บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้เป็นการวิจัย ซึ่งกำหนดแผนการวิจัยโดยจำลองการทดลองด้วยเทคนิคมอนติคาร์โลซิมูเลชั่น (Monte Carlo Simulation) เพื่อศึกษารูปแบบการแจกแจงของค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณยกกำลังสองที่ได้ปรับแก้ด้วยวิธีการต่างกัน และเปรียบเทียบประสิทธิภาพในการประมาณค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณยกกำลังสองของประชากร โดยมีลักษณะประชากรและกลุ่มตัวอย่างการวิจัยดังนี้

ลักษณะประชากรในการวิจัยครั้งนี้ เป็นลักษณะประชากรในแบบจำลอง (Model) ที่สร้างขึ้นด้วยโปรแกรมคอมพิวเตอร์ เพื่อใช้อ้างอิงในการประมาณค่าพารามิเตอร์ โดยกำหนดให้มีลักษณะการแจกแจงของข้อมูลเป็นแบบปกติหลายตัวแปร (Multivariate Normal Distribution) ที่มีค่า (μ) เท่ากับ 0 และความแปรปรวน (Σ) เท่ากับ 1 เมื่อมีจำนวนตัวแปรพยากรณ์เท่ากับ 3, 5, 7 และ 9 ตัว และมีขนาดของค่าสหสัมพันธ์พหุคูณ (ρ) เท่ากับ .20, .40, .60 และ .80 ตามลำดับโดยจำลองขึ้นเงื่อนไขละจำนวน 1000 ชุด

กลุ่มตัวอย่าง (Sample) ได้จากการสุ่มอย่างง่าย (Simple Random Sampling) โดยกำหนดให้มีขนาดของกลุ่มตัวอย่างที่สุ่มมาในแต่ละครั้งเท่ากับ 2, 5, 10, 15, 20, 25 และ 30 เท่าของตัวแปรที่ศึกษาทั้งหมด ขนาดละ 1000 ครั้ง โดยให้แต่ละครั้งเป็นอิสระจากกัน ดังมีรายละเอียดดังนี้

ตารางที่ 2 ลักษณะประชากรในแบบจำลองที่สร้างขึ้นตามเงื่อนไขต่าง ๆ เพื่อใช้ในการวิจัย เงื่อนไขละ 1000 ชุด

ระดับความสัมพันธ์ของประชากร	จำนวนตัวแปรพยากรณ์ 3	5	7	9
0.20	/	/	/	/
0.40	/	/	/	/
0.60	/	/	/	/
0.80	/	/	/	/

ตารางที่ 3 แผนการทดลองจำแนกตามจำนวนตัวแปรพยากรณ์ ขนาดของค่าสหสัมพันธ์พหุคูณในประชากรและขนาดของกลุ่มตัวอย่างที่ใช้ในการทดลอง

จำนวนตัวแปรพยากรณ์	ขนาดความสัมพันธ์ในประชากร	ขนาดกลุ่มตัวอย่าง (เป็นจำนวนเท่าของตัวแปร)						
		2	5	10	15	20	25	30
3	0.20							
	0.40							
	0.60							
	0.80							
5	0.20							
	0.40							
	0.60							
	0.80							
7	0.20							
	0.40							
	0.60							
	0.80							
9	0.20							
	0.40							
	0.60							
	0.80							

<u>ขั้นตอนดำเนินการวิจัย</u>

การสร้างและจำลองการทดลองครั้งนี้ ใช้เครื่องคอมพิวเตอร์ IBM 370/3031 ในระบบ OS/VSI ของสถาบันบริการคอมพิวเตอร์ จุฬาลงกรณ์มหาวิทยาลัย โดยการเขียนโปรแกรมด้วยภาษาฟอร์แทรน 4 (FORTRAN IV) รวมทั้งการใช้ Scientific Subroutine ในการสร้างการแจกแจงของประชากร ซึ่งมีลำดับขั้นในการทดลองดังนี้

1. การสร้างข้อมูลประชากร

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

$$R_{i+1} = (aR_i + C) \text{ Mod } T \tag{3.1}$$

เมื่อผลรวมของ $(aR_i + c)$ จะถูกหารด้วย T เศษที่เหลือจากการหารจะมีค่าเท่ากับตัวแปรสุ่ม R_{i+1} ที่มีการแจกแจงแบบสม่ำเสมอมีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1 [U (0,1)]โดยการวิจัยครั้งนี้เลือกใช้ค่าเริ่มต้น (R_i) เท่ากับ 973253 ค่า a = 16807 (7^5) และค่า T เท่ากับ 2147483647 (2^{31}) โดยมีโปรแกรม Subroutine ที่เลือกใช้คือ RANDU (IX, IY, RNN)

