บทที่ 2

ทฤษฎีและตัวสถิติที่เกี่ยวข้อง

ในการวิจัยครั้งนี้เป็นการเปรียบเทียบอำนาจการทดสอบของการทดสอบเทียบความกลมกลืนในการวิเคราะห์ความถดถอยของสถิติทดสอบ 3 ตัว ได้แก่ สถิติทดสอบเอฟ (F) สถิติทดสอบ Adaptive Neyman (AN) และสถิติทดสอบ Kuchibhatla และ Hart (KH) ซึ่งบทนี้จะกล่าวถึงรายละเอียดของสถิติทดสอบและสถิติที่ใช้ในการศึกษา

2.1 การวิเคราะห์ความถดถอย

การวิเคราะห์ความถดถอย (Regression Analysis) เป็นการศึกษาเกี่ยวกับฟังก์ชันหรือรูปแบบความสัมพันธ์ทางคณิตศาสตร์เพื่อใช้ทำนายค่าของตัวแปรที่ต้องการศึกษา ซึ่งความสัมพันธ์นี้จะประกอบด้วยตัวแปร 2 กลุ่ม ตัวแปรที่ต้องการศึกษาเรียกว่าตัวแปรตาม มักแทนด้วย Y การที่จะทำนายค่าของ Y ได้นั้นต้องอาศัยความรู้เกี่ยวกับค่าของตัวแปรอีกตัวหนึ่งหรือมากกว่า เรียกว่าตัวแปรอิสระ มักแทนด้วย X รูปแบบความสัมพันธ์เชิงคณิตศาสตร์ที่ได้เรียกว่า “ตัวแบบการถดถอย" (Regression Model) หรือ “สมการการถดถอย"(Regression Equation) เราสามารถเขียนตัวแบบการถดถอยในเทอมของเมทริกซ์ได้ โดยกำหนดเมทริกซ์ดังนี้

$$\underset{\sim}{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad X = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1,p-1} \\ 1 & X_{21} & X_{22} & \cdots & X_{2,p-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{n,p-2} \end{bmatrix}$$

$$\underset{\sim}{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix} \qquad \text{และ} \qquad \underset{\sim}{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

ดังนั้นตัวแบบการถดถอยคือ

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

เมื่อ $\underset{\sim}{Y}$ คือเวกเตอร์ที่มีขนาด n x 1

X คือเมทริกซ์ที่มีขนาด n x p โดยที่ p คือจำนวนของพารามิเตอร์

$\underset{\sim}{\beta}$ คือเวกเตอร์ของพารามิเตอร์สัมประสิทธิ์การถดถอยที่มีขนาด p x 1

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

โดยที่ $\underset{\sim}{\varepsilon} \sim N_n(\underset{\sim}{0}, \sigma^2 I_n)$, $E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0}$ และเมทริกซ์ความแปรปรวนร่วม (Covariance Matrix) ของ $\underset{\sim}{\varepsilon}$ คือ

$$Cov(\underset{\sim}{\varepsilon}) = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \sigma^2 \end{bmatrix} = \sigma^2 I_n$$

และเมทริกซ์ความแปรปรวนร่วมของ $\underset{\sim}{Y}$ จะมีค่าเหมือนกับ $\underset{\sim}{\varepsilon}$

การประมาณค่าสัมประสิทธิ์การถดถอย (Regression Coefficient) $\beta_0, \beta_1, ..., \beta_{p-1}$ โดยใช้ข้อมูลจากตัวแปรอิสระและตัวแปรตาม วิธีที่นิยมมากที่สุดได้แก่วิธีกำลังสองน้อยที่สุด(Method of Least Squares) เพราะค่าประมาณ $b_0, b_1, ..., b_{p-1}$ ของพารามิเตอร์ $\beta_0, \beta_1, ..., \beta_{p-1}$ จะทำให้ค่าความคลาดเคลื่อนจากการประมาณมีค่าเฉลี่ยเท่ากับ 0 และผลรวมกำลังสองของความคลาดเคลื่อน (Error Sum of Squares : SSE) มีค่าต่ำที่สุด ให้ $\hat{\underset{\sim}{Y}}$ เป็นเวกเตอร์ค่าประมาณของตัวแปรตาม

เวกเตอร์ค่าประมาณของตัวแปรตามกำหนดโดย

$$\hat{Y} = X \underset{\sim}{\beta}$$

เวกเตอร์ค่าคลาดเคลื่อนคือ

$$\begin{aligned} \underset{\sim}{\varepsilon} &= \underset{\sim}{Y} - \hat{\underset{\sim}{Y}} \\ &= \underset{\sim}{Y} - X \underset{\sim}{\beta} \end{aligned}$$

