บทที่ 3
วิธีดำเนินการวิจัย

1. รูปแบบการวิจัย

การวิจัยครั้งนี้เป็นการวิจัยเชิงจำลอง (Simulation Research) ทำการจำลองข้อมูลแต่ละสถานการณ์ด้วยเทคนิคมอนติคาร์โล บนเครื่องคอมพิวเตอร์ส่วนบุคคล โดยพัฒนาโปรแกรมในการจำลองแบบด้วยภาษาซีชาร์บ (C#) เพื่อหาผลสรุปในการศึกษาผลของข้อมูลที่มีค่าห่างผิดปกติจากกลุ่มต่อความแกร่งของสัมประสิทธิ์สหสัมพันธ์

2. ขั้นตอนการดำเนินการวิจัย

2.1 การกำหนดสถานการณ์ที่ศึกษา

2.1.1 ตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ที่ใช้ในการทดสอบความแกร่งของสัมประสิทธิ์สหสัมพันธ์ เมื่อตัวอย่างบางหน่วยมีค่าผิดปกติจากกลุ่มจะประมาณค่าสัมประสิทธิ์สหสัมพันธ์จำนวน 4 แบบ คือ

2.1.1.1 สหสัมพันธ์แบบเพียร์สัน

2.1.1.2 สหสัมพันธ์แบบสเปียร์แมน

2.1.1.3 สหสัมพันธ์แบบเคนดอลล์

2.1.1.4 สหสัมพันธ์แบบถ่วงน้ำหนัก

2.1.2 ตัวอย่างที่ศึกษาทำการจำลองขึ้นให้มีการแจกแจงแบบปกติสองตัวแปร (Bivariate Normal Distribution) โดยมีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1 (N (0,1))

2.1.3 ขนาดของตัวอย่างที่ทำการศึกษามี 3 ขนาด คือ 20, 50 และ100

2.1.4 ระดับความสัมพันธ์ (r) กำหนด 5 ระดับ คือ ไม่มีความสัมพันธ์ มีความสัมพันธ์น้อย มีความสัมพันธ์ปานกลาง มีความสัมพันธ์มากและมีความสัมพันธ์อย่างสมบูรณ์ โดยมีทิศทางความสัมพันธ์ในทางบวก ดังนั้น ค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0, 0.20, 0.50, 0.80, และ 1.0

2.1.5 กำหนดระดับนัยสำคัญในการทดสอบ (Level of Significance) คือ 0.05

2.1.6 กำหนดจำนวนของค่าผิดปกติจากกลุ่ม ซึ่งเป็นค่าผิดปกติจากกลุ่มด้านมีค่ามากในระดับปานกลาง ให้มีจำนวนที่แตกต่างกันตั้งแต่ 0%, 5%, 10%, 20% และ 30% ของขนาดตัวอย่าง ทั้งตัวในแปร X (Outliers in X-direction) และ Y (Outliers in Y-direction)

2.1.7 ทำการทดลองซ้ำ 1,000 รอบในแต่ละสถานการณ์

2.2 การสร้างข้อมูล

2.2.1 การสร้างเลขสุ่ม (Random Number) จำลองค่าของตัวแปรสุ่ม X และ Y ให้มีการแจกแจงความน่าจะเป็นแบบเอกรูป (Uniform Distribution; U (0,1)) ด้วยวิธีตัวแบบจำลองสมภาคแบบผสม (Mixed Congruential Simulator) (มานพ วราภักดิ์,2547) ดังรายละเอียดการจำลองในภาคผนวก ก

2.2.2 จำลองตัวแปรสุ่ม X จากข้อ 2.2.1 ให้เป็นข้อมูลที่มีการแจกแจงแบบปกติด้วยค่าพารามิเตอร์ μ = 0 และ σ^2 = 1 ซึ่งเขียนแทนด้วย X ~N(0,1) ด้วยวิธีรับ-ปฏิเสธ (มานพ วราภักดิ์,2547) ดังรายละเอียดในภาคผนวก ข และในกรณีจำลองตัวแปรสุ่ม Y ให้มีการแจกแจงแบบปกติ ก็ใช้วิธีการเดียวกัน

