บทที่ 3

วิธีดำเนินการวิจัย

ในการวิจัยครั้งนี้ต้องการศึกษาความสัมพันธ์ระหว่างสัมประสิทธิ์สหสัมพันธ์กับค่าทดสอบไคสแควร์ ในกรณีที่ประชากรมีการแจกแจงแบบปกติสองตัวแปร โดยจะประมาณช่วงความเชื่อมั่น (Confidence interval) พร้อมทั้งค่าคาดหวัง (Expected value) ของค่าสัมประสิทธิ์สหสัมพันธ์ ณ ค่าไคสแควร์ในแต่ละช่วง และศึกษาอิทธิพลของขนาดตัวอย่าง ขนาดตาราง และการจัดกลุ่มข้อมูล ที่มีต่อค่าไคสแควร์ ในกรณีที่ประชากรมีการแจกแจงแบบพหุนามสองตัวแปร จะศึกษาเฉพาะอิทธิพลของขนาดตัวอย่าง ขนาดตาราง ที่มีต่อค่าไคสแควร์ แต่จะไม่ประมาณช่วงความเชื่อมั่นพร้อมทั้งค่าคาดหวังของสัมประสิทธิ์สหสัมพันธ์ ทั้งนี้เพราะในการวิจัยครั้งนี้จะศึกษาเฉพาะสัมประสิทธิ์สหสัมพันธ์เชิงเส้นแบบเพียร์สัน (Pearson Product-Moment Correlation Coefficient) ซึ่งมีข้อตกลงเบื้องต้นว่าตัวแปรจะต้องมีการแจกแจงแบบต่อเนื่อง สำหรับข้อมูลที่ใช้ในการวิจัยจำเป็นที่จะต้องเตรียมข้อมูลประชากร (Population data) ที่มีสภาพการแจกแจงตามที่กำหนดเสียก่อน เพื่อให้ผลสรุปถูกต้องและมีความมั่นใจได้ว่าประชากรมีคุณสมบัติตามที่ต้องการศึกษา โดยจะใช้วิธีการที่เรียกว่า เทคนิคการจำลองแบบ (Simulation technique) มาผลิต (generate) ข้อมูลประชากร ซึ่งตัวแปรที่สร้างขึ้นคือค่า (X,Y) จากนั้นจะดำเนินการสุ่มตัวอย่างขนาดต่าง ๆ เพื่อคำนวณค่าไคสแควร์ในรูปแบบของขนาดตารางการณ์จรที่แตกต่างกัน และจะดำเนินการศึกษาในขั้นตอนต่อไปในแต่ละกรณีดังนี้

3.1 กรณีตัวแปรมีการแจกแจงแบบพหุนามสองตัวแปร (Bivariate multinomial)

การสร้างข้อมูลประชากรที่มีการแจกแจงแบบพหุนามสองตัวแปร จะใช้สับรูทีนแรนดู (Subroutine Randu) ผลิตตัวเลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอในช่วง (0 , 1) เลขสุ่มที่ได้จะถูกสร้างให้มีลักษณะสุ่ม (random) และเป็นอิสระต่อกัน โดยตัวเลขสุ่มแต่ละหมายเลขมีโอกาสเกิดขึ้นเท่าเทียมกันการใช้โปรแกรมย่อยจะใช้คำสั่ง CALL RANDU (IX, IY, YFL) พารามิเตอร์ในวงเล็บมีความหมายดังนี้

IX คือตัวเลขสุ่มตัวแรก จะต้องเป็นจำนวนเต็มบวกที่เป็นเลขคี่ และน้อยกว่า 2147483648 ซึ่ง IX นี้จะเป็นค่าเริ่มต้นที่จะให้สับรูทีนคำนวณ IY ออกมาให้ โดย IX จะเป็นค่าที่ส่งมาจากโปรแกรมหลัก (Main Program) IY คือตัวเลขสุ่มจำนวนเต็มบวกของสับรูทีนนี้ และจะใช้เป็นตัวคำนวณ IY ตัวต่อ ๆ ไป พิสัยของตัวเลขสุ่มจะอยู่ระหว่าง 0 กับ 2^{32} YFL คือเลขสุ่มที่คำนวณได้ และมีการแจกแจงแบบสม่ำเสมอในช่วง (0,1)

ในกรณีที่ประชากรมีการแจกแจงแบบพหุนามสองตัวแปร (bivariate multinomial) จะแบ่งการศึกษาออกเป็นสองกรณีคือ กรณีที่ตัวแปรทั้งสองเป็นอิสระต่อกัน และกรณีที่ตัวแปรทั้งสองไม่เป็นอิสระต่อกัน ดังนี้

3.1.1 <u>กรณีที่ตัวแปรทั้งสองเป็นอิสระต่อกัน</u>

ในกรณีที่ตัวแปรทั้งสองเป็นอิสระต่อกัน จะศึกษาขนาดตาราง 4x5 4x3 4x2 2x3 และ 2x2 ขนาดตัวอย่างเท่ากับ 50 100 และ 150 เพื่อศึกษาอิทธิพลของขนาดตัวอย่างและขนาดตารางที่มีต่อค่าไคสแควร์ ในแต่ละกรณีจะกระทำซ้ำเป็นจำนวน 300 รอบ โดยดำเนินการตามขั้นตอนต่อไปนี้

1. ผลิตตัวเลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอในช่วง (0,1) โดยใช้สับรูทีนแรนดู (Subroutine Randu) เรียกข้อมูลเลขสุ่มที่ได้ครั้งแรกว่า X_k $(k = 1,2,...,n)$ โดยที่ n คือขนาดตัวอย่าง จากนั้นจะผลิตตัวเลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอขึ้นมาอีกชุดหนึ่ง โดยใช้สับรูทีน แรนดู และเรียกข้อมูลเลขสุ่มที่ได้ว่า Y_k $(k = 1,2,---,n)$ ดังนั้นข้อมูลเลขสุ่มที่ผลิตขึ้นมาคือ(X_k , Y_k)