จากผลบวกกำลังสองของค่าคลาดเคลื่อน

$$\begin{aligned} SSE &= \underset{\sim}{\varepsilon}' \underset{\sim}{\varepsilon} \\ &= (\underset{\sim}{Y} - X \underset{\sim}{\beta})'(\underset{\sim}{Y} - X \underset{\sim}{\beta}) \\ &= \underset{\sim}{Y}' \underset{\sim}{Y} - \underset{\sim}{\beta}' X' \underset{\sim}{Y} - \underset{\sim}{Y}' X \underset{\sim}{\beta} + \underset{\sim}{\beta}' X'X \underset{\sim}{\beta} \end{aligned}$$

เนื่องจาก $(X \underset{\sim}{\beta})' = \underset{\sim}{\beta}' X'$ ดังนั้น $(\underset{\sim}{Y}' X \underset{\sim}{\beta})' = \underset{\sim}{\beta}' X' \underset{\sim}{Y}$ จะได้ว่า

$$SSE = \underset{\sim}{Y}' \underset{\sim}{Y} - 2 \underset{\sim}{\beta}' X' \underset{\sim}{Y} + \underset{\sim}{\beta}' X'X \underset{\sim}{\beta}$$

SSE จะมีค่าต่ำสุดเมื่อ $\dfrac{\partial}{\partial \underset{\sim}{\beta}}(SSE) = 0$

$$\therefore \quad \frac{\partial}{\partial \underset{\sim}{\beta}}(SSE) = -2X' \underset{\sim}{Y} + 2X'X \underset{\sim}{\beta} = 0$$

$$X'X \underset{\sim}{\beta} = X' \underset{\sim}{Y}$$

เราเรียกสมการข้างบนนี้ว่าสมการปกติ ดังนั้นเมื่อแทน $\underset{\sim}{\beta}$ ด้วย $\underset{\sim}{b}$ จะได้

$$X'X\underset{\sim}{b} = X'\underset{\sim}{Y}$$

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

2.2 ตัวแบบที่ใช้ในการวิจัย

การวิจัยครั้งนี้จะทำการศึกษาทดสอบสมมติฐานว่างว่าตัวแบบที่ศึกษานั้นเป็นตัวแบบเชิงเส้นในตัวแปรอิสระหรือไม่ โดยพิจารณาตัวแบบ 8 ตัวแบบดังนี้

<u>ตัวแบบสมมติฐานว่างกรณีตัวแปรอิสระ 1 ตัว</u> (H_0)

<u>ตัวแบบที่ 1.</u> ตัวแบบการถดถอยเป็นตัวแบบเชิงเส้นที่มีตัวแปรอิสระ 1 ตัว

$Y_i = 1 + 4X_{i1} + \varepsilon_i$ เมื่อ $1 \le i \le n$

<u>ตัวแบบสมมติฐานแย้งกรณีตัวแปรอิสระ 1 ตัว</u> (H_1)

<u>ตัวแบบที่ 2.</u> ตัวแบบการถดถอยเป็น

$$Y_i = 1 + \beta_1 X_{i1}^2 + \varepsilon_i$$

เมื่อ β_1 มีค่าเท่ากับ 0.2 , 0.4 , 0.6 , 0.8 และ 1.0 , $1 \le i \le n$

<u>ตัวแบบที่ 3.</u> ตัวแบบการถดถอยเป็น

$$Y_i = 1 + \cos(\beta_1 X_{i1}\pi) + \varepsilon_i$$

เมื่อ β_1 มีค่าเท่ากับ 0.2 , 0.4 , 0.6 , 0.8 และ 1.0 , $1 \le i \le n$

<u>ตัวแบบที่ 4.</u> ตัวแบบการถดถอยเป็น

$$Y_i = 1 + 4X_{i1} + \beta_2 X_{i1}^2 + \varepsilon_i$$

เมื่อ β_2 มีค่าเท่ากับ 0.2 , 0.4 , 0.6 , 0.8 และ 1.0 , $1 \le i \le n$

<u>ตัวแบบสมมติฐานว่างกรณีตัวแปรอิสระ 3 ตัว</u> (H_0)

<u>ตัวแบบที่ 5.</u> ตัวแบบการถดถอยเป็นตัวแบบเชิงเส้นพหุคูณที่มีตัวแปรอิสระสามตัว

$Y_i = 3 + X_{i1} + 4X_{i2} + 2X_{i3} + \varepsilon_i$ เมื่อ $1 \le i \le n$

<u>ตัวแบบสมมติฐานแย้งกรณีตัวแปรอิสระ 3 ตัว</u> (H_1)

