บทที่ 3
วิธีดำเนินการวิจัย

ข้อมูลที่ใช้ในการวิจัยครั้งนี้ ได้จากการจำลองด้วยเทคนิคมอนติคาร์โลซิมูเลชั่น (Monte carlo Simulation) เพื่อเปรียบเทียบค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์พหุคูณกำลังสองและร้อยละของตัวแปรทำนายที่เข้าสมการจากการคัดเลือกตัวแปรเข้าสมการถดถอยต่างวิธีกัน โดยมีลักษณะประชากรและกลุ่มตัวอย่างการวิจัยดังนี้

ลักษณะประชากร เป็นแบบจำลอง (Model) ที่สร้างขึ้นด้วยโปรแกรมคอมพิวเตอร์ โดยกำหนดให้ข้อมูลประชากรมีการแจกแจงเป็นแบบปกติ (Normal Distribution) ที่มีค่าเฉลี่ย (μ) เท่ากับ 0 และความแปรปรวน (Σ) เท่ากับ 1 เมื่อมีจำนวนตัวแปรทำนาย 5 ตัว โดยที่ตัวแปรทำนายแต่ละคู่มีความสัมพันธ์กันในระดับต่าง ๆ 3 ระดับ คือ (0.00-0.30), (0.30-0.70), (0.70-1.00) และตัวแปรทำนายแต่ละตัวมีความสัมพันธ์กับตัวแปรเกณฑ์อยู่ระหว่าง (0.30-1.00) โดยจำลองขึ้นเงื่อนไขละจำนวน 200 ชุด

กลุ่มตัวอย่าง (Sample) ได้จากการสุ่มอย่างง่าย (Simple Random Sampling) โดยกำหนดให้ขนาดของกลุ่มตัวอย่างเท่ากับ 10 เท่าของตัวแปรทำนาย ขนาดละ 200 ครั้ง โดยให้แต่ละครั้งเป็นอิสระจากกัน ดังมีรายละเอียดดังนี้

<u>ตารางที่ 2</u> แผนการสร้างข้อมูลจำแนกตามจำนวนตัวแปรทำนาย ความสัมพันธ์ระหว่างตัวแปรทำนายกับตัวแปรเกณฑ์ ระดับความสัมพันธ์ระหว่างตัวแปรทำนาย และขนาดของกลุ่มตัวอย่าง

จำนวนตัวแปรทำนาย	ความสัมพันธ์ระหว่างตัวแปรทำนายกับตัวแปรเกณฑ์	ระดับความสัมพันธ์ระหว่างตัวแปรทำนาย	ขนาดของกลุ่มตัวอย่าง
5	0.30-1.00	(0.00-0.30) (0.30-0.70) (0.70-1.00)	10 เท่าของตัวแปรทำนาย

<u>ขั้นตอนดำเนินการวิจัย</u>

การสร้างข้อมูลในการวิจัยครั้งนี้ ใช้เครื่องคอมพิวเตอร์ IBM 370/3031 ในระบบ Music ของสถาบันบริการคอมพิวเตอร์ จุฬาลงกรณ์มหาวิทยาลัย โดยการเขียนโปรแกรมใน การสร้างข้อมูลด้วยภาษาฟอร์แทรน 77 (Fortran 77) ใช้ Function Ran ในการสร้าง เลขสุ่มเพื่อสร้างการแจกแจงของประชากร และประมวลผลข้อมูลดังกล่าวด้วยโปรแกรมภาษา SPSSx ซึ่งมีลำดับขั้นการทำงานดังนี้

1. การสร้างข้อมูลประชากร

1.1 การสร้างเลขสุ่มให้มีการแจกแจงแบบสม่ำเสมอ (Uniform random number distribution) โดยใช้วิธี Prime Modulus Multiplicative (Averill M. Law and W. David Kelton 1982: 225-227) โดยใช้ Function Ran (IX) โดยสร้างเลขสุ่มได้ถึงจำนวน 2^{29} จำนวน ก่อนที่จะเกิดการซ้ำของชุดเลขสุ่มดังนี้

$$R_{i+1} = (aR_i) \text{ Mod } m$$

เมื่อ $a = 7^5$

$m = 2^{31} - 1$

$$\therefore R_{i+1} = (7^5 R_i) \text{ (Mod } 2^{31}-1)$$

เมื่อ $(7^5 R_i)$ ถูกหารด้วย $(2^{31} - 1)$ เศษที่เหลือจากการหารจะมีค่าเท่ากับ ตัวแปรสุ่ม R_{i+1} ที่มีการแจกแจงแบบสม่ำเสมอมีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1 โดยการวิจัยครั้งนี้ให้ R_i เป็นค่าเริ่มต้น (Seed) มีค่าเท่ากับ 973253 โดยค่าเริ่มต้นนี้ จะเป็นจำนวนคี่ที่มีค่าอยู่ระหว่าง $(\pm 2^{31}-1)$ หรือเป็นเลขไพม์ (Prime Number) ก็ได้ (Averill M. Law and W. David Kelton 1982: 223)