2. ถ้ากำหนดการแจกแจงโดยตารางแจกแจงความน่าจะเป็นร่วมดังนี้

ตารางที่ 3.1 แสดงตารางแจกแจงความน่าจะเป็นร่วมขนาดตาราง rxc

X \ Y	Y_1	Y_2	.....	Y_c		
X_1	f_{11}	f_{12}		f_{1c}	$f_{1.}$	
X_2	f_{21}	f_{22}		f_{2c}	$f_{2.}$	
$\vdots$						
X_r	f_{r1}	f_{r2}		f_{rc}	$f_{r.}$	
	$f_{.1}$	$f_{.2}$		$f_{.c}$	1.00	$f_{..}$

$$f_{i.} = \sum_{j=1}^{c} f_{ij}$$

$$f_{.j} = \sum_{i=1}^{r} f_{ij}$$

และ $$f_{i.} \times f_{.j} = f_{ij}$$

จากตารางแจกแจงความน่าจะเป็นร่วม

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

f_{ij} หมายถึงความน่าจะเป็นร่วมที่ตัวแปรพหุนามสองตัวแปร (X_k , Y_k) จะตกอยู่ในช่อง (i , j)

$f_{i.}$ หมายถึงความน่าจะเป็นที่ตัวแปรพหุนามสองตัวแปร (X_k , Y_k) จะมีค่า X_k ตกในแถวนอนที่ i Y_k ตกอยู่ในแถวตั้งที่ j ใด ๆ

$f_{.j}$ หมายถึงความน่าจะเป็นที่ตัวแปรพหุนามสองตัวแปร (X_k , Y_k) จะมีค่า Y_k ตกในแถวตั้งที่ j X_k ตกอยู่ในแถวนอนที่ i ใด ๆ

$$f_{..} = \sum_{i=1}^{r} f_{i.} = \sum_{j=1}^{c} f_{.j} = n$$

จากนั้น จะพิจารณาค่าของตัวแปร (X_k,Y_k) โดยจะพิจารณาตัวแปร X_k ดังนี้

ถ้า $0 \leq X_k < f_{1.}$ จะนับเป็นความถี่ในแถวนอนที่หนึ่ง

ถ้า $f_{1.} \leq X_k < f_{1.} + f_{2.}$ จะนับเป็นความถี่ในแถวนอนที่สอง

$\vdots$

ถ้า $f_{1.} + \cdots + f_{(r-1).} \leq X_k \leq 1.00$ จะนับเป็นความถี่ในแถวนอนที่ r

จะพิจารณาตัวแปร Y_k ในทำนองเดียวกันดังนี้

ถ้า $0 \leq Y_k < f_{.1}$ จะนับเป็นความถี่ในแถวตั้งที่หนึ่ง

ถ้า $f_{.1} \leq Y_k < f_{.1} + f_{.2}$ จะนับเป็นความถี่ในแถวตั้งที่สอง

$\vdots$

ถ้า $f_{.1} + f_{.2} + \cdots + f_{.(c-1)} \leq Y_k \leq 1.00$ จะนับเป็นความถี่ในแถวตั้งที่ c

ดังนั้น เมื่อพิจารณาข้อมูลพหุนามสองตัวแปร (X_k , Y_k) ถ้า X_k อยู่ในช่องตามแถวนอนที่ i และ Y_k อยู่ในแถวตั้งที่ j จะนับเป็นความถี่ในช่อง (i , j) นั้น และจะกระทำซ้ำตั้งแต่ $k = 1$ ถึง n ซึ่งจะได้ความถี่สะสมคือ O_{ij}

O_{ij} หมายถึง จำนวนความถี่ของข้อมูลพหุนามสองตัวแปรที่ผลิตขึ้นและตกในช่อง (i,j) ที่ได้จากการจำลองแบบ

3. เมื่อนับจำนวนความถี่บรรจุในตารางการณ์จรจนครบตามขนาดตัวอย่างที่กำหนด จากนั้นจะคำนวณจำนวนความถี่คาดหวัง (Expected frequency) ภายใต้สมมติฐานว่างดังนี้

$$E_{ij} = \frac{O_{i.} \times O_{.j}}{n}$$

เมื่อ E_{ij} หมายถึง ความถี่คาดหวังในช่อง (i , j)

$O_{i.}$ หมายถึง จำนวนความถี่รวมของทุกแถวตั้ง ในแถวนอน i

$O_{.j}$ หมายถึง จำนวนความถี่รวมของทุกแถวนอน ในแถวตั้ง j

n หมายถึง ขนาดตัวอย่าง

เมื่อหาจำนวนความถี่ที่คาดหวังจนครบทุกช่องในตารางการณ์จรแล้ว จะคำนวณหาค่าไคสแควร์ได้ดังนี้

$$W = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

จากสมมติฐานว่าง W จะมีการแจกแจงแบบไคสแควร์ มี d.f. = (r-1)(c-1) เช่น จะศึกษาขนาดตาราง 4x5 โดยมีตารางแจกแจงความน่าจะเป็นร่วมดังนี้

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 3.2 แสดงตารางแจกแจงความน่าจะเป็นร่วม ขนาดตาราง 4x5

		Y					
		Y_1	Y_2	Y_3	Y_4	Y_5	$f_{i.}$
X	X_1	.02	.04	.08	.04	.02	.20
	X_2	.03	.06	.12	.06	.03	.30
	X_3	.03	.06	.12	.06	.03	.30
	X_4	.02	.04	.08	.04	.02	.20
	$f_{.j}$	.10	.20	.40	.20	.10	1.00 $f_{..}$