<u>ตัวแบบที่ 6.</u> ตัวแบบพหุนามระดับขั้นเป็น 2

$$Y_i = 3 + X_{i1} + \beta_2 X_{i2}^2 + 2X_{i3} + \varepsilon_i$$

เมื่อ β_2 มีค่าเท่ากับ 0.2 , 0.4 , 0.6 , 0.8 และ 1.0 , $1 \le i \le n$

<u>ตัวแบบที่ 7.</u> ตัวแบบการถดถอยเป็น

$$Y_i = 3 + X_{i1} + \cos(\beta_2 X_{i2}\pi) + 2X_{i3} + \varepsilon_i$$

เมื่อ β_2 มีค่าเท่ากับ 0.2 , 0.4 , 0.6 , 0.8 และ 1.0 , $1 \le i \le n$

<u>ตัวแบบที่ 8.</u> ตัวแบบการถดถอยเป็น

$$Y_i = 3 + X_{i1} + \beta_2 X_{i1} X_{i2} + 2X_{i3} + \varepsilon_i$$

เมื่อ β_2 มีค่าเท่ากับ 0.2 , 0.4 , 0.6 , 0.8 และ 1.0 , $1 \le i \le n$

2.3 การแจกแจงที่เกี่ยวข้อง

การผลิตตัวแปรสุ่มที่ใช้ในการวิจัยเพื่อให้สอดคล้องกับตัวแบบข้างต้น จำเป็นจะต้องผลิตตัวแปรสุ่ม 2 กลุ่มได้แก่ ตัวแปรอิสระ และความคลาดเคลื่อน ซึ่งจะผลิตมาจากการแจกแจงปกติ การผลิตตัวแปรนั้นจะผลิตจากเลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอในช่วง (0,1) ซึ่งมีรายละเอียดของการแจกแจง การผลิตเลขสุ่ม และการผลิตตัวแปรสุ่มมีดังนี้

2.3.1 การแจกแจงแบบสม่ำเสมอ (Uniform Distribution)[1]

ตัวแปรสุ่ม X มีการแจกแจงแบบสม่ำเสมอในช่วง [a , b] เขียนแทนด้วยสัญลักษณ์ $X \sim U(a,b)$, $-\infty < a < b < \infty$ ถ้า X มีฟังก์ชันความหนาแน่น

$$f(x) = \frac{1}{b-a}$$

สำหรับ $X \sim U(a,b)$ จะได้ว่า X มีค่าเฉลี่ย E(X) และความแปรปรวน Var(X) ดังนี้

$$E(X) = \frac{a+b}{2}$$

$$Var(X) = \frac{(b-a)^2}{12}$$

และ X มีฟังก์ชันการแจกแจงสะสมเป็น

$$F(x) = \int_a^x \frac{1}{b-a} dy$$

$$= \frac{x-a}{b-a}$$

การแจกแจงแบบสม่ำเสมอในช่วง (0,1) จะนำมาใช้ในการผลิตเลขสุ่มเพื่อนำไปใช้ในการผลิตตัวแปรสุ่มที่มีการแจกแจงแบบอื่น ๆ

---

[1] มานพ วราภักดิ์. การจำลองเบื้องต้น (กรุงเทพฯ : ศูนย์ผลิตตำราเรียนสถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ,2547), หน้า 43.

2.3.2 การแจกแจงแบบปกติ (Normal Distribution)[2]

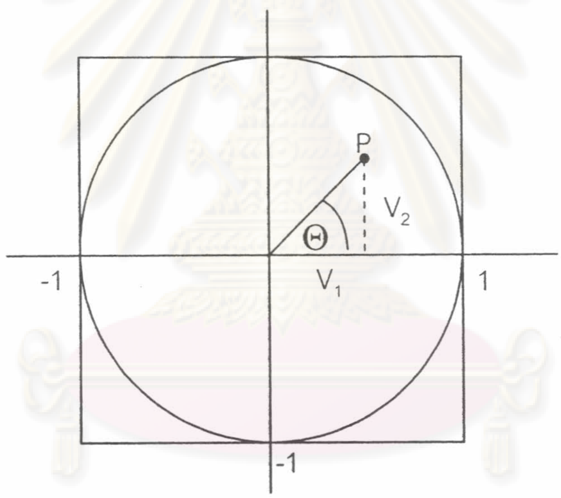
ถ้า X เป็นตัวแปรสุ่มที่มีการแจกแจงแบบปกติ X จะมีฟังก์ชันความน่าจะเป็นดังนี้

