/r_i = 105^2/3 + 171^2/4 + 264^2/4 + 442^2/4$

$= 105^2/3 + 1/4(171^2 + 264^2 + 442^2)$

$= 77250.25$ (4 df.)

....

4. SS(total) = (2) - (1)

$= 81162 - 64288.27 = 16873.73$ (14 df.)

5. SS(Treatments) = (3) - (1) = 12961.98 (3 df.)

6. SS(Error) = (4) - (5) = 3911.75

df. = N - t = 15 - 4 = 11

ANOVA

SOV.	df.	SS	MS	F-ratio
Treatments	3	12961.98	4320.66	12.15
Error	11	3911.75	355.61	
Total	14	16873.73		$f_{3,11(.01)} = 6.22$

ค่าสถิติตกอยู่ในเขตวิกฤต จึงสรุปว่าความทนทานเฉลี่ยของส่วนผสมยาง 4 ชนิดมีความแตกต่างกันอย่างมีนัยสำคัญยิ่ง

Linear additive model

$$X_{ij} = \mu + \tau_i + \varepsilon_{ij} \qquad i = 1,2,\ldots,t \quad j = 1,2,\ldots,r_i$$

μ = ค่าเฉลี่ยรวมยอด

τ_i = อิทธิพลของกรรมวิธีที่ i

ε_{ij} = อิทธิพลอื่นๆ ของหน่วยทดลองที่ j ซึ่งรับวิธีการ i

แบบจำลองนี้ยังแบ่งเป็น 2 อย่าง ขึ้นอยู่กับข้อสมมุติของกรรมวิธี ดังนี้

Model 1 (Fixed Model)

assumption $\sum_i \tau_i = 0$

Model 2 (Random Model)

assumption $E(\tau_i) = 0$, $var(\tau_i) = \sigma^2_\tau$

และทั้ง 2 Model มีข้อสมมุติของ ε_{ij} เหมือนกัน ดังนี้

ε_{ij} มีการแจกแจงแบบปกติ มีค่าเฉลี่ยคือ $E(\varepsilon_{ij}) = 0$ และความแปรปรวนเท่ากันไม่ว่าจะ มาจากกรรมวิธีใด คือ $var(\varepsilon_{ij}) = \sigma^2$

ความแตกต่างของแบบจำลอง 2 แบบนี้ คือ แบบจำลองที่ 1 กรรมวิธีที่ใช้ในการทดลองได้ถูกกำหนดไว้ล่วงหน้า ดังนั้น ข้อสรุปทั้งหลายที่ได้จากการทดลองจึงใช้ได้เฉพาะกรรมวิธีที่อยู่ในการทดลองนั้นเท่านั้น ส่วนแบบจำลองที่ 2 กรรมวิธีที่อยู่ในการทดลองทั้งหมดนั้นได้มาจากการสุ่มจากประชากรของกรรมวิธีซึ่งมีข้อสมมุติว่ามีการแจกแจงแบบปกติ ด้วยค่าเฉลี่ย 0 และความแปรปรวน σ_τ^2 ดังนั้นข้อสรุปที่ได้จากการทดลองจึงนำไปใช้กับประชากรของกรรมวิธีได้

การตั้งสมมติฐาน

เนื่องจากข้องสมมุติของกรรมวิธีของแบบจำลองทั้ง 2 แตกต่างกัน สมมติฐานจึงต่างกันด้วย

แบบจำลองที่ 1

$$Ho : \tau_i = 0 \; ; \; i = 1,2,\dots,t$$

$$Ha : \tau_i \neq 0$$

แบบจำลองที่ 2

$$Ho : \sigma_\tau^2 = 0$$

$$Ha : \sigma_\tau^2 \neq 0$$

การประมาณค่าพารามิเตอร์

จากตัวแบบ

$$X_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

μ = general mean

$\tau_i = \mu_i - \mu$ หรือ $\mu_i = \mu + \tau_i$

$$\begin{aligned}\varepsilon_{ij} &= X_{ij} - \mu - \tau_i \\ &= X_{ij} - \mu - (\mu_i - \mu) \\ &= X_{ij} - \mu_i\end{aligned}$$

จากการทดลอง จะได้ค่าประมาณของ μ, τ_i และ ε_{ij} ดังนี้

$$\hat{\mu} = \bar{X}_{..} = X_{..}/N$$

$$\hat{\tau}_i = \hat{\mu}_i - \hat{\mu} = \bar{X}_{i.} - \bar{X}_{..}$$

$$\hat{\varepsilon}_{ij} = X_{ij} - \hat{\mu}_i = X_{ij} - \bar{X}_{i.}$$

ข้อสังเกต

สำหรับตัวแบบกำหนด การทดสอบสมมติฐาน คือ

$$Ho : \tau_i = 0 \quad , \quad i = 1,2,\ldots,t$$

แทนค่า τ_i จะได้

$$Ho : \mu_i - \mu = 0 \quad , \quad i = 1,2,\ldots,t$$

$$= Ho : \mu_i = \mu$$

$$= Ho : \mu_1 = \mu_2 = \ldots = \mu_t = \mu$$

ส่วนความแปรปรวนของ τ_i คือ

$$V(\tau_i) = \sum_{i=1}^{t} (\tau_i - \bar{\tau})^2/(t-1)$$

จากข้อสมมุติของ Model 1 ว่า

$$\sum_{i=1}^{t} \tau_i = 0 \quad ,$$

ดังนั้น $\bar{\tau}_i = \sum_{i=1}^{t} \tau_i/t = 0$

นั่นคือ

$$V(\tau_i) = \sum_{i=1}^{t} (\tau_i - 0)^2/(t-1)$$

$$= \sum \tau_i^2/(t-1)$$

แต่ถ้าเป็น Model II จะมีข้อสมมุติว่า

$$\tau_i \sim NID\ (0,\ \sigma^2_\tau$$

นั่นคือ $V(\tau_i) = \frac{\Sigma(\tau_i - \bar{\tau})^2}{(t-1)} = \sigma^2_\tau$

เมื่อแทนค่าประมาณของ $\mu, \tau_i, \varepsilon_{ij}$ ในตัวแบบสถิติ จะได้

$$X_{ij} = \bar{X}_{..} + (\bar{X}_{i.} - \bar{X}_{..}) + (X_{ij} - \bar{X}_{i.})$$

นั่นคือค่าสังเกตทุกตัวจะแยกออกได้เป็น 3 ส่วน เช่น $X_{11} = 64$

$$X_{11} = \bar{X}_{..} + (\bar{X}_{1.} - \bar{X}_{..}) + (X_{11} - \bar{X}_{1.})$$

$$\begin{aligned} 64 &= 73.75 + (72-73.75) + (64-72) \\ &= 73.75 + (-1.75) + (-8) \\ &= 64 \end{aligned}$$

เทอมแรก $\bar{X}_{..}$ คือ ค่าคงที่

เทอมที่สอง $(\bar{X}_{i.} - \bar{X}_{..})$ คือ อิทธิพลของวิธีการ จะเปลี่ยนแปลงเมื่อเปลี่ยนวิธีการเท่านั้น ถ้าอยู่ในวิธีการเดียวกันจะต้องเหมือนกัน

เทอมสุดท้าย $X_{ij} - \bar{X}_{i.} = \hat{\varepsilon}_{ij}$ ความคลาดเคลื่อนจากการทดลอง จะเปลี่ยนเมื่อเปลี่ยนหน่วยทดลอง

อิทธิพลของวิธีการ และความคลาดเคลื่อน จะมีการเปลี่ยนแปลง เครื่องหมายบวกและลบแสดงทิศทางการเปลี่ยนแปลงว่ามากกว่าหรือน้อยกว่า ก็ตาม

$\sum_{i=1}^{t} (\bar{X}_{i.} - \bar{X}_{..}) = 0$ เพราะ $\bar{X}_{..}$ คือ ค่าเฉลี่ยของ $\bar{X}_{i.}$

$\sum_{j=1}^{r} (X_{ij} - \bar{X}_{i.}) = 0$ เพราะ $\bar{X}_{i.}$ คือ ค่าเฉลี่ยของ X_{ij}

ดังนั้นถ้าเราแยกค่าสังเกต X_{ij} ทุกตัว เป็น 3 เทอม แล้วนำรวมดังนี้

$$\begin{aligned}\sum_{i}^{t}\sum_{j}^{r} X_{ij} &= \sum_{i=1}^{t}\sum_{j=1}^{r}(\bar{X}_{..}) + \sum_{i=1}^{t}\sum_{j=1}^{r}(\bar{X}_{i.} - \bar{X}_{..}) + \sum_{i=1}^{t}\sum_{j=1}^{r}(X_{ij} - \bar{X}_{i.}) \\ &= tr\,\bar{X}_{..} + \underbrace{r\sum_{i=1}^{t}(\bar{X}_{i.} - \bar{X}_{..})}_{=0} + \sum_{i=1}^{t}\underbrace{[\sum_{j=1}^{r}(X_{ij} - \bar{X}_{i.})]}_{0} \\ &= tr\,\bar{X}_{..}\end{aligned}$$

จึงไม่ทราบค่ารวมอิทธิพล วิธีการและความคลาดเคลื่อน วิธีแก้ไขคือการศึกษาใน**รูปความผันแปร** คือไม่ต้องสนใจ**ทิศทาง**การเปลี่ยนแปลงหรือค่าเบี่ยงเบนจากค่าเฉลี่ย แต่จะดู**ขนาด**ของการเปลี่ยนแปลง นั่นคือค่ากำลังสองของค่าเบี่ยงเบน

จากตัวแบบเดิม

$$X_{ij} = \bar{X}_{..} + (\bar{X}_{i.} - \bar{X}_{..}) + (X_{ij} - \bar{X}_{i.})$$

