บทที่ 9
การวิเคราะห์ความแปรปรวน

ในการทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ยของประชากร 1 กลุ่ม และ 2 กลุ่ม ตัวสถิติทดสอบที่ใช้คือ Z หรือ T โดยการเลือกใช้ตัวสถิติทดสอบใดขึ้นอยู่กับว่าทราบความแปรปรวนของข้อมูลในประชากรนั้นหรือไม่ ตัวอย่างมีขนาดใหญ่ หรือเล็ก แต่ในกรณีที่ทำการศึกษาประชากรมากกว่า 2 กลุ่ม และต้องการทดสอบสมมติฐานว่าค่าเฉลี่ยของประชากรแต่ละกลุ่มนั้นแตกต่างกันหรือไม่จะต้องทดสอบสมมติฐานทีละคู่ เช่นในการทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ยของประชากร 3 กลุ่ม จะต้องทำการทดสอบสมมติฐานทีละคู่ จำนวน 3 ครั้ง ดังนี้

$$\mathrm{H}_0 : \mu_1 = \mu_2 \qquad \mathrm{H}_0 : \mu_1 = \mu_3 \qquad \mathrm{H}_0 : \mu_2 = \mu_3$$
$$\mathrm{H}_1 : \mu_1 \neq \mu_2 \qquad \mathrm{H}_1 : \mu_1 \neq \mu_3 \qquad \mathrm{H}_1 : \mu_2 \neq \mu_3$$

ซึ่งจะทำให้เสียเวลาในการทดสอบสมมติฐานที่ซ้ำซ้อนเป็นอย่างมาก และประการสำคัญคือเป็นการทำค่าระดับนัยสำคัญมีค่ามากเกินไป ดังนั้นจึงมีการนำเทคนิคการวิเคราะห์ความแปรปรวน (Analysis of Variance: ANOVA) ซึ่งเป็นการวิเคราะห์ข้อมูลที่ใช้ในการทดสอบสมมติฐานเกี่ยวกับความแตกต่างของค่าเฉลี่ยกรณีประชากรมากกว่า 2 กลุ่ม (k กลุ่ม) โดยทำการทดสอบเพียงครั้งเดียว เช่นกรณีประชากร 3 กลุ่ม สมมติฐานเชิงสถิติเป็นดังนี้

$$\mathrm{H}_0 : \mu_1 = \mu_2 = \mu_3$$
$$\mathrm{H}_1 : \mu_1 \neq \mu_2 \neq \mu_3 \text{ อย่างน้อย1 คู่}$$

ถ้าผลการทดสอบสมมติฐานปฏิเสธ H_0 หมายความว่ามีค่าเฉลี่ยอย่างน้อย 1 คู่ที่มีค่าแตกต่างกัน ซึ่งอาจจะเป็น $\mu_1 \neq \mu_2$ หรือ $\mu_1 \neq \mu_3$ หรือ $\mu_2 \neq \mu_3$ หรือ $\mu_1 \neq \mu_2 \neq \mu_3$ ก็ได้ ซึ่งการทดสอบว่าค่าเฉลี่ยของประชากรกลุ่มใดไม่เท่ากันนั้นเรียกว่าการเปรียบเทียบเชิงพหุ (Multiple Comparison) ซึ่งจะกล่าวรายละเอียดต่อไป

การวิเคราะห์ความแปรปรวนมีด้วยกันหลายประเภท ในเอกสารฉบับนี้จะกล่าวถึงการวิเคราะห์ความแปรปรวนเพียง 2 แบบ คือ

1. การวิเคราะห์ความแปรปรวนทางเดียว (One-way ANOVA)
2. การวิเคราะห์ความแปรปรวนสองทาง (Two-way ANOVA)

หลักการของการวิเคราะห์ความแปรปรวน

หลักเกณฑ์ที่สำคัญในการวิเคราะห์ความแปรปรวนคือแบ่งความแปรปรวนของข้อมูลทั้งหมดออกตามสาเหตุที่ทำให้ข้อมูลแตกต่างกัน คือความแปรปรวนภายในกลุ่ม (within group) และความแปรปรวนระหว่างกลุ่ม (between group) โดยที่

ความแปรปรวนทั้งหมด = ความแปรปรวนภายในกลุ่ม + ความแปรปรวนระหว่างกลุ่ม

การวิเคราะห์ความแปรปรวนทางเดียว

เป็นการศึกษาปัจจัยหรือแฟคเตอร์ (factor) ที่มีผลทำให้ข้อมูลแตกต่างกันเพียงปัจจัยเดียว โดยที่ปัจจัยนั้นอาจมีหลาย ๆ ระดับ เรียกระดับต่าง ๆ ของปัจจัยว่าทรีทเมนต์ (treatment) ดังนั้นจึงเป็นการเปรียบเทียบค่าเฉลี่ยของข้อมูลในระดับต่าง ๆ ของปัจจัยนั่นเอง นิยมเรียกข้อมูลว่าค่าสังเกต และหน่วยแจงนับที่ให้ข้อมูลว่าหน่วยทดลอง (experimental unit) เช่น

ตัวอย่าง 9.1 บริษัทผลิตถุงกระดาษที่ใช้ในร้านขายของชำพบว่าความเหนียวของถุงกระดาษขึ้นอยู่กับความเข้มข้นของเยื่อไม้ที่ใช้ทำเยื่อกระดาษ จึงทำการทดลองผลิตถุงกระดาษโดยใช้ความเข้มข้นของเยื่อไม้ต่างกัน คือ 5% 10% 15% และ 20% แล้วทำการวัดความเหนียวของถุงกระดาษที่เลือกจากแต่ละกลุ่ม ๆ ละ 6 ใบ

ค่าสังเกต คือความเหนียวของถุงกระดาษ

แฟคเตอร์ คือความเข้มข้นของเยื่อไม้

ทรีทเมนต์ คือความเข้มข้นของเยื่อไม้ต่างกัน คือ 5%, 10%, 15% และ 20%

หน่วยทดลอง คือถุงกระดาษ

ลักษณะของตารางข้อมูล

ความเข้มข้นของเยื่อกระดาษ			
5%	10%	15%	20%
9	12	13	19
10	11	15	23
8	13	15	19
11	13	17	20
8	14	17	21
8	15	17	21

1. ลักษณะของตารางข้อมูลในรูปทั่วไป

ลักษณะของข้อมูลที่ใช้ในการวิเคราะห์ความแปรปรวนทางเดียว เป็นดังนี้

	ทรีทเมนต์(treatment)					
	1	2	3	…	k	
	x_{11}	x_{21}	x_{31}	…	x_{k1}	
	x_{12}	x_{22}	x_{32}	…	x_{k2}	
	x_{13}	x_{23}	x_{33}	…	x_{k3}	
	:	:	:	…	:	
	x_{1n1}	x_{2n2}	x_{3n3}	…	x_{knk}	
รวม	T_1	T_2	T_3	…	T_k	T
ค่าเฉลี่ย	$\bar{x}_1$	$\bar{x}_2$	$\bar{x}_3$	…	$\bar{x}_k$	$\bar{x}$

เมื่อ x_{ij} แทนข้อมูลของทรีทเมนต์ที่ i หน่วยทดลองที่ j

$i = 1,2,3,\ldots,k$ และ $j = 1,2,3,\ldots,n_i$

T_i แทนผลรวมของข้อมูลทรีทเมนต์ที่ i

T แทนผลรวมข้อมูลทั้งหมด

$\bar{x}_i$ แทนค่าเฉลี่ยของข้อมูลทรีทเมนต์ที่ i

$\bar{x}$ แทนค่าเฉลี่ยของข้อมูลทั้งหมด

k แทนจำนวนทรีทเมนต์

n แทนจำนวนข้อมูลทั้งหมด เท่ากับ $n_1+n_2+n_3+\ldots+n_k$

เนื่องจากการวิเคราะห์ความแปรปรวนทางเดียวเป็นการศึกษาอิทธิพลของปัจจัยเดียวที่มีผลทำให้ค่าสังเกตแตกต่างกัน นั่นคือข้อมูลมีความแตกต่างเนื่องจากกลุ่มที่แตกต่างเท่านั้น ดังนั้นการวิเคราะห์จึงแบ่งความแปรปรวนของข้อมูลเป็นดังนี้

1. ความแปรปรวนระหว่างกลุ่ม (Between Groups Sum of Square) เขียนแทนด้วยสัญลักษณ SSB เป็นการพิจารณาความแปรปรวนที่เกิดจากการที่ค่าเฉลี่ยของตัวอย่างในแต่ละกลุ่มแตกต่างจากค่าเฉลี่ยรวม โดยที่

$$SSB = \sum_{i=1}^{k} n_i (\bar{x}_i - \bar{x})^2$$

2. ความแปรปรวนภายในกลุ่ม (Within Group Sum of Square) เขียนแทนด้วยสัญลักษณ์ SSE เป็นการพิจารณาความแปรปรวนที่เกิดขึ้นภายในกลุ่มแต่ละกลุ่มซึ่งไม่ทราบสาเหตุว่าเป็นความแปรปรวนที่เกิดจากสาเหตุใด ในบางครั้งจึงเรียกว่าความคลาดเคลื่อน (Error Sum of Square) โดยที่