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

$$Z_i = (-2\ln R_n)^{1/2} \quad \mathrm{Cos}\ 2\pi (R_{n+1})$$

R_n และ R_{n+1} เป็นค่าที่ได้มาจากลำดับของเลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอ โดยใช้ R ทีละ 2 ค่า ตามลำดับคี่และคู่ ทีละ 1 คู่ ในสูตรนี้ จะได้ค่าตัวเลขสุ่มที่มีการแจกแจงแบบปกติ (Z_i) มา 1 ค่า ที่มีค่าเฉลี่ยเท่ากับ 0 ความแปรปรวนเท่ากับ 1 โดยใช้ program Com Z.for

1.3 การสร้างตัวเลขสุ่มที่มีการแจกแจงเป็นแบบปกติหลายตัวแปร (multivariate normal random number distribution) ที่มีค่าเฉลี่ย (μ) เท่ากับ 0 และความแปรปรวน (Σ) เท่ากับ 1 เราสามารถสร้างตัวแปรสุ่มได้ดังนี้ (Scheuer 1962 : 278-281)

$$X = (X_1, x_2 \ldots. X_n)$$

โดยที่ X_i เป็นตัวแปรสุ่ม (Random Variable) และจะมีการแจกแจงเป็นแบบปกติ เมื่อมี p.d.f. (Probability density function) ดังนี้

$$f_x(x) = \frac{1}{(2\pi)^{P/2} (|\Sigma|)^{1/2}} \exp{-1/2}\, (X_i-\mu)' \, \Sigma^{-1} (X-\mu)$$

$$\text{โดยที่} \quad -\infty < x_i < \infty \;,\; i = 1, 2, \ldots\ldots, p$$

เขียนได้เป็น $x \sim n\, (\underset{\sim}{\mu}, \Sigma)$

$$\text{เมื่อ} \quad \underset{\sim}{\mu} = (u_1, u_2, \ldots\ldots u_p)$$

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \ldots\ldots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & & \sigma_{2p} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \sigma_{p1} & \sigma_{p2} & & \sigma_{pp} \end{bmatrix}$$

Σ เป็น p x p positive definite และสมมาตร (Symmetric)

ซึ่งคือเมตริกซ์ของความแปรปรวนร่วม (Covariance matrix)

สามารถสร้าง $X \sim n\ (0, \Sigma)$ ได้โดย

ให้ $Z \sim N\ (0, I_p)$

$$X = CZ$$

$$\begin{aligned} Var\ (X) &= Var\ (CZ) \\ &= C\ Var\ (Z)\ C' \\ &= C\ I_p\ C' \\ &= C\ C' \end{aligned}$$

ดังนั้น $X \sim N\ (0,\ C\ C')$

แต่เราต้องการสร้างให้ $X \sim N\ (0, \Sigma)$ นั้นก็หมายความว่า Σ จะต้องมีค่าเท่ากับ $C\ C'$ จึงจะทำให้ $X = C\ Z$ ดังนั้น จะต้องหาค่า C ซึ่งได้ว่า C เป็น lower triangular matrix

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

โดยที่ $C_{i1} = \sigma_{i1} / \sqrt{\sigma_{11}}$ $\quad 1 \leq i \leq p$

$$C_{ii} = \sigma_{ii} - \sum_{k=1}^{i-1} C_{ik} \qquad 1 < i \leq p$$

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

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

ซึ่งจะได้ $X_i = \sum_{j=1}^{i} C_{ij}Z_j, \; i=1, 2, \ldots\ldots, p$

เมื่อ $Z_1, Z_2, \ldots\ldots, Z_p \sim N(0, 1)$

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

เมื่อ $\sigma_i = \sigma_j = 1$ แล้ว

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

ซึ่งจะได้ $X_i = C_{ij} Z_j \qquad ; i=1, 2, \ldots P$

ตามหลักเกณฑ์ที่กล่าวมาข้างต้น สามารถนำมาสร้างข้อมูลที่มีการแจกแจงแบบปกติหลายตัวแปรได้ ซึ่งมีขั้นตอนในการสร้างโดยละเอียดดังนี้