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \quad ; \; -\infty < x < \infty$$

โดยที่ $\sigma > 0$

การผลิตตัวแปรสุ่มที่มีการแจกแจงแบบปกติ

ในการวิจัยครั้งนี้การผลิตตัวแปรสุ่มที่มีการแจกแจงแบบปกติจะใช้วิธีโพลาร์ ซึ่ง Marsaglia , MacLaren และ Bray (ค.ศ. 1964) ได้ดัดแปลงวิธีของ Box และ Muller เพื่อหลีกเลี่ยงการคำนวณค่า cosine และ sine ด้วยวิธีรับ-ปฏิเสธ โดยจะสุ่มจุด (V_1 , V_2) ในวงกลมรัศมี 1 ที่มีจุดศูนย์กลางที่จุด (0 , 0) โดยจำลอง V_1 จาก U(-1,1) และ V_2 จาก U(-1,1) อย่างอิสระกันการสุ่มจุด (V_1, V_2) ที่อยู่ในวงกลมจะทำให้ $V_1^2 + V_2^2 \leq 1$



จากรูปจุด (V_1, V_2) ซึ่งอยู่ในวงกลม จะแปลงเป็นจุด (P, Θ) ในพิกัดเชิงขั้ว ได้การแปลงเป็น

$$P = \sqrt{V_1^2 + V_2^2}$$

$$\Theta = \tan^{-1}\left(\frac{V_1}{V_2}\right)$$

P และ Θ มีฟังก์ชันการแจกแจงร่วมเป็น

---

[2] เรื่องเดียวกัน, หน้า 145 .

$$f_{P,\Theta}(\rho,\theta) = f_{V_1,V_2|C}(V_1,V_2 \mid C) \mid J \mid$$

$$= \frac{\rho}{\pi}$$

$$= \frac{1}{2\pi} 2\rho \quad ; 0 \le \theta \le 2\pi \text{ และ } 0 \le \rho \le 1$$

$$= f_\Theta(\theta) f_P(\rho)$$

โดยที่ $f_\Theta(\theta) = \dfrac{1}{2\pi}$, $0 \le \theta \le 2\pi$ และ $f_P(\rho) = 2\rho$, $0 \le \rho \le 1$ เพราะฉะนั้น P และ Θ เป็นอิสระกัน และได้ว่า $\Theta \sim U(0,2\pi)$ และแสดงได้ง่ายว่า $P^2 = V_1^2 + V_2^2 \sim (0,1)$ และเป็นอิสระกับมุม Θ ดังนั้นจะจำลอง $\cos\Theta$ และ $\sin\Theta$ ด้วยการจำลองในวงกลมรัศมี 1 และให้

$$\cos\Theta = \frac{V_1}{P} = \frac{V_1}{\sqrt{V_1^2 + V_2^2}}$$

$$\sin\Theta = \frac{V_2}{P} = \frac{V_2}{\sqrt{V_1^2 + V_2^2}}$$

ดังนั้นตัวแบบจำลองตัวแปรสุ่ม $Z_1 \sim N(0,1)$ และ $Z_2 \sim N(0,1)$ ที่อิสระกันคือ

$$Z_1 = \sqrt{-2\ln P^2}\,\frac{V_1}{\sqrt{V_1^2 + V_2^2}}$$

$$Z_2 = \sqrt{-2\ln P^2}\,\frac{V_2}{\sqrt{V_1^2 + V_2^2}}$$

หรือ
$$Z_1 = V_1\sqrt{\frac{-2\ln(V_1^2 + V_2^2)}{V_1^2 + V_2^2}}$$

$$Z_2 = V_2\sqrt{\frac{-2\ln(V_1^2 + V_2^2)}{V_1^2 + V_2^2}}$$

ดังนั้นเมื่อต้องการเลขสุ่มที่มีการแจกแจงแบบปกติ ที่มีค่าเฉลี่ยเท่ากับ μ และความแปรปรวนเท่ากับ σ^2 ทำได้โดยการแปลงเลขสุ่ม Z_1 , Z_2 โดยอาศัยฟังก์ชันต่อไปนี้

$$X_1 = \mu + \sigma Z_1$$

$$X_2 = \mu + \sigma Z_2$$

ซึ่งจะได้ X_1 และ X_2 ที่เป็นอิสระกัน

2.4 สถิติทดสอบที่ใช้ในการวิจัย

สถิติทดสอบที่ใช้ในการวิจัยครั้งนี้มีด้วยกัน 3 ตัวด้วยกันคือ

2.4.1 สถิติทดสอบเอฟ (F)