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2$$

3. ความแปรปรวนรวม (Total Sum of Square) เขียนแทนด้วยสัญลักษณ์ SST เป็นการพิจารณาความแปรปรวนที่เกิดจากค่าสังเกตแต่ละค่าแตกต่างจากค่าเฉลี่ยรวม โดยที่

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{x})^2 \quad \text{และ} \quad SST = SSB + SSE$$

การคำนวณ Sum of Square นอกจากจะคำนวณจากวิธีการข้างต้นแล้ว ยังมีวิธีการคำนวณที่ปรับให้ง่ายขึ้น ดังนี้

$$CM(\text{corrected of Mean}) = \frac{(\sum\sum x_{ij})^2}{n}$$

$$SST = \sum\sum(x_{ij} - \bar{x})^2$$

$$= \sum\sum x_{ij}^2 - CM$$

$$SSB = \sum n_i(\bar{x}_i - \bar{x})^2$$

$$= \sum\left(\frac{(\sum x_i)^2}{n_i}\right) - CM$$

$$SSE = SST - SSB$$

2. เงื่อนไขของการวิเคราะห์ความแปรปรวน

ในการทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ยของประชากร k กลุ่ม ด้วยเทคนิคการวิเคราะห์ความแปรปรวน มีเงื่อนไขดังนี้

1 ประชากร k กลุ่มมีการแจกแจงแบบปกติ

2 ความแปรปรวนของแต่ละประชากรเท่ากัน

3 ตัวอย่างสุ่มจากแต่ละประชากรเป็นอิสระต่อกัน

3. สมมติฐานในการทดสอบ

กำหนด μ_1 แทนค่าเฉลี่ยของประชากรกลุ่มที่ 1

μ_2 แทนค่าเฉลี่ยของประชากรกลุ่มที่ 2

.

.

.

μ_k แทนค่าเฉลี่ยของประชากรกลุ่มที่ k

สมมติฐานเชิงสถิติ

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

$H_1 : \mu_1 \neq \mu_2 \neq \mu_3 \neq \ldots \neq \mu_k$ อย่างน้อย1 คู่

หรือ

H_O : ค่าเฉลี่ยของประชากร k กลุ่มไม่แตกต่างกัน

H_1 : ค่าเฉลี่ยของประชากร k กลุ่มแตกต่างกันอย่างน้อย 1 คู่

4. ตัวสถิติทดสอบ และค่าวิกฤต

ตัวสถิติในการทดสอบคือ $F = \frac{MSB}{MSE}$ ซึ่งคำนวณจากตารางการวิเคราะห์ความแปรปรวน (Analysis of Variance Table) หรือเรียกว่า ANOVA ดังนี้

ตาราง ANOVA

แหล่งความแปรปรวน (Source of variation)	องศาอิสระ (df)	ผลรวมกำลังสอง (Sum of Square) (SS)	ผลรวมกำลังสองเฉลี่ย (Mean of Square) $(MS = \frac{ss}{df})$	ค่าตัวสถิติ (F)
ระหว่างกลุ่ม	k-1	SSB	$MSB = \frac{SSB}{k-1}$	$F = \frac{MSB}{MSE}$
ภายในกลุ่ม	n-k	SSE	$MSE = \frac{SSE}{n-k}$	
รวม	n-1	SST		

ค่าวิกฤต $f_{1-\alpha,k-1,n-k}$ และปฏิเสธ H_0 เมื่อค่าสถิติทดสอบ F มากกว่าค่าวิกฤต

ตัวอย่าง 9.2 ในการหาปริมาณฟอสฟอรัสจากฟางข้าวที่ใช้เทคนิคการสกัดที่แตกต่างกัน 4 เทคนิค ในการทดสอบจึงนำฟางข้าวที่มีอายุเท่ากัน และมาจากแหล่งเดียวกันมาทำการสกัดฟอสฟอรัสด้วยวิธีที่แตกต่างกัน ได้ผลดังนี้

ปริมาณฟอสฟอรัส			
เทคนิคที่ 1	เทคนิคที่ 2	เทคนิคที่ 3	เทคนิคที่ 4
34	37	39	36
31	41	37	35
34	38	38	37
30	43	37	38
34	37	36	36
32	42	42	38
34	40	40	36
32	44	41	35

อยากทราบว่าปริมาณฟอสฟอรัสจากฟางข้าวที่ได้จากเทคนิคการสกัดทั้ง 4 เทคนิคแตกต่างกันหรือไม่ที่ระดับนัยสำคัญ 0.05

วิธีทำ

$$n = n_1 + n_2 + n_3 + n_4 = 32$$
$$\sum x_{1j} = 34 + 31 + 34 + \ldots + 32 = 261$$
$$\sum x_{2j} = 37 + 41 + 38 + \ldots + 44 = 322$$
$$\sum x_{3j} = 39 + 37 + 38 + \ldots + 41 = 310$$
$$\sum x_{4j} = 36 + 35 + 37 + \ldots + 35 = 291$$
$$\sum\sum x_{ij} = 261 + 322 + 310 + 291 = 1184$$

$$CM = \frac{(\sum\sum x_{ij})^2}{n}$$
$$= \frac{1184^2}{32}$$
$$= 43808$$

$$SST = \sum\sum x_{ij}^2 - CM$$
$$= (34^2 + 31^2 + 34^2 + \ldots + 35^2) - 43808$$
$$= 44184 - 43808$$
$$= 376$$

$$SSB = \sum\left(\frac{(\sum x_i)^2}{n_i}\right) - CM$$
$$= \left(\frac{261^2}{8} + \frac{322^2}{8} + \frac{310^2}{8} + \frac{291^2}{8}\right) - 43808$$
$$= 44073.25 - 43808$$
$$= 265.25$$

$$SSE = SST - SSB$$
$$= 376 - 265.25$$
$$= 110.75$$

ตาราง ANOVA

แหล่งความแปรปรวน	df	SS	MS	F
ระหว่างกลุ่ม	3	SSB=265.25	MSB=88.41	F=22.38
ภายในกลุ่ม	28	SSE=110.75	MSE=3.95	
ผลรวม	31	SST=376.00		

สมมติฐานเชิงสถิติ

H_o : ปริมาณฟอสฟอรัสจากเทคนิคการสกัดทั้ง 4 เทคนิคไม่แตกต่างกัน

H_1 : ปริมาณฟอสฟอรัสจากเทคนิคการสกัดทั้ง 4 เทคนิคแตกต่างกันอย่างน้อย 2 เทคนิค

ตัวสถิติทดสอบ

จากตาราง ANOVA ตัวสถิติทดสอบ $F = \frac{MSB}{MSE}$=22.38

ค่าวิกฤต $f_{1-\alpha,k-1,n-k} = f_{0.95,3,28} = 2.95$

เนื่องจากค่าสถิติทดสอบ F=22.38 อยู่ในบริเวณปฏิเสธ H_o หมายความว่าปริมาณฟอสฟอรัสจากเทคนิคการสกัดทั้ง 4 เทคนิคแตกต่างกันอย่างน้อย 2 เทคนิค ที่ระดับนัยสำคัญ 0.05

ตัวอย่าง 9.3 อาจารย์ผู้สอนวิชาสถิติต้องการเปรียบเทียบผลการสอบย่อยของนักศึกษา 3 กลุ่ม ได้แก่ นักศึกษาชั้นปี 1, 2 และ 3 ที่ลงทะเบียนเรียน จึงทำการเลือกตัวอย่างนักศึกษาชั้นปี 1, 2 และ 3 มา กลุ่มละ 4, 6 และ 5 คน ตามลำดับ จากนั้นทำการทดสอบโดยใช้ข้อสอบเดียวกัน ซึ่งมีคะแนนเต็ม 10 คะแนน นักศึกษาได้คะแนนสอบ ดังนี้

	คะแนนสอบวิชาสถิติของนักศึกษาชั้นปีที่		
	1	2	3
	4	5	8
	7	1	6
	6	3	8
	6	5	9
		3	5
		4	
ผลรวม	23	21	36

ให้ทดสอบว่าคะแนนเฉลี่ยของนักศึกษา 3 ชั้นปีนี้แตกต่างกันหรือไม่ ที่ระดับนัยสำคัญ 0.05