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

$$X_n = (-2 \log R_n)^{1/2} \text{ Cos } 2 \; R_{n+1} \tag{3.2}$$

$$X_{n+1} = (-2 \log R_n)^{1/2} \text{ Sin } 2 \; R_{n+1} \tag{3.3}$$

เมื่อ R_n และ R_{n+1} เป็นค่าที่อยู่ในลำดับของเลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอ โปรแกรม Subroutine ที่เรียกใช้คือ NORMAL (IX, KK, RN) จากการใช้คำสั่งนี้ 1 ครั้ง จะได้ตัวเลขสุ่มที่มีการแจกแจงแบบปกติ 2 ค่าคือ X_n และ X_{n+1} มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1

1.3 การสร้างตัวเลขสุ่มที่มีการแจกแจงเป็นแบบปกติหลายตัวแปร (Maltivariate 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)

จาก $Y = (Y_1 , Y_2 , \ldots\ldots, Y_p)$

ซึ่งเป็นเลขสุ่มที่มีการแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวน 1 จะได้

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

จาก (3.4)

$$\begin{aligned} Var(X) &= Var(C\,Y) \\ &= C\,Var(Y)\,C' \\ &= C\,I_p\,C' \\ &= C\,C' \end{aligned} \tag{3.5}$$

ดังนั้น $X \sim N(0, C\,C')$

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

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

โดยที่

$$C_{ii} = \sigma_{ii}\sqrt{\sigma_{ii}} \qquad ;\ 1 \leq i \leq p,$$

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

$$C_{ij} = \left(\sigma_{ij} \sum_{t=1}^{k} C_{ik} C_{jk}\right) \Big/ C_{jj} \qquad ;\ 1 < j < i \leq p,$$

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

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

เมื่อ $\sigma_i = \sigma_j = 1$ แล้ว

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

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

ตามหลักเกณฑ์และขั้นตอนต่าง ๆ ที่กล่าวมาข้างต้นสามารถนำมาสร้างข้อมูลที่มีการแจกแจงแบบปกติหลายตัวแปรได้ โดยมี Main โปรแกรมที่เรียกใช้ในขั้นตอนนี้ คือ Subroutine MNORM (JK, KK, M, N, XB, SD, R) ซึ่ง Subroutine นี้เรียกโปรแกรม Subroutine SETUP (IDK, M, XB, SD, R) เพื่อสร้าง Lower triangular matrix (CC) และ Subroutine Normal (IX, KK, RN) เพื่อสร้างตัวแปรสุ่มที่มีการแจกแจงแบบปกติ (Y_i)

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

ตารางที่ 4 การเปรียบเทียบค่าเฉลี่ย ความแปรปรวน ความเบ้ และความโด่งของข้อมูล

จำนวนตัวแปรพยากรณ์	ระดับความสัมพันธ์ในประชากร (ρ)		Mean		Variance		Skewness		Kurtosis	
	ทฤษฎี	ปฏิบัติ	ทฤษฎี	ปฏิบัติ	ทฤษฎี	ปฏิบัติ	ทฤษฎี	ปฏิบัติ	ทฤษฎี	ปฏิบัติ
	.20	.208	0	0.000	1	0.993	0	0.022	3.0	3.037
3	.40	.408	0	0.000	1	0.993	0	0.022	3.0	3.037
	.60	.604	0	0.000	1	0.993	0	0.022	3.0	3.037
	.80	.804	0	0.000	1	0.993	0	0.022	3.0	3.037
	.20	.204	0	0.000	1	1.019	0	0.007	3.0	2.951
5	.40	.408	0	0.000	1	1.019	0	0.007	3.0	2.951
	.60	.609	0	0.000	1	1.019	0	0.007	3.0	2.951
	.80	.804	0	0.000	1	1.019	0	0.007	3.0	2.951

ตารางที่ 4 (ต่อ)

จำนวนตัวแปรพยากรณ์	ระดับความสัมพันธ์ในประชากร (ρ)		Mean		Variance		Skewness		Kurtosis	
	ทฤษฎี	ปฏิบัติ	ทฤษฎี	ปฏิบัติ	ทฤษฎี	ปฏิบัติ	ทฤษฎี	ปฏิบัติ	ทฤษฎี	ปฏิบัติ
	.20	.205	0	0.000	1	0.993	0	0.002	3.0	2.995
7	.40	.402	0	0.000	1	0.993	0	0.002	3.0	2.995
	.60	.605	0	0.000	1	0.993	0	0.002	3.0	2.995
	.80	.807	0	0.000	1	0.993	0	0.002	3.0	2.995
	.20	.205	0	0.000	1	0.975	0	0.002	3.0	2.983
9	.40	.403	0	0.000	1	0.975	0	0.002	3.0	2.983
	.60	.604	0	0.000	1	0.975	0	0.002	3.0	2.983
	.80	.807	0	0.000	1	0.975	0	0.002	3.0	2.983

3. สุ่มตัวอย่างชุดของตัวแปรพยากรณ์และตัวแปรเกณฑ์ตามเงื่อนไขต่าง ๆ ที่กำหนด โดยมีขนาดของกลุ่มตัวอย่างที่สุ่มมาแต่ละครั้งเท่ากับ 2, 5, 10, 15, 20, 25 และ 30 เท่าของตัวแปรขนาดละ 1000 ครั้ง โดยกำหนดให้แต่ละชุดของตัวแปรมีโอกาสได้รับเลือกเท่าเทียมกัน และให้มีความเป็นอิสระจากกันในการสุ่มแต่ละครั้ง โปรแกรม Subroutine ที่เรียกใช้ในขั้นตอนนี้คือ Subroutine UNF (IX, AL, BH, U) ซึ่งสร้างตารางเลขสุ่ม (Ravindran 1976 : 215) โดยมีการแจกแจงดังนี้

เมื่อ U มี p.d.f. (Probability density function)

$$F(u) = 1/(b - a) \qquad ; a \leq u \leq b \qquad (3.6)$$

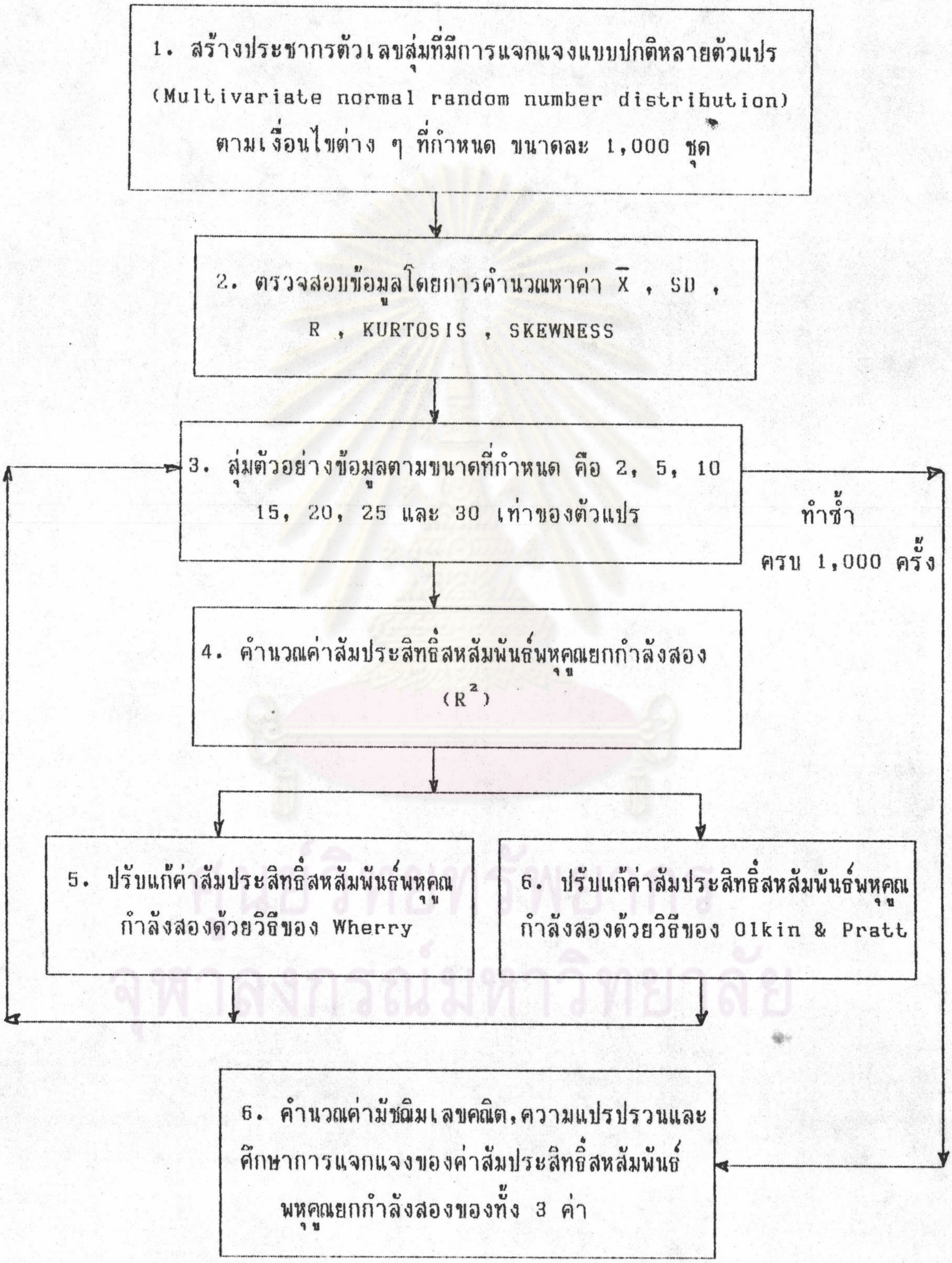
ค่า U จะเป็นค่าตัวเลขสุ่มที่มีค่าอยู่ระหว่าง a และ b (a = 1 และ b = ขนาดของประชากร) มีค่าเฉลี่ยเท่ากับ $(a + b)/2$ ความแปรปรวนเท่ากับ $(b - a)^2/12$

