บทที่ 1

บทนำ

ที่มาของปัญหาและความสำคัญของปัญหา

การค้นคว้าทดลองทางด้านวิทยาศาสตร์ในสาขาต่างๆ เช่น การเกษตร, ฟิสิกส์, ชีววิทยา เป็นต้น โดยทั่วๆ ไปการทดลองเหล่านี้มักจะมีการทำซ้ำหรือมีจำนวนซ้ำในแต่ละระดับของปัจจัยที่ทำการศึกษา ซึ่งปัจจัยเหล่านี้อาจจะเป็นปัจจัยเชิงคุณภาพ(qualitative factor) หรือปัจจัยเชิงปริมาณ(quantitative factor) การทดลองบางการทดลองต้องการศึกษารูปแบบความสัมพันธ์ของผลตอบสนอง (response)และปัจจัยบางปัจจัยที่ทำการศึกษาซึ่งเป็นปัจจัยเชิงปริมาณ เพื่อนำรูปแบบความสัมพันธ์ที่ได้มาใช้ในการประมาณค่าผลตอบสนอง ณ ระดับของปัจจัยนั้นๆ ซึ่งตัวแบบทางสถิติที่ใช้ในการวิเคราะห์ข้อมูล ก็คือ ตัวแบบการถดถอยเชิงเส้น (linear regression model) ซึ่งมีรูปแบบ

$$y_{ij} = \underset{\sim}{x}_i' \underset{\sim}{\beta} + e_{ij}, \qquad j = 1,2,\ldots,n_i \ , \ i = 1,2,\ldots,k$$

โดยที่ y_{ij} คือ ผลตอบสนองที่ j สอดคล้องกับ $\underset{\sim}{x}_i$

$\underset{\sim}{x}_i$ คือ เวกเตอร์ค่าคงที่ของปัจจัยที่ทำการศึกษา

$\underset{\sim}{\beta}$ คือ เวกเตอร์พารามิเตอร์สัมประสิทธิ์การถดถอย

e_{ij} คือ ค่าความคลาดเคลื่อน

ข้อกำหนดเบื้องต้น (assumption) ของตัวแบบข้างต้น คือ ความคลาดเคลื่อนมีค่าเฉลี่ยเท่ากับศูนย์ และความแปรปรวนคงที่ เท่ากับ σ^2 รวมทั้งความคลาดเคลื่อนภายในและระหว่างระดับของตัวแปรไม่สัมพันธ์กัน

แต่ในทางปฏิบัติข้อมูลที่เก็บรวบรวมได้นั้นอาจไม่เป็นไปตามข้อกำหนดเบื้องต้น ซึ่งทำให้การอนุมานเชิงสถิติ คือ การประมาณค่าและการทดสอบสมมุติฐานต่างๆ เกี่ยวกับพารามิเตอร์ให้ผลที่ไม่ถูกต้องและมีความเชื่อถือได้ต่ำ ดังนั้นหลังจาก

กำหนดรูปแบบความสัมพันธ์ของข้อมูลโดยใช้ตัวแบบการถดถอยเชิงเส้นแล้ว ควรทำการตรวจสอบเกี่ยวกับข้อกำหนดเบื้องต้นของตัวแบบด้วย ถ้าข้อมูลไม่เป็นไปตามข้อกำหนดเบื้องต้นอาจต้องทำการแก้ไขตามวิธีการทางสถิติ เช่น การแปลงข้อมูล (transformation), การถ่วงน้ำหนัก เป็นต้น เพื่อให้ข้อมูลที่นำมาวิเคราะห์เป็นไปตามข้อกำหนดเบื้องต้น

ปัญหาหนึ่งของตัวแบบข้างต้นที่อาจเกิดขึ้นได้บ่อยๆ ในทางปฏิบัติ คือ ความแปรปรวนของตัวแปรตาม y_{ij} (dependent variable) ณ ระดับคงที่ใดๆ ของตัวแปรอิสระ $\underset{-i}{x}$ (independent variable) ไม่คงที่ ซึ่งมีผลต่อการประมาณค่าและการทดสอบสมมุติฐานเกี่ยวกับพารามิเตอร์ โดยปกติการประมาณค่าสัมประสิทธิ์การถดถอย (regression coefficient) นิยมใช้วิธีกำลังสองน้อยที่สุด(least square method) ถ้าข้อมูลเป็นไปตามข้อกำหนดเบื้องต้นตัวประมาณที่ใช้ คือ ตัวประมาณกำลังสองน้อยที่สุดแบบสามัญ(ordinary least square estimator) แต่เมื่อเกิดปัญหาความแปรปรวนไม่คงที่ ตัวประมาณที่นิยมใช้แทนตัวประมาณกำลังสองน้อยที่สุดแบบสามัญ คือ ตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนัก(weighted least square estimator) เนื่องจากตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักจะมีประสิทธิภาพสูงกว่าตัวประมาณกำลังสองน้อยที่สุดแบบสามัญ ส่วนการทดสอบสมมุติฐานเกี่ยวกับสัมประสิทธิ์การถดถอยนั้นมีวิธีหนึ่งที่สามารถแก้ปัญหาได้ ซึ่งจะทำให้การทดสอบนั้นมีความแกร่ง(robustness) คือ การหาตัวประมาณของความแปรปรวนที่คงเส้นคงวา(consistent variance estimator) ของสัมประสิทธิ์การถดถอย