วิธีทำ

$$\begin{aligned}
\mathrm{CM} &= \frac{(\sum\sum x_{ij})^2}{n} \\
&= \frac{(4+7+6+\ldots+5)^2}{15} \\
&= 426.667
\end{aligned}$$

$$\begin{aligned}
\mathrm{SST} &= \sum\sum x_{ij}^2 - \mathrm{CM} \\
&= (4^2+7^2+6^2+\ldots+5^2) - 426.667 \\
&= 65.333
\end{aligned}$$

$$\begin{aligned}
\mathrm{SSB} &= \sum\left(\frac{(\sum x_i)^2}{n_i}\right) - \mathrm{CM} \\
&= \left(\frac{23^2}{4}+\frac{21^2}{6}+\frac{36^2}{5}\right) - 426.667 \\
&= 464.950 - 426.667 \\
&= 38.283
\end{aligned}$$

$$\begin{aligned} SSE &= SST - SSB \\ &= 65.333 - 38.283 \\ &= 27.05 \end{aligned}$$

$$\begin{aligned} MSB &= \frac{SSB}{k-1} \\ &= \frac{38.283}{2} \\ &= 19.142 \end{aligned}$$

$$\begin{aligned} MSE &= \frac{SSE}{n-k} \\ &= \frac{27.05}{12} \\ &= 2.254 \end{aligned}$$

$$\begin{aligned} F &= \frac{MSB}{MSE} \\ &= \frac{19.142}{2.254} \\ &= 8.49 \end{aligned}$$

ตาราง ANOVA

แหล่งความแปรปรวน	df	SS	MS	F
ระหว่างกลุ่ม	2	38.283	19.142	8.49
ภายในกลุ่ม	12	27.052	2.254	
ผลรวม	14	65.333		

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

$H_1 : \mu_i \neq \mu_j \;\; ; i \neq j$ อย่างน้อย 1 คู่

ตัวสถิติทดสอบ

จากตาราง ANOVA ตัวสถิติทดสอบ $F = \frac{MSB}{MSE}$ =8.49

ค่าวิกฤต $f_{1-\alpha,k-1,n-k} = f_{0.95,2,12} = 3.89$

เนื่องจากค่าสถิติทดสอบ F=8.49 อยู่ในบริเวณปฏิเสธ H_o หมายความว่าคะแนนเฉลี่ยของนักศึกษาอย่างน้อยหนึ่งกลุ่มจะแตกต่างไปจากกลุ่มอื่น ๆ ที่ระดับนัยสำคัญ 0.05

ตัวอย่าง 9.4 ในการศึกษาว่าวิธีการคั้นน้ำส้มที่แตกต่างกัน 4 วิธีมีอิทธิพลต่อปริมาณของวิตามินซีในน้ำส้มขนาด 8 ออนซ์ หรือไม่ที่ระดับนัยสำคัญ 0.05 จึงทำการทดลองแล้ววัดผลข้อมูลที่ได้ปรากฏดังตาราง

ปริมาณของวิตามินซี			
วิธีที่ 1	วิธีที่ 2	วิธีที่ 3	วิธีที่ 4
88	123	72	89
83	115	78	88
80	122	75	87
87	118	79	89
90	120	80	90
89	119	82	89
92	117	78	88
90	116	74	90
86	119	76	85
90	122	80	91

วิธีทำ

CM = (ผลรวมข้อมูลทั้งหมด)^2 / จำนวนตัวอย่าง

= $(3726)^2 / 40$

= 347076.9

SST = ผลบวก(ข้อมูลแต่ละตัว)^2 – CM

= $(88^2 + 83^2 + 80^2 + \ldots + 91^2) - 347076.9$

= 10039.1

SSB = ผลบวก((ผลรวมข้อมูลแต่ละกลุ่ม)^2 / จำนวนตัวอย่างแต่ละกลุ่ม) - CM

$$= \frac{857^2}{10} + \frac{1191^2}{10} + \frac{774^2}{10} + \frac{886^2}{10} - 347076.9$$

= 9740.9

$$SSE = SST - SSA$$
$$= 10039.1 - 9740.9$$
$$= 298.2$$

$$MSB = \frac{SSB}{k-1}$$
$$= \frac{9740.9}{3}$$
$$= 3246.97$$

$$MSE = \frac{SSE}{n-k}$$
$$= \frac{298.2}{36}$$
$$= 8.28$$

$$F = \frac{MSB}{MSE}$$
$$= \frac{3246.97}{8.28}$$
$$= 392.15$$

ตาราง ANOVA

แหล่งความแปรผัน (Source of variation)	องศาอิสระ (df)	ผลรวมกำลังสอง (SS)	ผลรวมกำลังสองเฉลี่ย (MS)	ค่าตัวสถิติ (F)
ระหว่างกลุ่ม	4 – 1 = 3	9740.9	3246.97	392.15
ภายในกลุ่ม	40 – 4 = 36	298.2	8.28	
รวม	40 – 1 = 39	100391.1		

μ_i แทนปริมาณวิตามินซีเฉลี่ยจากวิธีการคั้นที่ i เมื่อ i = 1,2,3,4

สมมติฐานเชิงสถิติ

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

$H_1 : \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$ อย่างน้อย 1 คู่

ค่าวิกฤต $f_{1-\alpha,k-1,n-k} = f_{0.95,3,36} = 2.866$

เนื่องจาก F = 392.15 มากกว่า 2.866 จึงปฏิเสธ H_0 หมายความว่าปริมาณวิตามินซีเฉลี่ยจากวิธีการคั้น 4 วิธี แตกต่างกัน อย่างมีนัยสำคัญที่ระดับ 0.05

การเปรียบเทียบพหุคูณ (Multiple comparison)

เทคนิคการวิเคราะห์ความแปรปรวนเป็นการทดสอบว่าจะมีค่าเฉลี่ยของประชากร k กลุ่มแตกต่างกันหรือไม่ ถ้าแตกต่างกันอย่างมีนัยสำคัญทางสถิติ (significant) ก็จะบอกเพียงว่ามีค่าเฉลี่ยอย่างน้อย 1 คู่ที่มีค่าแตกต่างกัน แต่จะไม่บอกว่าเป็นคู่ใด ซึ่งเราจะต้องทำการทดสอบหลังการวิเคราะห์ (Post hoc test) โดยวิธีการเปรียบเทียบพหุคูณ (Multiple comparison) ซึ่งมีหลายวิธีด้วยกัน ในเอกสารฉบับนี้จะขอกล่าวเพียงบ้างวิธีที่นิยมใช้

1. วิธี Least - Significant Different (LSD)

วิธีการเปรียบเทียบพหุคูณแบบ LSD หรือ Fisher's Least – Significant Different เป็นเทคนิคที่ R.A. Fisher ได้พัฒนาขึ้นหรือเปรียบเทียบค่าเฉลี่ยประชากรครั้งละหลายคู่ โดยใช้สูตร

$$\mathrm{LSD} = t_{\frac{\alpha}{2}}\sqrt{\mathrm{MSE}}\sqrt{\frac{1}{n_i}+\frac{1}{n_j}}$$

เมื่อ MSE แทนค่าความแปรปรวนจาก one way ANOVA

n_i แทนจำนวนข้อมูลกลุ่มที่ i

n_j แทนจำนวนข้อมูลกลุ่มที่ j

วิธี LSD มีขั้นตอนดังนี้

1. คำนวณค่า LSD
2. คำนวณความแตกต่างระหว่างค่าเฉลี่ย $\bar{x}_i - \bar{x}_j$
3. นำค่า $\left|\bar{x}_i - \bar{x}_j\right|$ เปรียบเทียบกับ ค่า LSD

   3.1 ถ้า ค่า $\left|\bar{x}_i - \bar{x}_j\right|$ > ค่า LSD แสดงว่า $\mu_i \neq \mu_j$

   3.2. ถ้า ค่า $\left|\bar{x}_i - \bar{x}_j\right| \leq$ ค่า LSD แสดงว่า $\mu_i = \mu_j$

2. วิธี Turkey's Honestly Significant Different (HSD)

เป็นวิธีการเปรียบเทียบภายใต้เงื่อนไขที่ว่าจำนวนกลุ่มตัวอย่างแต่ละกลุ่มมีขนาดเท่ากัน $(n_1 = n_2 = n_3 = \ldots = n_K = n)$ โดยมีสูตรของ Diekhoff ดังนี้

$$\mathrm{HSD} = q_{(\alpha,\mathrm{df},k)}\sqrt{\frac{\mathrm{MSE}}{n}}$$

เมื่อ q หาได้จากตารางค่าวิกฤตของ Studentized rough statistic

โดย $df = n - k$ จากตาราง ANOVA

MSE ได้จากการคำนวนหาค่าความแปรปรวน one way ANOVA

n จำนวนข้อมูลทั้งหมด

วิธี HDS มีขั้นตอนดังนี้

1. คำนวณค่า HSD
2. คำนวณค่า $\left|\bar{x}_i - \bar{x}_j\right|$
3. เปรียบเทียบค่า $\left|\bar{x}_i - \bar{x}_j\right|$ กับค่า HSD โดย
   - 3.1 ถ้า $\left|\bar{x}_i - \bar{x}_j\right|$ > HSD แสดงว่า $\mu_i \neq \mu_j$
   - 3.2 ถ้า $\left|\bar{x}_i - \bar{x}_j\right| \leq$ HSD แสดงว่า $\mu_i = \mu_j$