เปลี่ยนค่า $\bar{X}_{..}$ มาซ้ายมือ ด้านขวามือจะเหลือ 2 เทอม ด้านซ้ายจะอยู่ในรูปการค่าเปลี่ยนแปลง แล้วยกกำลังสองทั้ง 2 ด้าน

$$(X_{ij} - \bar{X}_{..})^2 = [(\bar{X}_{i.} - \bar{X}_{..}) + (X_{ij} - \bar{X}_{i.})]^2$$

รวมค่าสังเกตทั้งหมดจากการทดลอง

$$\sum_{i=1}^{t}\sum_{j=1}^{r}(X_{ij} - \bar{X}_{..})^2 = \sum_{i=1}^{t}\sum_{j=1}^{r}[(\bar{X}_{i.} - \bar{X}_{..}) + (X_{ij} - \bar{X}_{i.})]^2$$

$$= \sum_{i=1}^{t}\sum_{j=1}^{r}(\bar{X}_{i.} - \bar{X}_{..})^2 + \sum_{i=1}^{t}\sum_{j=1}^{r}(X_{ij} - \bar{X}_{i.})^2 + 2\sum_{i=1}^{t}\sum_{j=1}^{r}(\bar{X}_{i.} - \bar{X}_{..})(X_{ij} - \bar{X}_{i.})$$

เทอมสุดท้ายขวามือเป็นศูนย์เพราะ

$$2\sum_{i}^{t}\sum_{j}^{r}(\bar{X}_{i.} - \bar{X}_{..})(X_{ij} - \bar{X}_{i.}) = 2\sum_{i=1}^{t}[(\bar{X}_{i.} - \bar{X}_{..})\underbrace{\sum_{j=1}^{r}(X_{ij} - \bar{X}_{i.})}_{=0}]$$

จะได้สูตรนิยามสำหรับตารางวิเคราะห์ความแปรปรวน คือ

$$\sum_{i}^{t}\sum_{j}^{r}(X_{ij}-\bar{X}_{..})^2 = r\sum_{i=1}^{t}(\bar{X}_{i.}-\bar{X}_{..})^2 + \sum_{i}^{t}\sum_{j}^{r}(X_{ij}-\bar{X}_{i.})^2$$

SS(total) = SS(treatment)+SS(experimental error or residual)

การหาสูตรคำนวณ

1) SS(total)

$$\begin{aligned}
SS(\text{total}) &= \sum_{i}^{t}\sum_{j}^{r}(X_{ij}-\bar{X}_{..})^2 \\
&= \sum_{i}^{t}\sum_{j}^{r}(X_{ij}^2-2X_{ij}(\bar{X}_{..})+(\bar{X}_{..})^2) \\
&= \sum_{i}^{t}\sum_{j}^{r}X_{ij}^2 - 2\bar{X}_{..}\sum_{i}^{t}\sum_{j}^{r}X_{ij} + \sum_{i}^{t}\sum_{j}^{r}(\bar{X}_{..})^2 \\
&= \sum_{i}^{t}\sum_{j}^{r}X_{ij}^2 - \frac{2(X_{..})^2}{tr} + \frac{(X_{..})^2}{(tr)2} \\
&= \sum_{i}^{t}\sum_{j}^{r}X_{ij}^2 - \frac{2(X_{..})^2}{tr} + \frac{(X_{..})^2}{tr} \\
&\quad \sum_{i}^{t}\sum_{j}^{r}X_{ij}^2 - \frac{(X_{..})^2}{tr}
\end{aligned}$$

2) SS(treatment)

$$\begin{aligned}
SS(\text{treatment}) &= r\sum_{i=1}^{t}(\bar{X}_{i.}-\bar{X}_{..})^2 \\
&= r\sum_{i=1}^{t}\left(\frac{X_{i.}}{r}-\frac{X_{..}}{tr}\right)^2 \\
&= r\sum_{i=1}^{t}\frac{X_{i.}^2}{r^2} - 2r\sum_{i=1}^{t}\left(\frac{X_{i.}}{r}\right)\left(\frac{X_{..}}{tr}\right) + r\sum_{i=1}^{t}\frac{X_{..}^2}{t^2r^2} \\
&= \sum_{i=1}^{t}\frac{X_{i.}^2}{r} - 2(X_{..})\left(\frac{X_{..}}{tr}\right) + tr\left(\frac{X_{..}}{t^2r^2}\right)^2 \\
&= \sum_{i=1}^{t}\frac{X_{i.}^2}{r} - \frac{2(X_{..})^2}{tr} + \frac{X_{..}^2}{tr}
\end{aligned}$$

$$= \sum_{i=1}^{t} \frac{X_{i.}^2}{r} - \frac{(X_{..})^2}{tr}$$

3) SS(error) $= \sum_{i}^{t}\sum_{j}^{r} (X_{ij} - \bar{X}_{i.})^2$

$$= \sum_{i}^{t}\sum_{j}^{r} \left(X_{ij} - \frac{X_{i.}}{r}\right)^2$$

$$= \sum_{i}^{t}\sum_{j}^{r} \left(X_{ij}^2 - \frac{2(X_{i.})(X_{ij})}{r} + \frac{X_{i.}^2}{r^2}\right)$$

$$= \sum_{ij}^{t\,r} X_{ij}^2 - 2\sum_{i}^{t} \left[\left(\frac{X_{i.}}{r}\right)\left(\sum_{j}^{r} X_{ij}\right)\right] + \sum_{ij}^{tr} \frac{X_{i.}^2}{r^2}$$

$$= \sum_{i}^{t}\sum_{j}^{r} X_{ij}^2 - 2\sum_{i}^{t} \left[\left(\frac{X_{i.}}{r}\right)(X_{i.})\right] + r\sum_{i}^{t} \frac{X_{i.}^2}{r^2}$$

$$= \sum_{i}^{t}\sum_{j}^{r} X_{ij}^2 - 2\sum_{i}^{t} \frac{X_{i.}^2}{r} + \sum_{i}^{t} \frac{X_{i.}^2}{r}$$

$$= \sum_{i}^{t}\sum_{j}^{r} X_{ij}^2 - \sum_{i}^{t} \frac{X_{i.}^2}{r}$$

ค่าคาดหมายของกำลังสองเฉลี่ย (Expected Mean Square)

ค่าคาดหมายของกำลังสองเฉลี่ยสำหรับการจำแนกทางเดียวและมีจำนวนซ้ำเท่ากัน

SOV	df.	E(MS) Model 1	E(MS) Model 2
Treatments	$t - 1$	$\sigma^2 + r\Sigma\tau_i^2/(t - 1)$	$\sigma^2 + r\sigma_\tau^2$
Experimental error	$t(r - 1)$ $= N - t$	σ^2	σ^2

การหา สูตรนิยามของค่า SS ต่าง ๆ ใน ANOVA ของ CRD แบบมีตัวอย่างย่อย

$x_{ijk} = \mu + \tau_i + \varepsilon_{ij} + \delta_{ijk}$ $\quad i=1,..,t \quad ; j=1,..,r$

$\hat{\mu} = \bar{X}_{...}$ $\quad k=1,...,S$

$\hat{\tau}_i = \bar{X}_{i..} - \bar{X}_{...}$

$\hat{\varepsilon}_{ij} = \bar{X}_{ij.} - \bar{X}_{i..}$

$\hat{\delta}_{ijk} = X_{ijk} - \bar{X}_{ij.}$

แทนค่าประมาณ ลงใน Model ย้าย $\bar{X}_{...}$ มาซ้ายมือ ยกกำลังสองทั้ง 2 ด้าน และหาผลรวมทั้งหมด จะได้ค่าผลบวกกำลังสอง ดังนี้

$$\sum_i \sum_j \sum_k (X_{ijk} - \bar{X}_{...})^2 = \sum_i \sum_j \sum_k [(\bar{X}_{i..}-\bar{X}_{...})+(\bar{X}_{ij.}-\bar{X}_{i..})+(X_{ijk}-\bar{X}_{ij.})]^2$$

$$= \sum_i^t \sum_j^r \sum_k^s (\bar{X}_{i..} - \bar{X}_{...})^2 + \sum_i^t \sum_j^r \sum_k^s (\bar{X}_{ij.} - \bar{X}_{i..})^2 + \sum_i^t \sum_j^r \sum_k^s (X_{ijk} - \bar{X}_{ij.})^2$$

$\sum_i^t \sum_j^r \sum_k^s (X_{ijk} - \bar{X}_{...})^2 = SS(total)$ มี df $= trs - 1$

$\sum_i^t \sum_j^r \sum_k^s (\bar{X}_{i..} - \bar{X}_{...})^2 = rs\sum_i^t (\bar{X}_{i..} - \bar{X}_{...})^2 = SS(treatments)$ มี df $= t-1$

$\sum_{ijk}^{trs} (\bar{X}_{ij.} - \bar{X}_{i..})^2 = S\sum_{ij}^{tr} (\bar{X}_{ij.} - \bar{X}_{i..})^2 = SS(experimental\ error)$ มี df$=t(r-1)$

$\sum_i^t \sum_j^r \sum_k^s (X_{ijk} - \bar{X}_{ij.})^2 = SS(sampling\ error)$ มี df $= tr(S-1)$

สูตรคำนวณ ของ CRD แบบมีตัวอย่างย่อย

1. $SS(total) = \sum_{ij}^{tr} \sum_k^s (X_{ijk} - \bar{X}_{...})^2$

sampling unit

$$= \sum_{i}^{t}\sum_{j}^{r}\sum_{k}^{s} X_{ijk}^2 - \frac{X_{...}^2}{trs}$$

$$= \sum_{i}^{t}\sum_{j}^{r}\sum_{k}^{s} X_{ijk}^2 - CF$$

df $= trs - 1$

2. SS(วิธีการ) $= rs\sum_{i}^{t} (\bar{X}_{i..} - \bar{X}_{...})^2$

$$= \sum_{i}^{t} \frac{X_{i..}^2}{rs} - \frac{X_{...}^2}{trs}$$

df $= t - 1$

3. SS(experimental error) $= \sum_{i}^{t}\sum_{j}^{r} \frac{X_{ij.}^2}{s} - \frac{X_{i..}^2}{rs}$

df $= t(r-1)$

4. SS(sampling error) $= \sum_{i}^{t}\sum_{j}^{r}\sum_{k}^{s} X_{ijk}^2 - \frac{\Sigma\,\Sigma\, X_{ij.}^2}{S}$

df $= tr\ (s-1)$

ข้อสังเกต

1. SS(total)$_{\text{experimental unit}}$ = SS(วิธีการ)+SS(experimental error)

$$\sum_{ij}^{tr} \frac{X_{ij.}^2}{s} - \frac{X_{...}^2}{trs} = \left(\sum_{i}^{t} \frac{X_{i..}^2}{rs} - \frac{X_{...}^2}{trs}\right) + \left(\sum_{ij}^{tr} \frac{X_{ij.}^2}{s} - \frac{X_{i..}^2}{rs}\right)$$

df : [tr - 1] = [t - 1] + [t(r-1)]