ในปี ค.ศ. 1989 Jun Shao ได้เสนอการแจกแจงเมื่อใกล้อนันต์(asymptotic distribution) ของตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนัก ตัวสถิติทดสอบที่สอดคล้องกับวิธีหาตัวประมาณของความแปรปรวนของตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนัก เมื่อความคลาดเคลื่อนภายในและระหว่างระดับของตัวแปรอิสระไม่มีความสัมพันธ์กัน คือ t_w และเรียกวิธีนี้ว่า วิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนัก(Weighted Least square Method) และในปีเดียวกัน Jun Shao ยังได้เสนอวิธีปรับแก้แจคไนฟ์แบบตัดทีละค่าสังเกต(Modified Delete-one Jackknife Method) เพื่อหาตัวประมาณของความแปรปรวนของตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนัก เมื่อความคลาดเคลื่อนภายในและระหว่างระดับของตัวแปรอิสระไม่มีความสัมพันธ์กัน ตัวสถิติทดสอบที่สอดคล้องกับวิธีนี้ คือ t_M และในปี ค.ศ. 1993

Jun Shao และ J.N.K. Rao ได้เสนอวิธีแจคไนฟ์แบบตัดเป็นกลุ่ม(Delete-group Jaokknife Method) เพื่อหาตัวประมาณของความแปรปรวนของตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนัก เมื่อความคลาดเคลื่อนภายในระดับของตัวแปรอิสระมีความสัมพันธ์กัน ตัวสถิติทดสอบที่สอดคล้องกับวิธีนี้ คือ t_J

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบความสามารถในการควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบของตัวสถิติทดสอบทั้ง 3 ตัว ได้แก่

1) ตัวสถิติทดสอบ t_W

2) ตัวสถิติทดสอบ t_M

3) ตัวสถิติทดสอบ t_J

ตัวสถิติทดสอบทั้ง 3 ตัวนี้ขึ้นอยู่กับวิธีการหาตัวประมาณที่คงเส้นคงวาของความแปรปรวนของตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนัก

วัตถุประสงค์ของการวิจัย

เพื่อเปรียบเทียบอำนาจการทดสอบของตัวสถิติทดสอบ t_W, t_M และ t_J เมื่อความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ ทั้งกรณีความคลาดเคลื่อนภายในระดับของตัวแปรอิสระไม่มีความสัมพันธ์กันและมีความสัมพันธ์กัน

สมมุติฐานของการวิจัย

1) อำนาจการทดสอบของตัวสถิติทดสอบ t_W สูงสุด เมื่อความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ กรณีความคลาดเคลื่อนภายในระดับของตัวแปรอิสระไม่มีความสัมพันธ์กันและมีความสัมพันธ์กันน้อย

2) อำนาจการทดสอบของตัวสถิติทดสอบ t_J สูงสุด เมื่อความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ กรณีความคลาดเคลื่อนภายในระดับของตัวแปรอิสระมีความสัมพันธ์กันปานกลางและมาก

ข้อตกลงเบื้องต้น

1) ข้อมูลที่ศึกษามีความสัมพันธ์กันเชิงเส้น โดยตัวแบบการถดถอยเชิงเส้นที่ใช้ในการศึกษาครั้งนี้ คือ

$$y_{ij} = \underset{\sim}{x}_{i}' \underset{\sim}{\beta} + e_{ij}, \qquad j = 1,2,\ldots,n_i \ , i = 1,2,\ldots,k$$