ถ้าสุ่มข้อมูลเลขสุ่ม (X,Y) ขึ้นมา จะพิจารณาตัวแปร X ก่อน ดังนี้

ถ้า $0 \leq X < .2$ จะนับเป็นความถี่ในแถวนอนที่หนึ่ง

ถ้า $.2 \leq X < .5$ จะนับเป็นความถี่ในแถวนอนที่สอง

ถ้า $.5 \leq X < .8$ จะนับเป็นความถี่ในแถวนอนที่สาม

ถ้า $.8 \leq X \leq 1.00$ จะนับเป็นความถี่ในแถวนอนที่สี่

จากนั้นจะพิจารณาตัวแปร Y ดังนี้

ถ้า $0 \leq Y < .1$ จะนับเป็นความถี่ในแถวตั้งที่หนึ่ง

ถ้า $.1 \leq Y < .3$ จะนับเป็นความถี่ในแถวตั้งที่สอง

ถ้า $.3 \leq Y < .7$ จะนับเป็นความถี่ในแถวตั้งที่สาม

ถ้า $.7 \leq Y < .9$ จะนับเป็นความถี่ในแถวตั้งที่สี่

ถ้า $.9 \leq Y \leq 1.00$ จะนับเป็นความถี่ในแถวตั้งที่ห้า

เมื่อบันทึกความถี่ลงในตารางการณ์จรจนครบตามขนาดตัวอย่างที่กำหนดแล้วจะคำนวณความถี่คาดหวังภายใต้สมมติฐานว่าง ขั้นสุดท้ายจะคำนวณค่าไคสแควร์ จากนั้นจะสุ่มตัวอย่างชุดใหม่จนกระทั่งครบ 300 ครั้ง แล้วคำนวณหาค่าเฉลี่ย (mean) และส่วนเบี่ยงเบนมาตรฐาน (standard deviation) ของค่าไคสแควร์ พร้อมทั้งหาค่าสูงสุดและค่าต่ำสุดของค่าไคสแควร์ เพื่อนำไปคำนวณหาอันตรภาคชั้น สำหรับสร้างกราฟรูปฮิสโตแกรมแสดงการแจกแจงแบบไคสแควร์โดยประมาณ

ในแต่ละขนาดตัวอย่าง เมื่อสุ่มข้อมูลขึ้นมาชุดหนึ่ง จะคำนวณค่าไคสแควร์จนครบในทุกขนาดตารางที่ศึกษา จากนั้นจะกลับไปสุ่มข้อมูลชุดใหม่จนครบ 300 รอบ

3.1.2 กรณีที่ตัวแปรทั้งสองไม่เป็นอิสระต่อกัน

ตัวแปรพหุนามสองตัวแปรจะไม่เป็นอิสระต่อกันถ้า $f_{ij} \neq f_{i.} \times f_{.j}$ สำหรับค่าของ i , j ดังนั้น การผลิตข้อมูลพหุนามสองตัวแปร โดยทั่วไปการแจกแจงจะกำหนดโดยตารางแจกแจงความน่าจะเป็นร่วมดังนี้

<u>ตารางที่ 3.3</u> แสดงตารางแจกแจงความน่าจะเป็นร่วม กรณีตัวแปรทั้งสองไม่เป็นอิสระต่อกัน ขนาดตาราง rxc

X \ Y	Y_1	Y_2	....	Y_c	
X_1	f_{11}	f_{12}		f_{1c}	$f_{1.}$
X_2	f_{21}	f_{22}		f_{2c}	$f_{2.}$
⋮					
X_r	f_{r1}	f_{r2}		f_{rc}	$f_{r.}$
	$f_{.1}$	$f_{.2}$		$f_{.c}$	1.00 $f_{..}$

$$f_{i.} = \sum_{j=1}^{c} f_{ij}$$

$$f_{.j} = \sum_{i=1}^{r} f_{ij}$$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

การผลิตข้อมูลพหุนามสองตัวแปรตามลักษณะตารางแจกแจงความน่าจะเป็นร่วม จะดำเนินการดังนี้

1. ผลิตข้อมูลเลขสุ่มที่มีการแจกแจงสม่ำเสมอในช่วง (0,1) โดยใช้สับรูทีน RANDU เรียกข้อมูลเลขสุ่มที่ได้ว่า RN_1 , RN_2,---,RN_n โดยที่ n คือขนาดตัวอย่าง

2. จากตารางที่ 3.3 f_{ij} หมายถึงความน่าจะเป็นร่วมที่ตัวแปรพหุนามสองตัวแปร (X_k , Y_k) จะตกอยู่ในช่อง (i , j) เราจะพิจารณาค่าของ RN_k $(k = 1,2, \text{--------}, n)$ ดังนี้

ถ้า $0 \leq RN_k < f_{11}$ จะนับความถี่ลงในช่อง (1,1)

ถ้า $f_{11} \leq RN_k < f_{11} + f_{12}$ จะนับความถี่ลงในช่อง (1,2)

ถ้า $f_{11} + f_{12} \leq RN_k < f_{11} + f_{12} + f_{13}$ จะนับความถี่ลงในช่อง (1,3)