ในทางปฏิบัติ จึงไม่นิยมหา ค่า SS (experimental error) โดยตรงจากสูตรข้อ (3) แต่จะใช้ความสัมพันธ์ในข้อสังเกต (1) ว่า

SS(experimental error) = SS(total)$_{\text{experimental unit}}$ - SS(วิธีการ)

คือหา SS(total)$_{\text{experimental unit}}$ ในข้อสังเกต (1) ก่อน แล้วเอา SS(วิธีการ) ในข้อ (2) ไปหักออก

SS(sampling error)=SS(total)$_{\text{sampling unit}}$-SS(total)$_{\text{experimental unit}}$

$$= (\sum_i^t \sum_j^r \sum_k^s X_{ijk}^2 - CF) - (\sum_i^t \sum_j^r \frac{X_{ij.}^2}{s} - CF)$$

$$= \sum_i^t \sum_j^r \sum_k^s X_{ijk}^2 - \sum_i^t \sum_j^r \frac{X_{ij.}^2}{s}$$

= สูตรข้อ (4)

ในทางปฏิบัติ จะหา SS(sampling error) โดยการนำ SS(total)$_{\text{experimental unit}}$ ไปหักออกจาก SS(total)$_{\text{sampling unit}}$.

SOV	E(MS) ของ Model I	E(MS) ของ Model II
Treatment	$\sigma_\delta^2 + s\sigma_\varepsilon^2 + rs\frac{\Sigma\tau_i^2}{(t-1)}$	$\sigma_\delta^2 + s\sigma_\varepsilon^2 + rs\sigma_\tau^2$
Experimental error	$\sigma_\delta^2 + s\sigma_\varepsilon^2$	$\sigma_\delta^2 + s\sigma_\varepsilon^2$
Sampling error	σ_δ^2	σ_δ^2

ค่าประมาณของ Varlance component

1. $\hat{\sigma}_\delta^2$ = MS(sampling error = MSS

2. $\hat{\sigma}_\varepsilon^2$ = $\frac{MSE - MSS}{S}$

การวิเคราะห์ข้อมูลแบบจำแนกทางเดียว และมีตัวอย่างย่อย

(Hierarchical or Nested Design)

บางครั้งเมื่อทำการทดลองแล้ว การเก็บข้อมูลมาประเมินผลจากหน่วยทดลอง มีการเก็บข้อมูลหลายจำนวนจากหน่วยทดลองอันเดียวกัน เรียกว่าเก็บจากหน่วยตัวอย่าง หรือตัวอย่างย่อย ซึ่งดูได้จากตัวอย่างในตารางข้างล่าง เป็นการสุ่มต้นไม้มาปลูกในกระถาง ๆ ละ 4 ต้น ดังนั้น กระถาง 1 กระถางซึ่งประกอบด้วยต้นไม้ 4 ต้น จึงเป็น 1 หน่วยทดลอง มีทั้งหมด 18 กระถาง ส่วนกรรมวิธีมี 2 แฟคเตอร์คือ จำนวนชั่วโมง แสงสว่าง และ อุณหภูมิ รวมทั้งหมด 3 (2) = 6 วิธีการ ได้ทำการสุ่มกระถางเข้ารับวิธีการต่าง ๆ แบบสุ่มวิธีการละ 3 กระถาง ถ้าเก็บข้อมูลจากทุกกระถางมากระถางละ 1 รายการ ก็คือการวางแผนงานทดลองแบบสุ่มสมบูรณ์ที่มี 6 กรรมวิธี แต่ละกรรมวิธีมี 3 ซ้ำ แต่ข้อมูลที่เก็บมาคือความสูงของต้นไม้ทั้ง 4 ต้น จากแต่ละกระถาง จึงเป็นหน่วยตัวอย่างย่อย ข้อมูลมีดังนี้

ความสูงภายหลังรับวิธีการ 1 อาทิตย์

	กลางคืนมีอุณหภูมิต่ำ									กลางคืนมีอุณหภูมิสูง								
ชั่วโมงแสงสว่าง	8			12			16			8			12			16		
กระถาง	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
ต้นไม้ Tree																		
1	3.5	2.5	3.0	5.0	3.5	4.5	5.0	5.5	5.5	8.5	6.5	7.0	6.0	6.0	6.5	7.0	6.0	11.0
2	4.0	4.5	3.0	5.5	3.5	4.0	4.5	6.0	4.5	6.0	7.0	7.0	5.5	8.5	6.5	9.0	7.0	7.0
3	3.0	5.5	2.5	4.0	3.0	4.0	5.0	5.0	6.5	9.0	8.0	7.0	3.5	4.5	8.5	8.5	7.0	9.0
4	4.5	5.0	3.0	3.5	4.0	5.0	4.5	5.0	5.5	8.5	6.5	7.0	7.0	7.5	7.5	8.5	7.0	8.0
รวม ($X_{ij.}$)	15.0	17.5	11.5	18.0	14.0	17.5	19.0	21.5	22.0	32.0	28.0	28.0	22.0	26.5	29.0	33.0	27.0	35.0
ผลรวมกรรมวิธี ($X_{i..}$)	44.0			49.5			62.5			88.0			77.5			95.0		
ค่าเฉลี่ยกรรมวิธี ($\bar{X}_{i..}$)	3.7			4.1			5.2			7.3			6.5			7.9		

การคำนวณ

(1) $SS(Total) = \Sigma\,\Sigma\,\Sigma\, X_{ijk}^2 - X_{...}^2/trs$

$= (3.5^2 + 4.0^2 + \ldots + 8.0^2) - (416.5)^2/4(3)(6)$

$= 255.91 \quad (71 \text{ df.})$

2. SS(Treatments) = SS(ระหว่างชั่วโมงแสงสว่างและอุณหภูมิ)

$= \Sigma X_{i..}^2/sr \quad - \quad CF.$

$= 1/4(3) \quad \{44.0^2 + \dots + 95.0^2\} \quad - \quad CF.$

$= 179.64$, df. $= (6 - 1) = 5$

3. SS(Units) = SS(Cells) = SS(กระถางภายในวิธีการเดียวกัน)

$= \Sigma\Sigma X_{ij.}^2/s \quad - \quad CF.$

$= 1/4\ (15.0^2 + 17.5^2 + \dots + 35.0^2) - CF.$

$= 205.47$, df. $= (18 - 1) = 17$

SS(units) = SS(treatments) + SS(Experimental error)

4. SS(Experimental error) = SS(Units) - SS(treatments)

= (3) - (2) = 205.47 - 179.64, df. =17-5 = 12

5. SS(total) = SS(Units) + SS(Sampling error)

SS(Sampling error) = SS(Total) = SS(Units)

= 255.91 - 205.47

= 50.44 , df = (71 - 17) = 54

ANOVA

SOV.	df.	SS.	MS.	E(MS)
Treatments (ชั่วโมงแสงสว่าง)	5 (t - 1)	179.64	35.93	$\sigma_\delta^2+4\sigma_\varepsilon^2+12\sigma_\tau^2$ or $12\Sigma\tau_i^2/5$
Experimental error (ระหว่างกระถางภายในวิธีการเดียวกัน)	12 t(r-1)	25.83	2.15	$\sigma_\delta^2 + 4\sigma_\varepsilon^2$
Sampling error (ระหว่างต้นไม้ในกระถางเดียวกัน)	54 tr(s-1)	50.44	0..93	σ_δ^2
Total	71 trs - 1	255.91		

Estimate of variance components

$\hat{\sigma}^2 = S^2 = 0.93$; $\hat{\sigma}_\epsilon^2 = S_\epsilon^2 = (2.15 - 0.93)/4 = 0.30$

$\hat{\sigma}_\tau^2 = S_\tau^2$ (Model 2)$=\Sigma\tau_i^2/5$(Model 1)$=(35.93-2.15)/12 = 2.82$

$Ho : \tau_i = 0$, $Ha : \tau_i \neq 0$, $i = 1,2,\ldots,6$

$F = MST/MSE = 35.93/2.15$; $f_{5,12(.01)} = 5.06$

สรุปว่าการเจริญเติบโตของต้นไม้ภายใต้วิธีการต่าง ๆ มีความแตกต่างกันอย่างมีนัยสำคัญสูงยิ่ง ถ้าต้องการทดสอบความแตกต่างของหน่วยทดลองภายในวิธีการเดียวกัน เมื่อพิจารณาจากค่าคาดหมายของกำลังสองเฉลี่ย วิธีการทดสอบมีดังนี้

$Ho : \sigma_\varepsilon^2 = 0$, $Ha : \sigma_\varepsilon^2 \neq 0$

$F = MSE/MSS = 2.15/0.93 = 2.3$, $f_{12,54(.05)} = 1.92$

สรุปว่าต้นไม้จากกระถางที่ได้รับวิธีการปฏิบัติอย่างเดียวกัน มีการเจริญเติบโตแตกต่างกันด้วยความเชื่อมั่น 95%