โดยที่ y_{ij} เป็นค่าสังเกตุที่ j จากระดับที่ i

$\underset{\sim}{x}_{i}$ เป็นเวกเตอร์ของค่าคงที่ขนาด px1

$\underset{\sim}{\beta}$ เป็นเวกเตอร์พารามิเตอร์ขนาด px1

e_{ij} เป็นความคลาดเคลื่อนสุ่มของค่าสังเกตที่ j จากระดับที่ i

k เป็นจำนวนระดับของตัวแปรอิสระ

n_i เป็นจำนวนซ้ำในระดับที่ i

จากตัวแบบข้างต้นสามารถเขียนให้อยู่ในรูปเมทริกซ์ได้ดังนี้

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

โดยที่ $N = \sum_{i=1}^{k} n_i$

$$\underset{\sim}{y} = \left(y_{11},\ldots,y_{in_1},y_{21},\ldots,y_{2n_2},\ldots,y_{k1},\ldots,y_{kn_k}\right)'_{N\times 1}$$

$$X = \left(\underset{\sim}{x}_{1},\ldots,\underset{\sim}{x}_{1},\underset{\sim}{x}_{2},\ldots,\underset{\sim}{x}_{2},\ldots,\underset{\sim}{x}_{k},\ldots,\underset{\sim}{x}_{k}\right)'_{N\times p}$$

$$\underset{\sim}{\beta} = \left(\beta_0,\beta_1,\ldots,\beta_{p-1}\right)'_{p\times 1}$$

$$\underset{\sim}{e} = \left(e_{11},\ldots,e_{in_1},e_{21},\ldots,e_{2n_2},\ldots,e_{k1},\ldots,e_{kn_k}\right)'_{N\times 1}$$

2) เมทริกซ์ความแปรปรวน-ความแปรปรวนร่วมของความคลาดเคลื่อนมีรูปแบบดังนี้

2.1) ภายในกลุ่มไม่มีความสัมพันธ์กัน

$$\underset{\sim}{e} \sim N\left(\underset{\sim}{0},\sum\right)$$

โดยที่

$$\sum = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 & & & & & & & & \\ 0 & \sigma_1^2 & \cdots & 0 & & & & & & & & \\ \vdots & \vdots & \ddots & \vdots & & & & & & & & \\ 0 & 0 & \cdots & \sigma_1^2 & & & & & & & & \\ & & & & \sigma_2^2 & 0 & \cdots & 0 & & & & \\ & & & & 0 & \sigma_2^2 & \cdots & 0 & & & & \\ & & & & \vdots & \vdots & \ddots & \vdots & & & & \\ & & & & 0 & 0 & \cdots & \sigma_2^2 & & & & \\ & & & & & & & & \ddots & & & \\ & & & & & & & & & \sigma_k^2 & 0 & \cdots & 0 \\ & & & & & & & & & 0 & \sigma_k^2 & \cdots & 0 \\ & & & & & & & & & \vdots & \vdots & \ddots & \vdots \\ & & & & & & & & & 0 & 0 & \cdots & \sigma_k^2 \end{bmatrix}_{N\times N}$$

2.2) ภายในกลุ่มมีความสัมพันธ์กัน

$$\underline{e} \sim N\left(0, \sum^*\right)$$

โดยที่

$$\sum^* = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1^2 & \cdots & \rho\sigma_1^2 & & & & & & & & \\ \rho\sigma_1^2 & \sigma_1^2 & \cdots & \rho\sigma_1^2 & & & & & & & & \\ \vdots & \vdots & \ddots & \vdots & & & & & & & & \\ \rho\sigma_1^2 & \rho\sigma_1^2 & \cdots & \sigma_1^2 & & & & & & & & \\ & & & & \sigma_2^2 & \rho\sigma_2^2 & \cdots & \rho\sigma_2^2 & & & & \\ & & & & \rho\sigma_2^2 & \sigma_2^2 & \cdots & \rho\sigma_2^2 & & & & \\ & & & & \vdots & \vdots & \ddots & \vdots & & & & \\ & & & & \rho\sigma_2^2 & \rho\sigma_2^2 & \cdots & \sigma_2^2 & & & & \\ & & & & & & & & \ddots & & & \\ & & & & & & & & & \sigma_k^2 & \rho\sigma_k^2 & \cdots & \rho\sigma_k^2 \\ & & & & & & & & & \rho\sigma_k^2 & \sigma_k^2 & \cdots & \rho\sigma_k^2 \\ & & & & & & & & & \vdots & \vdots & \ddots & \vdots \\ & & & & & & & & & \rho\sigma_k^2 & \rho\sigma_k^2 & \cdots & \sigma_k^2 \end{bmatrix}_{N\times N}$$

<u>หมายเหตุ</u> ถ้าค่าสัมประสิทธิ์สหสัมพันธ์ (ρ) มีค่าเท่ากับศูนย์แสดงว่าภายในระดับไม่มีความสัมพันธ์กัน นั่นคือ $\sum = \sum^*$

3) ตัวประมาณของพารามิเตอร์สัมประสิทธิ์การถดถอย $\underset{\sim}{\beta}$ ของตัวแบบการถดถอยเชิงเส้นข้างต้น คือ

3.1) ตัวประมาณกำลังสองน้อยที่สุดแบบสามัญ (Ordinary Least Square Estimator : $\underset{\sim}{b}$) คือ

$$\underset{\sim}{b} = (X'X)^{-1}X'\underset{\sim}{y}$$

3.2) ตัวประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนัก (Weighted Least Square Estimator : $\underset{\sim}{b}^{W}$) คือ

$$\underset{\sim}{b}^{W} = (X'WX)^{-1}X'W\underset{\sim}{y}$$

โดยที่ $W = \text{block diag}\left(w_1 I_{n_1}, w_2 I_{n_2}, \ldots, w_k I_{n_k}\right)$

และ w_i คือ ตัวถ่วงน้ำหนักซึ่งเป็นส่วนกลับของตัวประมาณความแปรปรวน ณ ระดับของตัวแปรอิสระที่ i , $w_i > 0$

4) ตัวถ่วงน้ำหนัก (Weight) ที่ใช้ในการวิจัยครั้งนี้เป็นส่วนกลับของตัวประมาณความแปรปรวน ซึ่งตัวประมาณความแปรปรวนนี้เป็นตัวประมาณที่ไม่เอนเอียงและมีค่าประจำน้อยที่สุดของรูปแบบกำลังสอง (Minimum Norm Quadratic Unbiased Estimator : MINQUE) คือ

$$w_i = u_i^{-1} \qquad ,i = 1,2,\ldots,k$$

โดยที่ $u_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\left(y_{ij} - \underset{\sim}{x}_{i}'\underset{\sim}{b}\right)^2$

5) กำหนดให้ θ เป็นฟังก์ชันเชิงเส้นของ $\underset{\sim}{\beta}$ สามารถเขียนอยู่ในรูปของ $\theta = g\left(\underset{\sim}{\beta}\right)$ โดยที่ $g: R^p \rightarrow R$ สำหรับการวิจัยครั้งนี้ $\theta = \underset{\sim}{\ell}'\underset{\sim}{\beta}$, $\underset{\sim}{\ell}$ เป็นเวกเตอร์ใดๆ ที่มีสมาชิกเป็น 0 และ 1

6) สำหรับการวิจัยครั้งนี้ เกณฑ์ในการเลือกตัวสถิติทดสอบ คือ ความสามารถในการควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 (Type I error) และ อำนาจการทดสอบ (power of a test)

ขอบเขตการวิจัย

1) สมมุติฐานที่ทำการทดสอบ คือ

$$H_0 : \beta_1 = 0$$

$$H_1 : \beta_1 \neq 0$$

2) จำนวนตัวแปรอิสระ คือ 1 และ 3 ตัว

3) จำนวนระดับของตัวแปรอิสระ คือ 6, 9 และ 15 ระดับ

4) ค่าสัมประสิทธิ์สหสัมพันธ์ คือ 0.0, 0.1, 0.2, 0.3, 0.4, 0.5 และ 0.8

5) ระดับนัยสำคัญ คือ 0.10, 0.05, และ 0.01

6) จำนวนซ้ำในแต่ละระดับของตัวแปรอิสระ

6.1) กรณีทุกระดับมีจำนวนซ้ำเท่ากัน $(n_i = r)$ คือ 3, 5 และ 10

6.2) กรณีแต่ละระดับมีจำนวนซ้ำไม่เท่ากันพิจารณาจากตัวแบบของ Levy โดยกำหนดรูปแบบของจำนวนซ้ำเป็น $(n_i, \bar{n})$ ได้แก่ (3,5), (3,10) และ (7,10)

<u>หมายเหตุ</u> แบ่งกลุ่มทั้งหมดออกเป็น 3 ส่วนอย่างสุ่ม โดยแต่ละระดับในส่วนแรกมีจำนวนซ้ำเท่ากับ n_i แต่ละระดับในส่วนที่สองมีจำนวนซ้ำเท่ากับ $\bar{n}$ และแต่ละระดับในส่วนที่สามมีจำนวนซ้ำเท่ากับ $2\bar{n} - n_i$ โดยที่ $\bar{n} = \frac{1}{k}\sum_{i=1}^{k} n_i$

7)* อัตราส่วนของความแปรปรวนของความคลาดเคลื่อนแบ่งเป็น 3 กลุ่ม ได้แก่

7.1) อัตราส่วนที่แตกต่างกันน้อย คือ 0.9:1.0:1.1, 0.5:1.0:2.0 และ 0.3:1.0:1.7

---

* ϕ เป็นตัวแปรที่ใช้เพื่อแสดงค่าความแตกต่างของความแปรปรวนของความคลาดเคลื่อน โดยที่

$$\phi^2 = \frac{\frac{1}{s}\sum_{i=1}^{s}(\sigma_i^2 - \sigma^2)^2}{\sigma^2}$$

และ $\sigma^2 = \left(\prod_{i=1}^{s}\sigma_i^2\right)^{\frac{1}{s}}$, s คือ จำนวนชั้นในอัตราส่วนที่กำหนด

, σ^2 คือ ค่าเฉลี่ยเรขาคณิตของความแปรปรวน

ถ้า $\phi \in (0,1.5)$ แล้วจะได้ว่า อัตราส่วนของความแปรปรวนแตกต่างกันน้อย

$\phi \in [1.5,3.0)$ แล้วจะได้ว่า อัตราส่วนของความแปรปรวนแตกต่างกันปานกลาง

$\phi \in [3.0,\infty)$ แล้วจะได้ว่า อัตราส่วนของความแปรปรวนแตกต่างกันมาก

7.2) อัตราส่วนที่แตกต่างกันปานกลาง คือ 0.25:1.0:2.0, 0.1:1.0:1.5 และ 0.05:1.0:1.3

7.3) อัตราส่วนที่แตกต่างกันมาก คือ 0.25:1.0:4.0, 0.125:1.0:8.0 และ 0.0625:1.0:16.0

8) การวิจัยครั้งนี้ได้จำลองข้อมูลให้เป็นไปตามสถานการณ์ที่กำหนดไว้ข้างต้นโดยใช้เทคนิคการจำลองข้อมูลแบบมอนติคาร์โล (Monte Carlo Simulation Technique) และทำการทดลองซ้ำ 500 ครั้งในแต่ละสถานการณ์

เกณฑ์การตัดสินใจ

1) พิจารณาความสามารถในการควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 โดยใช้เกณฑ์ของ Bradley ซึ่งจะพิจารณาว่า ถ้าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 จากการทดลองอยู่ในช่วง (0.050,0.150), (0.025,0.075) และ (0.005,0.015) ณ ระดับ นัยสำคัญ 0.10, 0.05 และ 0.01 ตามลำดับ จะถือว่าการทดสอบนั้นสามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ได้

2) พิจารณาอำนาจการทดสอบเฉพาะตัวสถิติทดสอบที่สามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ได้เท่านั้น

คำจำกัดความ

1) ความคลาดเคลื่อนประเภทที่ 1 (Type I error) คือ ความผิดพลาดที่เกิดจากการปฏิเสธสมมุติฐานว่าง (null hypothesis) เมื่อสมมุติฐานว่างนั้นเป็นจริง

2) ความคลาดเคลื่อนประเภทที่ 2 (Type II error) คือ ความผิดพลาดที่เกิดจากการยอมรับสมมุติฐานว่าง (null hypothesis) เมื่อสมมุติฐานว่างนั้นไม่จริง

3) อำนาจการทดสอบ (power of a test) คือ ค่าความน่าจะเป็นที่จะปฏิเสธสมมุติฐานว่าง (null hypothesis) เมื่อสมมุติฐานว่างนั้นไม่จริง

4) ความแกร่ง (robustness) ของการทดสอบ หมายถึง คุณสมบัติของการทดสอบที่ไม่มีความไว(sensitive) ต่อการเปลี่ยนแปลงของปัจจัยอื่นที่ไม่ใช่ปัจจัยที่

ต้องการศึกษา และความแกร่งของการทดสอบจะพิจารณาที่ความสามารถในการควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1

ประโยชน์ที่คาดว่าจะได้รับ

เพื่อเป็นแนวทางในการศึกษาและเลือกใช้วิธีการแก้ปัญหาในการทดสอบสมมุติฐานเกี่ยวกับสัมประสิทธิ์การถดถอยในตัวแบบการถดถอยเชิงเส้น เมื่อความคลาดเคลื่อนมีความแปรปรวนไม่คงที่และมีความสัมพันธ์กันภายในระดับของตัวแปรอิสระ เพื่อให้ได้ตัวแบบที่เหมาะสมในการประมาณค่าและอธิบายความสัมพันธ์