$\vdots$

ดังนั้นถ้าให้

$$I_{ij} = \text{ช่วงกึ่งปิดกึ่งเปิด } [a_{i,j-1} , a_{i,j})$$

โดยที่

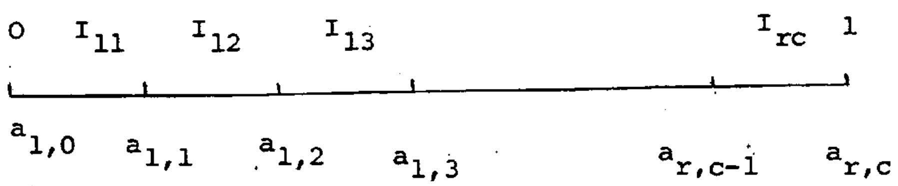
$$a_{1,0} = 0$$

$$a_{1,j} = \sum_{k=1}^{j} f_{1k} \qquad (1 \leq j \leq c)$$

$$a_{i,0} = \sum_{r=1}^{i} f_{r.} \qquad (2 \leq i \leq r)$$

$$a_{i,j} = \sum_{r=1}^{i} f_{r.} + \sum_{k=1}^{j} f_{ik} \qquad (2 \leq i \leq r , 1 \leq j \leq c)$$

$$a_{r,c} = 1$$



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

และให้ O_{ij} หมายถึง จำนวนความถี่ของข้อมูลพหุนามสองตัวแปรที่ผลิตขึ้นและตกในช่อง (i,j) ที่ได้จากการจำลองแบบ

เราจะพิจารณาค่าของ RN_1 , RN_2, ----, RN_n ถ้า $RN_k \in I_{ij}$ เราจะนับความถี่สะสมในช่อง (i,j) และจะกระทำซ้ำตั้งแต่ค่า k = 1 ถึง n ผลสุดท้ายจะได้ความถี่สะสมคือ O_{ij}

เมื่อนับจำนวนความถี่บรรจุในตารางการณ์จรจนครบตามขนาดตัวอย่างแล้ว จะหาความถี่ที่คาดหวัง (Expected frequency) ภายใต้สมมติฐานว่างดังนี้

$$E_{ij} = \frac{O_{i.} \times O_{.j}}{n}$$

จะคำนวณหาค่าไคสแควร์ จาก

$$W = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

W จะมีการแจกแจงแบบไคสแควร์ ที่มี d.f. = (r-1)(c-1) เช่น จะศึกษาตาราง ขนาด 2 x 2 ซึ่งมีตารางแจกแจงความน่าจะเป็นร่วม ดังนี้

		Y		
		Y_1	Y_2	$f_{i.}$
X	X_1	.1	.4	.5
	X_2	.4	.1	.5
	$f_{.j}$	.5	.5	1.00 $f_{..}$

จะพิจารณาค่าของ RN_1, ----, RN_n ถ้า $RN_k \in I_{ij}$ จะนับความถี่สะสมในช่อง (i,j) ดังนี้

$0 \leq RN_k < .1$ จะนับเป็นความถี่ในช่อง (1,1)

$.1 \leq RN_k < .5$ จะนับเป็นความถี่ในช่อง (1,2)

$.5 \leq RN_k < .9$ จะนับเป็นความถี่ในช่อง (2,1)

$.9 \leq RN_k \leq 1.00$ จะนับเป็นความถี่ในช่อง (2,2)

เมื่อบันทึกความถี่ลงในตารางการณ์จรจนครบตามขนาดตัวอย่างที่กำหนดแล้วจะคำนวณค่าไคสแควร์ จากนั้นจะสุ่มข้อมูลชุดใหม่จนครบ 300 รอบ และคำนวณค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานพร้อมทั้งสร้างกราฟรูปฮิสโตแกรมแสดงการแจกแจงแบบไคสแควร์โดยประมาณ เช่นเดียวกับกรณีตัวแปรทั้งสองเป็นอิสระต่อกัน โดยศึกษาขนาดตาราง 2x2 และขนาดตัวอย่างเท่ากับ 50 100 และ 150

3.2 <u>กรณีตัวแปรมีการแจกแจงแบบปกติสองตัวแปร</u>

ในกรณีตัวแปรมีการแจกแจงแบบปกติสองตัวแปร จะศึกษาอิทธิพลของขนาดตัวอย่าง ขนาดตารางและการแบ่งกลุ่มข้อมูลที่มีต่อค่าทดสอบไคสแควร์และประมาณช่วงความเชื่อมั่นพร้อมทั้งค่าคาดหวังของสัมประสิทธิ์สหสัมพันธ์ ณ ค่าไคสแควร์แต่ละช่วง โดยศึกษาขนาดตัวอย่างเท่ากับ 20 30 40 50 75 และ 100 ขนาดตาราง 2x2 2x3 2x4 2x5 3x3 3x4 3x5 4x4 4x5 และ 5x5 กำหนดค่าสัมประสิทธิ์สหสัมพันธ์ตั้งแต่ 0.00 (0.02) 0.40(0.01) 0.60(0.02) 0.98 (หมายความว่าตั้งแต่ 0.00 ถึง 0.40 มีความกว้างของช่วงเท่ากับ 0.02 จาก 0.40 ถึง 0.60 มีความกว้างของช่วงเท่ากับ 0.01 และจาก 0.60 ถึง 0.98 มีความกว้างของช่วงเท่ากับ 0.02) รวม 60 ระดับจะดำเนินการศึกษาตามขั้นตอนดังนี้

3.2.1 การสร้างข้อมูลประชากรที่มีการแจกแจงแบบปกติสองตัวแปร จะใช้สับรูทีนแรนดูผลิตตัวเลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอในช่วง (0 , 1) ก่อน โดยใช้คำสั่ง CALL RANDU (IX, IY, YFL) เช่นเดียวกับกรณีประชากรมีการแจกแจงแบบพหุนามสองตัวแปร จากนั้นจึงใช้สับรูทีนเกาส์ (Subroutine Gauss) ในการคำนวณหาตัวเลขสุ่มที่มีการแจกแจงแบบปกติ โดยใช้คำสั่ง CALL GAUSS (IX, S, AM, V) พารามิเตอร์ในวงเล็บมีความหมายดังนี้

