การวิเคราะห์ความแปรปรวน

แนวคิดของการวิเคราะห์ความแปรปรวน คือ การแยกความแปรปรวนของข้อมูลที่เกิดขึ้นทั้งหมดจาก

สาเหตุต่างๆที่ทำให้เกิดความแปรปรวน ตัวอย่างเช่น การที่จะสรุปว่าเสื้อผ้าที่ตัดเย็บด้วยมือ และ ตัดเย็บด้วยเครื่องจักรไม่มีความแตกต่างกัน อาจทำได้โดยการนำเสื้อหลายๆตัวที่ตัดเย็บด้วยมือและเสื้อหลายๆตัวที่ตัดเย็บด้วยเครื่องจักรมาวางปนกัน แล้วทำการแยก ถ้าผู้แยกสามารถทำการแยกได้ถูกต้องแสดงว่าเสื้อที่ตัดเย็บด้วยมือมีความแตกต่างจากเสื้อที่ตัดเย็บด้วยเครื่องจักร

ตัวอย่าง บริษัทขายรถยนต์แห่งหนึ่งต้องการเปรียบเทียบความสามารถของพนักงานขาย 3 คน ว่ามีความแตกต่างกันหรือไม่

เดือน	กิตติ	อนุชา	วันดี
มกราคม	5	8	10
กุมภาพันธ์	6	9	10
มีนาคม	5	8	9
เมษายน	8	7	8
พฤษภาคม	6	9	8
มิถุนายน	7	9	9
กรกฎาคม	6	10	10
สิงหาคม	5	8	9
กันยายน	6	8	8
ตุลาคม	7	9	9
รวม	61	85	90
เฉลี่ย	6.1	8.5	9.0

ข้อตกลงเบื้องต้นสำหรับการวิเคราะห์ความแปรปรวน

- ข้อมูลที่ได้จากตัวอย่างที่เลือกมาจากประชากรที่นำมาทดสอบความแตกต่างระหว่างค่าเฉลี่ย
- ข้อมูลแต่ละชุดที่นำมาทดสอบความแตกต่างระหว่างค่าเฉลี่ยจะต้องมีการแจกแจงปกติ
- ความแปรปรวนของข้อมูลจากประชากรแต่ละชุดที่นำมาทดสอบความแตกต่างระหว่างค่าเฉลี่ยจะต้องเท่ากัน

* ความถูกต้อง และความเชื่อถือได้ของผลการทดสอบสมมติฐานโดยใช้การวิเคราะห์ความแปรปรวนจะมีมากน้อยเพียงใดขึ้นอยู่กับคุณสมบัติทั้ง 3 ข้อ ในกรณีที่ผู้ทดสอบสมมติฐานไม่แน่ใจว่าข้อมูลของประชากรแต่ละชุดมีแจกแจงปกติหรือไม่ อาจทดสอบได้โดยใช้วิธีที่เรียกว่า การทดสอบภาวะสารูปสนิทดี (test for goodness of fit) ละถ้าไม่แน่ใจว่าความแปรปรวนของข้อมูลจากประชากรแต่ละชุดที่นำมาทดสอบสมมติฐานเท่ากันหรือไม่อาจทำการทดสอบได้โดยวิธีการทดสอบ ความเท่ากันของความแปรปรวนของข้อมูลในประชากรแต่ละชุด

การวิเคราะห์ความแปรปรวนแบบจำแนกทางเดียว

- ใช้ในการทดสอบความแตกต่างระหว่างค่าเฉลี่ยของประชากรตั้งแต่ 2 ชุดขึ้นไป โดยที่ประชากรมีความแตกต่างกัน เนื่องจากลักษณะเดียว เช่น การทดสอบความแตกต่างระหว่างราคาเฉลี่ยของยาสีฟันยี่ห้อหนึ่ง ในภาคเหนือ ภาคตะวันออกเฉียงเหนือ ภาคกลาง และภาคใต้ (ราคาเฉลี่ยของยาสีฟันมีความแตกต่างกันเนื่องจากลักษณะเดียว คือ ภาค)

ตัวอย่าง บริษัทขายรถยนต์แห่งหนึ่งต้องการเปรียบเทียบความสามารถของพนักงานขาย 3 คน ว่ามีความแตกต่างกันหรือไม่

