บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้เป็นการวิจัยซึ่งกำหนดแผนการวิจัยโดยการจำลองการทดลองด้วยเทคนิค มอนติคาร์โลซิมูเลชั่น (Monte Carlo Simulation) เพื่อหาผลสรุปเปรียบเทียบอัตราส่วนความคลาดเคลื่อนประเภทที่ 1 ของสถิติทดสอบ เอฟ กับอัตราส่วนความคลาดเคลื่อนที่ระบุ เมื่อสมมติฐานศูนย์ (Ho : $\rho = 0$) เป็นจริง และศึกษาอำนาจการทดสอบของสถิติทดสอบ เอฟ เมื่อใช้ทดสอบค่าสหสัมพันธ์พหุคูณ โดยมีลักษณะประชากรและกลุ่มตัวอย่างการวิจัยดังนี้

ลักษณะประชากรในการวิจัยครั้งนี้ เป็นลักษณะประชากรในแบบจำลอง (Model) ที่สร้างขึ้นด้วยโปรแกรมคอมพิวเตอร์ โดยกำหนดให้มีลักษณะการแจกแจงของข้อมูลเป็นแบบปกติหลายตัวแปร (Multivariate normal distribution) ที่มีค่าเฉลี่ย ($\underset{\sim}{\mu}$) เท่ากับ 0 และความแปรปรวน (Σ) เท่ากับ I เมื่อมีจำนวนตัวแปรพยากรณ์เท่ากับ 5, 7, 9 และ 11 ตัว และมีขนาดของค่าสหสัมพันธ์พหุคูณ (ρ) เท่ากับ 0.00, 0.20, 0.40, 0.60 และ 0.80 ตามลำดับ โดยจำลองขึ้นสถานการณ์ละ 2,000 ชุด

กลุ่มตัวอย่าง (Sample) ได้จากการสุ่มอย่างง่าย (Simple Random Sampling) โดยใช้ตารางเลขสุ่มแบบ Uniform Distribution โดยกำหนดให้มีขนาดของกลุ่มตัวอย่างที่สุ่มมาในแต่ละครั้งเท่ากับ 5, 10, 15 และ 20 เท่าของจำนวนตัวแปรทั้งหมดที่ศึกษา ขนาดละ 2,000 ครั้ง โดยให้แต่ละครั้งเป็นอิสระจากกัน ซึ่งมีรายละเอียดดังนี้

ตารางที่ 2 ลักษณะประชากรในแบบจำลองที่สร้างขึ้นตามสถานการณ์ต่างๆ เพื่อใช้ในการวิจัย สถานการณ์ละ 2,000 ชุด

ระดับความสัมพันธ์ของประชากร (ρ)	จำนวนตัวแปรพยากรณ์ 5	7	9	11
0.00	✓	✓	✓	✓
0.20	✓	✓	✓	✓
0.40	✓	✓	✓	✓
0.60	✓	✓	✓	✓
0.80	✓	✓	✓	✓

ตารางที่ 3 แผนการทดลอง จำแนกตามจำนวนตัวแปรพยากรณ์ ระดับของค่าสหสัมพันธ์พหุคูณในประชากรและขนาดของกลุ่มตัวอย่างที่ใช้ในการทดลอง

จำนวนตัวแปรพยากรณ์	ระดับความสัมพันธ์ในประชากร	ขนาดกลุ่มตัวอย่าง (เป็นจำนวนเท่าของตัวแปร) 5	10	15	20
5	0.00	✓	✓	✓	✓
	0.20	✓	✓	✓	✓
	0.40	✓	✓	✓	✓
	0.60	✓	✓	✓	✓
	0.80	✓	✓	✓	✓
7	0.00	✓	✓	✓	✓
	0.20	✓	✓	✓	✓
	0.40	✓	✓	✓	✓
	0.60	✓	✓	✓	✓
	0.80	✓	✓	✓	✓
9	0.00	✓	✓	✓	✓
	0.20	✓	✓	✓	✓
	0.40	✓	✓	✓	✓
	0.60	✓	✓	✓	✓
	0.80	✓	✓	✓	✓
11	0.00	✓	✓	✓	✓
	0.20	✓	✓	✓	✓
	0.40	✓	✓	✓	✓
	0.60	✓	✓	✓	✓
	0.80	✓	✓	✓	✓