IX คือตัวเลขสุ่มตัวแรก ซึ่งเป็นจำนวนเต็มบวกที่เป็นเลขคี่ที่น้อยกว่า 2147483648

S คือส่วนเบี่ยงเบนมาตรฐานของการแจกแจงแบบปกติ

AM คือค่าเฉลี่ยเลขคณิตของการแจกแจงแบบปกติ

V คือตัวเลขสุ่มที่มีการแจกแจงแบบปกติที่คำนวณได้

สำหรับ IX, S และ AM เป็นค่าที่ส่งมาจากโปรแกรมหลัก ส่วนผลลัพธ์คือ V จะเป็นตัวแปรสุ่มที่มีการแจกแจงแบบปกติ ที่มีค่าเฉลี่ย AM และส่วนเบี่ยงเบนมาตรฐานเท่ากับ S

ในการคำนวณหาเลขสุ่มที่มีการแจกแจงแบบปกติที่มีค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานตามที่กำหนด จะใช้ค่า X_i ที่ผลิตได้จากสับรูทีนแรนดู ซึ่งมีการแจกแจงแบบสม่ำเสมอ การประมาณจะใช้สูตร

$$Y = \frac{\sum_{i=1}^{k} X_i - \frac{K}{2}}{\frac{K}{12}}$$

เมื่อ X_i คือเลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอ และ $0 < X_i < 1$

K คือจำนวนค่า X_i ที่ใช้ในการประมาณค่า Y

Y คือเลขสุ่มที่มีการแจกแจงแบบปกติ

โดยปกติแล้ว ตัวเลขสุ่ม Y จะมีค่าเข้าใกล้เลขสุ่มที่มีการแจกแจงแบบปกติ เมื่อค่าของ K ใหญ่มากคือเข้าใกล้ค่าอนันต์ (infinity) สำหรับสับรูทีนเกาส์จะใช้ K = 12 เพื่อให้เครื่องลดเวลาในการคำนวณ (Excution) ดังนั้นจากสูตรในการคำนวณข้างต้นจะได้สูตรใหม่เป็น

$$Y = \sum_{i=1}^{12} X_i - 6.0$$

เพื่อให้เลขสุ่มที่สร้างขึ้นมาคือค่า Y มีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานตามที่ต้องการนั้น จะใช้สูตร

$$Y' = Y * S + AM$$

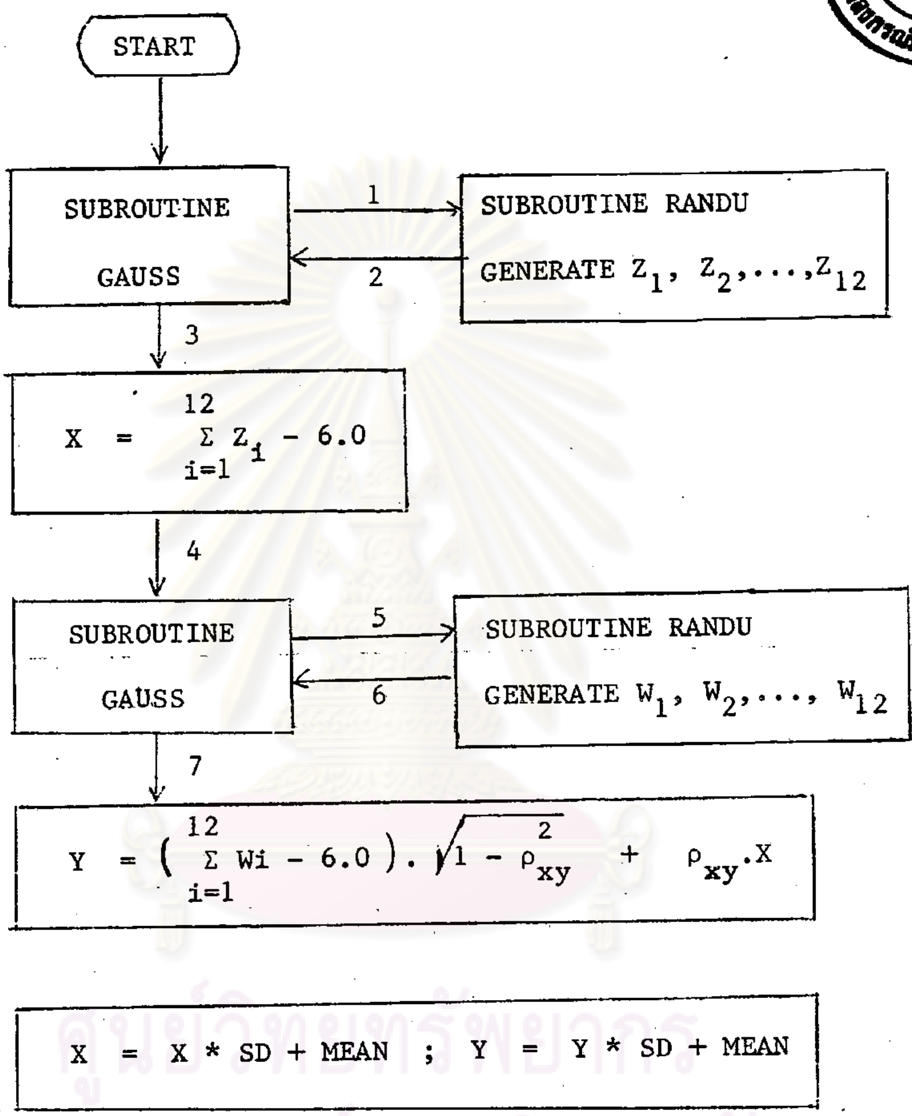
โดยที่ Y' คือตัวเลขสุ่มที่มีการแจกแจงแบบปกติที่มีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานตามที่กำหนด

