

วิธีดำเนินการวิจัย

3.1 ลักษณะข้อมูล

การวิจัยนี้ใช้ข้อมูลที่มีลักษณะเป็นแบบต่อเนื่อง และมีการแจกแจงแบบปกติสองตัวแปร (X,Y) ที่มีค่าเฉลี่ยเป็นศูนย์ และความแปรปรวนเป็น 1

ตัวอย่างลักษณะของข้อมูลที่ต้องการวิเคราะห์หาความสัมพันธ์โดยใช้ตัวสถิติที่เสนอในการวิจัยนี้ ได้แก่ การหาสัมประสิทธิ์สหสัมพันธ์ระหว่างข้อมูลที่เป็นรายได้กับข้อมูลที่เป็นค่านายหน้า (Commission) หรือ ข้อมูล รายได้กับข้อมูลที่เป็นค่าใช้จ่าย เป็นต้น

ข้อมูลที่เป็นค่าของตัวแปร Y จะมีลักษณะเป็นข้อมูลที่ปกปิด (Sensitive data) กล่าวคือ ค่าที่แท้จริงของข้อมูลแต่ละตัว จะไม่ถูกเปิดเผย แต่จะทราบค่าในรูปของค่าเฉลี่ยของข้อมูลในแต่ละกลุ่ม

	<u>ตัวแปร X</u>	<u>รูปแบบของข้อมูล</u>	<u>ตัวแปร Y</u>	
เรียงลำดับค่า จากน้อยไป มาก	X_{11}		Y	ค่าของตัวแปร Y ที่สัมพันธ์ กับค่าของตัวแปร X ที่
	X_{21}	ค่าของตัวแปร X	Y	กับค่าของตัวแปร X ที่
		จำนวน m ค่า		เรียงลำดับแล้วจำนวน m
	.	<u>กลุ่มที่ 1</u>	.	ค่า <u>กลุ่มที่ 1</u>
	.		.	
	X_{m1}		Y	
	X_{12}		Y	
	X_{22}	<u>กลุ่มที่ 2</u>	Y	<u>กลุ่มที่ 2</u>
	.		.	
	.		.	
	X_{m2}		Y	
	.		.	
	.		.	

หอสมุดกลาง สถาบันวิจัยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

X_{1n}		Y	
X_{2n}	<u>กลุ่มที่ n</u>	Y	<u>กลุ่มที่ n</u>
.		.	
.		.	
.		.	
X_{mn}		Y	

3.2 วิธีการแบ่งกลุ่มข้อมูล

ข้อมูลจะถูกจัดแบ่งเป็นกลุ่ม ๆ โดยมีจำนวนกลุ่มละเท่า ๆ กัน และเพื่อความสะดวกในการเปรียบเทียบ จะจำกัดจำนวนข้อมูลทั้งหมด (N) เป็น 4 ค่า ได้แก่ 100 200 500 และ 1,000 ทำการแบ่งจำนวนข้อมูลให้แต่ละกลุ่มเป็นร้อยละ 5 10 20 และ 25 ของจำนวนข้อมูลทั้งหมด ตามลำดับ วิธีแบ่งกลุ่มข้อมูลข้างต้นสามารถสรุปเป็นรูปแบบได้ดังนี้

กรณีศึกษา	N	ร้อยละ	m	n
1	100	5	5	20
		10	10	10
		20	20	5
		25	25	4
2	200	5	10	20
		10	20	10
		20	40	5
		25	50	4
3	500	5	25	20
		10	50	10
		20	100	5
		25	125	4
4	1,000	5	50	20
		10	100	10
		20	200	5
		25	250	4

3.3 วิธีการวิจัย

การวิจัยนี้ ใช้วิธีทดลอง โดยการจำลองข้อมูลให้มีสถานการณ์ตามที่ต้องการศึกษา ด้วยวิธีมอนติคาร์โลซิมูเลชัน* (Monte Carlo Simulation Technique) โดยเครื่องคอมพิวเตอร์ IBM 370/3031 และใช้โปรแกรมคอมพิวเตอร์** ที่มีส่วนประกอบของ

* ทุกรายละเอียดในภาคผนวก ก

** ดูโปรแกรม 1 ในภาคผนวก ข

Subprogram Gauss และ Subprogram Randu ตัวแปรที่สร้างขึ้น คือ ค่า (X, Y) จะเป็นตัวเลขสุ่ม* ที่มีการแจกแจงแบบปกติสองตัวแปร โดยกำหนดค่าพารามิเตอร์ คือ

$$\mu = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \rho \sigma_x \sigma_y \\ \rho \sigma_y \sigma_x & \sigma_y^2 \end{bmatrix} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

$\rho = 0.10 \ 0.25 \ 0.50 \ 0.75 \ 0.90 \ 1.00$ ตามลำดับ

โดยในแต่ละระดับของค่า ρ จะสร้างตัวแปรสุ่ม (X, Y) ที่มีขนาดเป็น 100 200 500 และ 1,000 ตามลำดับ และทำการแบ่งกลุ่มข้อมูล โดยให้มีข้อมูลในแต่ละกลุ่มเป็นร้อยละ 5 10 20 และ 25 ของจำนวนข้อมูลทั้งหมดตามลำดับ ดังนั้นในการวิจัยนี้จะได้สถานการณ์ที่ต้องศึกษารวมทั้งหมด 96 สถานการณ์ด้วยกัน ซึ่งสามารถสรุปเป็นขั้นตอนของการทดลองได้ดังนี้

3.3.1 สร้างแบบจำลองข้อมูลให้มีสถานการณ์ตามที่ต้องการศึกษาด้วยวิธีมอนติคาร์โล ซิมูเลชัน โดยเครื่องคอมพิวเตอร์ IBM 370/3031