2.2.3 สุ่มตัวแปร YและX ตามที่กำหนดจำนวนตัวอย่างในการศึกษา

2.2.4 สร้างตัวแปร YและX ให้มีค่าสัมประสิทธิ์สหสัมพันธ์กันตามกำหนด โดยใช้สูตรต่อไปนี้ (ทองดี แย้มสรวล, 2530)

$$Y = r*(X)+\sqrt{1-r^2}*X$$

เมื่อ r หมายถึง ค่าสัมประสิทธิ์สหสัมพันธ์ที่กำหนด

ในขั้นตอนนี้จะได้ตัวแปร X และ Y มีการแจกแจงแบบปกติ N(0,1) ที่มีขนาดความสัมพันธ์ และขนาดตัวอย่าง ตามที่กำหนด

2.2.5 เมื่อต้องการสร้างค่าผิดปกติจากกลุ่มในตัวแปร X จะทำการสุ่มค่าสังเกตจากตัวแปร X (x_i) ตามจำนวนเปอร์เซ็นต์ของค่าผิดปกติจากกลุ่ม แล้วปรับค่า x_i ที่สุ่มได้ให้เป็นค่าผิดปกติจากกลุ่มด้านบวก ซึ่งค่าผิดปกติจากกลุ่มที่ได้จะต้องมีค่ามากกว่า Q_3 + 1.5(IQR) (ศิริชัย พงษ์วิชัย,2544) โดยใช้สูตร

$$D_i = x_i + (Q_3 + L(IQR))$$

เมื่อ x_i = ค่าสังเกตจากตัวแปร X ที่สุ่มได้

D_i = ค่าสังเกตจากตัวแปร X ที่สุ่มได้ และปรับให้เป็นค่าผิดปกติจากกลุ่มแล้ว

Q_1 = ค่าควอไทล์ที่ 1 ของตัวแปร X

Q_3 = ค่าควอไทล์ที่ 3 ของตัวแปร X

IQR = $Q_3 - Q_1$

L = 2 เมื่อต้องการค่าผิดปกติจากกลุ่มระดับปานกลาง

ในการสร้างค่าผิดปกติจากกลุ่มในตัวแปร Y จะใช้วิธีการเดียวกันกับการสร้างค่าผิดปกติจากกลุ่มในตัวแปร X

2.2.6 ตรวจสอบค่าผิดปกติจากกลุ่มโดยใช้กราฟแบบ Box and Whisker และถ้าพบค่าสังเกตของตัวแปรมีค่ามากกว่า Q_3 + 1.5(IQR) จะถือว่าข้อมูลของตัวแปรที่สร้างขึ้นมีค่าผิดปกติจากกลุ่มด้านบวก

2.3. การประมาณค่าสัมประสิทธิ์สหสัมพันธ์และทดสอบสมมติฐาน

ในการศึกษาครั้งนี้ ได้ทำการประมาณค่าสัมประสิทธิ์สหสัมพันธ์และการทดสอบสมมติฐาน จากโปรแกรมจำลองแบบที่พัฒนาขึ้น ตามรายละเอียดดังนี้

2.3.1 ประมาณค่าสัมประสิทธิ์สหสัมพันธ์จากตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ทั้ง 4 แบบ ตามสถานการณ์ที่กำหนด

2.3.2 ทดสอบสมมติฐาน ตามสถานการณ์ที่กำหนด

2.3.3 นำผลการประมาณค่า และผลสรุปการทดสอบสมมติฐาน ไปเปรียบเทียบความแกร่งตามเกณฑ์ที่กำหนด