ขั้นตอนดำเนินการวิจัย

การสร้างและจำลองการทดลองครั้งนี้ ใช้เครื่องคอมพิวเตอร์ AMDAHL 5860 ในระบบ OS/VSI ของสถาบันบริการคอมพิวเตอร์จุฬาลงกรณ์มหาวิทยาลัยโดยการเขียนโปรแกรมภาษาฟอร์แทรน 77 (FORTRAN 77) รวมทั้งการใช้ Scientific Subroutine ในการสร้างการแจกแจงของประชากร ซึ่งมีลำดับขั้นในการทดลองดังนี้

1. การสร้างข้อมูลประชากร

1.1 สร้างเลขสุ่มในแบบให้มีการแจกแจงอย่างสม่ำเสมอ (Uniform random number distribution) โดยใช้วิธี Linear Multiplicative Congruential (Law 1982: 227) ซึ่งเป็น Scientific Subroutine ที่สามารถสร้างเลขสุ่มได้ถึงจำนวน 2^{29} จำนวนก่อนที่จะเกิดการซ้ำของชุดเลขสุ่มดังนี้

$$X_i = (aX_{i-1} + c) \text{ Mod } M \quad i = 1,2,3,\ldots \quad (3.1)$$

เมื่อผลรวมของ $(aX_{i-1} + c)$ จะถูกหารด้วย M เศษที่เหลือจากการหารจะมีค่าเท่ากับตัวแปรสุ่ม X_i ที่มีการแจกแจงอย่างสม่ำเสมอมีค่ามัชฌิมเลขคณิตเท่ากับ 0 และความแปรปรวนเท่ากับ 1 [U(0,1)] โดยการวิจัยครั้งนี้เลือกใช้ค่าเริ่มต้น (X_{i-1}) เท่ากับ 973253 ค่า a = 16 X 807 (7^5) และค่า M เท่ากับ 2147483647 $(2^{31} - 1)$ โดยมีโปรแกรม Subroutine ที่เลือกใช้คือ RAND (IX) และสามารถดูรายละเอียดได้ในภาคผนวก ข

1.2 การสร้างตัวเลขสุ่มที่มีการแจกแจงแบบปกติ (Normal random number distribution) โดยใช้วิธีของ Box และ Muller (Moder 1975: 550-556) ดังนี้

$$Z_1 = (-2\ln R_1)^{1/2} \cos (2\pi R_2) \quad (3.2)$$

$$Z_2 = (-2\ln R_1)^{1/2} \sin (2\pi R_2) \quad (3.3)$$

เมื่อ R_1 และ R_2 เป็นค่าที่อยู่ในลำดับของเลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอ โปรแกรม Subroutine ที่เรียกใช้คือ NORMAL (SD1, AMEAN, EX) จากการใช้คำสั่งนี้ 1 ครั้ง จะได้ตัวเลขสุ่มที่มีการแจกแจงแบบปกติ 2 ค่าคือ Z_1 และ Z_2 มีค่ามัชฌิมเลขคณิตเท่ากับ 0 และความแปรปรวนเท่ากับ 1

1.3 การสร้างตัวเลขสุ่มที่มีการแจกแจงแบบปกติหลายตัวแปร (Multivariate normal random number distribution) ที่มีค่าเฉลี่ย (μ) เท่ากับ 0 และความแปรปรวน (Σ) เท่ากับ 1 เราสามารถสร้างตัวแปรสุ่ม ($X = \mu_1, \mu_2, \ldots \mu_p$) ดังนี้ (Scheuer 1962: 278-281)

$$\text{จาก} \quad Y = (Y_1, Y_2, \ldots, Y_p)$$

ซึ่งเป็นเลขสุ่มที่มีการแจกแจงปกติ มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวน 1 จะได้

$$X = C\,Y \tag{3.4}$$

$$\begin{aligned}
\text{จาก (3.4)} \quad Var(X) &= Var(CY) \\
&= C\,Var(Y)\,C' \\
&= C\,I_p\,C' \\
&= C\,C'
\end{aligned}$$

$$\text{ดังนั้น} \quad X \sim N(0, C\,C')$$

แต่เราต้องการสร้าง $X \sim N(0,\Sigma)$ นั่นก็หมายความว่า Σ จะต้องมีค่าเท่ากับ $C\,C'$ เมื่อ C เป็น Lower triangular matrix

$$C = \begin{bmatrix} C_{11} & 0 & 0 & 0 \\ C_{21} & C_{22} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ C_{p1} & \cdot & \cdot & C_{pp} \end{bmatrix}$$