ตัวสถิติทดสอบเอฟนั้นมีข้อจำกัดในการใช้ นั่นคือการทดสอบนี้จะทดสอบได้เฉพาะกรณีเมื่ออย่างน้อยหนึ่งค่าของตัวแปรอิสระ X มีหลายค่าตัวแปรตาม Y ให้ c เป็นจำนวนระดับตัวแปรอิสระ X ที่แตกต่างกัน และ n_j เป็นจำนวนค่าของตัวแปรตาม Y ในแต่ละตัวแปรอิสระที่ j ที่มีค่าเท่ากัน โดยที่ j = 1,...,c นั่นคือ $n = \sum_{j=1}^{c} n_j$

ตัวสถิติทดสอบคือ

$$F = \frac{SSLF}{c-p} \div \frac{SSPE}{n-c}$$

$$= \frac{MSLF}{MSPE}$$

โดยที่ $$SSPE = \sum_{j=1}^{c}\sum_{i=1}^{n_j}(Y_{ij} - \overline{Y}_j)$$

$$SSLF = SSE - SSPE$$

เมื่อ SSLF คือ ผลรวมกำลังสองของการเปรียบเทียบความกลมกลืน (Lack of Fit Sum of Squares)

SSPE คือ ผลรวมกำลังสองของความคลาดเคลื่อนโดยตรง (Pure Error Sum of squares)

MSLF คือ กำลังสองเฉลี่ยของการเปรียบเทียบความกลมกลืน (Lack of Fit Mean Square)

MSPE คือ กำลังสองเฉลี่ยของความคลาดเคลื่อนโดยตรง (Pure Error Mean Square)

SSE คือ ผลรวมกำลังสองของความคลาดเคลื่อน (Error Sum of Squares)

จะปฏิเสธสมมติฐานว่างเมื่อ $F > F_{\alpha(c-p,n-c)}$ เมื่อ $F_{\alpha(c-p,n-c)}$ แทนค่าจากตารางการแจงแจงด้วยเอฟที่ระดับนัยสำคัญ α ที่มีระดับขั้นความเป็นอิสระ c-p และ n-c

2.4.2 ตัวสถิติทดสอบ Additive Neyman (AN)[3]

Jianqing Fan และ Li-Shan Haung (2001) ได้เสนอการทดสอบเทียบความกลมกลืนสำหรับการถดถอยด้วยสถิติทดสอบ Adaptive Neyman (AN) เพื่อใช้ในการทดสอบเทียบความกลมกลืนในการวิเคราะห์ความถดถอย ตัวสถิติทดสอบแบบ AN เป็นการทดสอบที่ใช้ค่าคลาดเคลื่อนในการทดสอบ จากตัวแบบ $\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ โดยที่ $\varepsilon \sim N(0,\sigma^2)$ เราให้ $\hat{\varepsilon}_i$ เป็นค่าคลาดเคลื่อนจากวิธีกำลังสองน้อยที่สุด นั่นคือ $\hat{\varepsilon}_i = Y_i - \hat{Y}_i$

ให้ $\underset{\sim}{\hat{\varepsilon}^*} = (\hat{\varepsilon}_1^*, \ldots, \hat{\varepsilon}_n^*)'$ เป็นเวกเตอร์ค่าคลาดเคลื่อนที่เกิดจากการแปลงฟูเรียร์แบบไม่ต่อเนื่อง (Discrete Fourier Transform) ของค่าคลาดเคลื่อน $\underset{\sim}{\hat{\varepsilon}}$

กำหนดให้

$$\hat{\varepsilon}^*_{2j-1} = (2/n)^{1/2} \sum_{i=1}^{n} \cos(2\pi ij/n)\hat{\varepsilon}_i$$

$$\hat{\varepsilon}^*_{2j} = (2/n)^{1/2} \sum_{i=1}^{n} \sin(2\pi ij/n)\hat{\varepsilon}_i$$

เมื่อ j = 1,…,[n/2] โดยที่ [n/2] คือจำนวนเต็มที่มากที่สุดที่น้อยว่าหรือเท่ากับ n/2

กรณีที่ n เป็นเลขคี่ พจน์ที่ n คือ $\hat{\varepsilon}^*_n = (1/\sqrt{n/2})\sum_{i=1}^{n}\hat{\varepsilon}_i$