2.4. เปรียบเทียบความแกร่งของสัมประสิทธิ์สหสัมพันธ์

การสรุปผลในแต่ละสถานการณ์จะทำการทดลองซ้ำ 1,000 รอบ แล้วนำผลที่ได้มาเปรียบเทียบความแกร่งหรือประสิทธิภาพวิธีการประมาณค่าพารามิเตอร์ในแต่ละวิธี ตามเกณฑ์ความแกร่งที่ต้องการทดสอบภายใต้สถานการณ์เดียวกัน ประกอบด้วย

2.4.1 ความแกร่งในการทดสอบทางสถิติ

2.4.1.1 ความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 (Type I Error) จะพิจารณาจากความถี่สัมพัทธ์ของการปฏิเสธสมมติฐานหลัก เมื่อสมมติฐานหลักเป็นจริง(ρ=0) แล้วนำค่าความถี่สัมพัทธ์ไปเปรียบเทียบกับช่วงของระดับนัยสำคัญที่กำหนด ถ้าตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์แบบใดมีความถี่สัมพัทธ์ของการปฏิเสธสมมติฐานหลัก เมื่อสมมติฐานหลักเป็นจริง อยู่ภายในช่วง 0.036 ถึง 0.064 แสดงว่าตัวประมาณค่าสัมประสิทธิ์นั้นมีความแกร่งหรือควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ โดยความถี่สัมพัทธ์ของการปฏิเสธสมมติฐานหลัก เมื่อสมมติฐานหลักเป็นจริง คำนวณจาก

$$\text{Type I Error} = \frac{\text{จำนวนครั้งที่ปฏิเสธสมมติฐานหลักเมื่อสมมติฐานหลักเป็นจริง}}{1{,}000}$$

2.4.1.2 อำนาจการทดสอบ (Power of Test) ในแต่ละสถานการณ์ที่ทดสอบ การพิจารณาอำนาจการทดสอบของตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ทั้ง 4 แบบ โดยจะพิจารณาเฉพาะตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ที่สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ตามเกณฑ์ที่กำหนดเท่านั้น อำนาจการทดสอบจะพิจารณาจากค่าความถี่สัมพัทธ์ของผลการทดสอบที่สรุปว่าปฏิเสธสมมติฐานหลักเมื่อสมมติฐานหลักไม่เป็นจริง ($\rho \neq 0$) ถ้าตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ใดมีค่าความถี่สัมพัทธ์ของผลการทดสอบสูงกว่าตัวประมาณค่าแบบอื่นๆ แสดงว่าตัวประมาณค่านั้นมีความแกร่งกว่าตัวประมาณค่าอื่นๆที่เหลือ และถ้าตัวประมาณค่า 2 แบบมีค่าความถี่สัมพัทธ์เท่ากัน แสดงว่าตัวประมาณค่าทั้งสองมีความแกร่งไม่แตกต่างกัน เมื่อจำนวนครั้งที่ทำการทดลองเพิ่มขึ้นมากๆ ค่าความถี่สัมพัทธ์ของการปฏิเสธสมมติฐานหลัก เมื่อสมมติฐานหลักไม่เป็นจริง จะมีค่าใกล้เคียงหรือเทียบเท่ากับค่าอำนาจของการทดสอบ โดยอำนาจการทดสอบคำนวณจาก

$$\text{Power of Test} = \frac{\text{จำนวนครั้งที่ปฏิเสธสมมติฐานหลักเมื่อสมมติฐานหลักไม่เป็นจริง}}{1{,}000}$$

2.4.1.3 ค่าประสิทธิภาพสัมพัทธ์ (Relative Efficiency; R.E.) จะพิจารณาจากการเปลี่ยนแปลงของค่าความถี่สัมพัทธ์ของการปฏิเสธสมมติฐานหลัก เมื่อสมมติฐานหลักไม่เป็นจริง เมื่อมีค่าผิดปกติจากกลุ่มในจำนวน 5%, 10%, 20% และ 30% ของขนาดตัวอย่าง เปรียบเทียบกับเมื่อไม่มีค่าผิดปกติจากกลุ่ม เพื่อพิจารณาว่าการมีค่าผิดปกติจากกลุ่มในตัวอย่างในปริมาณที่แตกต่างกันจะส่งผลกระทบต่อตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ในแต่ละแบบอย่างไร ค่าประสิทธิภาพสัมพัทธ์ มีสูตรในการคำนวณดังนี้

$$\text{R.E.} = \frac{f_2}{f_1}$$

เมื่อ f1 คือ ความถี่สัมพัทธ์ของผลการทดสอบที่สรุปว่าปฏิเสธสมมติฐานหลักเมื่อใช้ตัวประมาณค่านั้น และไม่มีค่าผิดปกติจากกลุ่ม

f2 คือ ความถี่สัมพัทธ์ของผลการทดสอบที่สรุปว่าปฏิเสธสมมติฐานหลักเมื่อใช้ตัวประมาณค่านั้น และมีค่าผิดปกติจากกลุ่มเท่ากับ 5%,10%, 20% และ 30% ของขนาดตัวอย่าง

การแปลความหมายของค่า R.E. เป็นดังนี้

R.E. > 1 แสดงว่า การมีค่าผิดปกติจากกลุ่มในตัวอย่างส่งผลกระทบต่อตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์นั้น ทำให้มีโอกาสปฏิเสธสมมติฐานหลักมากขึ้น แสดงตัวประมาณค่านั้นไม่มีความแกร่งในการทดสอบ

R.E. < 1 แสดงว่า การมีค่าผิดปกติจากกลุ่มในตัวอย่างส่งผลกระทบต่อตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์นั้น ทำให้มีโอกาสปฏิเสธสมมติฐานหลักน้อยลง แสดงตัวประมาณค่านั้นไม่มีความแกร่งในการทดสอบ

R.E. = 1 แสดงว่า การมีค่าผิดปกติจากกลุ่มในตัวอย่างไม่ส่งผลกระทบต่อตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์นั้น แสดงตัวประมาณค่านั้นมีความแกร่งในการทดสอบ

การพิจารณาผลกระทบของค่าผิดปกติจากกลุ่มต่อตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ เมื่อค่า R.E. $\neq$ 1 ถ้าตัวประมาณค่าแบบใดที่มีค่า R.E. ผันแปรจาก 1 น้อยที่สุดในแต่ละสถานการณ์ จะเป็นตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ที่ได้รับผลกระทบจากค่าผิดปกติจากกลุ่มน้อยที่สุด ซึ่งเป็นตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์มีความแกร่งที่สุดเมื่อเทียบกับตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์แบบอื่น

2.4.2 ความแกร่งในการประมาณค่า

เกณฑ์การตัดสินว่าตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์แบบใดมีความแกร่งในการประมาณค่า จะพิจารณาจากค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) ของตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ทั้ง 4 แบบ วิธีการประมาณค่าสัมประสิทธิ์สหสัมพันธ์ที่มีความแกร่งในการประมาณค่า ค่า MSE จะเท่ากับ 0 แต่ถ้าค่า MSE ไม่เท่ากับ 0 ตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์แบบใดให้ค่า MSE ต่ำที่สุด จะถือว่าเป็นตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ที่มีความแกร่งในการประมาณค่าเมื่อเทียบกับตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์แบบอื่น ๆ โดยค่า MSE คำนวณจากสูตร