โดยที่ $C_{i1} = \sigma_{i1}/\sqrt{\sigma_{i1}}$; $1 \leq i \leq p$

$$C_{ii} = \sqrt{\sigma_{ii} - \sum_{i=1}^{k} C^2_{ik}} \qquad ; 1 < i \leq p$$

$$C_{ij} = \left[\sigma_{ij} - \sum_{i=1}^{k} C_{ik}C_{jk}\right] / C_{jj} \qquad ; 1 < j < i \leq p$$

$$C_{ij} = 0 \qquad ; i < j \leq p$$

เนื่องจาก $= CC = \rho_{ij}\sigma_i\sigma_j$

เมื่อ $\sigma_i = \sigma_j = 1$ แล้ว

$\Sigma = CC' = \rho_{ij}$

ซึ่งจะได้ $X_i = C_{ij}Y_i$; $i = 1,\ldots,p$

ตามหลักเกณฑ์และขั้นตอนต่าง ๆ ที่กล่าวมาข้างต้นสามารถนำมาสร้างข้อมูลที่มีการแจกแจงแบบปกติหลายตัวแปรได้ โดยมี Main โปรแกรมที่เรียกใช้ในขั้นตอนนี้ คือ Subroutine Normal (SD1, AMEAN, EX) เพื่อสร้างข้อมูลประชากรก่อน จากนั้นเรียกใช้ Subroutine COMR (NS, MPP, YR, XRR, R) เพื่อคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ (R)

2. ตรวจสอบข้อมูลตามลักษณะการแจกแจงของประชากรแบบปกติหลายตัวแปร (X_1, ...X_p) โดยการคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณ (ρ) ค่าเฉลี่ยเลขคณิต (μ) ความแปรปรวน (Σ) ความเบ้และความโด่งตามลำดับ โดยทดสอบกับข้อมูลจำนวน 10,000 ชุด ดังแสดงในตารางที่ 4

ตารางที่ 4 การเปรียบเทียบค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ความเบ้ และความโด่งของข้อมูล

จำนวนตัวแปรพยากรณ์	ระดับความสัมพันธ์ในประชากร (ρ) ทฤษฎี	ปฏิบัติ	Mean ทฤษฎี	Mean ปฏิบัติ	S.D. ทฤษฎี	S.D. ปฏิบัติ	Skewnwss ทฤษฎี	Skewnwss ปฏิบัติ	Kurtosis ทฤษฎี	Kurtosis ปฏิบัติ
	.00	.0004	0	0.0000	1	0.9958	0	0.0420	3.0	2.9880
	.20	.2015	0	0.0000	1	0.9928	0	0.0106	3.0	2.9801
5	.40	.3997	0	0.0000	1	0.9871	0	0.0000	3.0	2.9357
	.60	.5999	0	0.0000	1	0.9583	0	0.0159	3.0	3.0350
	.80	.8010	0	0.0000	1	0.9263	0	0.0037	3.0	3.0152
	.00	.0013	0	0.0000	1	0.9976	0	0.0337	3.0	2.9680
	.20	.2014	0	0.0000	1	0.9953	0	0.0336	3.0	2.9679
7	.40	.3995	0	0.0000	1	0.9782	0	0.0134	3.0	3.0931
	.60	.6000	0	0.0000	1	0.9584	0	0.0460	3.0	2.9987
	.80	.8000	0	0.0000	1	0.9747	0	0.0131	3.0	2.9800
	.00	.0002	0	0.0000	1	0.9934	0	0.0107	3.0	3.0182
	.20	.2004	0	0.0000	1	0.9914	0	0.0235	3.0	3.0182
9	.40	.3992	0	0.0000	1	0.9848	0	0.0277	3.0	3.0099
	.60	.6003	0	0.0000	1	0.9732	0	0.0329	3.0	3.0016
	.80	.8005	0	0.0000	1	1.0089	0	0.0125	3.0	3.6630
	.00	.0014	0	0.0001	1	1.0024	0	0.0102	3.0	3.0238
	.20	.2006	0	0.0000	1	1.0005	0	0.0176	3.0	3.0234
11	.40	.3995	0	0.0000	1	0.9836	0	0.0174	3.0	2.9690
	.60	.6010	0	0.0000	1	0.9814	0	0.0163	3.0	3.0014
	.80	.8001	0	0.0000	1	1.0273	0	0.0107	3.0	3.0005