Linear Model for Subsampling

$$x_{ijk} = \mu + 7_i + \varepsilon_{ij} + 6_{ijk} , \quad i = 1,2,\ldots,t$$
$$j = 1,2,\ldots,r$$
$$k = 1,2,\ldots,s$$

μ = ค่าเฉลี่ยรวบยอด

τ_i = อิทธิพลของวิธีการที่ i

ε_{ij} = อิทธิพลอื่น ๆ นอกจากกรรมวิธีในหน่วยทดลองที่ j ซึ่งรับวิธีการที่ i

δ_{ijk} = อิทธิพลอื่น ๆ นอกจากกรรมวิธี ในหน่วยตัวอย่างย่อยที่ k ของหน่วยตัวอย่าง j ซึ่งรับวิธีการ i

X_{ijk} = ค่าสังเกตที่เก็บจากหน่วยตัวอย่างย่อยที่ k ในหน่วยตัวอย่างที่ j ของวิธีการ i

ทั้งนี้ยังคงมีข้อสมมุติของ τ_i และ ε_{ij} เช่นเดียวกับกรณีไม่มีตัวอย่างย่อย (ทั้ง 2 แบบจำลอง แต่มีข้อสมมุติของ δ_{ijk} เพิ่มเติม (เหมือนกันทั้ง 2 แบบจำลอง) ดังนี้

δ_{ijk} มีการแจกแจงจงแบบปกติด้วยค่าเฉลี่ย o และความแปรปรวน σ^2 (σ^2_δ) จึงสรุปข้อสมมุติได้ดังนี้

Model 1 : (1) $\Sigma\tau_i = 0$, (2) $\varepsilon_{ij} \sim NID(0, \sigma^2_\varepsilon)$

$i = 1,2,\ldots,t$

(3) $\delta_{ijk} \sim NID(0, \sigma^2_\delta)$

Model 2 : (1) $\tau_i \sim NID(0, \sigma^2_\tau)$, (2) $\varepsilon_{ij} \sim NID(0, \sigma^2_\varepsilon)$

(3) $\delta_{ijk} \sim NID(0, \sigma^2_\delta)$

การวิเคราะห์แบบจำแนกทางเดียว กรณีมีตัวอย่างย่อย และมีจำนวนตัวอย่างย่อยต่างกัน

การคำนวณใช้หลักการเดิมกับกรณีมีตัวอย่างย่อยเท่ากันทุกหน่วย แต่ในการวิเคราะห์ผลจะมีปัญหายุ่งยาก ซึ่งจะตรวจดูได้จากค่าคาดหมายของกำลังสองเฉลี่ย

ตัวอย่าง

ผลผลิตจากต้นไม้ที่สุ่มมาจากพื้นที่ 2 แห่ง ๆ ละ 3 ต้น และแห่งที่ 3 สุ่มมาเพียง 2 ต้น ได้ข้อมูลดังนี้

คุณภาพของผลผลิตจากต้นไม้ 8 ต้น ใน 3 ท้องที่

พื้นที่	A			B			C		
ต้นไม้	1	2	3	4	5	6	7	8	
ค่าสังเกต	6	6,8	6,7,8	5,7	6,7	6	7	7,9	
$X_{ij.}$	6	14	21	12	13	6	7	16	
$X_{i..}$		41			31		23		G = 95
r_{ij}	1	2	3	2	2	1	1	2	

$r_1 = 3,\ r_2 = 3,\ r_3 = 2,$ $N = 14$

(1) $SS(total) = (6^2+6^2+...+7^2+9^2) - (95)^2/14$

$= 14.36$, df. $=N-1=(14-1) = 13$

(2) SS(ห้องที่) = SS(treatments)

$= [(41)^2/6 + (31)^2/5 + (23)^2/3] - CF$

$= 4.07$, df. $= (t-1) = (3-1) = 2$

(3) $SS(Units) = \Sigma\Sigma X_{ij.}^2/r_{ij} - CF.$

$= [6^2+14^2/2+21^2/3+12^2/2+13^2/2+6^2+7^2 16^2/2]-CF.$

$= 5.86$, df.$= (\Sigma r_i-1) = (3+3+2)-1 = 7$ df.

(4) SS(Experimental error) = SS(Units) - SS(Treatments)

$= 5.86 = 4.07 = 1.79$

df. $= (7-2)= 5$ df หรือ $\sum_i^t (r_i-1) = 2+2+1 = 5$

(5) SS(Sampling error) = SS(Total) - SS(Units)

$= 14.36 - 5.86 = 8.50$

df. $= 13 - 7 = 6$

หรือ df. $= \sum_{ij}^{tr_i}(r_{ij}-1) = (0+1+2+1+1+0+0+1)$

$= 6$

ANOVA

SOV.	df.	SS.	MS.	E(MS.)
ระหว่างพื้นที่ (Treatments)	2	4.07	2.03	$\sigma^2+1.90\sigma_\varepsilon^2+4.50\sigma_\tau^2$
ระหว่างต้นไม้/พื้นที่ (Experimental error)	5	1.79	0.36	$\sigma^2+1.64\sigma_\varepsilon^2$
ระหว่างค่าสังเกต/ต้นไม้ (Sampling error)	6	8.50	1.42	σ^2
Total	13	14.36		

$\hat{\sigma}^2 = s^2 = 1.42$; $\hat{\sigma}^2_\varepsilon = (0.36 - 1.42)/1.64 < 0$

สัมประสิทธิ์ของ σ^2_E หาจากสูตร

$$[\, r_{..} - \sum_i(\sum_j r^2_{ij}/r_{i.})]\ /\ \text{df.(experimental error)}$$

$$=\{14- ((1^2+2^2+3^2)/6 + (2^2 + 2^2 + 1^2)/5 + (1^2 + 2^2)/3)\} /5$$

$$= 1.64$$

สัมประสิทธิ์ของ σ^2_ε (วิธีการ) หาจากสูตร

$$= \{\sum_i (\sum_j r^2_{ij}/r_{i.}) - (\sum_{i,j} r^2_{ij})/r_{..} \} \ /\ \text{df(Treatments)}$$

$$= \{(1^2+ 2^2+ 3^2)/6 + (2^2+ 2^2+ 1^2)/5 + (1^2+ 2^2)/3 - (1^2+...+ 2^2)/14\} /2$$

$$= (5.8\ 2)/2 = 1.90$$

สัมประสิทธิ์ของ σ^2_T หาจากสูตร

$$= (r_{..} - \sum_i r^2_{i.}/r_{..})/ \text{df.Treatments}$$

$$= \{ 14 - (6^2+ 5^2 + 3^2)/14 \} /2 = 4.50$$

จะเห็นว่าเมื่อใช้จำนวนตัวอย่างย่อยต่างกัน จะต้องคำนวณสัมประสิทธิ์ของ Variance Components ซึ่งนอกจากจะยุ่งยากแล้ว เมื่อพิจารณาจากช่อง E (MS.) จะเห็นได้ว่าไม่มีค่าสถิติสำหรับทดสอบความแตกต่างของวิธีการ ถ้าต้องการจะทดสอบจริง ๆ จะมีแต่วิธีทดสอบโดยประมาณ ซึ่งจะศึกษาเพิ่มเติมได้จากหนังสือ : Anderson and Bancroft : "Statistical Theory in Research", McGraw-Hill Book Company, Inc., New York, 1952 และ Snedecor, G.M. : Statistical Methods, 5th. ed., Iowa State College Press, Ames, Iowa, 1956.

วิธีพิจารณาอย่างง่ายวิธีหนึ่ง คือการหาค่าประมาณของ σ^2_T

$$\hat{\sigma}^2_T = \{ 2.03 - (1.42 - 1.90(0) \} /4.5 = 0.61/4.5 = 0.135$$

ค่าประมาณเป็นค่าที่น้อย จึงสรุปว่าไม่มีหลักฐานเพียงพอที่จะสนับสนุนความแตกต่างของวิธีการ (ความแตกต่างไม่มีนัยสำคัญ)

ข้อสังเกต เมื่อความคลาดเคลื่อนจากตัวอย่างสุ่มมีมากกว่าความคลาดเคลื่อนของหน่วยทดลอง ทำให้ค่าประมาณของ σ_ε^2 คือ $s_\varepsilon^2 < 0$ จึงให้สมมุติว่า $s_\varepsilon^2 = 0$

การพิจารณาเลือกจำนวนกรรมวิธีและจำนวนซ้ำสำหรับแบบจำลองที่ 2 (ไม่มีตัวอย่างย่อย)

ลองพิจารณาตัวอย่างต่อไปนี้ ซึ่งคือปริมาณแคลเซี่ยมที่วิเคราะห์ได้จากใบคะน้าที่สุ่มมา 4 ใบ แต่ละใบหาค่าวิเคราะห์ 4 ครั้ง (4 ซ้ำ)

ใบผัก	ปริมาณแคลเซียมเป็นเปอร์เซนต์				รวม	ค่าเฉลี่ย
1	3.28	3.09	3.03	3.03	12.43	3.11
2	3.52	3.48	3.38	3.38	13.76	3.44
3	2.88	2.80	2.81	2.76	11.25	2.81
4	3.34	3.38	3.23	3.26	13.21	3.30

SOV .	df.	MS.	Prameter Estimated
ระหว่างใบผัก	3	0.2961	$\sigma_\varepsilon^2 + 4\sigma_\tau^2$
ระหว่างค่าสังเกต (ความคลาดเคลื่อน)	12	0.0066	σ_ε^2

$$\hat{\sigma}_\varepsilon^2 = 0.0066 , \quad \hat{\sigma}_\tau^2 = (0.2961 - 0.0066)/4 = 0.0724$$

ปัญหาคือ ถ้าการหาค่าสังเกต (จำนวนซ้ำ) ต้องเสียค่าใช้จ่ายสูง จึงต้องการเปลี่ยนจำนวนใบผักที่สุ่มมาคือ t เป็น t' และจำนวนค่าสังเกต จาก r เป็น r'