3. วิธี The Sheffe's Post hoc Comparison (Sheffe')

การเปรียบเทียบพหุคูณโดยวิธี Sheffe นั้นสามารถใช้ได้กับกลุ่มตัวอย่างที่มีขนาดเท่ากันหรือไม่เท่ากันก็ได้ โดยใช้สูตรของ Byrkit

$$\mathrm{CV_d} = \sqrt{(\mathrm{k}-1)(\mathrm{F}^*)(\mathrm{MSE})\left(\frac{1}{\mathrm{n_i}} + \frac{1}{\mathrm{n_j}}\right)}$$

เมื่อ F^* คือ ค่าวิกฤต จากตาราง F โดยมี $df_1 = k-1, df_2 = n-k$

MSE ได้จากการคำนวนหาค่าความแปรปรวน one way ANOVA

n_i จำนวนข้อมูลกลุ่มที่ i

n_j จำนวนข้อมูลกลุ่มที่ j

วิธีของ Sheffe มีขั้นตอนดังนี้

1. คำนวณค่า CV_d
2. คำนวณค่า $\left|\bar{x}_i - \bar{x}_j\right|$
3. เปรียบเทียบ $\left|\bar{x}_i - \bar{x}_j\right|$ กับค่า CV_d โดย
   - 3.1 ถ้า $\left|\bar{x}_i - \bar{x}_j\right| \geq CV_d$ แสดงว่า $\mu_i \neq \mu_j$
   - 3.2 ถ้า $\left|\bar{x}_i - \bar{x}_j\right| < CV_d$ แสดงว่า $\mu_i = \mu_j$

ความสัมพันธ์ระหว่างตัวทดสอบสถิติเอฟ (F) และที (t)

ตัวสถิติทดสอบ t ที่ใช้ในการทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ยของประชากร 2 กลุ่มที่เป็นอิสระกัน ในกรณีที่ความแปรปรวนของข้อมูลทั้ง 2 กลุ่มเท่ากัน กับตัวสถิติทดสอบ F จากตาราง ANOVA ในกรณีที่มีประชากรเพียง 2 กลุ่ม ซึ่งก็มีเงื่อนไขว่าความแปรปรวนของข้อมูลในประชากรแต่ละกลุ่มต้องเท่ากันนั้น มีความสัมพันธ์กัน ดังนี้

$$F = t^2$$

เมื่อ F แทนค่าตัวสถิติทดสอบ F จากตาราง ANOVA

t แทนค่าตัวสถิติทดสอบ t

ตัวอย่าง 9.5 ในการเปรียบเทียบวิธีการสอน 2 แบบ โดยทำการเลือกนิสิตที่เรียนแบบ 1 มา 8 คน แบบ 2 มา 12 คน คำนวณคะแนนเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานได้ดังนี้

$$\bar{x}_1 = 76.9 \qquad \bar{x}_2 = 72.7$$
$$s_1 = 4.85 \qquad s_2 = 6.35$$

โดยที่คะแนนสอบของการสอนทั้ง 2 แบบมีการแจกแจงใกล้เคียงแบบปกติ และมีค่าความแปรปรวนเท่ากัน จงทดสอบว่าคะแนนสอบเฉลี่ยของการสอน 2 แบบเท่ากันหรือไม่ ที่ระดับนัยสำคัญ 0.05

การทดสอบสมมติฐานด้วยตัวสถิติทดสอบ t ดำเนินการตามขั้นตอนต่อไปนี้

สมมติฐานเชิงสถิติ

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

เนื่องไม่ทราบความแปรปรวนของประชากร แต่ทราบว่า $\sigma_1^2 = \sigma_2^2 = \sigma^2$ และตัวอย่างมีขนาดเล็ก ดังนั้นจะประมาณความแปรปรวนด้วยความแปรปรวนร่วม ดังนี้

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$
$$= \frac{(8-1)(4.85)^2 + (12-1)(6.35)^2}{8+12-2}$$
$$= 33.789$$

ตัวสถิติทดสอบ

$$T = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$= \frac{(76.9 - 72.7) - 0}{5.81\sqrt{\frac{1}{8} + \frac{1}{12}}}$$

$$= \frac{4.2}{2.65}$$

$$= 1.584$$

ค่าวิกฤต $t_{\frac{\alpha}{2}, n_1+n_2-2} = t_{0.025\ ,18} = 2.101$

เนื่องจากค่าสถิติทดสอบ T = 1.584 อยู่ในบริเวณยอมรับ $H_0 : \mu_1 = \mu_2$ หมายความว่าคะแนนเฉลี่ย จากวิธีการสอน 2 แบบไม่แตกต่างกันที่ระดับนัยสำคัญ 0.05

การทดสอบสมมติฐานด้วยตัวสถิติทดสอบ F ในตาราง ANOVA ดำเนินการตามขั้นตอนต่อไปนี้

สมมติฐานเชิงสถิติ
$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

คำนวณ $\bar{x}$ จาก
$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$$
$$= \frac{8(76.9) + 12(72.7)}{8 + 12}$$
$$= 74.38$$

SSB = $8(76.9)^2 + 12(72.7)^2 - (20)(74.38)^2$ = 84.672

SSE = $(8-1)(4.85)^2 + (12-1)(6.35)^2$ = 608.205

SST = 84.672 + 608.205 = 692.877

MSB = $\frac{84.672}{1} = 84.672$

MSE = $\frac{608.205}{18} = 33.789$

F = $\frac{84.672}{33.826} = 2.503$

ตาราง ANOVA

แหล่งความแปรปรวน	df	SS	MS	F
ระหว่างกลุ่ม	1	84.672	84.672	2.503
ภายในกลุ่ม	18	608.205	33.789	
รวม	19	692.877		

ค่าวิกฤต $f_{1-\alpha,k-1,n-k} = f_{0.95,1,18} = 4.41$

เนื่องจากค่าสถิติทดสอบ F=2.503 อยู่ในบริเวณยอมรับ H_0 หมายความว่าว่าคะแนนสอบเฉลี่ยของการสอน 2 แบบเท่ากัน

สำหรับความสัมพันธ์ระหว่างตัวสถิติทดสอบ F และ t ของตัวอย่างนี้

$$F = (1.584)^2 = 2.509$$

ดังนั้น $F = t^2$

การวิเคราะห์ความแปรปรวนสองทาง

การวิเคราะห์ความแปรปรวนสองทาง (Two–Way ANOVA) แตกต่างจากการวิเคราะห์ความแปรปรวนทางเดียวคือ ในการวิเคราะห์ความแปรปรวนทางเดียวหน่วยตัวอย่างภายในกลุ่มเดียวกันจะต้องมีความแตกต่างกันน้อยมาก เพื่อที่จะมั่นใจได้ว่าเมื่อเกิดความแปรปรวนในการทดลอง จะนำไปสู่ข้อสรุปได้ชัดเจนว่าเป็นความแปรปรวนระหว่างกลุ่ม แต่ในทางปฏิบัติอาจพบว่าการใช้หน่วยตัวอย่างที่เหมือนกันหรือมีความคล้ายคลึงกันจะเป็นไปได้ยากมาก เช่นถ้าจะเปรียบเทียบยอดขายประกันของบริษัทจากวิธีขายที่แตกต่างกัน 3 วิธี อาจเป็นได้ว่าความสามารถที่แตกต่างกันของพนักงานก็เป็นส่วนหนึ่งที่ทำให้ยอดขายแตกต่างกันได้ แม้จะใช้วิธีขายวิธีเดียวกัน ดังนั้นเมื่อเกิดความแปรปรวนของข้อมูล จึงทำให้สรุปได้ไม่ชัดเจนว่าเป็นเพราะวิธีขายที่ต่างกันหรือเป็นเพราะความสามารถของพนักงานที่แตกต่างกันที่เป็นสาเหตุทำให้ยอดขายแตกต่างกัน ดังนั้นจึงอาจจะแบ่งหน่วยทดลองออกเป็นกลุ่ม ๆ เรียกว่า บล็อค (block) โดยให้ภายในแต่ละบล็อคประกอบไปด้วยหน่วยตัวอย่างที่มีความคล้ายคลึงกัน ส่วนในต่างบล็อคก็จะเป็นหน่วยตัวอย่างที่แตกต่างกัน และจำนวนหน่วยทดลองภายในแต่ละบล็อกจะได้รับทรีทเมนต์ต่าง ๆ ครบชุด ตารางข้อมูลสำหรับการวิเคราะห์ความแปรปรวนสองทางเป็น ดังนี้