4. นำกลุ่มตัวอย่างที่สุ่มได้แต่ละขนาดและเงื่อนไขที่กำหนดมาคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณยกกำลังสองแล้วทำการปรับแก้ด้วยวิธีของ Wherry และวิธีของ Olkin กับ Pratt ซึ่งมีโปรแกรม Subroutine ที่เรียกใช้คือ Subroutine SSX (M, N, MM) และ Subroutine INVS (RM, RW, RO, M, N, MM)

5. ทำการเปรียบเทียบค่าสถิติที่คำนวณได้มานำเสนอในรูปกราฟ เพื่อให้เห็นแนวโน้มของการเปลี่ยนแปลงของค่าสถิติ เมื่อจำนวนตัวแปรพยากรณ์ ขนาดของกลุ่มตัวอย่าง และขนาดความสัมพันธ์ในประชากรเปลี่ยนไป พร้อมทั้งคำนวณหาค่ามัชฌิมเลขคณิต และความแปรปรวนของค่าสถิติที่คำนวณได้จากแต่ละวิธี

จากขั้นตอนการดำเนินการทดลองทั้งหมดสามารถสรุปได้ดังแผนภาพที่ 2

<u>แผนภาพที่ 2</u> แสดงขั้นตอนการดำเนินการทดลอง



<u>สถิติที่ใช้ในการวิเคราะห์ข้อมูล</u>

1. การคำนวณค่า R^2 (Pedhazur 1982: 73)

$$R^2_{Y.12...k} = \frac{b'X'Y - n\bar{y}^2}{Y'Y - n\bar{y}^2}$$

เมื่อ $R^2_{Y.12...k}$ คือ สหสัมพันธ์พหุคูณยกกำลังสองหรือสัมประสิทธิ์การตัดสินใจ (Determination Coefficient)

b' คือ เวคเตอร์สัมประสิทธิ์การถดถอย (Regression Coefficient)

$\bar{y}$ คือ ค่าเฉลี่ยของตัวแปรเกณฑ์

X คือ เมตริกของตัวแปรพยากรณ์

Y' คือ เวคเตอร์ของตัวแปรเกณฑ์

n คือ ขนาดของกลุ่มตัวอย่าง

2. การปรับแก้ค่า R^2 ด้วยวิธีของเวอร์รี่ (Wherry's Method) และวิธีของโอลกินกับแพรตต์ (Olkin and Pratt's Method)

$$R^2_{w2} = 1 - \frac{n-1}{n-p-1}(1-R^2)$$ (Herzberg: 1969: 3)

$$R^2_{OP} = 1 - \frac{n-3}{n-p-1}(1-R^2) - \frac{n-3}{n-p-1}\,\frac{2}{n-p+1}(1-R^2)^2$$

(Olkin & Pratt 1958: 201-211)

เมื่อ R^2_{w2} คือ ค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณยกกำลังสองที่ปรับแก้ด้วยวิธีของเวอร์รี่

R^2_{op} คือ ค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณยกกำลังสองที่ปรับแก้ด้วยวิธีของโอลกินกับแพรตต์

R^2 คือ ค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณยกกำลังสองที่คำนวณได้จากกลุ่มตัวอย่าง

P คือ จำนวนตัวแปรพยากรณ์

n คือ ขนาดของกลุ่มตัวอย่าง

3. การคำนวณค่าเฉลี่ยของค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณยกกำลังสอง

$$\overline{X}_R2 = \frac{1}{m}\sum_{i=1}^{m} R_i^2$$

เมื่อ $\overline{X}_R2$ คือ ค่าเฉลี่ยของค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณยกกำลังสอง

R_i^2 คือ ค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณยกกำลังสอง

m คือ จำนวนครั้งของการสุ่ม

4. การคำนวณค่าความแปรปรวนของค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณยกกำลังสอง

$$S^2_R2 = \frac{1}{m}\sum_{i=1}^{m} (R_i'^2 - \overline{X}_R2)^2$$

เมื่อ S^2_R2 คือ ความแปรปรวนของค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณยกกำลังสอง

R_i^2 คือ ค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณยกกำลังสอง

$\overline{X}_R2$ คือ ค่าเฉลี่ยสหสัมพันธ์พหุคูณยกกำลังสอง

m คือ จำนวนครั้งของการสุ่ม