วิธีการเปรียบเทียบว่าการทดลองใดจะดีกว่ากัน ใช้วิธีเปรียบเทียบความแปรปรวนของค่าเฉลี่ยรวมยอด คือ $V(\bar{X}..)$

โดยที่ $V(\bar{X}..) = MS(\text{Treatments})/tr = (\hat{\sigma}_\varepsilon^2 + r\hat{\sigma}_\tau^2)/tr$

ดังนั้น สำหรับงานทดลองใหม่ซึ่งใช้ t' และ r'

$$\hat{V}'(\bar{X}..) = (\hat{\sigma}^2_{\varepsilon} + r'\hat{\sigma}^2_{\tau})/t'r'$$

จะเห็นว่าจากสูตรที่ใช้คำนวณ ความแปรปรวนจะเล็กถ้าเพิ่มจำนวนกลุ่ม และลดจำนวนภายในกลุ่ม (จำนวนซ้ำ) เช่นสมมุติว่า ต้องเสียค่าใช้จ่ายสำหรับค่าวิเคราะห์เป็น 10 เท่าของใบผัก ดังนั้นจึงจะลดค่าวิเคราะห์ลงเป็น $r' = 1$ และเพิ่มจำนวนใบผักเป็น $t' = 15$ ซึ่งรวมแล้วจะเสียค่าใช้จ่ายเท่า ๆ กับการทดลองที่ได้สิ้นสุดลงไปแล้วนั้น อยากทราบว่าแผนใหม่จะดีหรือไม่ ต้องหาค่าประมาณของความแปรปรวนค่าเฉลี่ยรวบยอดมาเปรียบเทียบกัน

$$V(\bar{X}..) = 0.2961/16 = 0.0185$$

สำหรับงานทดลองใหม่

$$\hat{V}'(\bar{X}..) = (\hat{\sigma}^2_{\varepsilon} + \hat{\sigma}^2_{\tau})/15 = (0.0066 + 0.0724)/15 = 0.0053$$

จึงสรุปว่าเพิ่มจำนวนใบผักดีกว่า เพราะให้ความแปรปรวนน้อยกว่า

การพิจารณาจำนวนหน่วยทดลองและจำนวนตัวอย่างย่อย

เช่นเดียวกับการวิเคราะห์แบบจำแนกทางเดียวโดยใช้แบบจำลองที่ 2 คือผู้ทดลองอาจมีปัญหาว่าควรจะจัดสรรจำนวนเงินและเวลาอย่างไร ควรเน้นหนักโดยให้มีจำนวนหน่วยทดลองมาก ๆ และจำนวนตัวอย่างย่อยไม่มาก หรือควรให้มีจำนวนหน่วยทดลองน้อยหน่อยแต่ให้มีจำนวนตัวอย่างย่อยมาก ๆ สิ่งเหล่านี้ขึ้นอยู่กับขนาดของความคลาดเคลื่อนทั้งสอง และค่าใช้จ่ายต่อหน่วยทั้งสอง บางครั้งค่าใช้จ่ายสำหรับหน่วยตัวอย่างย่อยอาจสูง และเสียเวลามาก หรือบางครั้ง กลับกลายเป็นว่า การเพิ่มหน่วยทดลองต้องเพิ่มเครื่องใช้เครื่องมือด้วย แปลงทดลองหรือ สัตว์ทดลอง เป็นต้น แต่การเพิ่มตัวอย่างย่อยกลับไม่กระทบกระเทือน ข้อสังเกตที่พบจากการทดลองทั่ว ๆ ไป มักพบว่า การเพิ่มตัวอย่างย่อยจะเสียค่าใช้จ่ายน้อยกว่าการเพิ่มหน่วยทดลอง

นอกจากนี้ยังต้องพิจารณาขนาดของความแปรปรวนด้วย สมมุติจากการทดลองเรื่องการเจริญเติบโตของพืชภายใต้อุณหภูมิและชั่วโมงแสงสว่างต่าง ๆ กันนั้น สมมุติว่าไม่พิจารณาเรื่องค่าใช้จ่ายและเวลา จะพิจารณาเฉพาะเรื่องความแปรปรวน โดยสมมุติว่าถ้าทำการทดลองครั้งต่อไปจะได้ส่วนประกอบของความแปรปรวนแบบเดิม เพื่อจะได้ใช้ค่าประมาณที่หาได้จากการทดลองที่ได้สิ้นสุดลงไปแล้ว จากการทดลองเดิม ใช้หน่วยทดลอง 3 กระถางต่อวิธีการ และตัวอย่างย่อยคือต้นไม้ 4 ต้น/กระถาง สมมุติงานทดลองใหม่ จะเปลี่ยนเป็น 2 กระถางต่อวิธีการ และเพิ่มเป็นต้นไม้ 6 ต้น/กระถาง ซึ่งรวมแล้วจะได้ต้นไม้ 12 ต้น/วิธีการเท่าเดิม ลองพิจารณากำลังสองเฉลี่ยของความคลาดเคลื่อน ซึ่งเป็นค่าประมาณของ

$$\sigma^2_{\delta} + s\sigma^2_{\varepsilon} \quad ; \quad MSE = 2.15 \quad ; \quad \hat{\sigma}^2_{\varepsilon} = .30 \;, \quad \hat{\sigma}^2_{\delta} = 0.93$$

ดังนั้นสำหรับงานทดลองใหม่

$$\hat{MSE}' = \hat{\sigma}^2_{\delta} + s'\hat{\sigma}^2_{\varepsilon} = 0.93 + 6(0.30) = 2.73$$

เนื่องจากจำนวนค่าสังเกตต่อวิธีการเท่ากันทั้ง 2 แบบ คือ 12 ค่าสังเกต จึงเอา MSE เปรียบเทียบกันได้

relative efficiency = (2.73/2.15)100% = 127%

หมายความว่า สำหรับงานทดลองใหม่ ถ้าจะให้มีประสิทธิภาพเท่ากับอันเดิม จะต้องเพิ่มต้นไม้อีก 27% หรือมากกว่า จึงจะสามารถตรวจพบความแตกต่างได้ขนาดเดียวกับการทดลองเดิม พึง สังเกตว่า ความแปรปรวนของค่าเฉลี่ยกรรมวิธีสำหรับงานทดลองเดิม คือ

$$s^2_{\bar{x}} = 2.15/3(4) = .179$$

สำหรับงานทดลองใหม่ เมื่อเพิ่ม ต้นไม้อีก 27% คือ 27% ของ 12 = 3 เมื่อรวมกับของเดิมเป็น 12 + 3 = 15 ต้น/วิธีการ จะให้ความแปรปรวนของค่าเฉลี่ย

$$s^2_{\bar{x}} = 2.73/(12 + 3) = .182$$

ความแตกต่างของความแปรปรวนเนื่องจากการปัดเศษต้องคำนวณ 27% ของ 12 ซึ่งไม่ใช้ 3 พอดี สมมุติถ้าเปลี่ยนใหม่เป็นต้นไม้ 2 ต้น/กระถาง และ 6 กระถาง/วิธีการ

$$\hat{MSE}'' = .93 + 2(.30) = 1.53$$

Relative efficiency = (1.53/2.15) = 71 %

71(12) = 8.52 ต้น

นั่นคือ ถ้าใช้ต้นไม้ 8 ต้น/วิธีการก็จะให้ประสิทธิภาพด้อยกว่าแบบเดิมนิดหน่อย แต่ถ้าใช้ต้นไม้ 9 ต้น/วิธีการ ก็จะให้ประสิทธิภาพสูงกว่าแบบเดิมนิดหน่อย

ในกรณีที่จำนวนตัวอย่างย่อย/วิธีการต่างกัน ให้เปรียบเทียบประสิทธิภาพโดยเอาความแปรปรวนของค่าเฉลี่ยกรรมวิธีมาเปรียบเทียบกัน

เช่นสมมุติว่างานทดลองใหม่จะใช้ต้นไม้ 2 ต้น/กระถาง และ 5 กระถาง/วิธีการ

$$\hat{MSE}' = .93 + 2(0.30) = 1.53$$

$$s^{2'}_{\bar{x}} = \hat{MSE}'/r's' = 1.53/(5)(2) = 0.153$$

Relative efficiency = (0.153)/0.179 = 85.47 %

หมายความว่า ถ้างานทดลองเดิมใช้ต้นไม้ 100 ต้น งานทดลองใหม่จะใช้ต้นไม้เพียง 86 ต้น แต่จะให้ประสิทธิภาพเท่างานทดลองเดิม แสดงว่าแผนงานทดลองใหม่ดีกว่า จึงมีข้อสังเกตว่าถ้าไม่พิจารณาปัจจัยอื่น ๆ เช่นค่าใช้จ่ายและเวลา จะดูแต่ความแปรปรวนจะพบว่าการเพิ่มหน่วยทดลอง ดีกว่าการเพิ่มตัวอย่างย่อย ข้อนี้จะเห็นชัดจากสูตรที่ใช้คำนวณความแปรปรวนของค่าเฉลี่ยของวิธีการ

ข้อสังเกตสำหรับข้อสมมุติของการวิเคราะห์ความแปรปรวน

ในการวิเคราะห์ความแปรปรวน เมื่อต้องการทดสอบนัยสำคัญ จะต้องมีข้อสมมุติเบื้องต้นดังนี้

1. อิทธิพลของกรรมวิธีและสภาพแวดล้อมเป็นแบบเชิงบวก
2. ความคลาดเคลื่อนจากการทดลองต้องเป็นแบบสุ่ม เป็นอิสระกัน และมีการแจกแจงแบบปกติด้วยค่าเฉลี่ยศูนย์ และมีความแปรปรวนเป็นเอกภาพ

สำหรับคุณสมบัติว่าต้องมีการแจกแจงแบบปกตินั้นไม่จำเป็น นอกจากเมื่อจะทดสอบนัยสำคัญ แต่ถ้าประมาณส่วนประกอบของความแปรปรวนแล้วไม่จำเป็น ถ้าข้อสมมุติเหล่านี้ไม่จริง จะมีผลต่อความฉับไวในการทดสอบนัยสำคัญของค่าสถิติ F หรือ T