เมื่อ j มีค่ามากจะได้ว่า $\hat{\varepsilon}^*_j$ มีค่าเฉลี่ยเข้าใกล้ 0 และมีค่าความแปรปรวน σ^2

ตัวสถิติทดสอบคือ

$$T_{AN} = \max_{1\le m\le n} \frac{1}{\sqrt{m\hat{\sigma}_2^2}} \sum_{i=1}^{m}\left(\hat{\varepsilon}_i^{*2} - \hat{\sigma}_1^2\right)$$

เมื่อ

$$\hat{\sigma}_1^2 = \frac{1}{n-I_n}\sum_{i=I_n+1}^{n}\hat{\varepsilon}_i^{*2} - \left\{\frac{1}{n-I_n}\sum_{i=I_n+1}^{n}\hat{\varepsilon}_i^{*}\right\}^2$$

$$\hat{\sigma}_2^2 = \frac{1}{n-I_n}\sum_{i=I_n+1}^{n}\hat{\varepsilon}_i^{*4} - \left\{\frac{1}{n-I_n}\sum_{i=I_n+1}^{n}\hat{\varepsilon}_i^{*2}\right\}^2$$

$I_n = [n/4]$ โดยที่ [n/4] คือจำนวนเต็มที่มากที่สุดที่น้อยกว่าหรือเท่ากับ n/4

---

[3] Jiaqing Fan. and Li-Shan Haung. 2001. Goodness-of-fit test for parametric regression models. Journal of the American Statistical Association. ,96 , 642

เกณฑ์การตัดสินใจในการทดสอบสมมติฐานที่ระดับนัยสำคัญ α คือ จะปฏิเสธสมมติฐานว่างเมื่อ $T_{AN} > T_{AN(M,1-\alpha)}$ โดยที่ $T_{AN(M,1-\alpha)}$ คือค่าวิกฤตจากการสุ่มตัวอย่างหาค่าเปอร์เซ็นไทล์ที่ 100(1-α) จำนวน M รอบ ซึ่งในการวิจัยครั้งนี้กำหนดจำนวนรอบในการหาค่าวิกฤตเท่ากับ 50,000 รอบ

ตัวสถิติทดสอบ Adaptive Neyman เป็นตัวสถิติทดสอบที่ขึ้นอยู่กับสถิติอันดับของค่าคลาดเคลื่อน ดังนั้นจะต้องรู้อันดับของค่าคลาดเคลื่อนก่อนที่จะทำการทดสอบ

2.4.3 สถิติทดสอบ Kuchibhatla และ Hart (KH)[4]

Kuchibhatla และ Hart (1996) ได้เสนอสถิติทดสอบเทียบความกลมกลืนในการวิเคราะห์ความถดถอย สถิติทดสอบ KH มีการแปลงค่าคลาดเคลื่อนที่ได้จากวิธีกำลังสองน้อยที่สุดด้วยการแปลงแบบฟูเรียร์ นั่นคือ

$$\hat{\varepsilon}^*_{2j-1} = (2/n)^{1/2}\sum_{i=1}^{n}\cos(2\pi ij/n)\hat{\varepsilon}_i$$

$$\hat{\varepsilon}^*_{2j} = (2/n)^{1/2}\sum_{i=1}^{n}\sin(2\pi ij/n)\hat{\varepsilon}_i$$

กรณีที่ n เป็นเลขคี่ พจน์ที่ n คือ $\hat{\varepsilon}^*_n = (1/\sqrt{n/2})\sum_{i=1}^{n}\hat{\varepsilon}_i$

เมื่อ j = 1,…,[n/2] โดยที่ [n/2] คือจำนวนเต็มที่มากที่สุดที่น้อยว่าหรือเท่ากับ n/2

ถ้า j มีค่ามากจะได้ว่า $\hat{\varepsilon}^*_j$ มีค่าเฉลี่ยเข้าใกล้ 0 และมีค่าความแปรปรวน σ^2

ตัวสถิติทดสอบ KH คือ

$$S_n = \max_{1\le m\le n-p}\frac{1}{m}\sum_{i=1}^{m}\frac{2n\hat{\varepsilon}_i^{*2}}{\sigma_1^2}$$

เมื่อ $$\hat{\sigma}_1^2 = \frac{1}{n-I_n}\sum_{i=I_n+1}^{n}\hat{\varepsilon}_i^{*2} - \left\{\frac{1}{n-I_n}\sum_{i=I_n+1}^{n}\hat{\varepsilon}_i^{*}\right\}^2$$

เกณฑ์การตัดสินใจในการทดสอบสมมติฐานที่ระดับนัยสำคัญ α คือ จะปฏิเสธสมมติฐานว่างเมื่อ $S_n > S_{M,1-\alpha}$ โดยที่ $S_{M,1-\alpha}$ คือค่าวิกฤตจากการสุ่มตัวอย่างหาค่าเปอร์เซ็นไทล์ที่ 100(1-α)จำนวน M รอบ ซึ่งในการวิจัยครั้งนี้กำหนดจำนวนรอบในการหาค่าวิกฤตเท่ากับ 50,000 รอบ

---

[4] Ibid., p 645.