S คือส่วนเบี่ยงเบนมาตรฐานตามที่ต้องการ

AM คือค่าเฉลี่ยเลขคณิตที่กำหนด

สำหรับในการศึกษาครั้งนี้ จะต้องใช้ข้อมูลประชากรที่มีการแจกแจงแบบปกติสองตัวแปร ดังนั้นต้องสร้างตัวเลขสุ่มขึ้นมาก่อนโดยอาศัยสับรูทีนแรนดู จากนั้นจะนำมาสร้างตัวแปรปกติมาตรฐานโดยใช้สับรูทีนเกาส์ส์ดังกล่าวมาแล้วข้างต้น และเรียกค่าที่ได้ครั้งแรกว่า X ต่อไปจะสร้างตัวเลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอขึ้นมาอีกชุดหนึ่งโดยใช้สับรูทีนแรนดู แล้วจึงนำตัวเลขสุ่มนี้มาสร้างตัวแปรปกติมาตรฐาน เรียกค่าที่ได้นี้ว่า Y โดยค่า Y นี้ จะมีลักษณะการแจกแจงขึ้นอยู่กับค่า X ที่ได้จากครั้งแรก ซึ่งสรุปเป็นผังงาน (Flowchart) ได้ดังรูปที่ 3.1

รูปที่ 3.1 แสดงผังงานการสร้างข้อมูลที่มีการแจกแจงแบบปกติสองตัวแปร



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

3.2.2 การคำนวณค่าไคสแควร์

เมื่อสร้างข้อมูลประชากรที่มีการแจกแจงแบบปกติสองตัวแปร โดยกำหนดค่าพารามิเตอร์ คือ $\underset{\sim}{\mu} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$; $\rho = 0.1, \ldots, 0.9$ โดยวิธีการจำลองแบบด้วยโปรแกรมคอมพิวเตอร์ ซึ่งตัวแปรที่สร้างขึ้นคือค่า (X,Y) แล้ว จะดำเนินการสุ่มตัวอย่างตามขนาดตัวอย่างที่กำหนดแล้วปรับให้เป็นตัวแปรที่ไม่ต่อเนื่องเสียก่อน โดยการจัดกลุ่มข้อมูลให้สอดคล้องกับขนาดของตารางการณ์จร ซึ่งจะต้องกำหนดจุดแบ่งเพื่อเป็นเกณฑ์ในการจำแนกตัวแปร X และตัวแปรY ที่สุ่มขึ้นมานี้จะตกอยู่ในแถวนอน (row) แถวใดแถวหนึ่ง และแถวตั้ง (column) แถวใดแถวหนึ่ง ก็จะนับเป็นจำนวนความถี่ที่สังเกตได้ (Observed frequency) ในแถวนอนและแถวตั้งนั้น สำหรับการหาจุดแบ่งจะอาศัยพื้นที่ใต้โค้งปกติตามสัดส่วนในการจำแนกกลุ่มดังตารางที่ 3.1

ตารางที่ 3.4 แสดงค่าสัดส่วนพร้อมทั้งจุดแบ่งในแต่ละระดับ

จำนวนระดับ	สัดส่วนแต่ละระดับ	จุดแบ่ง (z)
2	.50, .50	0.000
3	.33, .33, .33	-0.430, 0.430
4	.25, .25, .25, .25	-0.675, 0.000, 0.675
5	.20, .20, .20, .20, .20,	-0.845, -0.225, 0.225, 0.845

สมมติว่ากำหนดขนาดตารางการณ์จรชนิด 2×3 นั่นคือมีแถวนอน 2 แถว และแถวตั้ง 3 แถว ถ้ากำหนดให้ตัวแปร X เป็นตัวแปรด้านแถวนอน และตัวแปร Y เป็นตัวแปรด้านแถวตั้ง ดังนั้น สัดส่วนที่ความถี่ควรจะตกในแต่ละช่องของแถวนอนประมาณ 50 % และมีจุดแบ่งที่ 0.000 และสัดส่วนที่ความถี่ควรจะตกในแต่ละช่องของแถวตั้งประมาณ 33 % มีจุดแบ่งที่ -0.430 และ 0.430 ถ้าสุ่มข้อมูล (X, Y) ขึ้นมา จะพิจารณาตัวแปร X ถ้าน้อยกว่าหรือเท่ากับ 0.000 จะตกในแถวนอนที่หนึ่ง ถ้ามากกว่า 0.000 จะตกในแถวนอนที่สอง ในทำนองเดียวกัน พิจารณาตัวแปร Y ถ้าน้อยกว่าหรือเท่ากับ -0.430 จะตกในแถวตั้งที่หนึ่ง ถ้าอยู่ระหว่าง -0.430 และ 0.430 จะตกในแถวตั้งที่สอง แต่ถ้ามากกว่าหรือเท่ากับ 0.430 จะตกในแถวตั้งที่สาม ดังนั้นข้อมูล (X, Y) คู่นี้จะตกในช่อง (Cell) ใด จะนับเป็นจำนวนความถี่เท่ากับหนึ่ง เมื่อจำแนกข้อมูลจนครบแล้ว จะคำนวณค่าความถี่ที่คาดหวัง (Expected frequency)

ภายใต้สมมติฐานว่าง (Ho) ถ้าความถี่ที่คาดหวังของช่วงใดช่วงหนึ่งเป็นศูนย์ จะไม่ใช้ข้อมูลชุดนี้ แต่จะกลับไปสุ่มข้อมูลชุดใหม่ขึ้นมา ขั้นสุดท้ายจะคำนวณค่าไคสแควร์ จากนั้นก็ย้อนกลับไปสุ่มตัวอย่างชุดใหม่ จนกระทั่งครบ 300 ครั้ง แล้วคำนวณหาค่าเฉลี่ยเลขคณิต (mean) และส่วนเบี่ยงเบนมาตรฐาน (standard deviation) ของค่าไคสแควร์ พร้อมทั้งหาค่าสูงสุดและค่าต่ำสุดของค่าไคสแควร์ เพื่อนำไปหาอันตรภาคชั้น (class interval) สำหรับสร้างกราฟระหว่างความถี่กับค่าไคสแควร์ รายละเอียดของโปรแกรมสร้างกราฟนี้อยู่ในภาคผนวก ค.