ในกรณีที่การแจกแจงไม่เป็นแบบปกติ จะทำให้ระดับนัยสำคัญที่ใช้ทดสอบสูงกว่าระดับจริงที่ตั้งไว้ จึงมีผลทำให้ปฏิเสธ Ho บ่อยกว่าความจริง นั่นคือความแตกต่างที่ไม่มีนัยสำคัญจะถูกสรุปว่ามีนัยสำคัญบ่อยครั้งขึ้น เช่นสมมุติว่า ตั้งระดับนัยสำคัญไว้ 5% แต่ความจริงกลายเป็น 7% หรือ 8%

สำหรับคุณสมบัติที่ว่า อิทธิพลของกรรมวิธีและสภาพแวดล้อมเป็นแบบเชิงบวก เพราะถ้าไม่เป็นแบบเชิงบวก มักพบว่าเป็นแบบทวิคูณ ดังตัวอย่างในตารางข้างล่าง

แบบจำลอง	แบบเชิงบวก		แบบทวิคูณ		$\log_{10}$ (ทำให้แบบทวิคูณเป็นเชิงบวก)	
บล็อค	1	2	1	2	1	2
วิธีการที่ 1	10	20	10	20	1.00	1.30
วิธีการที่ 2	30	40	30	60	1.48	1.78

จะเห็นว่าในแบบจำลองที่เป็นเชิงบวก การเพิ่มจากบล็อค 1 ไป บล็อค 2 เป็นจำนวนคงที่ไม่ว่าจะเป็นวิธีการใด จะเท่ากันตลอด แต่สำหรับแบบทวิคูณ การเพิ่มขึ้นจากบล็อคที่ 1 ไปบล็อค 2 เป็นการเพิ่มขึ้นในเปอร์เซนต์คงที่ไม่ว่าจะเป็นกรรมวิธีใด ส่วนการเพิ่มของวิธีการที่ 1 มาเป็นวิธีการที่ สำหรับแบบเชิงบวกจะเพิ่มในจำนวนคงที่ แต่สำหรับแบบทวิคูณจะเพิ่มในเปอร์เซนต์

คงที่ สำหรับแบบจำลองเชิงทวิคูณจะแปลงให้เป็นเชิงบวกได้โดยการใช้ค่า $\log_{10}$ เรียกว่า "การแปลงค่า" ข้อมูลที่แปลงค่าแล้วจะให้ความคลาดเคลื่อนเป็นอิสระและมีการแจกแจงแบบปกติ ถ้าข้อมูลไม่เป็นแบบเชิงบวก จะมีผลทั่วให้ MSE ไม่ใช่ค่าประมาณของ σ_E^2 (Pooled variance)

สำหรับข้อสมมุติว่าความคลาดเคลื่อนต้องเป็นอิสระกันนั้น เหตุที่จำเป็นต้องมีข้อสมมุติข้อนี้เพราะในการทดลองทางการเกษตร จะพบว่า แปลงทดลองที่อยู่ใกล้เคียงกันจะมีแนวโน้มที่จะให้ผลตอบสนองใกล้เคียงกันมากกว่าแปลงทดลองที่อยู่ห่างไกลออกไป หรือในการทดลองในห้องทดลองก็เช่นกัน จะพบว่า ค่าสังเกตที่ได้จากคนเดียวกันในเวลาที่ใกล้เคียงกันจะไม่ต่างกันมากในทางปฏิบัติเพื่อขจัดปัญหานี้อาจทำได้โดยการจัดกรรมวิธีให้หน่วยทดลอง**แบบสุ่ม** นั่นคือ **การสุ่มกรรมวิธีทำให้ความคลาดเคลื่อนเป็นอิสระกัน**

สำหรับข้อสมมุติว่า "ความคลาดเคลื่อนต้องมีการแจกแจงแบบปกติ" ซึ่งใช้เฉพาะเมื่อต้องการทดสอบนัยสำคัญ เพราะเมื่อการแจกแจงมีลักษณะเบ้จะมีผลให้ความคลาดเคลื่อนของแต่ละวิธีการมีความแปรปรวนไม่เป็นเอกภาพ (heterogeneity of the error term)

คุณสมบัติที่ว่า "ความคลาดเคลื่อนต้องมีความแปรปรวนร่วมกัน (common variance) เพราะถ้าความแปรปรวนแตกต่างกัน จะทดสอบความแตกต่างของวิธีการไม่ได้ เช่นการทดลองยาฆ่าแมลงหลายสูตร ถ้าวิธีการหนึ่งคือ "ไม่ใช้ยาฆ่าแมลง" ซึ่งเท่ากับทำหน้าที่เป็น "Control" ผลการทดลองจะพบว่าค่าสังเกตในกลุ่ม "ควบคุม" จะมีความแปรปรวนสูงกว่าวิธีการอื่น ๆ (และมีค่าเฉลี่ยสูงกว่าด้วย) กรณีนี้ความคลาดเคลื่อนอาจไม่เป็นเอกภาพ เมื่อจะทดสอบ จะต้องแบ่งความคลาดเคลื่อนสำหรับการทดสอบแต่ละอัน (contrast) หรือวิธีแก้อีกวิธีคือให้ตัดค่าเฉลี่ยที่สูงมากออกไป (ตัดออกทั้งวิธีการ แต่ไม่ควรเกิน 2 วิธีการ)

พลังการทดสอบ (Power of the Analysis of Variance)

เมื่อมีการทดสอบสมมติฐาน ผู้ทดลองควรสนใจพลังการทดสอบของแบบทดสอบนั้นด้วย เราจะทราบพลังการทดสอบโดยการศึกษาจากเส้น operating characteristic curves หรือเรียกสั้น ๆ ว่า เส้น OC เพื่อเป็นแนวทางให้ผู้ทดลองทราบว่าควรจะให้วิธีการมีจำนวนซ้ำเท่าใดจึงจะเพียงพอ ซึ่งจะต้องขึ้นอยู่กับตัวแบบที่ใช้ด้วย

1. พลังการทดสอบเมื่อใช้ตัวแบบกำหนดและทุกวิธีการมีจำนวนซ้ำเท่ากัน (power of the fixed effects model)

พลังการทดสอบ (power of the test) คือ

$$1 - \beta = P\{\text{ปฏิเสธ } Ho/Ho \text{ เป็นเท็จ}\}$$

$$= P\{F_o > f_{\alpha,t-1,N-t} \;/\; Ho \text{ เป็นเท็จ}\}$$

ทั้งนี้เพราะ เมื่อ Ho เป็นเท็จ ตัวสถิติ

$$F_o = \frac{MS\ (\text{วิธีการ})}{MS_E}$$ จะมีการแจกแจงแบบ **noncentral F**

ด้วย $\nu_1 = (t-1)$, $\nu_2 = N-t$ และมี noncentrality parameter δ
ถ้า $\delta = 0$ การแจกแจงแบบ noncentral F จะกลายเป็นการแจกแจงแบบ (central) F (ตามปกติ)

เราจะดูเส้น OC ทั้งหลายได้จาก Chart V ในตารางท้ายเล่มเส้น OC จะพล็อต ความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 2 ที่เรียกว่า type II error (β) กับ พารามิเตอร์ Φ ในเมื่อ

$$\Phi^2 = \frac{r\sum_{i=1}^{t}\tau_i^2}{t\sigma^2}$$ **noncentrality parameter**

Φ^2 จะมีความสัมพันธ์กับ noncentrality parameter δ การใช้เส้น OC ต้องเลือก α ด้วยระหว่าง $\alpha = .05$ และ $\alpha = .01$ และต้องทราบค่า ν_1 , ν_2 โดยที่ $\nu_1 = t-1$, $\nu_2 = N-t$ และยังต้องกำหนดขนาดความแตกต่างระหว่างค่าเฉลี่ยของวิธีการที่ต้องการจะตรวจจับ (detecting) ใน รูปของ $\sum_{i=1}^{t}\tau_i^2$ ส่วน σ^2 นั้นเรามักไม่ทราบค่าที่แท้จริง ดังนั้นเราอาจกำหนดค่าของมันในรูปอัตราส่วน คือกำหนดอัตราส่วน $\sum_{i=1}^{t}\tau_i^2/\sigma^2$ ที่ต้องการจะตรวจจับ หรือจะใช้ค่าประมาณของ σ^2 คือ MS_E ก็ได้

ตัวอย่าง จะเปรียบเทียบค่าเฉลี่ย 5 จำนวน โดยใช้ ANOVA, $\alpha = .01$ ต้องการทราบจำนวนซ้ำต่อวิธีการ เพื่อจะได้ปฏิเสธ Ho ด้วยความน่าจะเป็นอย่างน้อยที่สุด .90 ถ้า

$$\sum_{i=1}^{5}\tau_i^2/\sigma^2 = 5.0$$

$$\Phi^2 = \frac{r\sum^{t}\tau_i^2}{t\sigma^2} = \frac{r}{5}(5) = r$$

$$\nu_1 = t-1 = 5-1 = 4$$

$$\nu_2 = N-t = t(r-1) = 5(r-1)$$

เปิด Chart V โดยสมมุติค่า r หลาย ๆ ค่า เพื่อหาค่า ν_2 เช่นเริ่มจาก $r=4$, $\Phi^2 = 4$ $\Phi = 2$, $\nu_2 = 5(3) = 15$ จากเส้น OC จะได้ $\beta \simeq .38$ ดังนั้น พลังการทดสอบหรือ $1 - \beta = 1 - .38 = .62$ ซึ่งยังน้อยกว่า .90 แสดงว่าการใช้ $r = 4$ ยังไม่เพียงพอ ต้องเพิ่มค่า r อีก จะได้ผลดังนี้

r	Φ^2	Φ	t(r-1)	β	$(1 - \beta)$ = พลังการทดสอบ
4	4	2.00	15	.38	.62
5	5	2.24	20	.18	.82
6	6	2.45	25	.06	.94