3. สุ่มตัวอย่างชุดของตัวแปรพยากรณ์และตัวแปรเกณฑ์ตามเงื่อนไขต่าง ๆ ที่กำหนด โดยมีขนาดของกลุ่มตัวอย่างที่สุ่มมาแต่ละครั้งเท่ากับ 5, 10, 15 และ 20 เท่าของตัวแปรทั้งหมด ขนาดละ 2000 ครั้ง โดยกำหนดให้แต่ละชุดของตัวแปรมีโอกาสได้รับเลือกเท่าเทียมกัน และให้มีความอิสระจากกันในการสุ่มแต่ละครั้ง

4. นำกลุ่มตัวอย่างที่สุ่มได้แต่ละขนาดและเงื่อนไขที่กำหนดมาคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณ

5. นำค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณที่ได้มาทดสอบสมมติฐาน โดยทุกสถานการณ์ถูกทดสอบด้วยสมมติฐานว่า

$$H_0 : \rho_{y.123..p} = 0$$

$$H_1 : \rho_{y.123..p} \neq 0$$

ที่ระดับนัยสำคัญ .05 และ .01 โดยมีขั้นตอนการทดสอบดังนี้

5.1 ที่แต่ละสถานการณ์ เปิดตารางเพื่อหาค่าวิกฤตของ $F_{v,n,.05}$ และ $F_{v,n,.01}$ เพื่อไว้เป็นเกณฑ์ในการเปรียบเทียบ

5.2 นำค่าสัมประสิทธิ์สหสัมพันธ์ที่คำนวณได้ในแต่ละครั้งของแต่ละสถานการณ์มาแทนในสูตร

$$F = \frac{R^2/p}{(1-R^2)/(N-p-1)}$$

เมื่อ p คือ จำนวนตัวแปรพยากรณ์

และ N คือ ขนาดกลุ่มตัวอย่าง

5.3 นำค่า F ที่ได้จาก 5.2 ไปเปรียบเทียบกับค่าวิกฤตของ F จาก 5.1 ถ้าค่า F ที่คำนวณได้มากกว่า ค่าวิกฤตของ F ก็แสดงว่าค่าสัมประสิทธิ์สหสัมพันธ์ที่ได้ครั้งนั้น Significant นั่นหมายถึง ไม่เกิดความคลาดเคลื่อนประเภทที่ 1 ในทางตรงกันข้าม ถ้าค่า F

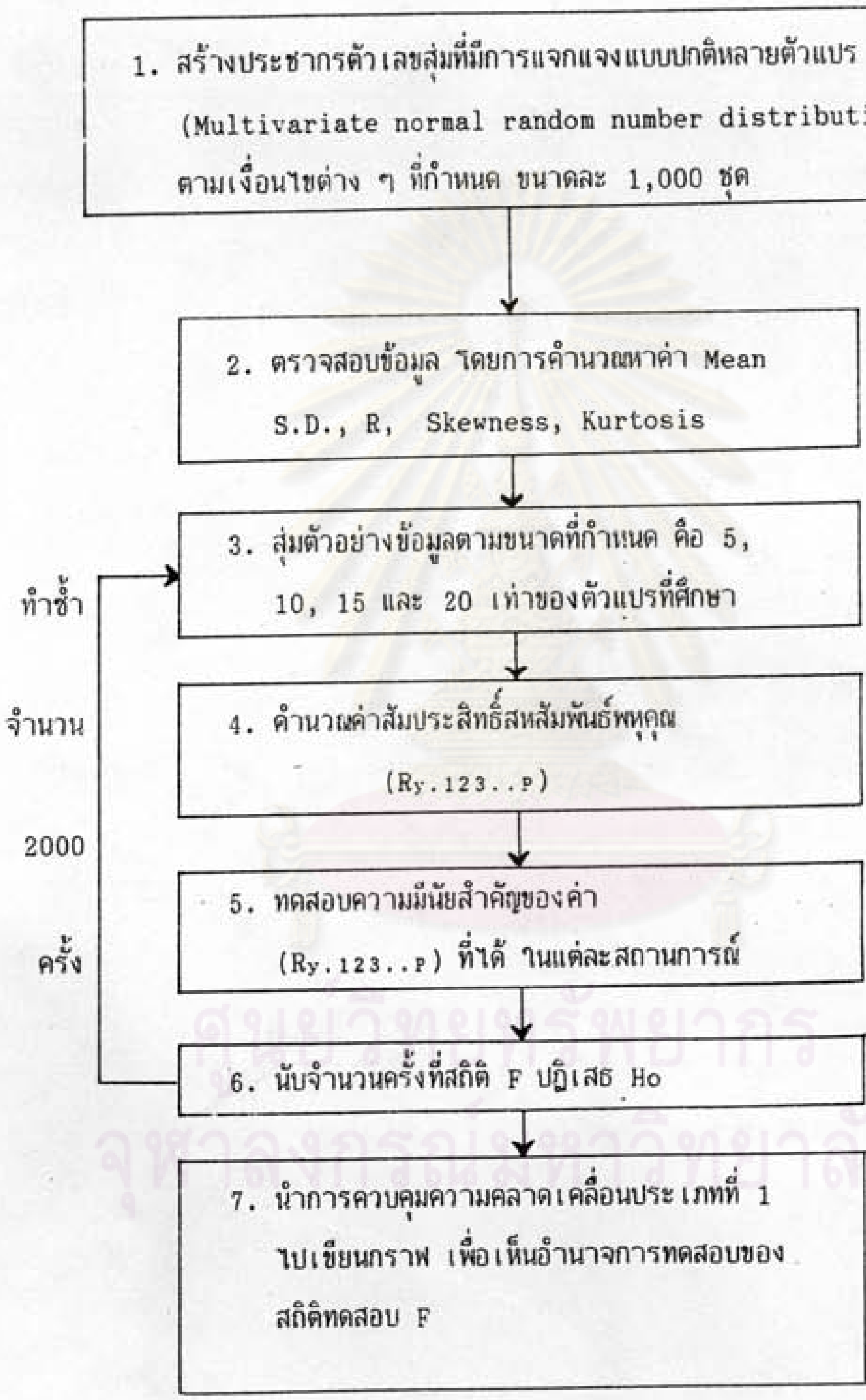
ที่คำนวณได้มีค่าน้อยกว่าค่าวิกฤตของ F ก็แสดงว่า ค่าสัมประสิทธิ์สหสัมพันธ์ที่ได้ในครั้งนั้นไม่ Significant นั่นก็หมายถึง เกิดความคลาดเคลื่อนประเภทที่ 1 ขึ้นแล้ว