เดือน	กิตติ	อนุชา	วันดี
มกราคม	5	8	10
กุมภาพันธ์	6	9	10
มีนาคม	5	8	9
เมษายน	8	7	8
พฤษภาคม	6	9	8
มิถุนายน	7	9	9
กรกฎาคม	6	10	10
สิงหาคม	5	8	9
กันยายน	6	8	8
ตุลาคม	7	9	9
รวม	61	85	90
เฉลี่ย	6.1	8.5	9.0

ตัวอย่าง บริษัทขายรถยนต์แห่งหนึ่งต้องการเปรียบเทียบความสามารถของพนักงานขาย 3 คน ว่ามีความแตกต่างกันหรือไม่

เดือน	กิตติ	อนุชา	วันดี	
มกราคม	5(X_{11})	8(X_{21})	10 (X_{31})	
กุมภาพันธ์	6(X_{12})	9(X_{22})	10(X_{32})	
มีนาคม	5(X_{13})	8(X_{23})	9(X_{33})	
เมษายน	8(X_{14})	7(X_{24})	8 (X_{34})	
พฤษภาคม	6(X_{15})	9(X_{25})	8(X_{35})	
มิถุนายน	7(X_{16})	9(X_{26})	9(X_{36})	
กรกฎาคม	6(X_{17})	10(X_{27})	10(X_{37})	
สิงหาคม	5(X_{18})	8 (X_{28})	9(X_{38})	
กันยายน	6(X_{19})	8(X_{29})	8(X_{39})	
ตุลาคม	7(X_{110})	9 (X_{210})	9 (X_{310})	
รวม	61 (T_1)	85(T_2)	90(T_3)	T=61+85+90
เฉลี่ย	6.1	8.5	9.0	X= 7.87

สัญลักษณ์ที่ใช้ในการวิเคราะห์ความแปรปรวน

- k หรือ N แทนจำนวนประชากรทั้งหมดที่นำมาทดสอบ
- n_i แทนจำนวนตัวอย่างที่เลือกมาจากประชากรชุดที่ i
- n แทนจำนวนตัวอย่างทั้งหมดที่เลือกมาจากประชากรทุกชุด $= n_1 + n_2 + n_3 + n_k$
- x_{ij} แทนข้อมูลซึ่งได้จากตัวอย่างที่ j ที่เลือกจากประชากรชุดที่ i
- T_i แทนผลรวมของข้อมูลจากทุกตัวอย่างที่เลือกมาจากประชากรชุดที่ i
- T แทนผลรวมของข้อมูลจากทุกตัวอย่างที่เลือกมาจากประชากรทุกชุด
- x_i แทนค่าเฉลี่ยของข้อมูลจากทุกตัวอย่างที่เลือกมาจากประชากรชุดที่ i
- x แทนค่าเฉลี่ยของข้อมูลจากทุกตัวอย่างที่เลือกมาจากประชากรทุกชุด
- μ_i แทนค่าเฉลี่ยของข้อมูลจากประชากรชุดที่ i

ผลรวมกำลังสองเฉลี่ยของยอดรวม (Total sum of square)

- เขียนแทนด้วยสัญลักษณ์ MS (T) ซึ่งมีสูตรที่ใช้ดังนี้

$$\mathrm{MS(T)} = \frac{SS(T)}{df\,(T)} = \frac{1}{\sum n_i - 1}\left[\sum\sum x^2{}_{ij} - \frac{T^2}{n}\right]$$

ผลรวมกำลังสองเฉลี่ยระหว่างประชากร

- เขียนแทนสัญลักษณ์ MS(B) ซึ่งมีสูตรที่ใช้ดังนี้

$$\mathrm{MS(B)} = \frac{SS(B)}{df(B)} = \frac{1}{k-1}\left[\sum \frac{T_i^{\,2}}{n_i} - \frac{T^2}{n}\right]$$

ผลรวมกำลังสองเฉลี่ยประชากร

- เขียนแทนด้วยสัญลักษณ์ MS (W) ซึ่งมีสูตรที่ใช้ดังนี้

$$\mathrm{MS(W)} = \frac{1}{\sum n_i - k}\left[\sum \sum x^2{}_{ij} \;-\; \sum \frac{T_i{}^2}{n_i}\right]$$

