การวิเคราะห์ความแปรปรวน

Analysis of Variance :ANOVA

(การทดสอบความแตกต่างของค่าเฉลี่ยมากกว่า 2 กลุ่ม)

ผู้ช่วยศาสตราจารย์กัญญารัตน์ บุษบรรณ

การวิเคราะห์ความแปรปรวน

Analysis of Variance :ANOVA

- ทดสอบความแตกต่างระหว่างค่าเฉลี่ยของประชากรตั้งแต่ 3 กลุ่มขึ้นไป
- การทดสอบจะใช้การวิเคราะห์ความผันแปรทั้งหมดโดยจะแยกความผันแปรออกเป็นส่วนๆตามสาเหตุที่เกิด
- ทดสอบความต่างของข้อมูลที่ได้รับปัจจัยต่างระดับกันโดยการสร้างตารางการวิเคราะห์ความแปรปรวน
- ตัวแปรที่ใช้วิเคราะห์ต้องเป็นตัวแปรเชิงปริมาณ เช่น ปริมาณเชื้อรา คะแนนสอบ น้ำหนัก เป็นต้น
- ปัจจัยเป็นตัวแปรเชิงกลุ่ม เช่น ความชื้น วิธีสอน สูตรอาหาร เป็นต้น

ข้อมูล

- เป็นอิสระต่อกัน
- การแจกแจงแบบปกติที่ค่าเฉลี่ยใดๆและมีความแปรปรวนเท่ากันหมด

$$X_i \sim N(\mu_i, \sigma^2)$$

สมมติฐาน

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

H_1 : ประชากรอย่างน้อย 2 กลุ่ม มีค่าเฉลี่ยไม่เท่ากัน

ส่วนประกอบต่างๆของการทดลอง

กรรมวิธี (treatment)

คือสิ่ง หรือวิธีการที่จะปฏิบัติต่อหน่วยทดลอง เพื่อวัดผลที่ต้องการศึกษาเปรียบเทียบตามวัตถุประสงค์ของการทดลอง

หน่วยทดลอง (experimental unit)

สิ่งที่ได้รับการกระทำจากกรรมวิธีต่าง ๆที่ต้องการศึกษา

การซ้ำ (replication)

การที่กรรมวิธีปรากฏมากกว่าหนึ่งครั้งในการทดลอง

ความคลาดเคลื่อนของการทดลอง

Experimental Error

- ความแตกต่างระหว่างหน่วยทดลอง 2 หน่วย ซึ่งได้รับกรรมวิธีเดียวกัน
- หน่วยทดลองเป็นสิ่งที่ใช้วัดความคลาดเคลื่อนการทดลอง
- ความคลาดเคลื่อน(Error)

$$e_i \sim N(0,\sigma^2)$$

การวางแผนการทดลอง

Experimental Design

- วิธีการจัดหน่วยทดลองให้แก่ กรรมวิธีโดยวิธีสุ่ม(randomization)
- **แผนการทดลองแบบสุ่มตลอด(Completely Random Design : CRD)**
  - เป็นการสุ่มเพียงครั้งเดียว โดยหน่วยทดลองจะต้องเหมือนกันหรือคล้ายคลึงกันทุกหน่วย
- **แผนการทดลองแบบบล็อกสุ่ม((Randomized (completely) Block Design : RBD)**
  - จำแนกหน่วยทดลองเป็นบล็อกแล้วสุ่มภายในบล็อก

การวิเคราะห์ความแปรปรวนจำแนกทางเดียว(One-way Analysis)
แผนการทดลองแบบสุ่มตลอด(CRD)

มีนักศึกษา 20 คน สุ่มนักศึกษาขึ้นมาเพื่อทดลองสอนด้วยวิธีสอน 4 วิธี สุ่มให้วิธีละ 5 คน เมื่อสิ้นสุดการสอนจึงวัดผลใช้ข้อสอบชุดเดียวกัน เพื่อตรวจสอบดูว่าวิธีสอนแต่ละวิธีให้ผลแตกต่างกันหรือไม่