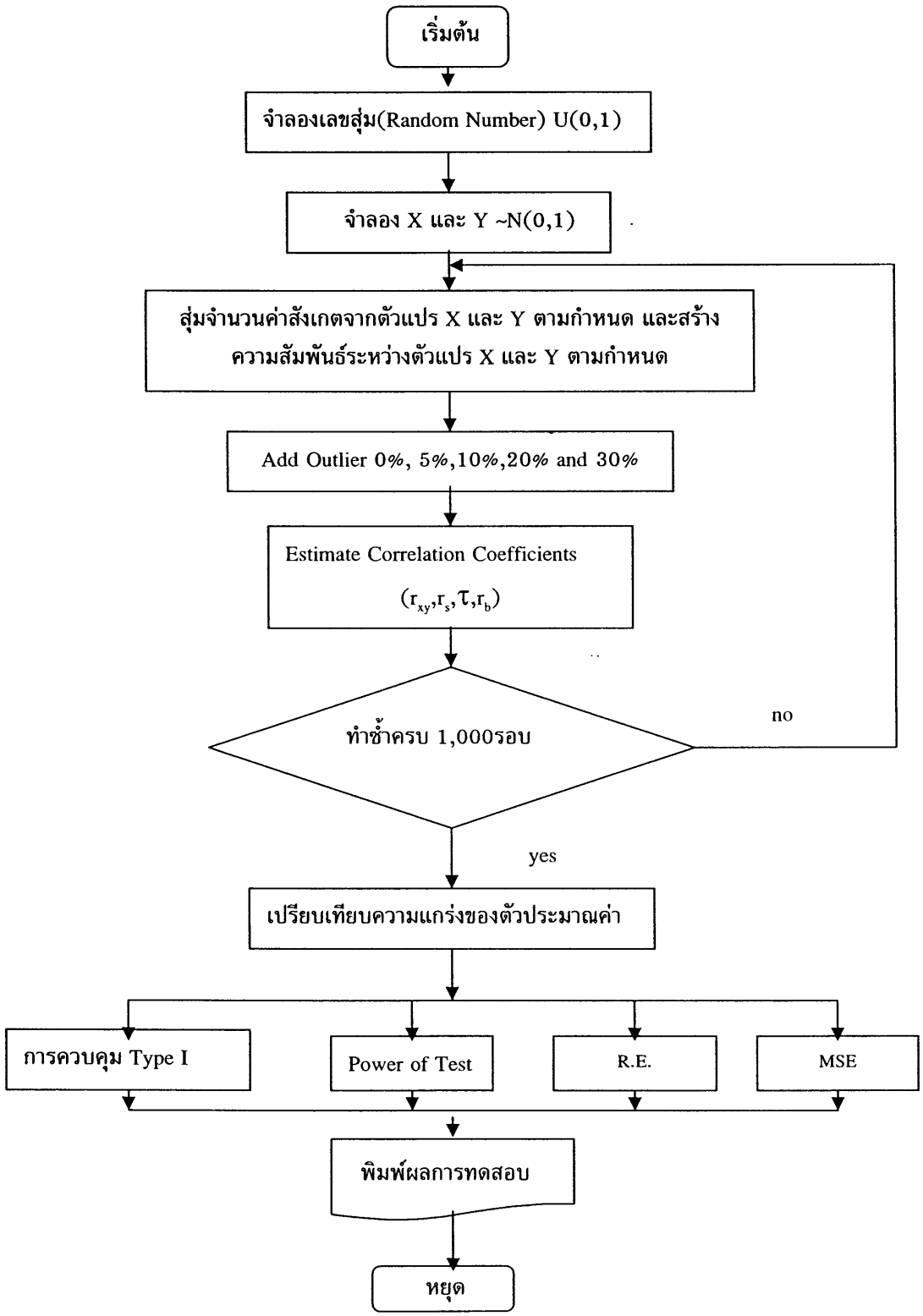
$$MSE = \frac{1}{m}\sum_{k=1}^{m}(r_k - \rho_j)^2$$

เมื่อ r_k คือ ค่าประมาณสัมประสิทธิ์สหสัมพันธ์ในการทำซ้ำรอบที่ k ในแต่ละสถานการณ์

ρ_j คือ ค่าพารามิเตอร์ที่กำหนดในแต่ละสถานการณ์ j (ρ_j = 0.00, 0.20, 0.50, 0.80 และ 0.10)

m คือ จำนวนรอบที่ทำซ้ำ

3. ขั้นตอนการทำงานโปรแกรมจำลองแบบ



ภาพที่ 14 ขั้นตอนการทำงานโปรแกรมจำลองแบบ