แถว (block)	คอลัมภ์ (treatment) 1	2	...	j	...	c	รวม	เฉลี่ย
1	x_{11}	x_{12}	...	x_{1j}	...	x_{1c}	$\mathbf{T_{1.}}$	$\bar{x}_{1.}$
2	x_{21}	x_{22}	...	x_{2j}	...	x_{2c}	$\mathbf{T_{2.}}$	$\bar{x}_{2.}$
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
i	x_{i1}	x_{i2}	...	x_{ij}	...	x_{ic}	$\mathbf{T_{i.}}$	$\bar{x}_{i.}$
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
r	x_{r1}	x_{r2}	...	x_{rj}	...	x_{rc}	$\mathbf{T_{r.}}$	$\bar{x}_{r.}$
รวม	$\mathbf{T_{.1}}$	$\mathbf{T_{.2}}$	**...**	$\mathbf{T_{.j}}$	**...**	$\mathbf{T_{.c}}$	**T**	
เฉลี่ย	$\bar{x}_{.1}$	$\bar{x}_{.2}$	**...**	$\bar{x}_{.j}$	**...**	$\bar{x}_{.c}$		$\bar{x}$

เมื่อ x_{ij} แทนค่าสังเกตแถวที่ i คอลัมภ์ที่ j

$T_{i.}$ แทนผลรวมค่าสังเกตแถวที่ i

$T_{.j}$ แทนผลรวมค่าสังเกตคอลัมภ์ที่ j

T แทนผลรวมทั้งหมด

$\bar{x}_{i.}$ แทนค่าเฉลี่ยของค่าสังเกตแถวที่ i

$\bar{x}_{.j}$ แทนค่าเฉลี่ยของค่าสังเกตคอลัมภ์ที่ j

$\bar{x}$ แทนค่าเฉลี่ยของข้อมูลทั้งหมด

r แทนจำนวนแถว

c แทนจำนวนคอลัมภ์

n แทนจำนวนตัวอย่างทั้งหมด = $r \times c$

ในกรณีนี้จะแยกแหล่งความแปรปรวนทั้งหมดออกได้เป็น

ความแปรปรวนรวม (SST) = ความแปรปรวนระหว่างทรีทเมนต์ (SSA) + ความแปรปรวนระหว่างบล็อก (SSB) + ความแปรปรวนอื่น ๆ (SSE)

หรือ SST = SSA + SSB + SSE

การคำนวณหาค่า Sum of Square เริ่มต้นจากการหาค่า

$$CM(\text{corrected of Mean}) = \frac{(\sum\sum x_{ij})^2}{n}$$

SST แทนความแปรปรวนรวมคำนวณได้โดย

$$SST = \sum_{i=1}^{r}\sum_{j=1}^{c} x_{ij}^2 - CM$$

SSA แทนความแปรปรวนระหว่างทรีทเมนต์ในแต่ละคอลัมน์ คำนวณได้โดย

$$SSA = \sum_{j=1}^{c}\frac{T_j^2}{n_j} - CM$$

SSB แทนความแปรปรวนระหว่างบล็อคในแต่ละแถว คำนวณได้โดย

$$SSB = \sum_{i=1}^{r}\frac{T_i^2}{n_i} - CM$$

SSE แทนความผันแปรภายในอื่น ๆ ที่ไม่ทราบสาเหตุ คำนวณได้โดย

$$SSE = SST - SSA - SSB$$

1. สมมติฐานการทดสอบ

กรณีทดสอบว่าทรีทเมนต์มีผลทำให้ค่าเฉลี่ยแตกต่างกันหรือไม่

สมมติฐานเชิงสถิติ

H_0 : ค่าเฉลี่ยของข้อมูลที่ได้รับทรีทเมนต์แต่ละทรีทเมนต์ไม่แตกต่างกัน

H_1 : ค่าเฉลี่ยของข้อมูลที่ได้รับทรีทเมนต์แต่ละทรีทเมนต์แตกต่างกันอย่างน้อย 1 คู่

หรือ

$$H_0 : \mu_{.1} = \mu_{.2} = \mu_{.3} = \ldots = \mu_{.c}$$
$$H_1 : \mu_{.1} \neq \mu_{.2} \neq \mu_{.3} \neq \ldots \neq \mu_{.c}$$

กรณีทดสอบว่าบล็อคมีผลทำให้ค่าเฉลี่ยแตกต่างกันหรือไม่

H_0 : ค่าเฉลี่ยของข้อมูลที่อยู่ในบล็อคแต่ละบล็อคไม่แตกต่างกัน

H_1 : ค่าเฉลี่ยของข้อมูลที่อยู่ในบล็อคแต่ละบล็อคแตกต่างกันอย่างน้อย 1 คู่

หรือ

$$H_0 : \mu_{1.} = \mu_{2.} = \mu_{3.} = \ldots = \mu_{r.}$$
$$H_1 : \mu_{1.} \neq \mu_{2.} \neq \mu_{3.} \neq \ldots \neq \mu_{r.}$$

2. ตัวสถิติทดสอบ และค่าวิกฤต

ตัวสถิติในการทดสอบอิทธิพลของทรีทเมนต์คือ $F = \frac{MSA}{MSE}$ และตัวสถิติทดสอบอิทธิพลของบล็อกคือ $F = \frac{MSB}{MSE}$ ซึ่งคำนวณจากตารางการวิเคราะห์ความแปรปรวน (Analysis of Variance) หรือที่เรียกว่า ANOVA ดังนี้

ตาราง ANOVA

แหล่งความแปรปรวน (Source of variation)	องศาอิสระ (df)	ผลรวมกำลังสอง (Sum of Square) (SS)	ผลรวมกำลังสองเฉลี่ย (Mean of Square) $(MS = \frac{ss}{df})$	ค่าตัวสถิติ (F)
ระหว่างทรีทเมนต์	c-1	SSA	$MSA = \frac{SSA}{c-1}$	$F = \frac{MSA}{MSE}$
ระหว่างบล็อค	r-1	SSB	$MSB = \frac{SSB}{r-1}$	$F = \frac{MSB}{MSE}$
ความคลาดเคลื่อน	(c-1)(r-1)	SSE	$MSE = \frac{SSE}{(c-1)(r-1)}$	
รวม	n-1	SST		

ค่าวิกฤตในการทดสอบอิทธิพลของทรีทเมนต์คือ $f_{1-\alpha, c-1, (c-1)(r-1)}$

ค่าวิกฤตในการทดสอบอิทธิพลของบล็อคคือ $f_{1-\alpha, r-1, (c-1)(r-1)}$

และปฏิเสธ H_0 เมื่อค่าสถิติทดสอบ F มากกว่า ค่าวิกฤต

ตัวอย่าง 9.6 ร้านค่าแห่งหนึ่งมีสาขาอยู่หลายแห่ง ต้องการทดสอบว่าสาขามีผลทำให้ยอดขายขนมแตกต่างกันหรือไม่ จึงเลือกร้านค้ามา 5 สาขา เพื่อทำการเก็บรวบรวมข้อมูล แต่เนื่องจากในร้านค้าแต่ละสาขานั้นมีลักษณะการจัดวางขนมที่แตกต่างกัน 3 แบบ ดังนั้นจึงเก็บรวบรวมข้อมูลยอดขายขนม (ร้อยบาท) จากร้านค้า 5 แห่ง และตำแหน่งที่วางขนม 3 ตำแหน่ง ได้ผลดังนี้

ตำแหน่งที่วางขนม	ร้านค้า					รวม
	A	B	C	D	E	
ชั้นบน	7	8	9	10	11	45
ชั้นกลาง	9	9	9	9	12	48
ชั้นล่าง	10	10	12	12	14	58
รวม	26	27	30	31	37	151

ที่ระดับนัยสำคัญ 0.05 จงทดสอบว่ายอดขายขนมของแต่ละร้าน และตำแหน่งที่วางขนมแต่ละระดับแตกต่างกันหรือไม่