2.5 สถิติอันดับ

สถิติทดสอบ AN และ KH จะต้องหาตัวสถิติอันดับของค่าคลาดเคลื่อนเสียก่อน ซึ่งสถิติอันดับของค่าคลาดเคลื่อนที่ใช้จะขึ้นอยู่กับสถิติอันดับของตัวแปรอิสระ ดังนั้นก่อนที่จะใช้สถิติทดสอบทั้งสองตัวจะต้องหาสถิตอันดับของค่าคลาดเคลื่อน ซึ่งจะมีค่าตามสถิติอันดับของตัวแปรอิสระ

2.5.1 สถิติอันดับกรณีมีตัวแปรอิสระ 1 ตัว

ถ้า X_i ; i = 1 , 2 , ... , n มีการแจกแจงที่เหมือนกันและเป็นอิสระต่อกัน (Independent and Identically Distributed) และมีฟังก์ชันการแจกแจงสะสม (F)

กำหนด $Y_i =$ ค่าของ X ที่มีค่าน้อยลำดับที่ i

$$Y_1 = \min(X_1, X_2, ..., X_n)$$

$$Y_n = \max(X_1, X_2, ..., X_n)$$

กล่าวคือ $Y_1 \le Y_2 \le ... \le Y_n$ เราเรียก Y_i , i = 1 , 2 , ... , n ว่าสถิติอันดับ

2.5.2 สถิติอันดับกรณีมีตัวแปรอิสระมากกว่า 1 ตัว

กรณีที่มีตัวแปรอิสระมากกว่า 1 ตัว การหาตัวสถิติอันดับจะยุ่งยากขึ้น ซึ่งการหาค่าสถิติอันดับที่ใช้ในการวิจัยครั้งนี้ มี 2 แบบด้วยกัน คือ การใช้ค่าองค์ประกอบหลักที่ 1 (The First Principai Component) และวิธีหาค่าคะแนนของตัวอย่าง (Sample Score) โดยทั้งสองวิธีนี้จะใช้หลักการของการวิเคราะห์องค์ประกอบหลัก (Principal Component Analysis)

1. การหาค่าสถิติอันดับโดยการหาค่าคะแนนตัวอย่าง (Sample Score)

สถิติทดสอบ AN จะใช้วิธีนี้ในการหาค่าสถิติอันดับ ซึ่งการหาสถิติอันดับโดยวิธีหาค่าคะแนนตัวอย่างจะต้องหาค่าคะแนน z (z-score) ก่อน จากนั้นจึงนำค่าคะแนน z ไปหาค่าคะแนนของตัวอย่างอีกขั้นหนึ่ง ขั้นตอนมีดังต่อไปนี้

กำหนด X คือเมทริกซ์ของตัวแปรอิสระ

$$X = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1,p-1} \\ X_{21} & X_{22} & \cdots & X_{2,p-1} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{n,p-1} \end{bmatrix}$$

S คือเมทริกซ์ความแปรปรวนร่วมของ X

$$S = \begin{bmatrix} S_1^2 & S_{12} & \cdots & S_{1,p-1} \\ S_{21} & S_2^2 & \cdots & S_{2,p-1} \\ \vdots & \vdots & \ddots & \vdots \\ S_{p-1,2} & S_{p-1,2} & \cdots & S_{p-1}^2 \end{bmatrix}$$

$\underset{\sim}{\lambda}$ คือ ค่าเจาะจง (Eigenvalues) ของเมทริกซ์ S คือค่า $\underset{\sim}{\lambda}$ ที่ทำให้ $|S - \underset{\sim}{\lambda} I| = 0$

$$\underset{\sim}{\lambda} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_{p-1} \end{bmatrix}$$

โดยที่ $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{p-1}$

$\underset{\sim}{a}$ คือเวกเตอร์เจาะจง (Eigenvectors) ของเมทริกซ์ S หาได้จาก $(S - \lambda_j I)\underset{\sim j}{a} = 0$