ข้อสังเกต

เนื่องจาก $SS(T) = SS(B) + SS(W)$ และ $df(T)= df(B) + df(W)$

จึงนิยมหา $SS(W)$ และ $df(W)$ จาก $SS(W) = SS(T)-SS(B)$

$df(W) = df(T)-df(B)$

การทดสอบสมมติฐานโดยการวิเคราะห์ความแปรปรวน

- นิยมเขียนในรูปตารางวิเคราะห์ความแปรปรวน

สาเหตุของความแปรปรวน	df	SS	MS	F
ระหว่างประชากร	k-1	$SS(B)=\sum\frac{T_i^2}{n_i}-\frac{T^2}{n}$	$MS(B)=\frac{SS(B)}{k-1}$	$\frac{MS(B)}{MS(W)}$
ภายในประชากร	n-k	$SS(W)=\sum\sum x^2_{ij}-\sum\frac{T_i^2}{n_i}$	$MS(W)=\frac{SS(W)}{n-k}$	
รวม	n-1	$SS(T)=\sum\sum x^2_{ij}-\sum\frac{T^2}{n}$		

ตัวอย่าง บริษัทขายรถยนต์แห่งหนึ่งต้องการเปรียบเทียบความสามารถของพนักงานขาย 3 คน ว่ามีความแตกต่างกันหรือไม่

เดือน	กิตติ	อนุชา	วันดี	
มกราคม	$5(X_{11})$	$8(X_{21})$	$10\ (X_{31})$	
กุมภาพันธ์	$6(X_{12})$	$9(X_{22})$	$10(X_{32})$	
มีนาคม	$5(X_{13})$	$8(X_{23})$	$9(X_{33})$	
เมษายน	$8(X_{14})$	$7(X_{24})$	$8\ (X_{34})$	
พฤษภาคม	$6(X_{15})$	$9(X_{25})$	$8(X_{35})$	
มิถุนายน	$7(X_{16})$	$9(X_{26})$	$9(X_{36})$	
กรกฎาคม	$6(X_{17})$	$10(X_{27})$	$10(X_{37})$	
สิงหาคม	$5(X_{18})$	$8\ (X_{28})$	$9(X_{38})$	
กันยายน	$6(X_{19})$	$8(X_{29})$	$8(X_{39})$	
ตุลาคม	$7(X_{110})$	$9\ (X_{210})$	$9\ (X_{310})$	
รวม	$61\ (T_1)$	$85(T_2)$	$90(T_3)$	T=60+85+90
เฉลี่ย	6.1	8.5	9.0	X= 7.87

ตารางวิเคราะห์ความแปรปรวนสำหรับข้อมูลของตัวอย่างพนักงานขายรถยนต์

สาเหตุของความแปรปรวน	df	SS	MS	F
ระหว่างพนักงาน	k-1=2	SS(B)=48.07	MS(B)=$\frac{SS(B)}{k-1}$ = 24.04	$\frac{MS(B)}{MS(W)}$ = 30.43
ภายในพนักงานขายคนเดียว	n-k=27	SS(W)=21.40	MS(W)=$\frac{SS(W)}{n-k}$=0.79	
รวม	n-1=29	SS(T)= 69.47		

บริษัทรับส่งสินค้าแห่งหนึ่งต้องการซื้อรถบรรทุกสินค้าที่กินน้ำมันน้อยมาใช้ในกิจการของบริษัท มีตัวแทนจำหน่ายรถบรรทุก 4 ยี่ห้อ คือ **A B C D** มาเสนอขายรถกับบริษัทจากการวัดระยะทางที่รถบรรทุกทั้ง 4 ยี่ห้อนี้วิ่งได้เมื่อใช้น้ำมัน 1ลิตร โดยที่ทดลองกับรถแต่ละยี่ห้อจำนวน 5 คันในสถาพถนนที่ไม่แตกต่างกันปรากฏว่า

ระยะทางที่รถวิ่งได้ (km)				
A	B	C	D	
14	15	10	13	
11	13	12	11	
12	12	11	12	
11	16	10	15	
11	12	12	13	
T1=59	T2=68	T3=55	T4=64	T=246