วิธีทำ การคำนวณเริ่มต้นจากการหาค่า

$$CM = \frac{(\sum\sum x_{ij})^2}{n}$$

$$= \frac{(151)^2}{15}$$

$$= 1520.07$$

$$SST = \sum_{i=1}^{r}\sum_{j=1}^{c} x_{ij}^2 - CM$$

$$= (7^2 + 9^2 + 10^2 + \ldots + 14^2) - 1520.07$$

$$= 1567 - 1520.07$$

$$= 46.93$$

$$SSA = \sum_{j=1}^{c} \frac{T_j^2}{n_j} - CM$$

$$= \left(\frac{26^2}{3} + \frac{27^2}{3} + \frac{30^2}{3} + \frac{31^2}{3} + \frac{37^2}{3}\right) - 1520.07$$

$$= 1545 - 1520.07$$

$$= 24.93$$

$$SSB = \sum_{i=1}^{r} \frac{T_i^2}{n_i} - CM$$
$$= \left(\frac{45^2}{5} + \frac{48^2}{5} + \frac{58^2}{5} \right) - 1520.07$$
$$= 1538.60 - 1520.07$$
$$= 18.53$$

$$SSE = SST - SSA - SSB$$
$$= 46.93 - 24.93 - 18.53$$
$$= 3.47$$

ตาราง ANOVA

แหล่งความแปรปรวน	df	SS	MS	F
ระหว่างทรีทเมนต์	4	24.93	6.23	14.49
ระหว่างบล็อค	2	18.53	9.265	21.55
ความคลาดเคลื่อน	8	3.47	0.43	
รวม	14	46.93		

กรณีทดสอบว่าสาขามีผลทำให้ยอดขายขนมเฉลี่ยแตกต่างกันหรือไม่

H_0 : ยอดขายขนมเฉลี่ยของร้านค้าแต่ละสาขาไม่แตกต่างกัน

H_1 : ยอดขายขนมเฉลี่ยของร้านค้าแต่ละสาขาแตกต่างกันอย่างน้อย 1 คู่

หรือ

$$H_0 : \mu_{.1} = \mu_{.2} = \mu_{.3} = \mu_{.4} = \mu_{.5}$$
$$H_1 : \mu_{.1} \neq \mu_{.2} \neq \mu_{.3} \neq \mu_{.4} \neq \mu_{.5}$$

ตัวสถิติทดสอบคือ $F = \frac{MSA}{MSE}$ จากตาราง ANOVA จะได้ F = 14.49

ค่าวิกฤตคือ $f_{1-\alpha, c-1, (c-1)(r-1)} = f_{0.95, 4, 8} = 3.84$

เนื่องจาก F=14.49 อยู่ในบริเวณปฏิเสธ H_0 หมายความว่ายอดขายขนมเฉลี่ยของร้านค้าแต่ละสาขาแตกต่างกันอย่างน้อย 1 คู่ หรือสาขาของร้านค้ามีผลทำให้ยอดขายขนมเฉลี่ยแตกต่างกันอย่างมีนัยสำคัญที่ระดับ 0.05

กรณีทดสอบว่าตำแหน่งการวางมีผลทำให้ยอดขายขนมเฉลี่ยแตกต่างกันหรือไม่

H_0 : ยอดขายขนมเฉลี่ยที่วางอยู่ในตำแหน่งแต่ละตำแหน่งไม่แตกต่างกัน

H_1 : ยอดขายขนมเฉลี่ยที่วางอยู่ในตำแหน่งแต่ละตำแหน่งแตกต่างกันอย่างน้อย 1 คู่

หรือ

$$H_0 : \mu_{1.} = \mu_{2.} = \mu_{3.}$$
$$H_1 : \mu_{1.} \neq \mu_{2.} \neq \mu_{3.}$$

ตัวสถิติทดสอบคือ $F = \dfrac{MSB}{MSE}$ จากตาราง ANOVA จะได้ F = 21.55

ค่าวิกฤตคือ $f_{1-\alpha,r-1,(c-1)(r-1)} = f_{0.95,2,8} = 4.46$

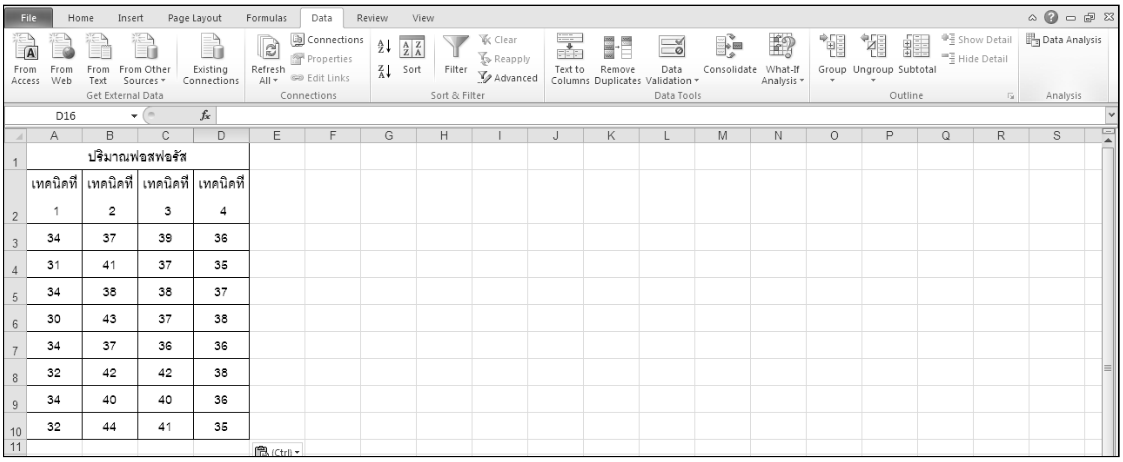
เนื่องจาก F=21.55 อยู่ในบริเวณปฏิเสธ H_0 หมายความว่ายอดขายขนมเฉลียที่วางอยู่ในตำแหน่งแต่ละตำแหน่งแตกต่างกันอย่างน้อย 1 คู่ หรือตำแหน่งการวางขนมมีผลทำให้ยอดขายขนมเฉลี่ยแตกต่างกันอย่างมีนัยสำคัญที่ระดับ 0.05

การใช้โปรแกรมสำเร็จรูป

1. การวิเคราะห์ความแปรปรวนทางเดียว

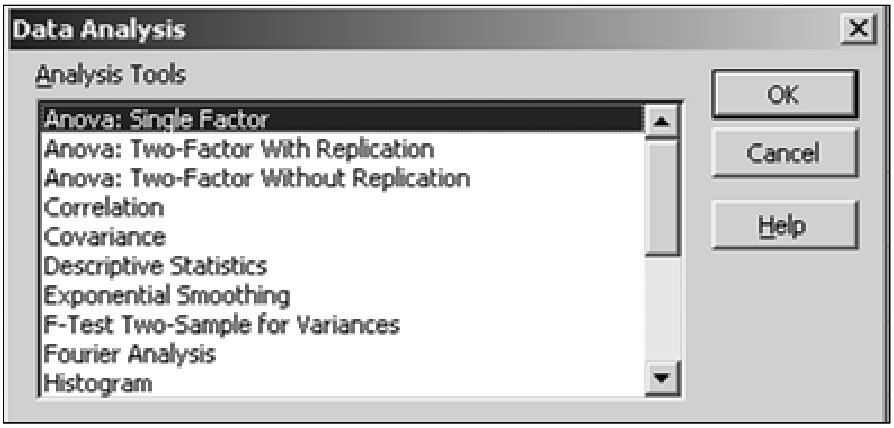
ตัวอย่าง 9.7 จากตัวอย่าง 9.2 จงวิเคราะห์ข้อมูลด้วย Microsoft Excel

ขั้นตอนที่ 1 ใส่ข้อมูลดังรูป เลือกเมนู Data เลือก Data Analysis ดังนี้



ปริมาณฟอสฟอรัส			
เทคนิคที่ 1	เทคนิคที่ 2	เทคนิคที่ 3	เทคนิคที่ 4
34	37	39	36
31	41	37	35
34	38	38	37
30	43	37	38
34	37	36	36
32	42	42	38
34	40	40	36
32	44	41	35