3.3.2 จากข้อมูลที่สร้างขึ้นทำการจัดเรียงลำดับค่าของตัวแปร X จากน้อยไปมาก และทำการจัดแบ่งจำนวนกลุ่ม และจำนวนข้อมูลในแต่ละกลุ่มตามที่กำหนด

* ดูรายละเอียดในภาคผนวก ก

- 3.3.3 จัดกลุ่มข้อมูลของตัวแปร Y ที่สัมพันธ์กับค่าของตัวแปร X ที่เรียงลำดับแล้ว
- 3.3.4 คำนวณหาค่าตัวสถิติที่ใช้วัดความสัมพันธ์ระหว่างตัวแปรทั้ง 5 วิธี
- 3.3.5 กระทำซ้ำตั้งแต่ 1 ถึง 4 จำนวน 50 ครั้ง
- 3.3.6 คำนวณหาค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน ของตัวสถิติทั้ง 5 วิธี
- 3.3.7 คำนวณหาค่าเอนเอียง (Bias)
- 3.3.8 คำนวณหาค่าคลาดเคลื่อนกำลังสองเฉลี่ย (mean square error) ของตัวสถิติทั้ง 5 วิธี คือ $\hat{\rho}_{\text{ungrouped}}$, $\hat{\rho}_{\text{grouped}}$, $\hat{\rho}_{\text{LS}}$, $\hat{\rho}_{\text{BW}}$ และ $\hat{\rho}_{\text{Cramer}}$ ตามลำดับ
- 3.3.9 คำนวณหาค่าอัตราส่วนของความคลาดเคลื่อนกำลังสองเฉลี่ย ระหว่างตัวสถิติที่ได้จากการจัดกลุ่มข้อมูล กับตัวสถิติที่ไม่ได้จัดกลุ่มข้อมูล ($\hat{\rho}_{\text{ungrouped}} = r$) เพื่อใช้เปรียบเทียบประสิทธิภาพของตัวสถิติที่ได้จากวิธีต่าง ๆ

3.4 วิธีวิเคราะห์ข้อมูล

การวิจัยนี้ ศึกษา เรื่อง การเปรียบเทียบตัวสถิติที่ใช้วัดความสัมพันธ์โดยการจัดกลุ่มข้อมูล ซึ่งเล่นอไว้ 4 วิธี เปรียบเทียบกับวิธีของ Pearson ($\hat{\rho}_{\text{ungrouped}} = r$) การเปรียบเทียบประสิทธิภาพของตัวประมาณที่ได้จากวิธีต่าง ๆ จะพิจารณาจากความคลาดเคลื่อนกำลังสองเฉลี่ย ที่มีค่าต่ำสุด และอัตราส่วนความคลาดเคลื่อนกำลังสองเฉลี่ย ระหว่างตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ที่ได้จากการจัดกลุ่มข้อมูล กับตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ที่ไม่ได้จัดกลุ่มข้อมูล

แนวทางการวิเคราะห์ข้อมูล สามารถสรุปเป็นข้อ ๆ ได้ดังนี้

- 3.4.1 เปรียบเทียบกรณีที่จัดแบ่งจำนวนข้อมูลในแต่ละกลุ่มแบบต่าง ๆ ที่ระดับประชากรขนาดเดียวกัน จำนวนกลุ่ม จำนวนเท่าใด ที่ทำให้การประมาณค่าสัมประสิทธิ์สหสัมพันธ์มีประสิทธิภาพที่สุด

3.4.2 เมื่อจำนวนประชากรเพิ่มขึ้น ตัวประมาณวิธีใดให้ผลดี และมีประสิทธิภาพ ณ แต่ละระดับของการจัดแบ่งจำนวนข้อมูลในแต่ละกลุ่ม

3.4.3 เปรียบเทียบกรณีที่ใช้จำนวนกลุ่ม และจำนวนประชากรเท่ากัน ตัวประมาณวิธีใด ที่ให้ประสิทธิภาพที่สูงสุด ณ แต่ละระดับของค่า ρ ประชากร

3.4.4 การประมาณค่าสัมประสิทธิ์สหสัมพันธ์ โดยการจัดกลุ่มข้อมูลจะใช้ได้ดี และมีประสิทธิภาพ เมื่อค่า ρ ประชากร ควรจะอยู่ช่วงใด ณ แต่ละระดับของจำนวนกลุ่ม

3.5 โปรแกรมที่ใช้ในการวิจัย

โปรแกรมที่ใช้ในการวิจัยทั้งหมด เขียนด้วยภาษาฟอร์แทรน IV โดยใช้กับเครื่อง IBM 370/3031 ผู้วิจัยได้เขียนโปรแกรม ซึ่งมีแผนผังแสดงลำดับขั้นตอนในการคำนวณค่าสถิติต่าง ๆ ที่ต้องการเพื่อนำไปวิเคราะห์และสรุปผลการวิจัยต่อไป ซึ่งมีรายละเอียดแสดงไว้ในหน้าถัดไป

สำหรับโปรแกรมน้อยที่เรียกใช้ ได้แก่

3.5.1 โปรแกรมสร้างตัวเลขสุ่ม

3.5.2 โปรแกรมสร้างลักษณะการแจกแจงของประชากร ให้มีการแจกแจงแบบปกติ
สองตัวแปร

3.5.3 โปรแกรมตัดเรียงลำดับของค่าสังเกตจากน้อยไปมาก

สำหรับขั้นตอนในการหาค่าสถิติต่าง ๆ เพื่อใช้ในการวิจัย สรุปลงเป็นผังงาน (Flowchart) ดังนี้

รูปที่ 2 ผังงานของโปรแกรมที่ใช้ในการวิจัย