1.3.1 สร้างค่าความแปรปรวนร่วม (Covariance) ของกลุ่มที่ 1, 2, 3 โดยใช้ program rand 4. for โดยให้มีค่าเท่ากับระดับความสัมพันธ์ ระหว่างตัวแปรที่กำหนดไว้ในแต่ละกลุ่ม

1.3.2 สุ่ม (Random) ความแปรปรวนร่วมที่ได้จาก 1.3. 1 ลงในเมตริกส์ความแปรปรวนและความแปรปรวนร่วม (Variance-Covariance Matrix) โดยให้มีความสัมพันธ์ระหว่างตัวแปรตามที่กำหนดไว้ในแต่ละกลุ่ม โดยใช้ program rand 6. for ซึ่งมีแนวทะแยงเป็น 1 และ สมมาตร (Symmetric) กัน $\sigma_{ij} = \sigma_{ji}$ ซึ่งจะได้เมตริกส์ความแปรปรวนและความแปรปรวนร่วม (Variance-Covariance Matrix) ทั้งหมด 3 กลุ่ม ดังนี้

<u>กลุ่มที่ 1</u> มีค่า $r_{x_i x_j} = (0.00 - 0.30)$, $r_{x_i y} = (0.30-1.00)$
มีตัวแปรทำนาย (X_i) 5 ตัว และ Y เป็นตัวแปรเกณฑ์

$$
\Sigma_1 = \begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ Y \end{array}
\begin{array}{c}
\begin{array}{cccccc} X_1 & X_2 & X_3 & X_4 & X_5 & Y \end{array} \\
\begin{bmatrix}
1.0000000 & .1466423 & .1222862 & .2648697 & .1515013 & .7189047 \\
.1466423 & 1.0000000 & .1098033 & .0135721 & .1058245 & .4134787 \\
.1222862 & .1098033 & 1.0000000 & .0137088 & .2905896 & .7670849 \\
.2648697 & .0135721 & .0137088 & 1.0000000 & .2924712 & .3995662 \\
.1515013 & .1058245 & .2905896 & .2924712 & 1.0000000 & .6821264 \\
.7189047 & .4134787 & .7670849 & .3995662 & .6821264 & 1.0000000
\end{bmatrix}
\end{array}
$$

<u>กลุ่มที่ 2</u> มีค่า $r_{x_i x_j} = (0.30-0.70)$, $r_{x_i y} = (0.30-1.00)$
มีตัวแปรทำนาย (X_i) 5 ตัว และ Y เป็นตัวแปรเกณฑ์

$$
\Sigma_2 = \begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ Y \end{array}
\begin{array}{c}
\begin{array}{cccccc} X_1 & X_2 & X_3 & X_4 & X_5 & Y \end{array} \\
\begin{bmatrix}
1.0000000 & .6170374 & .5469237 & .6168449 & .3121131 & .7189047 \\
.6170374 & 1.0000000 & .6650531 & .5474232 & .5423642 & .4134787 \\
.5469237 & .6650631 & 1.0000000 & .6966920 & .3022640 & .7670849 \\
.6168449 & .5474232 & .6966920 & 1.0000000 & .4770990 & .3995662 \\
.3121131 & .5423642 & .3022640 & .4770990 & 1.0000000 & .6821264 \\
.7189047 & .4134787 & .7670849 & .3995662 & .6821264 & 1.0000000
\end{bmatrix}
\end{array}
$$

<u>กลุ่มที่ 3</u> มีค่า $r_{x_i x_j}$ = (0.70-1.00), $r_{x_i y}$ = (0.30-1.00)

มีตัวแปรทำนาย (X_i) 5 ตัว และ Y เป็นตัวแปรเกณฑ์

$$
\Sigma_y = \begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ Y \end{array}
\begin{array}{c}
\begin{array}{cccccc} X_1 & X_2 & X_3 & X_4 & X_5 & Y \end{array} \\
\begin{bmatrix}
1.0000000 & .8496150 & .9803595 & .9024184 & .9452838 & .7189047 \\
.8496150 & 1.0000000 & .8264656 & .7536374 & .8670074 & .4134787 \\
.9803595 & .8264656 & 1.0000000 & .8734861 & .9070458 & .7670849 \\
.9024184 & .7536374 & .8734861 & 1.0000000 & .7670849 & .3995662 \\
.9452838 & .8670074 & .9070458 & .7670849 & 1.0000000 & .6821264 \\
.7189047 & .4134787 & .7670849 & .3995662 & .6821264 & 1.0000000
\end{bmatrix}
\end{array}
$$