รถบรรทุกทั้ง 4 ยี่ห้อกินน้ำมันแตกต่างกันหรือไม่ที่ระดับนัยสำคัญ 0.05

สรุปขั้นตอนการทดสอบสมมติฐาน

ขั้นตอนที่ 1 ตั้งสมมติฐานเพื่อการทดสอบ

เป็นขั้นตอนในการกำหนด H_0 และ H_1

ขั้นตอนที่ 2 กำหนดระดับนัยสำคัญ

โดยทั่วไปมักใช้ α = 0.10 , 0.05 ,0.01

ขั้นตอนที่ 3 กำหนดสถิติทดสอบ

F test

ขั้นตอนที่ 4 คำนวณค่าสถิติทดสอบ

คำนวณค่าสถิติทดสอบจากสูตรที่เลือกให้เหมาะสมกับเงื่อนไขใช้ตารางวิเคราะห์ความแปรปรวน

สาเหตุของความแปรปรวน	df	SS	MS	F
ระหว่างประชากร	k-1	$SS(B)=\sum\frac{T_i^2}{n_i}-\frac{T^2}{n}$	$MS(B)=\frac{SS(B)}{k-1}$	$\frac{MS(B)}{MS(W)}$
ภายในประชากร	n-k	$SS(W)=\sum\sum x^2_{ij}-\sum\frac{T_i^2}{n_i}$	$MS(W)=\frac{SS(W)}{n-k}$	
รวม	n-1	$SS(T)=\sum\sum x^2_{ij}-\sum\frac{T^2}{n}$		

ขั้นตอนที่ 5 การสร้างเขตปฏิเสธสมมติฐาน/เปรียบเทียบค่าคำนวณกับค่าจากตาราง

หาค่าวิกฤติ และกำหนดขอบเขตของการปฏิเสธ

ขั้นตอนที่ 6 สรุปผลการทดสอบ

สรุปผลในการยอมรับ H_0 หรือ H_1

ในการทดสอบความแตกต่างระหว่างราคาขายปลีกเฉลี่ยของสินค้าชนิดหนึ่งในจังหวัด โคราช อุดร นครสวรรค์ เชียงใหม่ และสงขลา ผู้ทดสอบได้เลือกสินค้าชนิดนี้มาจากแต่ละจังหวัดข้อมูลเป็นดังนี้

โคราช	อุดร	นครสวรรค์	เชียงใหม่	สงขลา
78	84	80	75	80
70	80	79	73	80
72	82	82	70	
80		86	70	
		83		

ราคาขายปลีกเฉลี่ยของสินค้าทั้ง 5 จังหวัดแตกต่างกันหรือไม่ ที่ระดับนัยสำคัญ 0.05

สรุปขั้นตอนการทดสอบสมมติฐาน

ขั้นตอนที่ 1 ตั้งสมมติฐานเพื่อการทดสอบ

เป็นขั้นตอนในการกำหนด H_0 และ H_1

ขั้นตอนที่ 2 กำหนดระดับนัยสำคัญ

โดยทั่วไปมักใช้ α = 0.10 , 0.05 ,0.01

ขั้นตอนที่ 3 กำหนดสถิติทดสอบ

F test

ขั้นตอนที่ 4 คำนวณค่าสถิติทดสอบ

คำนวณค่าสถิติทดสอบจากสูตรที่เลือกให้เหมาะสมกับเงื่อนไขใช้ตารางวิเคราะห์ความแปรปรวน

สาเหตุของความแปรปรวน	df	SS	MS	F
ระหว่างประชากร	k-1	$SS(B)=\sum \frac{T_i^2}{n_i} - \frac{T^2}{n}$	$MS(B)=\frac{SS(B)}{k-1}$	$\frac{MS(B)}{MS(W)}$
ภายในประชากร	n-k	$SS(W)=\sum\sum x_{ij}^2 - \sum \frac{T_i^2}{n_i}$	$MS(W)=\frac{SS(W)}{n-k}$	
รวม	n-1	$SS(T)=\sum\sum x_{ij}^2 - \sum \frac{T^2}{n}$		