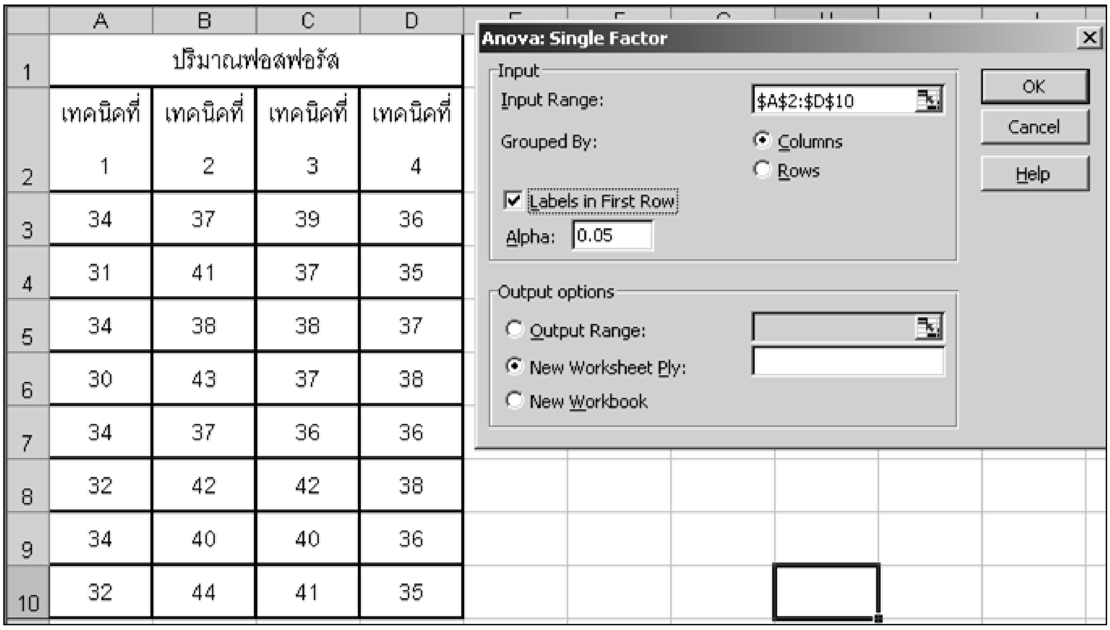
ขั้นตอนที่ 2 ในหน้าต่าง Data Analysis เลือก Anova: Single Factor ดังนี้



ขั้นตอนที่ 3 ในส่วน Input Range คือระบุขอบเขตของข้อมูลทั้งหมด

Grouped By คือระบุแบ่งกลุ่มของข้อมูลตาม row หรือ column

Alpha คือค่าระดับนัยสำคัญ



ขั้นตอนที่ 4 ได้ผลลัพธ์ ดังนี้

	A	B	C	D	E	F	G
1	**Anova: Single Factor**						
2							
3	**SUMMARY**			**ค่าเฉลี่ย**	**ความแปรปรวน**		
4	***Groups***	***Count***	***Sum***	***Average***	***Variance***		
5	**เทคนิคที่ 1**	**8**	**261**	**32.625**	**2.553571429**		
6	**เทคนิคที่ 2**	**8**	**322**	**40.25**	**7.357142857**		
7	**เทคนิคที่ 3**	**8**	**310**	**38.75**	**4.5**		
8	**เทคนิคที่ 4**	**8**	**291**	**36.375**	**1.410714286**		
9							
10			**องศาแห่ง**				
11	**ANOVA**		**ความเป็นอิสระ**		**ค่าสถิติทดสอบ F**	**P-value**	**ค่าวิกฤต**
12	***Source of Variation***	***SS***	***df***	***MS***	***F***	***P-value***	***F crit***
13	**Between Groups**	**265.25**	**3**	**88.4166667**	**22.35364936**	**1.36529E-07**	**2.946685266**
14	**Within Groups**	**110.75**	**28**	**3.95535714**			
15							
16	**Total**	**376**	**31**				
17							

สมมติฐานเชิงสถิติ

H_O : ปริมาณฟอสฟอรัสจากเทคนิคการสกัดทั้ง 4 เทคนิคไม่แตกต่างกัน

H_1 : ปริมาณฟอสฟอรัสจากเทคนิคการสกัดทั้ง 4 เทคนิคแตกต่างกันอย่างน้อย 2 เทคนิค

หรือ สมมติฐานเชิงสถิติ

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

$H_1 : \mu_i \neq \mu_j \quad ; i \neq j$ อย่างน้อย 1 คู่

ตัวสถิติทดสอบ

จากตาราง ANOVA ตัวสถิติทดสอบ $F = \frac{MSB}{MSE} = 22.38$

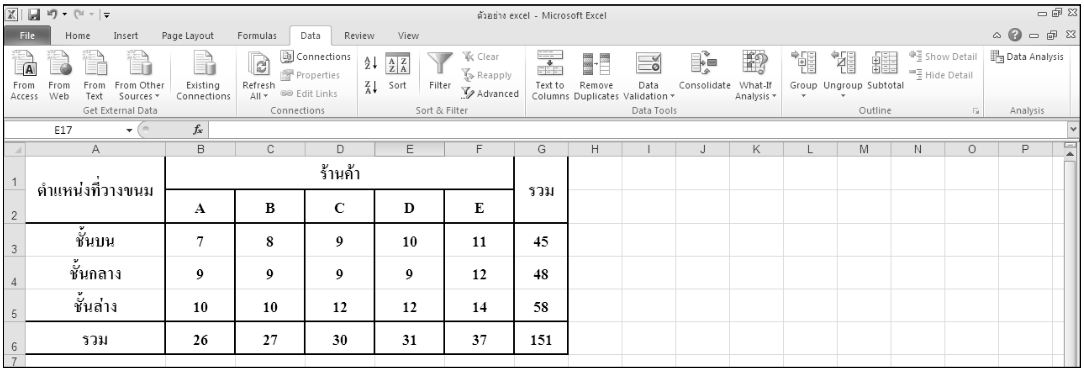
ค่าวิกฤต $f_{1-\alpha,k-1,n-k} = f_{0.95,3,28} = 2.95$

เนื่องจากค่าสถิติทดสอบ F=22.38 อยู่ในบริเวณปฏิเสธ H_O หมายความว่าปริมาณฟอสฟอรัสจากเทคนิคการสกัดทั้ง 4 เทคนิคแตกต่างกันอย่างน้อย 2 เทคนิค ที่ระดับนัยสำคัญ 0.05

2. การวิเคราะห์ความแปรปรวนสองทาง

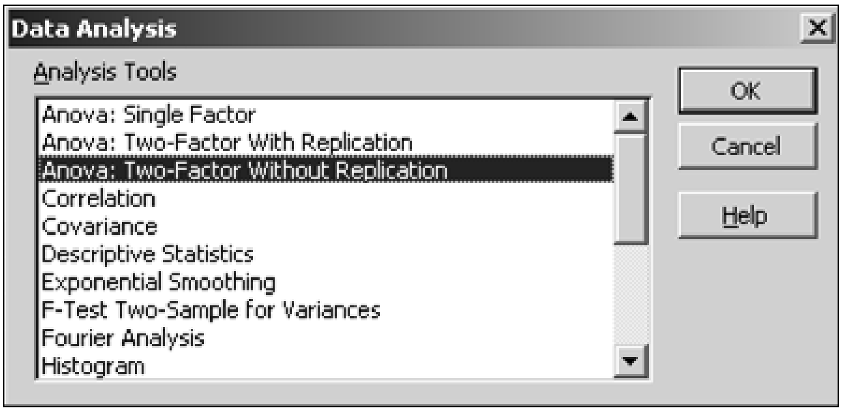
ตัวอย่าง 9.8 จากตัวอย่าง 9.6 จงวิเคราะห์ข้อมูลด้วย Microsoft Excel

ขั้นตอนที่ 1 ใส่ข้อมูลดังรูป เลือกเมนู Data เลือก Data Analysis

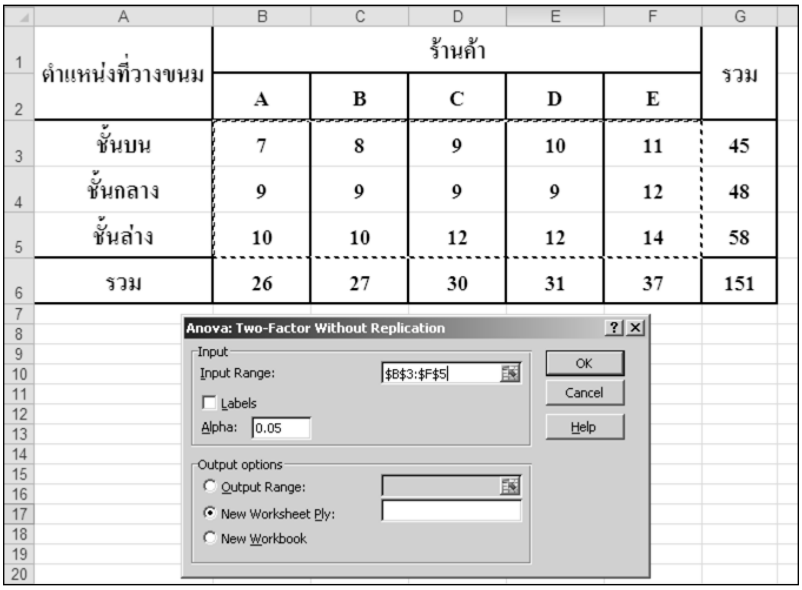


ตำแหน่งที่วางขนม	ร้านค้า					รวม
	A	B	C	D	E	
ชั้นบน	7	8	9	10	11	45
ชั้นกลาง	9	9	9	9	12	48
ชั้นล่าง	10	10	12	12	14	58
รวม	26	27	30	31	37	151