1.3.3 แปลงเมตริกส์ความแปรปรวนและความแปรปรวนร่วม (Variance-Covariance Matrix) ให้เป็น Lower triangular Matrix (C) ทั้ง 3 กลุ่ม โดยใช้ program Comxy.for และใช้สูตรในข้อ 1.3 ซึ่งจะได้เมตริกส์ C ดังนี้

<u>กลุ่มที่1</u>

$$
C_1 = \begin{bmatrix}
1.0000000 & .0000000 & .0000000 & .0000000 & .0000000 & .0000000 \\
.1466423 & .9891896 & .0000000 & .0000000 & .0000000 & .0000000 \\
.1222862 & .0928750 & .9881399 & .0000000 & .0000000 & .0000000 \\
.2648697 & -.0255452 & -.0165043 & .9638045 & .0000000 & .0000000 \\
.1515013 & .0845217 & .2673844 & .2686387 & .9089787 & .0000000 \\
.7189047 & .3114235 & .6580539 & .2365275 & .3381769 & .4659961
\end{bmatrix}
$$

กลุ่มที่ 2

$$C_2 = \begin{bmatrix} 1.0000000 & .0000000 & .0000000 & .0000000 & .0000000 & .0000000 \\ .6170374 & .7869338 & .0000000 & .0000000 & .0000000 & .0000000 \\ .5469237 & .4162748 & .7263537 & .0000000 & .0000000 & .0000000 \\ .6168449 & .2119706 & .3732161 & .6597580 & .0000000 & .0000000 \\ .3121131 & .4444830 & -.0736076 & .3301632 & .7685014 & .0000000 \\ .7189047 & -.0382655 & .5366915 & -.3578240 & .8229004 & .7820031 \end{bmatrix}$$

กลุ่มที่ 3

$$C_3 = \begin{bmatrix} 1.0000000 & .0000000 & .0000000 & .0000000 & .0000000 & .0000000 \\ .8496150 & .5274034 & .0000000 & .0000000 & .0000000 & .0000000 \\ .9803595 & -.0122535 & .1968378 & .0000000 & .0000000 & .0000000 \\ .9024184 & -.0247833 & -.0584846 & .4261530 & .0000000 & .0000000 \\ .9452838 & .1211220 & -.0924010 & -.2073406 & .2005989 & .0000000 \\ .7189047 & -.3741224 & .2932137 & -.5662523 & -.2115762 & .3288952 \end{bmatrix}$$

1.3.4 หาค่าของตัวแปรทำนาย (X_1, $X_2 \ldots X_5$) และตัวแปรเกณฑ์ (Y) โดยใช้ program Created. for จากสมการ

$$X_i = C_{ij} Z_j$$ ดังนี้

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ \vdots \\ X_5 \\ Y \end{bmatrix} = \begin{bmatrix} C_{11} & 0 & 0 & 0 & 0 & 0 \\ C_{21} & \ddots & 0 & 0 & 0 & 0 \\ C_{31} & & \ddots & 0 & 0 & 0 \\ C_{41} & & & \ddots & 0 & 0 \\ C_{51} & & & & \ddots & 0 \\ C_{61} & \cdots & \cdots & \cdots & \cdots & C_{66} \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ \vdots \\ \vdots \\ Z_6 \end{bmatrix}$$

จะได้ตัวแปร X_i, Y มา 3 กลุ่ม โดยมีความสัมพันธ์ระหว่างตัวแปรตามที่ต้องการ จากนั้น generate ตัวแปร X_i, Y ของแต่ละกลุ่ม ๆ ละ 200 ชุด ๆ ละ 50 ค่า (10 เท่าของตัวแปรทำนาย

2. ตรวจสอบข้อมูลตามลักษณะการแจกแจงแบบปกติหลายตัวแปร (Multivariate normal random number distribution) โดยใช้โปรแกรม Statall เพื่อหาค่าสหสัมพันธ์ระหว่างตัวแปร (intercorrelation) ค่าเฉลี่ยเลขคณิต (μ) ความแปรปรวน (Σ) ความเบ้และความโด่งตามลำดับ ดังนี้

ค่าสหสัมพันธ์ระหว่างตัวแปรที่ได้จากการตรวจสอบ

<u>กลุ่มที่ 1</u> มีค่า $r_{x_i x_j}$ = (0.00 - 0.30), $r_{x_i y}$ = (0.30-1.00)

มีตัวแปรทำนาย (X_i) 5 ตัว และ Y เป็นตัวแปรเกณฑ์

$$\Sigma_1 = \begin{array}{c|cccccc|} & X_1 & X_2 & X_3 & X_4 & X_5 & Y \\ \hline X_1 & 1.0000 & & & & & \\ X_2 & .1487^{**} & 1.0000 & & & & \\ X_3 & .1292^{**} & .1131^{**} & 1.0000 & & & \\ X_4 & .2618^{**} & .0138^{**} & .0129^{**} & 1.0000 & & \\ X_5 & .1505^{**} & .1065^{**} & .3013^{**} & .2994^{**} & 1.0000 & \\ Y & .7206^{**} & .4127^{**} & .7653^{**} & .3998^{**} & .6801^{**} & 1.0000 \end{array}$$

* P < .05

** P < .01

<u>กลุ่มที่ 2</u> มีค่า $r_{x_ix_j}$ = (0.30-0.70), r_{x_iy} = (0.30-1.00)

มีตัวแปรทำนาย (X_i) 5 ตัว และ Y เป็นตัวแปรเกณฑ์

$$
\begin{array}{ll}
 & \begin{array}{cccccc} \quad X_1 \quad & \quad X_2 \quad & \quad X_3 \quad & \quad X_4 \quad & \quad X_5 \quad & \quad Y \quad \end{array} \\
\begin{array}{r} X_1 \\ X_2 \\ \Sigma_2 = X_3 \\ X_4 \\ X_5 \\ Y \end{array} &
\left[\begin{array}{cccccc}
1.0000 & & & & & \\
.6170^{**} & 1.0000 & & & & \\
.5549^{**} & .6741^{**} & 1.0000 & & & \\
.6163^{**} & .5342^{**} & .6966^{**} & 1.0000 & & \\
.3121^{**} & .5633^{**} & .3022^{**} & .4770^{**} & 1.0000 & \\
.7189^{**} & .4150^{**} & .7670^{**} & .3991^{**} & .6830^{**} & 1.0000
\end{array} \right]
\end{array}
$$

$*\ P<.05$

$**\ P<.01$

<u>กลุ่มที่ 3</u> มีค่า $r_{x_ix_j}$ = (0.70-1.00), r_{x_iy} = (0.30-1.00)

มีตัวแปรทำนาย (X_i) 5 ตัว และ Y เป็นตัวแปรเกณฑ์

$$
\begin{array}{ll}
 & \begin{array}{cccccc} \quad X_1 \quad & \quad X_2 \quad & \quad X_3 \quad & \quad X_4 \quad & \quad X_5 \quad & \quad Y \quad \end{array} \\
\begin{array}{r} X_1 \\ X_2 \\ \Sigma_3 = X_3 \\ X_4 \\ X_5 \\ Y \end{array} &
\left[\begin{array}{cccccc}
1.0000 & & & & & \\
.8495^{**} & 1.0000 & & & & \\
.9803^{**} & .8264^{**} & 1.0000 & & & \\
.9024^{**} & .7536^{**} & .8742^{**} & 1.0000 & & \\
.9453^{**} & .8670^{**} & .9070^{**} & .7681^{**} & 1.0000 & \\
.7189^{**} & .4136^{**} & .7670^{**} & .3995^{**} & .6821^{**} & 1.0000
\end{array} \right]
\end{array}
$$

$*\ P < .05$

$**\ P < .01$

ตารางที่ 3 เปรียบเทียบค่าเฉลี่ย ความแปรปรวน ความเบ้และความโด่งของข้อมูล

ระดับความสัมพันธ์ระหว่างตัวแปรทำนาย	Mean		Variance		Skewness		Kurtosis	
	ทฤษฎี	ปฏิบัติ	ทฤษฎี	ปฏิบัติ	ทฤษฎี	ปฏิบัติ	ทฤษฎี	ปฏิบัติ
(0.00-0.30)	0	0.000	1	0.996	0	0.005	3	2.971
(0.30-0.70)	0	0.000	1	1.019	0	0.013	3	3.026
(0.07-1.00)	0	0.000	1	0.984	0	0.002	3	3.015

3. นำข้อมูลที่สร้างได้ในแต่ละกลุ่ม มาคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณยกกำลังสอง (R^2) ด้วยวิธีการคัดตัวแปรแบบไปข้างหน้า (Forward Selection) การกำจัดตัวแปรแบบถอยหลัง (Backward Elimination) และการถดถอยแบบขั้นบันได (Stepwise Regression) จะได้ R^2 มา 200 ค่าในแต่ละวิธี (ชุดละ 1 ค่า)โดยใช้โปรแกรม SPSSx

4. คำนวณหาค่าเฉลี่ยและความแปรปรวนของ R^2 ในแต่ละวิธีของแต่ละกลุ่มเพื่อเปรียบเทียบค่าเฉลี่ยและความแปรปรวนของค่า R^2 ที่ได้ในแต่ละวิธี เมื่อระดับความสัมพันธ์ระหว่างตัวแปรทำนายอยู่กลุ่มเดียวกัน และเปรียบเทียบค่าเฉลี่ยและความแปรปรวนของ R^2 ในวิธีเดียวกัน เมื่อระดับความสัมพันธ์ระหว่างตัวแปรทำนายเปลี่ยนไป

5. คำนวณร้อยละของตัวแปรทำนายที่ถูกตัดเข้าสมการในอันดับต่าง ๆ ของวิธีคัดตัวแปรแบบไปข้างหน้า (Forward Selecton) และวิธีถดถอยแบบขั้นบันได (Stepwise Regsession) เมื่อความสัมพันธ์ระหว่างตัวแปรทำนายอยู่ในระดับเดียวกัน

จากขั้นตอนการดำเนินการทดลองทั้งหมดสามารถสรุปได้ดังแผนภาพต่อไปนี้

แผนภาพที่ 4 แสดงขั้นตอนการดำเนินการทดลอง

1. สร้างเลขสุ่มให้มีการแจกแจงแบบสม่ำเสมอ (Uniform random number distribution) โดยใช้สูตร

$$R_{i+1} = (7^5 R_i) \text{ (Mod } 2^{31}-1)$$

↓

2. สร้างเลขสุ่มที่มีการแจกแจงแบบปกติ (Normal random number distribution) โดยใช้สูตร

$$Z = \sqrt{-2\ln R_n} \text{ Cos } 2\pi \, (R_{n+1})$$

โดยใช้ R จาก 1. แทนค่าในสูตรทีละ 2 ค่า ตามลำดับคี่และคู่ จะได้ Zi มา 1 กลุ่ม

↓

3. สร้างเลขสุ่มที่มีการแจกแจงแบบปกติหลายตัวแปร (Maltivariate normal random number distribution) โดย Random ค่าความแปรปรวนร่วมเข้าเมตริกส์ความแปรปรวนร่วม (Σ) โดยให้มีความสัมพันธ์ระหว่างตัวแปรในแต่ละกลุ่มตามที่กำหนด แปลงเมตริกส์ความแปรปรวนร่วมให้เป็นเมตริกส์ C และหาตัวแปร X_i, Y จากสมการ

$$X_i = C_{ij} Z_j$$

และ generate ตัวแปรมาชุดละ 50 ค่า

↓

4. ทำการคัดเลือกตัวแปรด้วยวิธี forward Selection Backward Elimination และ Stepwise Regression และคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณยกกำลังสอง (R^2) ของแต่ละวิธี

↓

5. คำนวณหาค่ามัชฌิมเลขคณิตและความแปรปรวนของสัมประสิทธิ์สหสัมพันธ์พหุคูณยกกำลังสอง (R^2) ในแต่ละวิธี

ทำซ้ำครบ 200 ครั้ง ในแต่ละกลุ่ม

↓

6. คำนวณหาร้อยละของตัวแปรทำนายที่ถูกคัดเข้าสมการในอันดับต่าง ๆ ของวิธีคัดตัวแปรแบบไปข้างหน้า (Forward Selection) และ วิธีถดถอยแบบขั้นบันได (Stepwise Regression)

สถิติที่ใช้ในการวิเคราะห์ข้อมูล

1. การเลือกตัวแปรแบบไปข้างหน้า (Forward Selection)

1.1 พิจารณาสัมประสิทธิสหสัมพันธ์บางส่วนระหว่างตัวแปรเกณฑ์กับตัวแปรทำนายแต่ละตัว (r_{xy}) เลือกตัวแปรทำนายที่ให้ค่าสัมประสิทธิ์สหสัมพันธ์บางส่วนสูงที่สุดสมมุติได้ X_j สมการเป็น $\hat{Y} = a + b_j X_j$

1.2 หาค่าสัมประสิทธิ์สหสัมพันธ์บางส่วนระหว่างตัวแปรเกณฑ์กับตัวแปรทำนายแต่ละตัวที่ยังไม่อยู่ในสมการ โดยถือว่าได้รวมตัวแปรเกณฑ์ X_j ไว้ในสมการแล้ว นั่นคือหาค่า

$$r_y(lj) = \frac{r_{yl} - r_{yj}\, r_{lj}}{\sqrt{(1-r^2_{lj})}}$$

เมื่อ $l = 1, 2, \ldots\ldots, j-1, j+1, \ldots\ldots k$

k = จำนวนตัวแปรทำนายทั้งหมดที่จะพิจารณา

1.3 เลือกค่าสัมประสิทธิ์สหสัมพันธ์บางส่วนที่สูงที่สุด สมมุติให้ $r_{y(1.j)}$ จึงควรรวม X_1 ไว้ในสมการเป็นตัวแปรใหม่

1.4 พิจารณา partial F ของตัวแปรทำนายใหม่ X_1 ในข้อ 1.3 ถ้ามีค่าสูงกว่า $F_{\infty}(1, n-m-1)$ แสดงว่าเป็นการสมควรที่จะรวม X_1 ไว้ในสมการ ในที่นี้ m คือจำนวนตัวแปรทำนายในสมการใหม่ และ n คือจำนวนค่าสังเกตุ (observation)

1.5 ทำตามข้อ 1.2, 1.3 และ 1.4 อีก โดยถือว่าสมการได้รวมตัวแปรทำนายไว้แล้ว 2 ตัว 3 ตัว ฯลฯ ตามลำดับ จน partial F ที่ได้จากตัวแปรทำนายมีค่าน้อยกว่า $F_{\infty}(1, n-m-1)$ ตัวแปรทำนายใหม่จึงไม่ควรรวมอยู่ในสมการ ซึ่งแสดงว่าได้สมการที่เหมาะสมที่สุดแล้ว

ในกรณีที่รวมตัวแปรทำนายไว้แล้ว 2 ตัว สมมุติเป็นตัวที่ 1 และ 2 จะหาสัมประสิทธิ์สหสัมพันธ์บางส่วนของ Y และตัวแปรทำนายตัวที่ 3 ได้ดังนี้

$$r_y(3.12) = \frac{r_{y(3.1)} - r_{y(2.1)}\, r_{3(2.1)}}{\sqrt{(1-r^2_{3(2.1)})}}$$

และในกรณีที่รวมตัวแปรทำนายไว้แล้ว 3 ตัว สัมประสิทธิสหสัมพันธ์บางส่วนของ Y และตัวแปรทำนายตัวที่ 4 คือ

$$r_{y(4.123)} = \frac{r_{y(4.12)} - r_{y(3.12)} r_{4(3.12)}}{\sqrt{(1-r^2_{4(3.12)})}}$$

2. การกำจัดตัวแปรแบบถอยหลัง (Backward Elimination)

2.1 ตั้งสมการที่รวมตัวแปรทำนายที่ควรจะพิจาณาทั้งหมด สมมุติมี K ตัว สมการจะเป็นดังนี้

$$\hat{Y} = a + b_1X_1 + b_2X_2 + \ldots\ldots\ldots + b_kX_k$$

2.2 คำนวณ partial F ของตัวแปรทำนายแต่ละตัว ในจำนวน partial F นี้ตัวเลือกที่น้อยที่สุด สมมุติได้ F_j แล้วนำ F_j เปรียบเทียบกับค่า F จากตารางคือ $F_{\alpha}(1,n-k-1)$ ถ้า F_j น้อยกว่า F จากตารางให้กำจัดตัวแปรทำนาย X_j ออกจากสมการ

2.3 ตั้งสมการใหม่โดยไม่รวม X_j ในสมการ แล้วทำตามข้อ 2.2 อีก โดยที่ค่า K จะเปลี่ยนเป็น K-1 ทำเช่นนี้จนในที่สุด partial F ทุก ๆ ตัว มีค่ามากกว่า F ที่กำหนดจากตาราง

3. การถดถอยแบบขั้นบันได (Stepwise Regression)

3.1 พิจารณาสัมประสิทธิ์สหสัมพันธ์บางส่วนระหว่างตัวแปรเกณฑ์กับตัวแปรทำนายแต่ละตัว (r_{xy}) เลือกตัวแปรทำนายที่ให้ค่าสัมประสิทธิ์สหสัมพันธ์บางส่วนสูงที่สุด สมมุติให้ X_j สมการจะเป็น $\hat{Y} = a + b_j X_j$

3.2 หาค่าสัมประสิทธิ์สหสัมพันธ์บางส่วนระหว่างตัวแปรเกณฑ์กับตัวแปรทำนายแต่ละตัวที่ยังไม่อยู่ในสมการ โดยถือว่าได้รวมตัวแปรทำนาย X_j ไว้ในสมการแล้ว และเลือกตัวแปรทำนายที่ให้ค่าสัมประสิทธิ์สหสัมพันธ์บางส่วนสูงสุด สมมุติให้ X_1 สมการจะเป็น

$$\hat{Y} = a + b_j X_j + b_1X_1$$

3.3 พิจารณา partial F ของทั้ง X_1 และ X_j ถ้ามีค่ามากกว่า $F_{\alpha}(1,n-3)$ ทั้ง 2 ตัวก็รวม X_1 และ X_j ไว้ในสมการ

3.4 ทำตามข้อ 3.2 และ 3.3 โดยที่จะมีตัวแปรทำนายรวมอยู่ในสมการแล้ว 2 ตัว 3 ตัว ฯลฯ ตามลำดับ ในแต่ละขั้นต้องพิจารณาค่า partial F ของตัวแปรทำนายทุกตัว ถ้าตัวใดมีค่าน้อยกว่า $F_{\infty}(1, n-m-1)$ ก็จะตัดตัวแปรทำนายนั้นออกจากสมการ ทำเช่นนี้จนไม่มีตัวแปรทำนายตัวใดที่จะรวมในสมการจะไม่มีตัวใดที่จะถูกตัดออกจากสมการ (วิชิต หล่อจีระชุณห์กุล 2524)

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

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

เมื่อ $R^2_{y.12....k}$ คือสหสัมพันธ์พหุคูณยกกำลังสองหรือสัมประสิทธิ์การทำนาย (Determination Coefficient)

b' คือเวคเตอร์สัมประสิทธิ์การถดถอย (Regression Coettiecient)

$\bar{Y}$ คือ ค่าเฉลี่ยของตัวแปรเกณฑ์

X คือ เมตริกส์ของตัวแปรทำนาย

Y' คือ เวคเตอร์ของตัวแปรเกณฑ์

N คือ ขนาดของกลุ่มตัวอย่าง

5. การคำนวนค่าเฉลี่ยของค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณยกกำลังสอง (R^2)

$$\bar{X}_{R^2} = \frac{1}{m}\sum_{i=1}^{m} R_i^2$$

เมื่อ $\bar{X}_{R^2}$ คือ ค่าเฉลี่ยของ R^2

$R^2{}_i$ คือ ค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณยกกำลังสอง

m คือ จำนวนครั้งของการสุ่ม

6. การคำนวนค่าความแปรปรวนของสัมประสิทธิ์สหสัมพันธ์พหุคูณกำลังสอง (R^2)

$$S^2_{R^2} = \frac{\sum_{i=1}^{N} (R^2_i - \bar{R}^2)^2}{N-1}$$

เมื่อ

$S^2_{R^2}$ คือค่าความแปรปรวนของสัมประสิทธิ์สหสัมพันธ์พหุคูณกำลังสอง (R^2)

R^2_i คือค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณกำลังสอง (R^2)

$\bar{R}^2$ คือค่าเฉลี่ยของสัมประสิทธิ์สหสัมพันธ์พหุคูณกำลังสอง (R^2)

N คือจำนวนครั้งของการสุ่ม

7. ทดสอบความมีนัยสำคัญของ $\bar{R}_F{}^2$, $\bar{R}_B{}^2$, $\bar{R}_S{}^2$, โดยใช้ t-test ทำการทดสอบทีละคู่

$$t = \frac{(\bar{R}_1{}^2 - \bar{R}_2{}^2) - (\mu 1 - \mu 2)}{\sqrt{\frac{S^2}{n1} + \frac{S^2}{n2}}}$$

เมื่อ $\bar{R}_1{}^2$ คือ ค่าเฉลี่ยของค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณกำลังสองของกลุ่มตัวอย่างของวิธีคัดตัวแปรวิธีที่ 1

$\bar{R}_2{}^2$ คือ ค่าเฉลี่ยของค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณกำลังสองของกลุ่มตัวอย่างของวิธีคัดตัวแปรวิธีที่ 2

$\mu 1$, $\mu 2$ คือ ค่าเฉลี่ยของค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณกำลังสองในประชากรของกลุ่มตัวอย่างของวิธีคัดตัวแปรวิธีที่ 1 และวิธีที่ 2

S^2_1, S^2_2 คือ ค่าความแปรปรวนของค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณกำลังสองของกลุ่มตัวอย่างของวิธีคัดตัวแปรวิธีที่ 1 และวิธีที่ 2

n1, n2 คือ คำนวนชุดของกลุ่มตัวอย่างของวิธีคัดตัวแปรวิธีที่ 1 และวิธีที่ 2