ขั้นตอนที่ 5 การสร้างเขตปฏิเสธสมมติฐาน/เปรียบเทียบค่าคำนวณกับค่าจากตาราง

หาค่าวิกฤติ และกำหนดขอบเขตของการปฏิเสธ

ขั้นตอนที่ 6 สรุปผลการทดสอบ

สรุปผลในการยอมรับ H_0 หรือ H_1

การวิเคราะห์ความแปรปรวนแบบจำแนกสองทาง

- ใช้ในการทดสอบความแตกต่างระหว่างค่าเฉลี่ยของประชากรตั้งแต่ 2 ชุดขึ้นไป โดยที่ประชากรมีความแตกต่างกันเนื่องจาก 2 ลักษณะ ตัวอย่างเช่น ผู้จัดการฝ่ายขายต้องการทราบว่าระดับการศึกษาและประสบการณ์ในการขายของพนักงานขาย มีผลต่อยอดขายที่แตกต่างกันหรือไม่

	ประสบการณ์	ระดับการศึกษา		
		ประถม	มัธยม	อุดมศึกษา
แถวที่ 1	ไม่มี	5	7	6
แถวที่ 2	1	8	8	5
แถวที่ 3	2	4	9	7
แถวที่ 4	มากกว่า 2	2	12	4
		สดมภ์	สดมภ์	สดมภ์

สัญลักษณ์ที่ใช้ในการวิเคราะห์ความแปรปรวน

- n แทนจำนวนประชากร/ตัวอย่างทั้งหมดที่นำมาทดสอบ $= r * c$
- r แทนจำนวนแถว
- c แทนจำนวนสดมภ์
- x_{ij} แทนข้อมูลซึ่งได้จากตัวอย่างแถวที่ i และสดมภ์ที่ j
- T_i แทนผลรวมของข้อมูลจากทุกตัวอย่างที่เลือกมาจากประชากรจากแถวที่ i
- T_j แทนผลรวมของข้อมูลจากทุกตัวอย่างที่เลือกมาจากประชากรจากสดมภ์ที่ j
- T แทนผลรวมของข้อมูลจากทุกตัวอย่างที่เลือกมาจากประชากรทุกแถวและทุกสดมภ์
- x_i แทนค่าเฉลี่ยของข้อมูลจากทุกตัวอย่างที่เลือกมาจากประชากรชุดที่ i
- x_i แทนค่าเฉลี่ยของข้อมูลจากทุกตัวอย่างที่เลือกมาจากประชากรชุดที่ i
- x_j แทนค่าเฉลี่ยของข้อมูลจากทุกตัวอย่างที่เลือกมาจากประชากรชุดที่ j
- x แทนค่าเฉลี่ยของข้อมูลจากทุกตัวอย่างที่เลือกมาจากประชากรทุกแถวและทุกสดมภ์
- μ_i แทนค่าเฉลี่ยของข้อมูลจากประชากรชุดที่ i

การทดสอบสมมติฐานโดยการวิเคราะห์ความแปรปรวน

- นิยมเขียนในรูปตารางวิเคราะห์ความแปรปรวน

สาเหตุของความแปรปรวน	df	SS	MS	F
ระหว่างแถว	r-1	SS(R) $=\sum \frac{T^2{}_i}{c} - \frac{T^2}{rc}$	MS(R)$=\frac{SS(R)}{r-1}$	$\frac{MS(R)}{MS(W)}$
ระหว่างสดมภ์	c-1	SS(c) $=\sum \frac{T^2{}_i}{c} - \frac{T^2}{rc}$	MS(c)$=\frac{SS(c)}{c-1}$	$\frac{MS(C)}{MS(W)}$
ภายในแถวและสดมภ์เดียวกัน	(r-1)(c-1)	SS(w) =SS(T)-SS(R)-SS(C)	MS(w)$=\frac{SS(w)}{(r-1)(c-1)}$	
รวม	rc-1	SS(T) $=\sum\sum x^2{}_{ij} - \frac{T^2}{rc}$		

จากตัวอย่าง

ประสบการณ์	ระดับการศึกษา			รวม	เฉลี่ย
	ประถม	มัธยม	อุดมศึกษา		
ไม่มี	5 (X_{11})	7(X_{12})	6(X_{13})	$T_{1.}=18$	6
1	8(X_{21})	8(X_{22})	5 (X_{23})	$T_{2.}=21$	7
2	4(X_{31})	9(X_{32})	7(X_{33})	$T_{3.}=20$	6.67
มากกว่า 2	2(X_{41})	12(X_{42})	4(X_{43})	$T_{4.}=18$	6
รวม	$T_{.1}=19$	$T_{.2}=36$	$T_{.3}=22$	77(T)	
เฉลี่ย	4.75	9	5.5		6.42