ในแต่ละขนาดตัวอย่าง จะคำนวณค่าไคสแควร์จนครบทุก ๆ ขนาดตาราง และครบทุกระดับสัมประสิทธิ์สหสัมพันธ์ และจะเปลี่ยนขนาดตัวอย่างไปจนครบทุกรูปแบบของการศึกษา

3.2.3 <u>การประมาณช่วงแห่งความเชื่อมั่นและค่าคาดหวังของสัมประสิทธิ์สหสัมพันธ์</u>

เมื่อสร้างลักษณะการแจกแจงของประชากรและคำนวณค่าไคสแควร์ได้แล้ว ขั้นต่อไปจะคำนวณหาช่วงแห่งความเชื่อมั่น (confidence interval) พร้อมทั้งค่าคาดหวัง (expected value) ของค่าสัมประสิทธิ์สหสัมพันธ์ ณ ค่าไคสแควร์แต่ละช่วง เริ่มจากการสร้างตารางแจกแจงความถี่ (frequency distribution) ระหว่างค่าไคสแควร์กับค่าสัมประสิทธิ์สหสัมพันธ์ โดยแบ่งค่าไคสแควร์จากค่าต่ำสุดถึงค่าสูงสุดออกเป็นช่วง ๆ กำหนดความกว้างของช่วงในแต่ละระดับดังตารางที่ 3.5

<u>ตารางที่ 3.5</u> แสดงความกว้างของช่วงในแต่ละระดับของค่าไคสแควร์

ค่าไคสแควร์	ความกว้างของช่วง
0.00 - 0.99	0.1
1.00 - 2.99	0.2
3.00 - 9.99	0.5
10.00 - 19.99	1.00
20.00 - 49.99	2.00
50.00 - 99.99	5.00
100.00 - 199.99	10.00

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ดังนั้น จะมีจำนวนขั้นรวมทั้งหมด 79 ขั้น ซึ่งจะครอบคลุมค่าไคสแควร์ทุกค่าในทุกกรณี ซึ่งจะจัดไว้ในแถวนอนของตาราง สำหรับแถวตั้งจะเป็นค่าสัมประสิทธิ์สหสัมพันธ์ จะศึกษาตั้งแต่ระดับ 0.00 (0.02) 0.40 (0.01) 0.60 (0.02) 0.98 จะมีจำนวนทั้งหมด 60 ระดับ จะใช้สัญญลักษณ์แทนว่า ρ_i จากนั้นจะพิจารณาว่าค่าไคสแควร์ที่คำนวณได้ในแต่ละระดับของสัมประสิทธิ์สหสัมพันธ์จะตกอยู่ในช่วงใดของค่าไคสแควร์ จะนับเป็นความถี่ (frequency ใช้สัญญลักษณ์ f_i) ในช่วงนั้น เมื่อได้จำแนกความถี่ลงในตารางจนครบแล้ว จากนั้นจะคำนวณหาค่าคาดหวังของค่าสัมประสิทธิ์สหสัมพันธ์ โดยคำนวณได้จากสูตร

$$E(\rho|W) = \frac{\Sigma f_i \rho_i}{\Sigma f_i}$$

ในที่นี้ ρ_i คือค่าสัมประสิทธิ์สหสัมพันธ์

$$W = \sum_{\ell=1}^{i} \sum_{k=1}^{j} \frac{\left(O_{\ell k} - E_{\ell k}\right)^2}{E_{\ell k}} \quad , W \sim \chi^2$$

$E(\rho|W)$ คือค่าคาดหวังของค่าสัมประสิทธิ์สหสัมพันธ์ เมื่อกำหนดค่าไคสแควร์ในแต่ละขั้น

$\Sigma f_i \rho_i$ คือผลรวมของผลคูณระหว่างความถี่กับค่าสัมประสิทธิ์สหสัมพันธ์ในแต่ละขั้นของค่าไคสแควร์

Σf_i คือผลรวมของความถี่ทั้งหมดในแต่ละขั้นของค่าไคสแควร์

และคำนวณส่วนเบี่ยงเบนมาตรฐานได้จากสูตรดังนี้

$$Var(\rho|W) = E(\rho^2|W) - (E(\rho|W))^2$$

$$= \frac{\Sigma f_i \rho_i^2}{\Sigma f_i} - \frac{(\Sigma f_i \rho_i)^2}{\Sigma f_i}$$

$$SD(\rho|W) = \sqrt{Var(\rho|W)}$$

$Var(\rho|W)$ คือความแปรปรวนของค่าสัมประสิทธิ์สหสัมพันธ์ เมื่อกำหนดค่าไคสแควร์ในแต่ละขั้น

$SD(\rho|W)$ คือส่วนเบี่ยงเบนมาตรฐานของค่าสัมประสิทธิ์สหสัมพันธ์เมื่อกำหนดค่าไคสแควร์ในแต่ละขั้น

i 17397480

$\Sigma f_i \rho_i^2$ คือผลรวมของผลคูณระหว่างความถี่กับค่าสัมประสิทธิ์สหสัมพันธ์ยกกำลังสอง