แต่ละสถานการณ์หาค่า $R_{y.123..p}$ 2000 ครั้ง แล้วนับจำนวนครั้งของการปฏิเสธสมมติฐานศูนย์ กรณีที่ค่า $\rho = 0.00$ หารด้วย 2,000 ก็จะเป็นอัตราส่วนความคลาดเคลื่อนประเภทที่ 1 ส่วนค่าอำนาจการทดสอบนั้น คำนวณจากจำนวนครั้งของการปฏิเสธสมมติฐานศูนย์ หารด้วย 2,000 เฉพาะกรณีที่ค่า ρ ของทุกกรณีไม่เท่ากับ 0

6. นำอัตราการควบคุมความคลาดเคลื่อนประเภทที่ 1 มาเขียนกราฟ เพื่อให้เห็นถึงแนวทางของอำนาจการทดสอบของสถิติทดสอบตามขนาดของจำนวนตัวแปรพยากรณ์ p และขนาดของกลุ่มตัวอย่าง

ขั้นตอนการดำเนินการทั้ง 6 ขั้นตอน นำมาเขียนเป็นแผนภาพได้ดังนี้

แผนภาพที่ 7 ขั้นตอนการดำเนินการทดลอง



สถิติที่ใช้ในการวิเคราะห์ข้อมูล

การคำนวณค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณ ($R_{y.123...p}$) (Pedhazur 1982 : 73)

$$R^2_{y.123...p} = \frac{b'X'Y - n\bar{y}^2}{Y'Y - n\bar{y}^2} = \frac{b'X'Y - (\Sigma y)^2 / N}{Y'Y - (\Sigma y)^2 / N}$$

$$R_{y.123...p} = \sqrt{\frac{b'X'Y - n\bar{y}^2}{Y'Y - n\bar{y}^2}}$$

เมื่อ $R_{y.123...p}$ คือ สัมประสิทธิ์สหสัมพันธ์พหุคูณ

b' คือ เวคเตอร์สัมประสิทธิ์การถดถอย (Regression Coefficient) ซึ่ง $b = (x'x)^{-1} x'y$

$\bar{y}$ คือ ค่าเฉลี่ยของตัวแปรเกณฑ์

X คือ เมตริกของตัวแปรพยากรณ์

Y' คือ เวคเตอร์ของตัวแปรเกณฑ์

n คือ ขนาดของกลุ่มตัวอย่าง

p คือ จำนวนตัวแปรพยากรณ์