การทดสอบสมมติฐาน

- ขั้นตอนที่ 1 ตั้งสมมติฐานเพื่อการทดสอบ

  เป็นขั้นตอนในการกำหนด H0 และ H1

- ขั้นตอนที่ 2 กำหนดระดับนัยสำคัญ

  โดยทั่วไปมักใช้ $\alpha = 0.10$, 0.05 ,0.01

- ขั้นตอนที่ 3 กำหนดสถิติทดสอบ

  F test

- ขั้นตอนที่ 4 คำนวณค่าสถิติทดสอบ

  คำนวณค่าสถิติทดสอบจากสูตรที่เลือกให้เหมาะสมกับเงื่อนไขใช้ตารางวิเคราะห์ความแปรปรวน

สาเหตุของความแปรปรวน	df	SS	MS	F
ระหว่างแถว	r-1	$SS(R) = \sum \frac{T^2_i}{c} - \frac{T^2}{rc}$	$MS(R) = \frac{SS(R)}{r-1}$	$\frac{MS(R)}{MS(W)}$
ระหว่างสดมภ์	c-1	$SS(c) = \sum \frac{T^2_j}{c} - \frac{T^2}{rc}$	$MS(c) = \frac{SS(c)}{c-1}$	$\frac{MS(C)}{MS(W)}$
ภายในแถวและสดมภ์เดียวกัน	(r-1)(c-1)	SS(w) =SS(T)-SS(R)-SS(C)	$MS(w) = \frac{SS(w)}{(r-1)(c-1)}$	
รวม	rc-1	$SS(T) = \sum \sum x^2_{ij} - \frac{T^2}{rc}$		

- ขั้นตอนที่ 5 การสร้างเขตปฏิเสธสมมติฐาน/เปรียบเทียบค่าคำนวณกับค่าจากตาราง

  หาค่าวิกฤติ และกำหนดขอบเขตของการปฏิเสธ

- ขั้นตอนที่ 6 สรุปผลการทดสอบ

  สรุปผลในการยอมรับ H0 หรือ H1

การทดสอบไคสแควร์

- ข้อมูลที่อยู่ในรูปความถี่ สามารถนำไปใช้ทดสอบสมมติฐานเกี่ยวกับอัตราส่วนของลักษณะที่ผู้วิเคราะห์สนใจศึกษา หรือ ใช้ทดสอบความเป็นอิสระระหว่างข้อมูลหรือลักษณะที่ผู้วิเคราะห์สนใจศึกษา
- ใช้เมื่อต้องการทดสอบว่าอัตราส่วนของข้อมูลที่มีลักษณะต่างๆเป็นไปตามที่คาดไว้หรือไม่
- ตัวสถิติที่ใช้ทดสอบ คือ $\lambda^2 = \frac{(O_i - E_i)^2}{E_i}$

ซึ่งมีการแจกแจงไคสแควร์ที่ df เท่ากับ k-1

เมื่อ O_i แทนความถี่ที่ได้จากการทดลองหรือการสังเกต

E_i แทนความถี่ที่คาดว่าควรจะเป็นหรือตามทฤษฎี

k แทนจำนวนพวกหรือกลุ่มที่ทำการทดสอบ

ตัวอย่าง

- ในการทดสอบความรู้ของผู้เข้ารับการอบรมในหลักสูตรการขายสินค้าของบริษัทแห่งหนึ่งจำนวน 150 คน ผู้จัดการอบรม**คาดว่าผลการทดสอบความรู้ของผู้เข้ารับการอบรมซึ่งแบ่งเป็น 3 ระดับ คือ ดีมาก ดี และพอใช้ จะมีอัตราส่วนเท่ากับ 2:1:2** แต่ภายหลังการทดสอบความรู้ปรากฏว่าอัตราส่วนที่ได้เป็น 7:3:5 อัตราส่วนที่ได้จากการทดสอบความรู้กับอัตราส่วนที่ผู้จัดการอบรมคาดไว้มีความแตกต่างกันที่ระดับนัยสำคัญ 0.05 หรือไม่

การทดสอบสมมติฐาน

- ขั้นตอนที่ 1 ตั้งสมมติฐานเพื่อการทดสอบ

  เป็นขั้นตอนในการกำหนด H_0 และ H_1

- ขั้นตอนที่ 2 กำหนดระดับนัยสำคัญ

  โดยทั่วไปมักใช้ $\alpha = 0.10 , 0.05 , 0.01$

- ขั้นตอนที่ 3 กำหนดสถิติทดสอบ

  λ^2

- ขั้นตอนที่ 4 คำนวณค่าสถิติทดสอบ

  คำนวณค่าสถิติทดสอบจากสูตรที่เลือกให้เหมาะสมกับเงื่อนไข

$$\lambda^2 = \frac{(O_i - Ei)^2}{E_i}$$

- ขั้นตอนที่ 5 การสร้างเขตปฏิเสธสมมติฐาน/เปรียบเทียบค่าคำนวณกับค่าจากตาราง

  หาค่าวิกฤติ และกำหนดขอบเขตของการปฏิเสธ

- ขั้นตอนที่ 6 สรุปผลการทดสอบ

  สรุปผลในการยอมรับ H_0 หรือ H_1

ผู้จัดการโรงแรมแห่งหนึ่งคาดว่าโรงแรม C จะมีอัตราส่วนผู้ใช้บริการเป็น 2 เท่าของโรงแรมอื่นๆ โดยข้อมูลที่ได้จากการเก็บรวบรวมจริงเป็นดังนี้

โรงแรม	A	B	C	D
จำนวนผู้ใช้บริการ	60	80	190	70

จงทดสอบว่าการคาดการณ์อัตราส่วนผู้ใช้บริการโรงแรมจากการเลือกตัวอย่างลูกค้าจำนวน 20 คน ของผู้จัดการคนนี้เป็นจริงหรือไม่ที่ระดับนัยสำคัญ 0.1

การทดสอบไคสแควร์กับข้อมูลที่มีสองลักษณะ

- ในกรณีที่ลักษณะของข้อมูลที่ต้องการศึกษาแบ่งออกได้เป็นสองลักษณะ คือ แบ่งตามสดมภ์(c)และแถว(r) เช่นในการทดสอบเกี่ยวกับความสามารถในการขายสินค้าของพนักงานขายชายและพนักงานหญิง ทางด้านสดมภ์อาจแบ่งความสามารถในการขายของพนักงานขายออกเป็น 3 ระดับ คือ สูง ปานกลาง และต่ำ ส่วนทางด้านแถวแบ่งเป็นเพศ คือ หญิง และชาย

เพศ	ความสามารถในการขายสินค้า		
	สูง	กลาง	ต่ำ
ชาย	3	2	3
หญิง	4	2	3

- ตัวสถิติที่ใช้ทดสอบ คือ $\lambda^2 = \sum \sum \frac{(O_{ij}-E_{ij})^2}{E_{ij}}$

เมื่อ O_{ij} แทนความถี่ที่ได้จากการทดลองของระดับที่ iของลักษณะแรกและระดับที่jของลักษณะที่สอง

E_i แทนความถี่ที่คาดว่าควรจะเป็นหรือตามทฤษฎีของระดับที่ iของลักษณะแรกและระดับที่jของลักษณะที่สอง

การทดสอบไคสแควร์กับข้อมูลที่มีสองลักษณะ

- โดยที่ $E_{ij} = n_{i.}n_{.j}/n_{..}$
- เมื่อ $n_{i.}$ แทนความถี่รวมของทุกระดับที่ iของลักษณะแรกและทุกระดับของลักษณะที่สอง
- เมื่อ $n_{\cdot j}$ แทนความถี่รวมของทุกระดับของลักษณะแรกและระดับที่ j ของลักษณะที่สอง
- เมื่อ $n_{..}$ แทนความถี่รวมทั้งหมด

ตัวอย่าง

- บริษัทผลิตยาสีฟันแห่งหนึ่งต้องการทราบว่าความชอบของยาสีฟันมีความสัมพันธ์กับเพศของผู้ใช้หรือไม่ จึงได้เลือกตัวอย่างชายและหญิงที่มีอายุตั้งแต่ 20 ปี ขึ้นไปมาเพศละ 500 คน จากการสอบถามถึงยาสีฟันที่ชอบใช้ได้ผลดังนี้

เพศ	สีของยาสีฟันที่ชอบ		
	ขาว	สีอื่นๆ	รวม
ชาย	328	172	500
หญิง	138	362	500
รวม	466	534	1000