เพราะฉะนั้นจะประมาณช่วงแห่งความเชื่อมั่น $(1-\alpha)$ 100 % ได้จากสูตร

$$E(\rho|W) \pm Z_{\frac{\alpha}{2}} \cdot SD(\rho|W)$$

โดย $Z_{\frac{\alpha}{2}}$ คือค่าที่ได้จากตารางพื้นที่ใต้โค้งปกติ เมื่อกำหนดระดับนัยสำคัญชนิด 2 หางเท่ากับ α สำหรับในการวิจัยครั้งนี้จะใช้ระดับนัยสำคัญ 2 ระดับ คือ $\alpha = 0.05$ จะได้ $Z_{\frac{\alpha}{2}} = 1.96$ และ $\alpha = 0.10$ จะได้ $Z_{\frac{\alpha}{2}} = 1.645$

ดังนั้น ขีดจำกัดความเชื่อมั่น (Confidence limit) คือ

$E(\rho|W) - Z_{\frac{\alpha}{2}} \cdot SD(\rho|W)$ เป็นขีดจำกัดล่าง (lower limit)

$E(\rho|W) + Z_{\frac{\alpha}{2}} \cdot SD.(\rho|W)$ เป็นขีดจำกัดบน (upper limit)

จากนั้น จะคำนวณหาความน่าจะเป็นที่ค่าสัมประสิทธิ์สหสัมพันธ์จะอยู่ระหว่างขีดจำกัดล่าง และขีดจำกัดบนของช่วงแห่งความเชื่อมั่นจากการประมาณค่า ดังนี้

$$P(\rho_1 \leq \rho \leq \rho_2|W) = \frac{m}{n}$$

เมื่อ ρ_1 คือขีดจำกัดล่างของช่วงแห่งความเชื่อมั่นจากการคำนวณ

ρ_2 คือขีดจำกัดบนของช่วงแห่งความเชื่อมั่นจากการคำนวณ

m คือผลรวมของความถี่ที่อยู่ระหว่าง ρ_1 และ ρ_2

n คือความถี่รวมทั้งหมดในแต่ละช่วงของค่าไคสแควร์

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

3.2.4 การวิเคราะห์ความถดถอยอย่างง่ายแบบเชิงเส้น (Simple linear regression analysis)

หลังจากประมาณช่วงแห่งความเชื่อมั่นพร้อมทั้งค่าคาดหวังของค่าสัมประสิทธิ์ในแต่ละช่วงของค่าไคสแควร์แล้ว ขั้นต่อไปจะพยากรณ์ค่าคาดหวังของสัมประสิทธิ์สหสัมพันธ์ ซึ่งกำหนดให้เป็นตัวแปรไม่อิสระ (dependent variable) โดยอาศัยค่าของตัวแปรอิสระ (independent variable) คือ ค่าจุดกลาง (mid point) ของค่าไคสแควร์ในแต่ละช่วง ซึ่งสมการที่จะใช้พยากรณ์หรือประมาณค่าคาดหวังของสัมประสิทธิ์สหสัมพันธ์ เป็นดังนี้

$$E(\rho|W = w_{ij}) = a + b.w_{ij}$$

โดย $E(\rho|W = w_{ij})$ คือค่าคาดหวังของสัมประสิทธิ์สหสัมพันธ์ ณ จุดกลางของค่าไคสแควร์ (w_{ij})

a คือค่าคงที่ (constant) หรือค่าของ $E(\rho|W)$ เมื่อ $w_{ij} = 0$

b คือสัมประสิทธิ์ความถดถอย (regression coefficient) หรือค่าความชัน (slope)

การหาค่าประมาณ a, b สามารถทำได้หลายวิธี แต่ในการศึกษาครั้งนี้จะใช้วิธีกำลังสองน้อยที่สุด (Method of least square) โดยในแต่ละขนาดตัวอย่างและแต่ละขนาดตารางจะได้สมการความถดถอยหนึ่งสมการ จากนั้นจะหาค่าประมาณค่าคาดหวังของสัมประสิทธิ์สหสัมพันธ์ โดยแทนค่าจุดกลางในแต่ละสมการ แล้วจะนำค่าประมาณที่ได้นี้ของทุกขนาดตัวอย่าง ในแต่ละขนาดตารางมาสร้างสมการความถดถอยอีกครั้งหนึ่ง กำหนดให้ค่าประมาณค่าคาดหวังเป็นตัวแปรไม่อิสระ และขนาดตัวอย่างเป็นตัวแปรอิสระ เพื่อสามารถพยากรณ์ค่าคาดหวังของสัมประสิทธิ์สหสัมพันธ์ เมื่อกำหนดขนาดตัวอย่าง ดังสมการ

$$E(\hat{\rho}|W = w_{ij}) = a + b.n.$$

$E(\hat{\rho}|W = w_{ij})$ คือ ค่าประมาณค่าคาดหวังของสัมประสิทธิ์สหสัมพันธ์ในแต่ละค่าจุดกลางของค่าไคสแควร์

a คือค่าของ $E(\hat{\rho}|W)$ เมื่อกำหนด $n = 0$

b คือสัมประสิทธิ์ความถดถอย หรือค่าความชัน (slop)

n คือขนาดตัวอย่าง

การหาค่าประมาณ a, b จะใช้วิธีกำลังสองน้อยที่สุดเช่นเดียวกัน ซึ่งในแต่ละขนาดตารางจะได้สมการความถดถอย ณ ค่าจุดกลางของค่าไคสแควร์ในแต่ละช่วง

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย