บทที่ 1

บทนำ

ที่มาและความสำคัญของปัญหา

ในปัจจุบันการวิจัยในด้านต่าง ๆ อาทิเช่น วิทยาศาสตร์ สังคมศาสตร์ และในด้านธุรกิจต่าง ๆ จำเป็นต้องอาศัยความรู้ทางสถิติเข้ามาช่วยในการวิจัยค้นคว้าโดยเฉพาะ เพื่อการคาดคะเนสิ่งที่จะเกิดขึ้นในอนาคตหรือที่เรียกว่าการพยากรณ์ ซึ่งอาจจะจำแนกได้เป็นสองประเภทใหญ่คือ ประเภทที่ 1 เป็นวิธีการที่มีแนวความคิดว่าพฤติกรรมในอดีตของสิ่งที่จะพยากรณ์ควรจะเพียงพอที่จะพยากรณ์พฤติกรรมในอนาคตของตนเองได้ เช่น อนุกรมเวลา ส่วนประเภทที่ 2 เป็นวิธีการที่มีแนวความคิดว่าพฤติกรรมของสิ่งที่จะพยากรณ์ถูกกำหนดขึ้นโดยสิ่งอื่น ๆ ซึ่งมีความสัมพันธ์บางลักษณะกับสิ่งที่จะพยากรณ์ เช่น การวิเคราะห์การถดถอย (วิชิต หล่อจีระชุณห์กุล, 2539: 1) ในปัจจุบันนิยมใช้สมการการถดถอยเข้ามาช่วยในการพยากรณ์และหาคำตอบ ในการวิเคราะห์การถดถอยจะประกอบไปด้วยตัวแปรอิสระ (independent variables) และตัวแปรตาม (dependent variable) แต่ในความเป็นจริงการใช้ตัวแปรอิสระเพียงตัวเดียวอาจจะไม่สามารถอธิบายตัวแปรตามได้ดีจึงควรใช้ตัวแปรอิสระมากกว่าหนึ่งตัวขึ้นไปมาช่วยในการอธิบายตัวแปรตามให้มีประสิทธิภาพในการพยากรณ์มากขึ้น ซึ่งตัวแปรอิสระที่มากกว่าหนึ่งตัวขึ้นไปที่เข้ามาอยู่ในสมการถดถอยนี้เรียกว่า การถดถอยพหุคูณ (multiple regression) และมีรูปแบบดังนี้

$$\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$$

เมื่อ $\underset{\sim}{y}$ เป็นเวกเตอร์ของตัวแปรตามขนาด n x 1

X เป็นเมทริกซ์ของตัวแปรอิสระขนาด n x (p+1)

$\underset{\sim}{\beta}$ เป็นเวกเตอร์ของสัมประสิทธิ์การถดถอยขนาด (p+1) x 1

$\underset{\sim}{\varepsilon}$ เป็นเวกเตอร์ของความคลาดเคลื่อนขนาด n x 1

n เป็นขนาดตัวอย่างของตัวแปรอิสระ

และ p เป็นจำนวนตัวแปรอิสระ

โดยที่ $E(\underset{\sim}{\varepsilon}) = 0$ และ $Cov(\underset{\sim}{\varepsilon}) = \sigma^2 I_n$

ในการประมาณสัมประสิทธิ์การถดถอยพหุคูณวิธีที่นิยมใช้กันมากคือวิธีกำลังสองน้อยสุด ซึ่งจะได้ตัวประมาณสัมประสิทธิ์การถดถอยพหุคูณอยู่ในรูปของ $\hat{\underset{\sim}{\beta}} = (X'X)^{-1}X'\underset{\sim}{y}$ ตัวประมาณที่ได้ดังกล่าวเป็นตัวประมาณไม่เอนเอียง

การหาสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีกำลังสองน้อยสุดนี้ใช้งานได้ดีเมื่อตัวแปรอิสระไม่มีความสัมพันธ์กัน แต่กรณีนี้ในทางปฏิบัติเป็นไปได้ยากและไม่สามารถทราบได้ว่าตัวแปรอิสระใดบ้างที่มีหรือไม่มีความสัมพันธ์กัน เพราะตัวแปรอิสระบางตัวอาจมีความสัมพันธ์กันอยู่โดยที่ตัวแปรอิสระบางตัวอาจเป็นฟังก์ชันของตัวแปรอิสระตัวอื่นๆ กล่าวคือ ตัวแปรอิสระมีพหุสัมพันธ์ (multicollinearity) แต่ถ้าใช้สมการถดถอยเมื่อตัวแปรเกิดพหุสัมพันธ์กันช่วยในการพยากรณ์จะทำให้การประมาณค่าที่ได้อาจไม่เหมาะสมและขาดความเที่ยงตรง (accuracy)

ในปี คศ. 1970 โฮเอิล (Hoel) และเคนนาร์ด (Kennard) ได้เสนอวิธีแก้ปัญหาในกรณีที่เกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ ซึ่งจะได้ค่าความคลาดเคลื่อนกำลังสองน้อยสุดซึ่งเรียกว่า วิธีริดจ์รีเกรชชัน (Ridge regression method) และมีหลักการทำงานจากระยะทางระหว่างตัวประมาณสัมประสิทธิ์กำลังสองน้อยสุด($\hat{\underset{\sim}{\beta}}$) และค่าสัมประสิทธิ์การถดถอยพหุคูณ($\underset{\sim}{\beta}$) เพื่อพิจารณาค่าเฉลี่ยยกกำลังสองของระยะทางระหว่าง $\hat{\underset{\sim}{\beta}}$ และ $\underset{\sim}{\beta}$ มีค่าเท่ากับ $\sigma^2 trace(\mathrm{X}'\mathrm{X})^{-1}$ ซึ่งจะมีค่าสมมูล (equivalent) กับค่าเฉลี่ยของ $\hat{\underset{\sim}{\beta}}$ และ $\underset{\sim}{\beta}$ ซึ่งสามารถเขียนให้อยู่ในรูปของ

$$E(\hat{\underset{\sim}{\beta}}'\hat{\underset{\sim}{\beta}}) = \underset{\sim}{\beta}'\underset{\sim}{\beta} + \sigma^2 trace(\mathrm{X}'\mathrm{X})^{-1}$$

ดังนั้น

$$E(\hat{\underset{\sim}{\beta}}'\hat{\underset{\sim}{\beta}}) = \underset{\sim}{\beta}'\underset{\sim}{\beta} + \left(\frac{\sigma^2}{\lambda_1} + \frac{\sigma^2}{\lambda_2} + \ldots + \frac{\sigma^2}{\lambda_p}\right)$$

ซึ่งผู้อ่านจะเห็นได้ว่าผลบวกกำลังสองของตัวประมาณ $\hat{\underset{\sim}{\beta}}$ มากกว่าผลบวกกำลังสองของ $\underset{\sim}{\beta}$ อยู่เท่ากับ $\sigma^2 trace(\mathrm{X}'\mathrm{X})^{-1}$ เพราะฉะนั้นการที่จะทำให้ผลบวกกำลังสองของ $\hat{\underset{\sim}{\beta}}$ ลดลงก็โดยการเพิ่มค่า λ ของเมทริกซ์ $(\mathrm{X}'\mathrm{X})$ และทำการบวกค่าคงที่ k ที่มากกว่าศูนย์บนเส้นทแยงมุมของเมทริกซ์ $(\mathrm{X}'\mathrm{X})$ ซึ่งจะได้ค่าประมาณสัมประสิทธิ์ของสมการถดถอยริดจ์จากรูปแบบดังนี้

$$\hat{\underset{\sim}{\beta}}_R(k) = (\mathrm{X}'\mathrm{X} + kI)^{-1}\mathrm{X}'\underset{\sim}{y} \qquad ; k > 0$$

ในการหาค่าประมาณสัมประสิทธิ์ของสมการการถดถอยริดจ์ที่ได้กล่าวมานี้พบว่ามีการประมาณค่าพารามิเตอร์ k เพื่อช่วยในการหาค่าสัมประสิทธิ์ (เจษฎาพร ยุทธนวิบูลย์ชัย, 2533: 3) ซึ่งได้มีผู้เสนอวิธีการประมาณค่าพารามิเตอร์ k หลายวิธี แต่ยังไม่มีข้อสรุปว่าวิธีไหนที่ให้ค่าพารามิเตอร์ k ที่มีความเหมาะสมและเที่ยงตรงมากที่สุด ในที่นี้ผู้วิจัยได้เสนอวิธีการประมาณค่าพารามิเตอร์ k ดังนี้

ในปี คศ. 1995 บลีแมน (Breiman) ได้เสนอให้มีประมาณค่าพารามิเตอร์ k จากค่าของ $trace(X'X(X'X+kI)^{-1})$ จนกระทั่งได้ค่าของผลรวมเท่ากับจำนวนตัวแปรอิสระ ซึ่งวิธีที่บลี-แมนได้เสนอมาใหม่นี้มีช่วยในการหาสัมประสิทธิ์ของตัวประมาณริดจ์ได้สะดวกรวดเร็วยิ่งขึ้น โดยการบวกค่าคงที่เข้าไปในเทอมของเมทริกซ์ $(X'X)$ และแก้สมการของผลรวมของเส้นทแยงมุมเพื่อหาค่า k วิธีนี้ค่อนข้างจะสะดวกในการคำนวณและคาดคะเนแต่เป็นวิธีใหม่ที่เพิ่งเสนอจึงยังไม่เป็นที่นิยมใช้แพร่หลายมากนัก

ในปีเดียวกันนั้นบลีแมนยังได้เสนอสมการเชิงเส้นใหม่ขึ้นมาเพื่อใช้ช่วยในการพยากรณ์เมื่อเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ ซึ่งสมการเชิงเส้นใหม่นี้ค่อนข้างง่ายในการพิจารณาได้แก่วิธีของสมการถดถอยเชิงเส้นการ์ล็อต (Garrote linear regression method (GAR)) โดยมีหลัก-การในการทำงานดังนี้คือ ทำการหาตัวประมาณสัมประสิทธิ์การถดถอยจากวิธีกำลังสองน้อยสุดเพื่อนำมาช่วยในการหาตัวประมาณใหม่ โดยจะต้องทำให้สมการ $\sum_{i=1}^{n}(y_i - \sum_{k=1}^{p+1} c_k \hat{\beta}_k x_{ki})^2$ ให้มีค่าต่ำสุดเพื่อหาค่า $\{c_k\}$ ภายใต้เงื่อนไขที่ว่า $\sum_{k=1}^{p+1} c_k^2 \le (p+1)^2$ เมื่อ p คือจำนวนตัวแปรอิสระ และ n คือขนาดตัวอย่าง ซึ่งจะได้สัมประสิทธิ์การถดถอยใหม่จากสมการดังนี้

$$\underset{\sim}{\hat{\beta}_G}(k) = \underset{\sim}{\hat{c}_k}\, \underset{\sim}{\hat{\beta}_k} \qquad ; k = 1,\dots,p+1$$

วิธีนี้ได้พัฒนามาจากวิธีกำลังสองน้อยสุด โดยการเพิ่ม $\{c_k\}$ ให้เป็นตัวถ่วงน้ำหนักของตัวประมาณสัมประสิทธิ์การถดถอยของวิธีกำลังสองน้อยสุดและได้อาศัยหลักการคล้ายกับตัวประมาณสไตน์ (Stein estimator) โดยเป็นการนำเอาค่าคงที่คูณกับสัมประสิทธิ์ตัวประมาณกำลังสองน้อยสุด โดยค่าคงที่ซึ่งนำมาคูณหามาจากวิธีกำลังสองน้อยสุดซึ่งเป็นวิธีที่ใช้กันอย่างแพร่หลาย ดังนั้นในการคำนวณหาค่า c ที่เหมาะสมจึงไม่ค่อยยุ่งยากและมีความเที่ยงตรงมากกว่าวิธีกำลังสองน้อยสุดในกรณีเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ

จากที่กล่าวมาข้างต้นผู้วิจัยจะทำการประมาณสัมประสิทธิ์การถดถอย 3 วิธีคือ วิธีกำลังสองน้อยสุด วิธีที่ได้จากสมการถดถอยริดจ์โดยใช้วิธีของบลีแมน และวิธีของสมการถดถอยเชิงเส้นการ์ล็อต ซึ่งทั้งสามนี้วิธีเป็นการหาตัวประมาณสัมประสิทธิ์การถดถอยในกรณีที่เกิดพหุ-สัมพันธ์ระหว่างตัวแปรอิสระ โดยจะพิจารณาว่าวิธีไหนที่จะให้สมการการพยากรณ์ที่มีประสิทธิ-ภาพและความเที่ยงตรงในการพยากรณ์มากที่สุดในการวิจัยครั้งนี้

วัตถุประสงค์ของการวิจัย

การวิจัยครั้งนี้เพื่อประมาณสัมประสิทธิ์การถดถอยเมื่อเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ ซึ่งในการวิจัยครั้งนี้ได้แบ่งออกเป็น 3 วิธีดังนี้

1. วิธีกำลังสองน้อยสุด (Ordinary Least Square Method) : OLS
2. วิธีที่ได้จากสมการถดถอยริดจ์โดยใช้วิธีของบรีแมน (Ridge Regression by Breiman Method) : RID
3. วิธีของสมการถดถอยเชิงเส้นการ์ล็อด (Garrote Linear Regression Method) : GAR

สมมติฐานของการวิจัย

เมื่อตัวแปรอิสระเกิดพหุสัมพันธ์กัน วิธี RID จะให้ค่าสัมประสิทธิ์การถดถอยพหุคูณมีความถูกต้องมากกว่าวิธีอื่นภายใต้จำนวนตัวแปรอิสระ ขนาดตัวอย่างและร้อยละของสัมประสิทธิ์การแปรผันของความคลาดเคลื่อนเดียวกัน โดยที่วิธี OLS และ GAR จะมีประสิทธิภาพดีเมื่อตัวแปรอิสระเกิดพหุสัมพันธ์กันสูงมากและการแจกแจงเข้าสู่การแจกแจงแบบปกติ

ข้อตกลงเบื้องต้น

ในการวิจัยครั้งนี้ใช้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเป็นตัวตัดสินใจในการเลือกวิธีที่จะใช้ในการประมาณสัมประสิทธิ์การถดถอยเชิงเส้นพหุคูณเมื่อเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ

ขอบเขตของการวิจัย

1.<u>การแจกแจงของความคลาดเคลื่อนมี 4 การแจกแจงคือ</u>

1.1 <u>ความคลาดเคลื่อนมีการแจกแจงปกติ (normal distribution)</u>

ฟังก์ชันความหนาแน่นของ X อยู่ในรูปของ

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} \quad , -\infty < x < \infty$$

เมื่อ μ คือพารามิเตอร์ค่าเฉลี่ยซึ่งมีค่าเท่ากับ 1

และ σ^2 คือพารามิเตอร์ความแปรปรวนซึ่งมีค่าเท่ากับ 0.0025 และ 0.0225 ถ้าความแปรปรวนมีค่ามากไปกว่านี้ความคลาดเคลื่อนของการแจกแจงจะไม่ใช่การแจกแจงแบบปกติ (จิรายุส พู่มนตรี, 2534: 6)

1.2 ความคลาดเคลื่อนมีการแจกแจงปกติปลอมปน (scale-contaminated normal distribution)

ฟังก์ชันการแจกแจงอยู่ในรูปของ

$$F = (1-p)N(\mu,\sigma^2) + pN(\mu,c^2\sigma^2)$$

เมื่อ μ คือพารามิเตอร์ค่าเฉลี่ยซึ่งมีค่าเท่ากับ 1

σ^2 คือพารามิเตอร์ความแปรปรวนซึ่งมีค่าเท่ากับ 0.0025 และ 0.0225

c คือสเกลแฟคเตอร์ (scale factor) ซึ่งมีค่าเท่ากับ 3 และ 10 ถ้าค่า c มีค่าสูงจะทำให้ค่าสังเกตที่ผิดปกติมีค่าสูงด้วย

และ p คือเปอร์เซ็นต์การปลอมปน (percent of contamination) ซึ่งมีค่าเท่ากับ 5 และ 10 ถ้าค่า p มีค่าสูงจะทำให้ค่าสังเกตที่ผิดปกติมีค่าสูงด้วย

1.3 เมื่อความคลาดเคลื่อนมีการแจกแจงไวบูลล์ (Weibull distribution)

ฟังก์ชันความหนาแน่นของ X อยู่ในรูปของ

$$f(x) = \frac{\beta}{\alpha}\left(\frac{x}{\alpha}\right)^{\beta-1}\exp\left\{-\left(\frac{x}{\alpha}\right)^{\beta}\right\} \quad , \quad x>0$$

เมื่อ α คือพารามิเตอร์มาตราส่วน (scale parameter) ซึ่งมีค่าเท่ากับ 1

และ β คือพารามิเตอร์สัณฐาน (shape parameter) ซึ่งมีค่าเท่ากับ 1, 2 และ 5 ถ้าพารามิเตอร์สัณฐานมีค่าเพิ่มขึ้นไปมากกว่านี้กราฟของการแจกแจงจะเข้าสู่การแจกแจงแบบปกติมากขึ้น

1.4 ความคลาดเคลื่อนมีการแจกแจงลอกนอร์มอล (lognormal distribution)

ฟังก์ชันความหนาแน่นของ X อยู่ในรูปของ

$$f(x) = \frac{1}{\sigma x\sqrt{2\pi}}\exp\left\{-\frac{1}{2\sigma^2}(\ln x-\mu)^2\right\} \quad , \quad x>0$$

เมื่อ μ คือพารามิเตอร์ค่าเฉลี่ยซึ่งมีค่าเท่ากับ 0

และ σ^2 คือพารามิเตอร์ส่วนความแปรปรวนซึ่งมีค่าเท่ากับ 0.05, 0.3 และ 0.7 ถ้าค่าความแปรปรวนมีค่าลดลงไปมากกว่านี้กราฟของการแจกแจงจะเข้าสู่การแจกแจงปกติมากขึ้น (เจษฎาพร ยุทธนวิบูลย์ชัย, 2533: 7)

2. <u>จำนวนตัวแปรอิสระ</u>

จำนวนตัวแปรอิสระที่ศึกษามี 2 ระดับ คือ 3 ตัว และ 5 ตัว

3. <u>ขนาดตัวอย่าง</u>

ขนาดตัวอย่างที่ศึกษามี 4 ระดับคือ 10, 30, 50 และ 100

4. <u>การกำหนดค่าสัมประสิทธิ์การถดถอยพหุคูณ ($\underset{\sim}{\beta}$)</u>

การกำหนดค่าสัมประสิทธิ์การถดถอยพหุคูณ ($\underset{\sim}{\beta}$) เพื่อสร้างค่า $\underset{\sim}{y}$ ขึ้นจากตัวแบบ $\underset{\sim}{y} = X\underset{\sim}{\beta}+\underset{\sim}{\varepsilon}$ ผู้วิจัยกำหนดค่า $\underset{\sim}{\beta}$ จากเวกเตอร์เฉพาะที่สอดคล้องกับค่าเฉพาะที่มีค่ามากที่สุดของเมทริกซ์$(X'X)$ เพราะจะให้ค่ากำลังสองของความแตกต่างระหว่าง $\hat{\underset{\sim}{\beta}}$ และ $\underset{\sim}{\beta}$ มีค่าน้อยที่สุด

5. <u>ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ</u>

ระดับความสัมพันธ์ระหว่างตัวแปรอิสระแต่ละตัวจะคำนวณจากสูตร $\rho_{ij} = \rho^{|i-j|}$ เมื่อ i และ j เป็นตำแหน่งของตัวแปรอิสระ โดยกำหนดให้ค่าสัมประสิทธิ์สหสัมพันธ์ (ρ) เริ่มต้นมีค่าเท่ากับ 0.1 ,0.3, 0.5 ,0.7, 0.9 และ 0.99 ซึ่งผู้วิจัยได้แบ่งระดับสัมประสิทธิ์สหสัมพันธ์เป็น 4 ระดับดังนี้

- ระดับต่ำที่ $\rho = 0.1$ และ 0.3
- ระดับปานกลางที่ $\rho = 0.5$
- ระดับสูงที่ $\rho = 0.7$ และ 0.9
- ระดับสูงมากที่ $\rho = 0.99$

ผู้วิจัยได้ใช้วิธีการจำลองของ วิเชิน(Wichern) และ เชอร์ชิลล์ (Churchill) ทำให้สามารถสร้างตัวแปรอิสระที่มีความสัมพันธ์กันในระดับต่าง ๆ ดังนี้

1. <u>กรณีที่ตัวแปรอิสระเท่ากับ 3 ตัว</u>

สูตรที่ใช้ในการทำการวิจัยคือ

$$x_{ij} = (1-\alpha_1^2)^{\frac{1}{2}} Z_{ij} + \alpha_2 Z_{i4} \qquad ; i = 1,\ldots,n \quad j = 1,\ldots,3$$

เมื่อ Z_{ij} เป็นตัวแปรอิสระที่สร้างจากการแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับศูนย์และค่าความแปรปรวนเท่ากับหนึ่ง

α_1 เป็นความสัมพันธ์ระหว่างตัวแปร (x_1กับ x_2) และ (x_2กับ x_3)

α_2 เป็นความสัมพันธ์ระหว่างตัวแปร (x_1กับ x_3)

และ n เป็นขนาดตัวอย่าง

2. <u>กรณีที่ตัวแปรอิสระเท่ากับ 5 ตัว</u>

สูตรที่ใช้ในการทำการวิจัยคือ

$$x_{ij} = (1-\alpha_1^2)^{\frac{1}{2}} Z_{ij} + \alpha_2 Z_{i6} \quad ; i = 1,\dots,n \quad j = 1,2,3$$

$$x_{ij} = (1-\alpha_3^2)^{\frac{1}{2}} Z_{ij} + \alpha_4 Z_{i6} \quad ; i = 1,2,\dots,n \quad j = 4,5$$

เมื่อ Z_{ij} เป็นตัวแปรอิสระที่สร้างจากการแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับศูนย์และค่าความแปรปรวนเท่ากับหนึ่ง

α_1 เป็นความสัมพันธ์ระหว่างตัวแปร (x_1กับ x_2),(x_2กับ x_3),(x_3กับ x_4) และ (x_4กับ x_5)

α_2 เป็นความสัมพันธ์ระหว่างตัวแปร (x_1กับ x_3),(x_2กับ x_4) และ (x_3กับ x_5)

α_3 เป็นความสัมพันธ์ระหว่างตัวแปร (x_1กับ x_4) และ (x_2กับ x_5)

α_4 เป็นความสัมพันธ์ระหว่างตัวแปร (x_1กับ x_5)

และ n เป็นขนาดตัวอย่าง

(Wichern and Churchill, 1978: 304)

โดยมีหลักเกณฑ์ในการพิจารณาจากที่ตัวแปรอิสระแต่ละตัวไม่สามารถมีความสัมพันธ์เท่ากันได้ทุกตัวดังนั้นในการวิจัยครั้งนี้กำหนดให้ตัวแปรอิสระที่อยู่ใกล้กันมีค่าสัมประสิทธิ์สหสัมพันธ์กันมากและตัวแปรอิสระที่อยู่ห่างกันมีค่าสัมประสิทธิ์สหสัมพันธ์น้อยลงไปเรื่อย ๆเพราะตัวแปรอิสระที่อยู่ใกล้กันอาจมีความเกี่ยวข้องมากและตัวแปรที่อยู่ห่างกันอาจจะมีความเกี่ยวข้องกันน้อย โดยพิจารณาในทุกกรณีที่สามารถเกิดขึ้นได้ ซึ่งแบ่งเป็น 2 กรณี ดังนี้

1. <u>กรณีที่มีตัวแปรอิสระเท่ากับ 3 ตัว</u>

ค่าสัมประสิทธิ์สหสัมพันธ์เริ่มต้นระหว่าง

(x_1กับ x_2) และ (x_2กับ x_3) = 0.1,0.3,0.5,0.7,0.9 และ 0.99*

ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่าง

(x_1กับ x_3) = 0.01,0.09,0.25,0.49,0.81 และ 0.9801

---

* เมื่อตัวแปรอิสระเท่ากับ 3 ตัว ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่าง (x_1กับ x_2) , (x_2กับ x_3) = 0.1 และ (x_1กับ x_3) = $(0.1)^2$ = 0.01 เนื่องจากตัวแปรอิสระที่มี subscript ใกล้กันจะมีความสัมพันธ์กันมากและตัวแปรอิสระที่มี subscript ห่างกันมากขึ้นจะมีความสัมพันธ์กันน้อยลง โดยค่าสัมประสิทธิ์สหสัมพันธ์เริ่มต้นจะเพิ่มขึ้นไปเป็น 0.3,0.5,0.7,0.9 และ 0.99 ตามลำดับ

2. <u>กรณีที่มีตัวแปรอิสระเท่ากับ 5 ตัว</u>

ค่าสัมประสิทธิ์สหสัมพันธ์เริ่มต้นระหว่าง

(x_1กับ x_2) , (x_2กับ x_3) , (x_3กับ x_4) และ (x_4กับ x_5) = 0.1,0.3,0.5,0.7,0.9 และ 0.99*

ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่าง

(x_1กับ x_3) , (x_2กับ x_4) และ (x_3กับ x_5) = 0.01,0.09,0.25,0.49,0.81 และ 0.9801

ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่าง

(x_1กับ x_4) และ (x_2กับ x_5) = 0.001,0.027,0.125,0.343,0.729 และ 0.9701

ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่าง

(x_1กับ x_5) = 0.0001,0.0081,0.0625,0.2401,0.6561 และ 0.9650

วิธีดำเนินการวิจัย

1. สร้างข้อมูลของตัวแปรอิสระโดยใช้เทคนิคมอนติคาร์โลตามระดับความสัมพันธ์ต่าง ๆ ที่กำหนดไว้ในขอบเขตการวิจัย

2. สร้างข้อมูลของตัวแปรตามจากตัวแปรอิสระที่สร้างขึ้นที่สัมประสิทธิ์สหสัมพันธ์ที่กำหนดขึ้นมา และค่าความคลาดเคลื่อนตามการแจกแจงต่าง ๆที่กำหนดไว้

3. ประมาณค่าพารามิเตอร์ตามวิธีที่กำหนดไว้

4. คำนวณค่าเฉลี่ยของค่าความเฉลี่ยความคลาดเคลื่อนยกกำลังสองโดยพิจารณาเปรียบ-เทียบแต่ละวิธี

5. ทำการเปรียบเทียบค่าต่าง ๆของวิธีที่กำหนดไว้และสรุปผล

ประโยชน์ที่คาดว่าจะได้รับ

1. เพื่อให้ทราบวิธีที่เหมาะสมสำหรับใช้ในการประมาณสัมประสิทธิ์ของสมการถดถอยเมื่อเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ

2. สามารถนำเอาวิธีที่เหมาะสมไปประยุกต์กับเหตุการณ์จริง

---

* เมื่อตัวแปรอิสระเท่ากับ 5 ตัว ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่าง (x_1กับ x_2) , (x_2กับ x_3) , (x_3กับ x_4) และ (x_4กับ x_5) = 0.1 , (x_1กับ x_3) , (x_2กับ x_4) และ (x_3กับ x_5) = $(0.1)^2$ = 0.01 , (x_1กับ x_4) และ (x_2กับ x_5) = $(0.1)^3$ = 0.001 และ (x_1กับ x_5) = $(0.1)^4$ = 0.0001 เนื่องจากตัวแปรอิสระที่มี subscript ใกล้กันจะมีความสัมพันธ์กันมากและตัวแปรอิสระที่มี subscript ห่างกันมากขึ้นจะมีความสัมพันธ์กันน้อยลง โดยค่าสัมประสิทธิ์สหสัมพันธ์เริ่มต้นจะเพิ่มขึ้นไปเป็น 0.3,0.5,0.7,0.9 และ 0.99 ตามลำดับ