คะแนนสอบ

สมมติฐานทางสถิติสำหรับการทดสอบ

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

$: \mu_i$ คือคะแนนสอบเฉลี่ยจากวิธีสอนที่ i

H_1: ค่าเฉลี่ยของคะแนนสอบจากวิธีสอนอย่างน้อย 2 วิธีไม่เท่ากัน

ตารางค่าสังเกต

กรรมวิธี				
1	2	...		k
X_{11}	X_{12}			X_{1k}
X_{21}	X_{22}			X_{2k}
$\vdots$	$\vdots$		X_{ij}	$\vdots$
$x_{n_1,1}$	$x_{n_2,2}$			$X_{n_k k}$
$T_{.1}$	$T_{.2}$	...	$T_{.j}$	$T_{..k}$
$\bar{x}_{.1}$	$\bar{x}_{.2}$	...	$\bar{x}_{.j}$	$\bar{x}_{.k}$

T = ผลรวมทั้งหมด

$$= \sum_{j=1}^{k}\sum_{i=1}^{n_j} x_{ij}$$

$T_{.j}$ = ผลรวมของกรรมวิธีที่ j

$$= \sum_{i=1}^{n_j} x_{ij}$$

$\bar{x}_{.j}$ = ค่าเฉลี่ยของกรรมวิธีที่ j

$$= \frac{T_{\bullet j}}{n_j}$$

$\bar{x}$ = ค่าเฉลี่ยรวม

$$= \frac{T}{n}$$

การแยกความผันแปร(ผลบวกกำลังสอง)

Total Variation

$$SST = \sum_{j=1}^{k}\sum_{i=1}^{n}\left(X_{ij} - \bar{X}\right)^2$$

Summation of Square of Treatment

$$SSTr = \sum_{j=1}^{k}\left(\bar{X}_{.j} - \bar{X}\right)^2$$

Summation of Square of Error

$$SSE = \sum_{J=1}^{k}\sum_{i=1}^{n}\left(X_{ij} - \bar{X}_{.j}\right)^2$$

กรรมวิธี				
1	2	...		k
X_{11}	X_{12}			X_{1k}
X_{21}	X_{22}			X_{2k}
⋮	⋮		X_{ij}	⋮
$X_{n_1,1}$	$X_{n_2,2}$			$X_{n_k k}$
$T_{.1}$	$T_{.2}$	...	$T_{.j}$	$T_{..k}$
$\bar{X}_{.1}$	$\bar{X}_{.2}$	...	$\bar{X}_{.j}$	$\bar{X}_{.k}$

การคำนวณ SST = SSTr + SSE

$$SST=\sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(x_{ij}-\bar{x}\right)^2=\sum_{j=1}^{k}\sum_{i=1}^{n_j}x_{ij}^2-\frac{T^2}{n}$$

$$SSTr=\sum_{j=1}^{k}\left(\bar{x}_j-\bar{x}\right)^2=\sum_{j=1}^{k}\frac{T_{.j}^2}{n_j}-\frac{T^2}{n}$$

$$SSE=\sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(x_{ij}-\bar{x}_{.j}\right)^2=\sum_{j=1}^{k}\sum_{i=1}^{n_j}x_{ij}^2-\sum_{j=1}^{k}\frac{T_{.j}^2}{n_j}$$

ความผันแปรเฉลี่ยของการทดลอง

(mean of square of total)

$$MST = \frac{SST}{n-1}$$

ความแปรผันเฉลี่ยของกรรมวิธี

(mean of square of treatment)

$$MSTr = \frac{SSTr}{k-1}$$

ความแปรผันเฉลี่ยของความคลาดเคลื่อน

(mean of square of error)

$$MSE = \frac{SSE}{n-k}$$

ตารางวิเคราะห์ความแปรปรวน

แหล่งของความแปรผัน	d.f	SS	MS	F-ratio
กรรมวิธี ความคลาดเคลื่อน	k-1 n – k	SSTr SSE	MSTr MSE	$\frac{MSTr}{MSE}$
รวม	n -1	SST		

$$\frac{SSTr}{k-1}$$

คะแนนสอบนักศึกษา 20 คน

วิธีสอน			
1	2	3	4
61	46	75	59
68	53	64	73
59	59	68	70
66	56	78	63
51	61	75	75
305	275	360	340

Sum of Square:

SST = 83404 - 81920 = 1484

SSTr= 82770 - 81920 = 850

SSE = 1484 - 850 = 634

ตารางวิเคราะห์ความแปรปรวน

แหล่งของความแปรผัน	d.f	SS	MS	F-ratio
วิธีสอน	3	850	283.33	F = 7.15
ความคลาดเคลื่อน	16	634	39.625	
รวม	19	1484		

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

H_1 :ค่าเฉลี่ยคะแนนสอบจากวิธีสอนอย่างน้อย 2 วิธีไม่เท่ากัน

ถ้า $\mathbf{H_1}$ จริง(กรรมวิธีมีผล) MSTr จะมีค่ามาก บริเวณวิกฤต คือ $F > F_{\alpha,k-1,n-k}$

$F_{0.05,(3,16)} = 3.24$

(เปิดตารางสถิติ F)

เนื่องจาก **7.15>3.24** ดังนั้น สรุปผลการทดสอบ คือ ปฏิเสธ H_0

นั่นคือค่าเฉลี่ยคะแนนสอบจากวิธีสอนอย่างน้อย 2 วิธีไม่เท่ากัน ที่ระดับนัยสำคัญ 0.05

การเปรียบเทียบเชิงพหุ (Multiple Comparisons)

(จะกระทำเมื่อผลการทดสอบความแปรปรวนมีนัยสำคัญ: ปฏิเสธ H_0)

- Least Significant Difference(LSD) : $LSD = t_{\frac{\alpha}{2}, n-k} \sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$

- เปรียบเทียบผลต่างค่าเฉลี่ยของ 2 ประชากร $(\bar{X}_{(1)} - \bar{X}_{(2)}) \Leftrightarrow LSD$

- ถ้า $(\bar{X}_{(1)} - \bar{X}_{(2)}) > LSD$ แล้ว $\mu_{(1)} \neq \mu_{(2)}$

จากการทดสอบค่าเฉลี่ยของคะแนนสอบจากวิธีสอน 4 วิธี พบว่ามีค่าเฉลี่ยของคะแนนจากวิธีสอนอย่างน้อย 2 วิธีที่แตกต่างกัน

(ผลการทดสอบความแปรปรวนมีนัยสำคัญ)

ตรวจสอบว่าค่าเฉลี่ยของคะแนนจากวิธีสอนใดบ้างที่แตกต่างการจาก การเปรียบเทียบเชิงพหุ (Multiple Comparisons)

$$LSD = t_{\frac{\alpha}{2}, n-k} \sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

จำนวนตัวอย่างของทุกวิธีสอนเท่ากันดังนั้น ค่า LSD ของการทดสอบแต่ละคู่จะเป็นค่าเดียวกัน

$$LSD = t_{0.025,16} \sqrt{39.625\left(\frac{1}{5} + \frac{1}{5}\right)}$$

$$LSD = 2.12 \sqrt{39.625\left(\frac{1}{5} + \frac{1}{5}\right)}$$

$$LSD \approx 8.0$$

Multiple Comparisons: $\overline{X}_2$ $\overline{X}_1$ $\overline{X}_4$ $\overline{X}_3$

LSD=**8.0**

55	61	68	72

$$\overline{X}_3-\overline{X}_4=72-68=4.0$$

$$\boxed{\overline{X}_3-\overline{X}_1=72-61=11.0^*}$$

$$\boxed{\overline{X}_3-\overline{X}_2=72-55=17.0^*}$$

$$\overline{X}_4-\overline{X}_1=68-61=7.0$$

$$\boxed{\overline{X}_4-\overline{X}_2=68-55=13.0^*}$$

$$\overline{X}_1-\overline{X}_2=61-55=6.0$$

สรุปผล

วิธีสอนที่ 1 ให้ผลแตกต่างจากวิธีสอนที่ 3

วิธีสอนที่ 2 ให้ผลแตกต่างจากวิธีสอนที่ 3 และ

วิธีสอนที่ 4

ตัวอย่างที่ 1 เพื่อศึกษาว่าระดับความชื้นต่าง ๆ ในห้องเก็บรักษาจะมีอิทธิพลต่อปริมาณเชื้อราที่จะเกิดขึ้นบนแผ่นยางหรือไม่ จึงทำการทดลองโดยใช้แผ่นยางขนาดและคุณภาพยางตลอดจนอายุการผลิตอย่างเดียวกัน เก็บไว้ในห้องเก็บรักษาโดยมีความชื้นต่าง ๆ กัน 5 ระดับ บันทึกปริมาณเชื้อราที่เกิดขึ้นบนแผ่นยาง ได้ดังนี้

ระดับความชื้นในห้องเก็บรักษา					
1	2	3	4	5	
7.3	5.4	8.1	7.9	7.1	
8.3	7.4	6.4	9.5	7.3	
7.6	7.0		10.0		
8.7					
31.9	19.8	14.5	27.4	14.4	108

จงวิเคราะห์ข้อมูลเพื่อสรุปว่าระดับความชื้นในห้องเก็บรักษาทั้ง 5 ระดับ มีอิทธิพลต่อปริมาณเชื้อราต่างกันหรือไม่ ที่ระดับนัยสำคัญ .01

ตัวอย่างที่ 2 สถาปนิกคนหนึ่งต้องการทดสอบระยะเวลาที่สีจะแห้งของสี 4 ชนิด จึงทดลองทาสีทั้ง 4 ชนิด บนวัสดุอย่างเดียวกัน ขนาดหนึ่งตารางเมตร ชนิดละ 5 ตารางเมตร บันทึกระยะเวลาที่สีแห้งภายหลังจากทาเสร็จ (หน่วย : นาที) ได้ข้อมูลดังตารางข้างล่างนี้ จงวิเคราะห์ข้อมูลด้วยระดับนัยสำคัญ 5 %

	ชนิดของสี				
	A	B	C	D	
	73	74	68	71	
	73	74	69	71	
	73	74	69	72	
	75	74	69	72	
	75	75	70	73	
รวม	369	371	345	359	1,444

การวิเคราะห์ความแปรปรวนสองทาง (two-way Analysis) แผนการทดลองแบบบล็อกสุ่ม

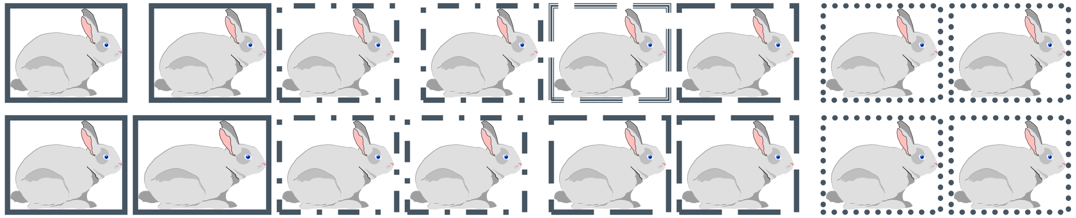
- หน่วยทดลองแตกต่างกัน
- จำแนกหน่วยทดลองเป็นกลุ่ม เรียกว่า บล็อก
- หน่วยทดลองในบล็อกเดียวกันมีความคล้ายคลึงกันมากที่สุด
- มีแหล่งความผันแปรเนื่องจากบล็อก

แผนการทดลองแบบ บล็อกสุ่ม (RBD)

ต้องการทดสอบสูตรอาหารเลี้ยงกระต่าย4 สูตรกับกระต่าย 4 สายพันธ์

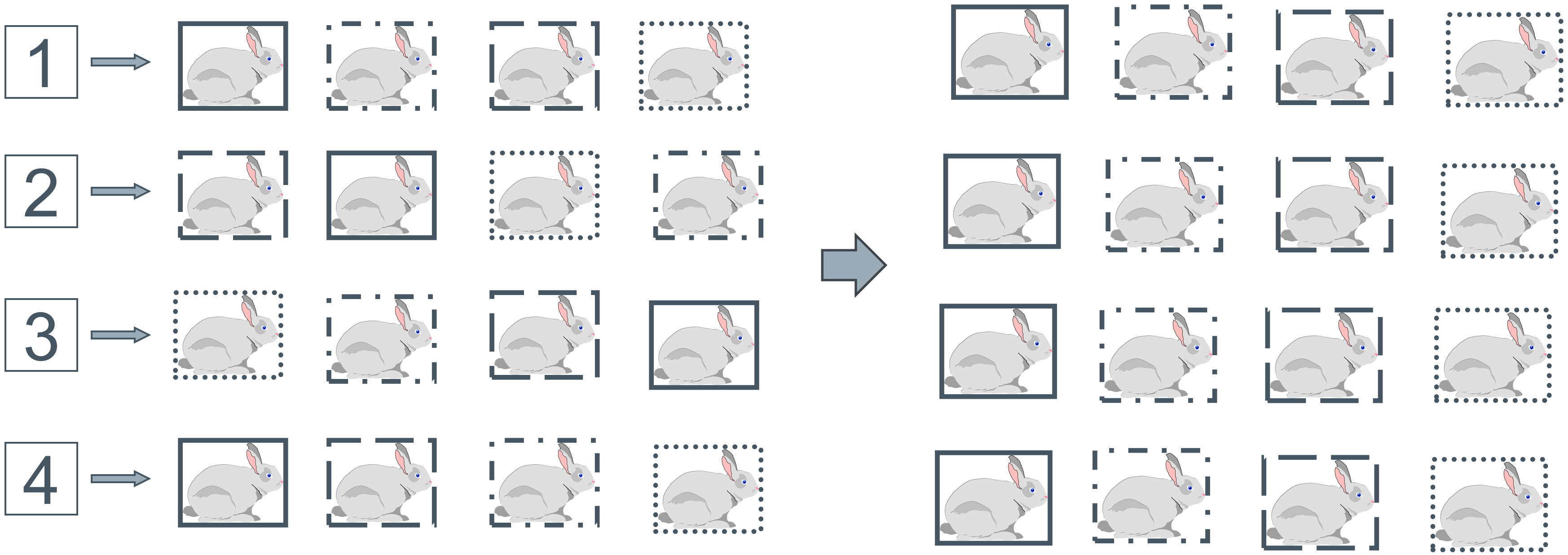
(กรรมวิธี)สูตรอาหาร : 1 2 3 4

(บล็อก) สายพันธ์ :

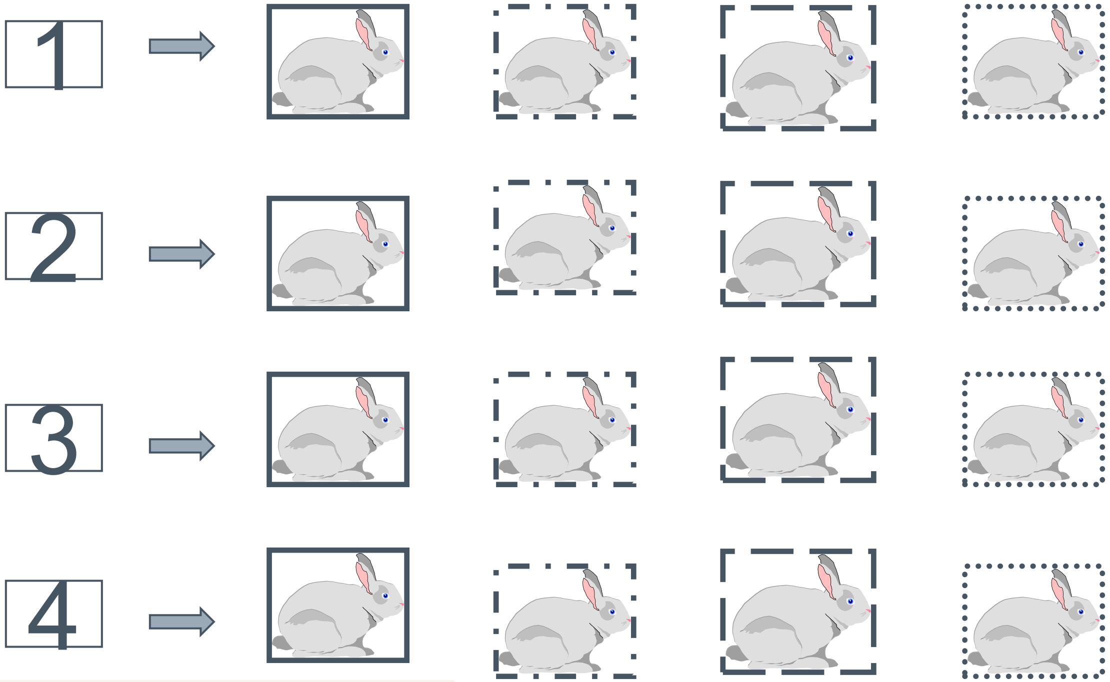


Treatment

Block



Treatment(กรรมวิธี:สูตรอาหาร Block



สมมติฐานที่จะทดสอบ

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

H_1: ค่าเฉลี่ยจากสองกรรมวิธีอย่างน้อย 2 กรรมวิธีไม่เท่ากัน

ส่วนประกอบจองผลรวมกำลังสอง : SST = SSTr + SSB + SSE

สูตรใช้คำนวณ

$$SST = \sum\sum x_{ij}^2 - \frac{T^2}{kn}$$

$$SSTr = \sum_j \frac{T_{.j}^2}{n_j} - \frac{T^2}{kn}$$

$$SSB = \sum_i \frac{T_{i.}^2}{n_i} - \frac{T^2}{kn}$$

SSE = SST - SSTr - SSB

ตารางวิเคราะห์ความแปรปรวน

แหล่งของความแปรผัน	d.f	SS	MS	F-ratio
กรรมวิธี	k-1	SSTr	MSTr	$\frac{MSTr}{MSE}$
บล็อก	n-1	SSB		
ความคลาดเคลื่อน	(k-1)(n-1)	SSE	MSE	
รวม	kn -1	SST		

ตัวอย่างที่ 4 จากการศึกษาประสิทธิภาพของสูตรอาหารเลี้ยงวัวกับวัวสายพันธ์ต่างๆ ผลการทดลองเป็นระยะเวลา 6 เดือน ได้น้ำหนักดังตาราง

สายพันธ์	สูตรอาหาร				รวม
	1	2	3	4	
A	550	450	700	500	2200
B	300	350	550	400	1600
C	350	550	400	600	1900
รวม	1200	1350	1650	1500	5700

จงทดสอบว่าประสิทธิภาพของอาหารเลี้ยงวัว 4 สูตรแตกต่างกันหรือไม่ที่ระดับนัยสำคัญ .01

สมมติฐานที่จะทดสอบ

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ $: \mu_i$ คือน้ำหนักเฉลี่ยของวัวจากการเลี้ยงด้วยอาหารสูตรที่ i

H_1: ค่าเฉลี่ยของน้ำหนักวัวจากการเลี้ยงด้วยอาหารอย่างน้อย 2 สูตรแตกต่างกัน

ตารางวิเคราะห์ความแปรปรวน

แหล่งของความแปรผัน	d.f	SS	MS	F-ratio
สูตรอาหาร	3	37,500	12,500	F=1.0
สายพันธ์	2	45,000	22,500	
ความคลาดเคลื่อน	6	75,000	12,500	
รวม	11			

$F_{0.01,(3,6)}$=9.78 (เปิดตาราง)

F_{cal}=1.0<9.78 **ยอมรับ H_0** นั่นคือสูตรอาหารทั้ง 4 สูตรมีประสิทธิภาพไม่แตกต่างกัน

ตัวอย่างที่ 5 ผู้บริหารโรงแรมแห่งหนึ่งซึ่งมี 3 สาขา อยู่ที่เชียงใหม่ พัทยา และหัวหิน ได้สำรวจจำนวนผู้เข้าพักในโรงแรม ในแต่ละฤดูกาลของปีที่ผ่านมา ได้ข้อมูลดังตาราง ต่อไปนี้

ฤดูกาล	สาขาของโรงแรม		
	เชียงใหม่	พัทยา	หัวหิน
ฤดูร้อน	200	220	160
ฤดูฝน	180	210	140
ฤดูหนาว	250	280	150

ที่ระดับนัยสำคัญ .05 จะสรุปได้หรือไม่ว่าจำนวนผู้เข้าพักในโรงแรมทั้ง 3 แห่งไม่แตกต่างกัน