ขั้นตอนที่ 2 ในหน้าต่าง Data Analysis เลือก Anova: Two-Factor Without Replication ดังนี้



ขั้นตอนที่ 3 Input Range คือ กลุ่มของข้อมูล



ตำแหน่งที่วางขนม	ร้านค้า					รวม
	A	B	C	D	E	
ชั้นบน	7	8	9	10	11	45
ชั้นกลาง	9	9	9	9	12	48
ชั้นล่าง	10	10	12	12	14	58
รวม	26	27	30	31	37	151

ขั้นตอนที่ 4 จะได้ผลลัพธ์ ดังนี้

	A	B	C	D	E	F	G
1	Anova: Two-Factor Without Replication						
2							
3	*SUMMARY*	*Count*	*Sum*	*Average*	*Variance*		
4	Row 1	5	45	9	2.5		
5	Row 2	5	48	9.6	1.8		
6	Row 3	5	58	11.6	2.8		
7							
8	Column 1	3	26	8.666667	2.333333		
9	Column 2	3	27	9	1		
10	Column 3	3	30	10	3		
11	Column 4	3	31	10.33333	2.333333		
12	Column 5	3	37	12.33333	2.333333		
13							
14							
15	ANOVA						
16	*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
17	Rows	18.53333	2	9.266667	21.38462	0.000617	4.45897
18	Columns	24.93333	4	6.233333	14.38462	0.001002	3.837853
19	Error	3.466667	8	0.433333			
20							
21	Total	46.93333	14				

กรณีทดสอบว่าสาขามีผลทำให้ยอดขายขนมเฉลี่ยแตกต่างกันหรือไม่

H_0 : ยอดขายขนมเฉลี่ยของร้านค้าแต่ละสาขาไม่แตกต่างกัน

H_1 : ยอดขายขนมเฉลี่ยของร้านค้าแต่ละสาขาแตกต่างกันอย่างน้อย 1 คู่

หรือ

$$H_0 : \mu_{.1} = \mu_{.2} = \mu_{.3} = \mu_{.4} = \mu_{.5}$$
$$H_1 : \mu_{.1} \neq \mu_{.2} \neq \mu_{.3} \neq \mu_{.4} \neq \mu_{.5}$$

ตัวสถิติทดสอบคือ $F = \frac{MSA}{MSE}$ จากตาราง ANOVA F = 14.49

ค่าวิกฤตคือ $f_{1-\alpha, c-1, (c-1)(r-1)} = f_{0.95,4,8} = 3.84$

เนื่องจาก F=14.49 อยู่ในบริเวณปฏิเสธ H_0 หมายความว่ายอดขายขนมเฉลี่ยของร้านค้าแต่ละสาขาแตกต่างกันอย่างน้อย 1 คู่ หรือสาขาของร้านค้ามีผลทำให้ยอดขายขนมเฉลี่ยแตกต่างกันอย่างมีนัยสำคัญที่ระดับ 0.05

กรณีทดสอบว่าตำแหน่งการวางมีผลทำให้ยอดขายขนมเฉลี่ยแตกต่างกันหรือไม่

H_0 : ยอดขายขนมเฉลี่ยที่วางอยู่ในตำแหน่งแต่ละตำแหน่งไม่แตกต่างกัน

H_1 : ยอดขายขนมเฉลี่ยที่วางอยู่ในตำแหน่งแต่ละตำแหน่งแตกต่างกันอย่างน้อย 1 คู่

หรือ

$$H_0 : \mu_{1.} = \mu_{2.} = \mu_{3.}$$
$$H_1 : \mu_{1.} \neq \mu_{2.} \neq \mu_{3.}$$

ตัวสถิติทดสอบคือ $F = \dfrac{MSB}{MSE}$ จากตาราง ANOVA F = 21.55

ค่าวิกฤตคือ $f_{1-\alpha,r-1,(c-1)(r-1)} = f_{0.95,2,8} = 4.46$

เนื่องจาก F=21.55 อยู่ในบริเวณปฏิเสธ H_0 หมายความว่ายอดขายขนมเฉลี่ยที่วางอยู่ในตำแหน่งแต่ละตำแหน่งแตกต่างกันอย่างน้อย 1 คู่ หรือตำแหน่งการวางขนมมีผลทำให้ยอดขายขนมเฉลี่ยแตกต่างกันอย่างมีนัยสำคัญที่ระดับ 0.05

สรุปท้ายบท

การวิเคราะห์ความแปรปรวนเป็นวิธีการทางสถิติอ้างอิงที่ใช้ในการทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ยกรณีประชากรมากกว่า 2 กลุ่ม โดยใช้หลักการของความแปรปรวน และคำนวณค่าสถิติทดสอบ F ในรูปของตารางการวิเคราะห์ความแปรปรวน หรือตาราง ANOVA เพื่อให้เป็นขั้นตอนที่ง่ายและสะดวก ประเภทของการวิเคราะห์ความแปรปรวนขึ้นอยู่กับว่ามีปัจจัยที่เกี่ยวข้องกี่ปัจจัย ดังนั้นควรระมัดระวังในการเลือกวิธีการวิเคราะห์ความแปรปรวนให้เหมาะสมเพื่อผลการวิเคราะห์ข้อมูลที่ถูกต้อง

แบบฝึกหัดท้ายบท

1. ผู้จัดการฝ่ายการตลาดต้องการเปรียบเทียบยอดขายของกระดาษชำระทีมีลักษณะการบรรจุ 3 แบบ จึงทำการสุ่มตัวอย่างร้านที่ขายกระดาษชำระชนิดนี้มา 30 ร้าน และเก็บรวบรวมข้อมูลยอดขาย (ม้วน) ในเวลา 1 วันของแต่ละแบบมาจาก 10 ร้านเท่า ๆ กัน ได้ผลดังนี้

	ยอดขาย (ม้วน)		
ร้านที่	กล่องสี่เหลี่ยมจัตุรัส	กล่องสี่เหลี่ยมผืนผ้า	ห่อกลม
1	52	28	15
2	48	35	14
3	43	34	23
4	50	32	21
5	43	34	14
6	44	27	20
7	46	31	21

	ยอดขาย (ม้วน)		
ร้านที่	กล่องสี่เหลี่ยมจัตุรัส	กล่องสี่เหลี่ยมผืนผ้า	ห่อกลม
8	46	27	16
9	43	29	20
10	49	25	14

จากผลการวิเคราะห์ข้อมูลสรุปได้หรือไม่ว่า ลักษณะการบรรจุการดาษชำระ 3 แบบมีผลทำให้ยอดขายแตกต่างกันที่ระดับนัยสำคัญ 0.05

2. บริษัทABC จำกัด ผลิตฮาร์ดดิสก์สำหรับคอมพิวเตอร์ ขั้นตอนการติดขดลวดทองแดงเข้ากับแขนของหัวอ่านแผ่นดิสก์เป็นขั้นตอนการผลิตขั้นตอนหนึ่ง ในการติดขดลวดทองแดงเข้ากับแขนของหัวอ่านแผ่นดิสก์จะใช้กาว Epoxy 1140 หลังจากติดขดลวดทองแดงเข้ากับแขนของหัวอ่านแผ่นดิสก์แล้ว จะทำส่วนประกอบนี้เข้าเตาอบเพื่ออบที่อุณหภูมิ 180 °F เป็นเวลา 50 นาที วิศวกรฝ่ายผลิตต้องการศึกษาว่าอุณหภูมิกับระยะเวลาที่ใช้อบมีผลต่อแรงเฉือน (shear strength) ณ ตำแหน่งที่ติดกาวยึดส่วนประกอบทั้งสองอย่างไร เขาจึงทำการทดลองแบบ factorial design ข้อมูลของแรงเฉือนที่ได้จากการทดลองมีหน่วยเป็น psi แสดงดังตารางข้างล่างนี้

เวลาที่ใช้อบ(นาที)	อุณหภูมิที่ใช้อบ (°F)				
	150	180	200	250	300
30	20.3	19.5	22.1	17.6	23.6
	19.8	18.6	23	18.3	24.5
	21.4	18.9	22.4	18.2	25.1
40	21.6	20.1	20.1	19.5	17.6
	22.4	19.9	21	19.2	18.3
	21.3	20.5	19.8	20.3	18.1
50	19.8	19.6	22.3	19.4	22.1
	18.6	18.3	22	18.5	24.3
	21	19.8	21.6	19.1	23.8

จงวิเคราะห์ความแปรปรวนเพื่อทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ยของแรงฉือนที่ได้รับอิทธิพลจากอุณหภูมิกับระยะเวลาที่ใช้อบ โดยที่ระดับนัยสำคัญเท่ากับ 0.05 พร้อมทั้งสรุปผลที่ได้จากการวิเคราะห์