กำหนดให้ $$\underset{\sim j}{a}^{*} = \frac{1}{\sqrt{\underset{\sim j}{a}' \underset{\sim j}{a}}} \underset{\sim j}{a}$$

ดังนั้นค่าคะแนน z (z-scores) ของการวิเคราะห์องค์ประกอบหลักคือ

$$z_{i,j} = \underset{\sim j}{a}^{*'} (\underset{\sim i}{x} - \bar{x})$$

เมื่อ $\underset{\sim i}{x}$ เป็นเวกเตอร์ของค่าสังเกตที่ i

$\bar{x}$ เป็นค่าเฉลี่ยของแต่ละเวกเตอร์ของตัวแปรอิสระ

นั่นคือ $\underset{\sim i}{x} = \begin{bmatrix} X_{i1} & X_{i2} & \ldots & X_{i(p-1)} \end{bmatrix}'$

ดังนั้น $\underset{\sim i}{x} - \bar{x} = \begin{bmatrix} X_{i1} - \bar{X}_{\cdot 1} & X_{i2} - \bar{X}_{\cdot 2} & \cdots & X_{i(p-1)} - \bar{X}_{\cdot (p-1)} \end{bmatrix}'$

จากค่าคะแนน z ทำให้ได้ค่าคะแนน (S_i) เพื่อหาตัวสถิติอันดับของค่าคลาดเคลื่อนดังนี้

$$S_i = \frac{1}{n-1} \sum_{j=1}^{p-1} \lambda_j z_{i,j}^2$$

ค่าคลาดเคลื่อนที่ใช้ในการหาคำนวณค่าสถิติทดสอบ AN จะเรียงตามค่าคะแนนของตัวอย่าง (S_i) ที่ได้

2. วิธีค่าองค์ประกอบหลักที่ 1 (The First Principal Component)

ตัวสถิติ KH จะใช้วิธีองค์ประกอบหลักที่ 1 ในการหาค่าสถิติอันดับ ค่าองค์ประกอบหลักที่ 1 หาได้จาก

$$D_i = \underset{\sim 1}{a}^{*'} \underset{\sim i}{x}$$

เมื่อ $\underset{\sim 1}{a}^{*}$ คือเวกเตอร์เจาะจงที่ 1

และ $\underset{\sim i}{x} = \left[X_{i1} \quad X_{i2} \quad \ldots \quad X_{i(p-1)}\right]'$

ค่าคลาดเคลื่อนที่ใช้ในการหาคำนวณค่าสถิติทดสอบ KH จะเรียงตามค่าขององค์ประกอบหลักที่ 1 (D_i)

2.6 การหาค่าวิกฤตของสถิติทดสอบ AN และ KH

ในการวิจัยครั้งนี้การหาค่าวิกฤตที่ระดับนัยสำคัญ (α) ต่าง ๆ จะใช้หลักการหาค่าเปอร์เซ็นไทล์ (Percentile) ที่ 100 (1-α)

โดยขั้นตอนในการทำงานมีดังนี้

2.6.1 กำหนดให้ประชากรมีการแจกแจงตาม H_0 การสุ่มตัวอย่างค่าคลาดเคลื่อนจะทำการสุ่มใหม่ทุกครั้ง ซึ่งจะสุ่มตัวอย่างทั้งหมด 50,000 รอบ

2.6.2 นำตัวอย่างแต่ละชุดที่ได้จากการสุ่มตัวอย่างมาอนุมานค่าพารามิเตอร์ที่เราสนใจ ซึ่งในการวิจัยครั้งนี้คือตัวสถิติทดสอบ AN และ KH

2.6.3 นำค่าสถิติที่ได้ทั้งหมดมาเรียงลำดับจากน้อยไปหามาก จากนั้นหาค่าเปอร์เซ็นไทล์ของค่าสถิติทดสอบในตำแหน่งนั้น ๆ กล่าวคือ

ค่าวิกฤตที่เปอร์เซ็นไทล์ที่ 90 คือค่าวิกฤตที่ระดับนัยสำคัญ 0.1 ซึ่งก็คือค่าสถิติทดสอบลำดับที่ 45,000

ค่าวิกฤตที่เปอร์เซ็นไทล์ที่ 95 คือค่าวิกฤตที่ระดับนัยสำคัญ 0.05 ซึ่งก็คือค่าสถิติทดสอบลำดับที่ 47,500

ค่าวิกฤตที่เปอร์เซ็นไทล์ที่ 99 คือค่าวิกฤตที่ระดับนัยสำคัญ 0.01 ซึ่งก็คือค่าสถิติทดสอบลำดับที่ 49,500