ดังนั้น จะต้องใช้ $r = 6$ จึงจะได้พลังการทดสอบตามต้องการ

2. พลังการทดสอบเมื่อใช้ตัวแบบเชิงสุ่ม

(power of the test for the random effects model)

$$1 - \beta = P\{\text{ปฏิเสธ } H_o \,/\, H_o \text{ เป็นเท็จ}\}$$

$$= P\{F_o > f_{\alpha, t-1, N-t} \,/\, \sigma_\tau^2 > 0\}$$

โดยที่ $F_o = MS\,(\text{วิธีการ}) \,/\, MS_E$

ถ้า H_1 เป็นจริง $\sigma_\tau^2 > 0, F_o$ จะมีการแจกแจงแบบ central F ด้วย $\nu_1 = t-1$, $\nu_2 = N-t$

จะเห็นว่า พลังการทดสอบของ random model ขึ้นอยู่กับการแจกแจงแบบ central F ดังนั้น เรายังคงใช้ตารางการแจกแจงของ F ตามปกติ อย่างไรก็ตามถ้าใช้เส้น OC จะง่ายกว่า วิธีการใช้ก็เหมือนกับกรณี fixed effects model แต่เปลี่ยนเป็น Chart VI และต้องหาพารามิเตอร์

$$\lambda = \sqrt{1 + \frac{r\sigma_\tau^2}{\sigma^2}}$$

ถ้าไม่มีค่าประมาณของ σ^2 ให้กำหนดในรูป σ_τ^2/σ^2

ตัวอย่าง สมมุติมี 5 วิธีการซึ่งเลือกมาแบบสุ่ม แต่ละกลุ่มมีค่าสังเกต 6 ค่า $\alpha = .05$ พลังการทดสอบ ถ้า $\sigma_\tau^2 = \sigma^2$

$$a = 5 \ , \quad r = 6 \ , \quad \sigma_\tau^2 = \sigma^2$$

$$\lambda = \sqrt{1 + 6(1)} = 2.646$$

จาก Chart VI , $\nu_1 = t-1 = 4$, $\nu_2 = n-t = 25$, $\alpha = .05$

จะได้ $\beta \simeq .20$

ดังนั้น พลังการทดสอบจะประมาณ .80

แบบฝึกหัด

1. จงวิเคราะห์ข้อมูลของกรรมวิธีต่าง ๆ ต่อไปนี้

ก	ข	ค
27	22	37
45	24	38
44	42	25
31	41	47
38	31	23

MSTr. = 31.67

MSE = 82.67

2. สุ่มหนู 20 ตัวให้เข้ารับการทดลอง 2 วิธี (วิธีละ 20 ตัว) หนูกลุ่มหนึ่งให้เป็น "กลุ่มควบคุม" คือทิ้งให้อดอาหารระยะหนึ่ง อีกกลุ่มเป็นกลุ่ม "เข้ารับการทดลอง" คือให้อยู่ในกรงซึ่งกั้นทางเดินไว้พิเศษเหมือนเขาวงกต จะมีทางออกได้ทางเดียว ให้กลุ่มทดลองอยู่ในกรงดังกล่าวก่อน 12 ชั่วโมงและให้อดอาหารระยะหนึ่งด้วย เมื่อปล่อยหนูทั้ง 2 กลุ่ม ในกรงดังกล่าว แล้วนับจำนวนครั้งที่วิ่งพบจุดบอดจนถึงครั้งสุดท้ายที่วิ่งรวดเดียวออกประตูได้ จงทดสอบความแตกต่างด้วยการทดสอบแบบ "ที" และ การวิเคราะห์ความแปรปรวน พึงสังเกตว่า (T = 1.74)

Control							Experimental						
10	7	9	6	8	6	10	12	7	9	6	5	9	9
13	9	7	12	12	15	6	9	6	4	8	4	9	10
9	11	9	13	4	9		11	6	9	7	10	7	

3. ในการศึกษาคุณภาพของ molasses โดยวัดดีกรีของค่า Brix จาก molasses ที่สุ่มมาจาก 3 ท้องที่

ก. จงทดสอบว่าท้องที่ทั้ง 3 ให้ค่าดีกรี Brix เท่ากันหรือไม่

ข. ท่านจะยอมรับสมมุติฐานว่าดีกรี Brix มีค่าเป็น 82 สำหรับ

1. ทุกท้องที่รวมกัน
2. แต่ละท้องที่

ห้องที่

1	2	3
81.6	81.8	82.1
81.3	84.7	79.6
82.0	82.0	83.1
79.6	85.6	80.7
78.4	79.9	81.8
81.8	83.2	79.9
80.2	84.1	82.6
80.7	85.0	81.9

(MSTr. = 14.14 , MSE. = 2.30 , $S_{\bar{x}} = \sqrt{2.30/8} = 0.54$ Brir degree

location	1	2	3
estimated mean	80.70	83.29	81.46)

4. กำหนดให้ข้อมูลจากตัวอย่างที่สุ่มจากประชากร 4 กลุ่ม มีข้อมูลสรุปในตารางข้างล่าง จงทดสอบว่ามาจากประชากรเดียวกันหรือไม่ ?

	1	2	3	4
mean ($\bar{X}_i$)	0.261	0.296	0.312	0.135
S_i	0.21	0.17	0.19	0.08
replicate	44	44	44	44

5. เปอร์เซนต์การดูดซึมของแผ่นคอนกรีตที่ผลิตจากส่วนผสมต่าง ๆ 5 สูตร มีดังนี้ จงทดสอบว่าความแตกต่างของการดูดซึมมีนัยสำคัญหรือไม่ ?

สูตร	ก	ข	ค	ง	จ	รวม
	6.7	5.1	4.4	6.7	6.5	
	5.8	4.7	4.9	7.2	5.8	
	5.8	5.1	4.6	6.8	4.7	
	5.5	5.2	4.5	6.3	5.9	

(F = 2.79/0.213
= 13.1)

6. บริษัทขายยาแก้ปวด 4 บริษัทต่างโฆษณาว่ายาแก้ปวดของตนสามารถแก้ปวดได้รวดเร็วกว่ายาชนิดอื่น ได้มีการนำยาทั้ง 4 ชนิดมาทดสอบกับผู้ทดลองจำนวนหนึ่งแล้วบันทึกเวลาเป็นวินาทีที่ทำให้อาการปวดหายไป 50% จงทดสอบว่ายาแก้ปวดเหล่านี้มีอิทธิพลไม่ต่างกันอย่างมีนัยสำคัญ

บริษัท (ก)	บริษัท (ข)	บริษัท (ค)	บริษัท (ง)
28	34	29	22
19	23	24	31
30	20	33	18
25	16	21	24

7. สมาคมผู้บริโภคเครื่องอบผ้าด้วยแกสจาก 3 ยี่ห้อมายี่ห้อละ 5 เครื่อง และบันทึกเวลาที่ใช้อบผ้า 1 กิโลกรัม จนแห้ง จงทดสอบประสิทธิภาพของเครื่องอบ

ก	ข	ค
42	52	38
36	48	44
47	43	33
43	49	35
38	51	32

8. สุ่มยางรถยนต์จาก 6 ยี่ห้อมา ๆ ละ 4 เส้น แล้วบันทึกระยะทางห้ามล้อเป็นฟุตในขณะขับความเร็ว 25 ไมล์/ชั่วโมง จงทดสอบความแตกต่างของยางเหล่านี้

ก	ข	ค	ง	จ	ฉ
22	25	17	21	27	20
20	23	19	24	29	14
24	26	15	25	24	17
18	22	18	23	25	15

9. คะแนนสอบของนักเรียน 2 กลุ่ม มีดังนี้

กลุ่มที่ 1 :	12	26	31	12	14	16	10	;	$\bar{X}_1 = 17.29$
กลุ่มที่ 2	8	14	29	7	14	6		;	$\bar{X}_2 = 13.00$

จงทดสอบ Ho : $\mu_1 = \mu_2$

10. ผู้ผลิตสินค้าได้สุ่มสินค้าตัวอย่างทีละ 100 ชิ้นจากวิธีการผลิต 3 วิธี แล้วนับสินค้าที่ชำรุดได้ข้อมูลข้างล่าง จงทดสอบว่าข้อมูลเหล่านี้มาจากประชากรที่มีค่าเฉลี่ยเท่ากันหรือไม่

process 1 :	2	2	7	2	5	4	3	$\Sigma X_1 = 25$, $\Sigma X_1^2 = 111$
process 2 :	7	3	7	9	4	5	3	$\Sigma X_2 = 38$, $\Sigma X_2^2 = 238$
process 3 :	8	4	5	9	10	11	9	$\Sigma X_3 = 56$, $\Sigma X_3^2 = 488$

(F = 34.62/5.19 = 6.67)

11. สุ่มอิฐที่ผลิตได้มา 20 ก้อน แล้วนำไปเก็บในที่เก็บ 4 ประเภท แบบสุ่มในระหว่างการทดลองปรากฏว่าเกิดความผิดพลาดทำให้อิฐเสียไป 6 ก้อน เมื่อเก็บไว้ครบ 1 สัปดาห์ในนำอิฐเหล่านั้นมาหาปริมาณน้ำคิดเป็นเปอร์เซนต์ ดังนี้

วิธีเก็บ	ปริมาณน้ำ				
1	7.3	8.3	7.6	8.4	8.3
2	5.4	7.4	7.1		
3	8.1	6.4			
4	7.9	9.5	10.0	7.1	

ก. จงเขียนแบบจำลองที่เหมาะกับข้อมูลนี้

ข. จงทดสอบว่าปริมาณน้ำจากวิธีเก็บต่าง ๆ มีความแตกต่างอย่างมีนัยสำคัญหรือไม่

$(X_{ij} = \mu + \tau_i + \varepsilon_{ij})$, F = 2.53/1,03 = 2.46 ns

12. ข้อมูลต่อไปนี้ได้จากการทดลองแบบสุ่มสมบูรณ์เพื่อเปรียบเทียบคุณภาพของสี 4 ชนิด ในด้านการสะท้อนแสง ได้สุ่มแผ่นทดลองให้กับสีต่าง ๆ ได้ข้อมูลจากเครื่องวัดโดยเฉพาะ ดังนี้

สี	1	2	3	4
	195	45	230	110
	150	40	115	55
	205	195	235	120
	120	65	225	50
	160	145		80
		195		

ก. จงหาค่าประมาณของ residual variance

ข. คุณภาพของสีเหล่านี้แตกต่างกันหรือไม่

ค. จงสร้าง 95% ช่วงเชื่อมั่นของค่าเฉลี่ยของสีชนิดที่ 1

13. กำหนดตารางวิเคราะห์ความแปรปรวนให้

ก. จงเขียนแบบจำลอง และอธิบายความหมายของแต่ละเทอม

ข. จงทดสอบว่ากรรมวิธีทั้ง 6 มีค่าเฉลี่ยของประชากรไม่ต่างกัน

ค. จงหาความแปรปรวนของค่าเฉลี่ยของกรรมวิธี

ง. ถ้าค่าเฉลี่ยของวิธีการที่ 3 เป็น 193.7 จงสร้าง 95% ช่วงเชื่อมั่นของค่าเฉลี่ยของประชากรของวิธีการที่ 3

จ. สมมุติว่าส่วนประกอบของความแปรปรวนคงเดิม ถ้าจะเปลี่ยนไปใช้ 9 หน่วยทดลอง/วิธีการ และ 4 ตัวอย่างย่อย/หน่วยทดลอง แผนใหม่นี้จะดีกว่าหรือด้อยกว่าแผนเดิม

SOV	df.	MS.	E(MS)
Treatments	5	12489	$\sigma_\delta^2 + 3\sigma^2 + 30/5\ \Sigma\tau_i^2$
Exp.units within treatments	54	3349	$\sigma_\delta^2 + 3\sigma^2$
Determination per exp.unit	120	627	σ_δ^2

14. ผู้ทดลองปลูกข้าว 5 สายพันธุ์โดยสุ่มปลูกในแปลงทดลองที่มีอยู่ 30 แปลง ชนิดละ 6 แปลง แต่เขาเก็บเกี่ยวข้าวจากพื้นที่ที่สุ่มมาขนาด 3 คูณ 8 ฟุต จำนวน 8 แห่ง/แปลง และนำข้าวจากพื้นที่ตัวอย่างมาหาค่าวิเคราะห์เคมี 3 ค่าวิเคราะห์/พื้นที่ตัวอย่าง จงแสดงการแบ่ง df. และแสดงค่าคาดหมายของกำลังเฉลี่ย และหาอัตราส่วนของ F เพื่อใช้ทดสอบความแตกต่างของวิธีการ

15. จากตารางวิเคราะห์ข้างล่าง

ก. จงหาค่าคาดหมายของกำลังสองเฉลี่ย โดยสมมุติว่าสนใจเฉพาะกลุ่มที่อยู่ในการทดลองแต่หน่วยทดลอง และค่าวิเคราะห์เป็นแบบสุ่ม

ข. ทดสอบความแตกต่างของค่าเฉลี่ยประชากร และอธิบายความหมาย

ค. หาความแปรปรวนของค่าเฉลี่ยของวิธีการ

SOV	df.	MS.
Groups	3	600
Experimental units within groups	36	120
Determination per experimental unit	80	12

16. จงหาค่าประมาณของส่วนประกอบของความแปรปรวน (Components of variance) และอธิบายความหมายของแต่ละเทอมในแบบจำลอง $Y_{ijk} = \mu + \tau_i + \varepsilon_{ij} + \delta_{ijk}$ พร้อมทั้งแสดงข้อสมมุติของแบบจำลองที่ 2

SOV.	df.	MS.
Among treatments	4	20
Experimental units with trts.	15	15
Determination per experimental unit	20	4

17. จากตารางวิเคราะห์ข้างล่าง

ก. จงหาความแปรปรวนของค่าเฉลี่ยกรรมวิธี

ข. จงทดสอบ Ho : $\sigma_\tau^2 = 0$ และอธิบายข้อสรุป

ค. ถ้าค่าเฉลี่ยของวิธีการที่ 1 เป็น 80 จงหา 95% ช่วงเชื่อมั่นของค่าเฉลี่ยของประชากร

วิธีการที่ 1

SOV.	df.	SS	MS	E(MS)
Treatments	3	1800	600	$\sigma_\delta^2 + 3\sigma^2 + 30\sigma_\tau^2$
Experimental units/trts	36	3600	100	$\sigma_\delta^2 + 3\sigma^2$
Det. per expt.units	80	960	12	σ_δ^2

18. จากตารางวิเคราะห์ข้างล่าง

ก. จงทดสอบความแตกต่างของวิธีการและอธิบายข้อสรุป

ข. หาค่าคาดหมายของกำลังสองเฉลี่ย

ค. หาความแปรปรวนของค่าเฉลี่ยกรรมวิธี

ง. เปรียบเทียบประสิทธิภาพกับแผนทดลองใหม่ซึ่งจะใช้ 10 หน่วยทดลอง/วิธีการ และค่าวิเคราะห์ 1 ค่าวิเคราะห์/หน่วยทดลอง

SOV.	df.	ms.
Treatments	4	960
Experimental units within trts.	35	320
Determination per experimental unit	40	20

19. ถ้าแบ่งประชากรในกรุงเทพฯ เป็น 5 ภาค แต่ละภาคสุ่ม 5 อำเภอ แต่ละอำเภอสุ่มมา 4 บล็อค (ชุมรุมอาคาร) และแต่ละบล็อคสุ่มบ้านตัวอย่าง (ครัวเรือนตัวอย่าง) มา 3 ครัวเรือน เพื่อใช้ศึกษาเรื่องรายได้ของหัวหน้าครอบครัว จงแสดงการแบ่ง df. ในตารางวิเคราะห์

20. จงแสดงการแบ่ง df. ของงานทดลองต่อไปนี้

ฉีดฮอร์โมนสเปรย์ให้แก่ต้นผลไม้ในสวนซึ่งมีทั้งหมด 100 ต้นแบบสุ่มสมบูรณ์ ทุก ๆ วิธีการเวลาฉีดใช้ฉีดพร้อมกันที่ละ 2 ต้น (4 วิธีการใช้กับต้นไม้ 8 ชุด ที่เหลืออีก 2 วิธีการใช้กับต้นไม้ 9 ชุด) แล้วเก็บผลผลิตตัวอย่างมาต้นละ 4 ตัวอย่าง

21. ในการเปรียบเทียบพื้นที่เกษตร 8 ห้องที่ ได้สุ่มอำเภอตัวอย่างมาท้องที่ละ 5 อำเภอ แต่ละอำเภอสุ่มมา 20 ฟาร์มเพื่อศึกษา "รายได้ของฟาร์ม" ได้ข้อมูลสรุปดังนี้ SS (Total)

corrected = 8183000 , SS (ระหว่างอำเภอ/ท้องที่) = 352000 MS (ระหว่างพื้นที่เกษตร) = 33000

ก. จงสร้างตารางวิเคราะห์ความแปรปรวนของ "รายได้ของฟาร์ม"

ข. จงหาความแปรปรวนของค่าเฉลี่ยของพื้นที่เกษตร

ค. สมมุติว่าเลือกสุ่มมาเพียง 2 อำเภอ/ท้องที่ และ 50 ฟาร์ม/อำเภอ จะให้ประสิทธิภาพเมื่อเทียบกับแผนทดลองเดิมอย่างไร

ง. ถ้าพื้นที่เกษตร 8 ท้องที่นั้นได้มาแบบเจาะจง อยากทราบว่า "รายได้ของฟาร์ม" แตกต่างตามท้องที่หรือไม่

22. กำหนดตารางวิเคราะห์ความแปรปรวนของการทดลองแบบสุ่มสมบูรณ์ที่มี 4 วิธีการ 8 หน่วยทดลอง/วิธีการ 3 ตัวอย่างย่อย/หน่วยทดลอง และ 2 ค่าวิเคราะห์/ตัวอย่างย่อย ดังนี้

SOV.	df.	MS.
Treatments	3	19200
Among experimental units alike	28	4800
Among samples per experimental unit	64	2400
Between determinations per sample	96	1200

จงพิจารณาว่า ถ้าเปลี่ยนไปใช้ 12 หน่วยทดลอง/วิธีการ 2 ตัวอย่างย่อย/หน่วยทดลอง และ 1 ค่าวิเคราะห์/ตัวอย่างย่อย จะมีผลในการประมาณค่าอิทธิพลของวิธีการอย่างไรเมื่อเทียบกับแบบเดิม

23. นำแบตเตอรี่ที่คุณภาพเหมือนกัน 30 ลูกมาสุ่มเพื่อนำไปใช้งานในระดับอุณหภูมิที่ต่างกัน 5 ระดับ แล้วบันทึกอายุการใช้งานเป็นวินาทีได้ดังนี้

0°c.	25°c.	50°c.	75°c.	100°c.
55	60	70	72	65
55	61	72	72	66
57	60	73	72	60
54	60	68	70	64
54	60	77	68	65
56	60	77	69	65

จงวิเคราะห์ข้อมูลข้างบนนี้

24. ถ้าประชากรปกติ 4 ประชากร มีความแปรปรวนเท่ากัน คือ $\sigma^2 = 25$ และมี $\mu_1 = 50$, $\mu_2 = 60$, $\mu_3 = 50$, $\mu_4 = 60$ จะต้องใช้ค่าสังเกตกลุ่มละกี่จำนวน จึงทำให้ความน่าจะเป็นที่ปฏิเสธ $Ho : \tau_i = 0$ มีค่าอย่างน้อยที่สุด .90 เมื่อใช้ $\alpha = .05$
