บทที่ 2

ตัวสถิติและผลงานวิจัยที่เกี่ยวข้อง

การศึกษาความถดถอยเชิงเส้นแบบง่ายในการวิจัยครั้งนี้ สิ่งที่สนใจศึกษาแบ่งได้เป็น 2 เรื่องใหญ่ ๆ คือ เรื่องแรกเป็นการศึกษาถึงวิธีการประมาณค่าพารามิเตอร์ ส่วนเรื่องที่สองจะเป็นการศึกษาถึงวิธีการทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์ ซึ่งแบ่งการทดสอบออกเป็นการทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์ β_1 และการทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์ β_o, β_1 ซึ่งในบทนี้จะกล่าวถึงรายละเอียดของวิธีการประมาณค่า และวิธีการทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์แต่ละวิธี ส่วนในตอนท้ายของบทจะนำเสนอผลงานวิจัยที่เกี่ยวข้อง โดยมีรายละเอียดต่าง ๆ เป็นดังนี้

2.1 การประมาณค่าพารามิเตอร์

ในกรณีที่ผู้วิจัยต้องการตัวสถิติสำหรับประมาณค่าพารามิเตอร์นั้น ตัวสถิติแบบพาราเมตริกที่ใช้คือ ตัวประมาณกำลังสองต่ำสุด ซึ่งหาได้จากวิธีกำลังสองต่ำสุด สำหรับตัวสถิติแบบนอนพาราเมตริก ที่ใช้ในการประมาณค่าพารามิเตอร์ จะหาได้จากหลายวิธี เช่น วิธีของบราวน์และมู้ด วิธีของเซ็นและทิลล์ และวิธีของซีเวอร์ ซึ่งรายละเอียดของแต่ละวิธีเป็นดังนี้

2.1.1 วิธีกำลังสองต่ำสุด (Least Squares Method)

วิธีการหาตัวประมาณของพารามิเตอร์วิธีนี้ เป็นวิธีที่มีรากฐานมาจากทฤษฎีการประมาณเชิงเส้น (Theory of Linear Estimation) เป็นวิธีการที่คิดขึ้นโดย คาร์ล เฟรดริก เกาส์ (Karl Friedrich Gauss 1777-1855) และอังเดร แอนดรีวิช มาร์คอฟ (Andrei Andreevich Markov 1856-1922) โดยมีหลักเกณฑ์ดังนี้คือ ให้หาตัวประมาณค่าพารามิเตอร์ที่ทำให้ผลบวกกำลังสองของผลต่างของค่าที่สังเกตได้จากค่าคาดหวังของตัวแปรมีค่าต่ำสุด ซึ่งแสดงรายละเอียดในการหาดังนี้

2.1.1.1 การหาตัวประมาณกำลังสองต่ำสุด จากหลักเกณฑ์ดังกล่าวตั้งเป็นนิยามการหาตัวประมาณกำลังสองต่ำสุด ในการประมาณค่าพารามิเตอร์ของความถดถอยเชิงเส้นแบบง่ายได้ดังนี้

<u>นิยาม 2.1</u> ให้ (x_i , y_i) , $i = 1,2,\ldots,n$ เป็นจุด n จุดที่เป็นไปตามรูปแบบความถดถอยเชิงเส้นแบบง่าย $y_i = \beta_o + \beta_1 x_i + \varepsilon_i$ เมื่อ ε_i ไม่มีสหสัมพันธ์ต่อกัน มีค่าเฉลี่ยเป็น 0 ค่าความแปรปรวนเป็น σ^2 ตัวประมาณกำลังสองต่ำสุดของ β_o และ β_1 คือ $\hat{\beta}_o$ และ $\hat{\beta}_1$ ที่ทำให้ผลบวกกำลังสองความคลาดเคลื่อน (Sum Square Errors (SSE)) หรือ $\sum_{i=1}^{n} \varepsilon_i^2$ มีค่าต่ำสุด

จากนิยาม 2.1 จะทำการหาตัวประมาณกำลังสองต่ำสุด ได้ดังนี้

เนื่องจาก $$SSE = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_o - \beta_1 x_i)^2$$

การหาค่าต่ำสุดของผลบวกกำลังสองความคลาดเคลื่อน ทำได้โดยการหาอนุพันธ์ของผลบวกกำลังสองความคลาดเคลื่อน เทียบกับ β_o และ β_1 แล้วกำหนดให้เท่ากับ 0 ดังนี้

$$\frac{\partial}{\partial \beta_o} \sum_{i=1}^{n} (y_i - \beta_o - \beta_1 x_i)^2 = 0 \qquad \text{---------} \ (1)$$

$$\frac{\partial}{\partial \beta_1} \sum_{i=1}^{n} (y_i - \beta_o - \beta_1 x_i)^2 = 0 \qquad \text{---------} \ (2)$$

จาก (1) และ (2) จะได้สมการปกติ 2 สมการ คือ

$$\sum_{i=1}^{n} (y_i - \beta_o - \beta_1 x_i) = 0$$

และ $$\sum_{i=1}^{n} x_i (y_i - \beta_o - \beta_1 x_i) = 0$$

ซึ่งเมื่อแก้สมการทั้งสอง จะได้ตัวประมาณดังนี้คือ

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \quad \text{และ} \quad \hat{\beta}_o = \bar{y} - \hat{\beta}_1 \bar{x}$$

2.1.1.2 คุณสมบัติที่สำคัญของตัวประมาณกำลังสองต่ำสุด

2.1.1.2.1 ความไม่เอนเอียง (Unbiasedness) เมื่อ $y_i = \beta_o + \beta_1 x_i + \varepsilon_i$ และ $E(\varepsilon_i) = 0$ ตัวประมาณกำลังสองต่ำสุดของ β_o และ β_1 จะเป็นตัวประมาณที่ไม่เอนเอียง นั่นคือ $E(\hat{\beta}_o) = \beta_o$ และ $E(\hat{\beta}_1) = \beta_1$

การพิสูจน์ว่า $E(\hat{\beta}_1) = \beta_1$ แสดงได้ดังนี้

จาก $$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$= \sum_{i=1}^{n} \left\{ \frac{x_i - \bar{x}}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right\} y_i$$

กำหนดให้ $$w_i = \frac{x_i - \bar{x}}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \quad , \quad \sum_{i=1}^{n} w_i = 0 \quad , \quad \sum_{i=1}^{n} w_i x_i = 1$$

จะได้ $$\hat{\beta}_1 = \sum_{i=1}^{n} w_i \, (\beta_o + \beta_1 x_i + \varepsilon_i)$$

$$= \beta_1 + \sum_{i=1}^{n} w_i \varepsilon_i \qquad \text{----------- (3)}$$

$$E(\hat{\beta}_1) = E\,(\beta_1 + \sum_{i=1}^{n} w_i \varepsilon_i)$$

$$= E(\beta_1) + \sum_{i=1}^{n} w_i \, E(\varepsilon_i)$$

ดังนั้น $E(\hat{\beta}_1) = \beta_1$ แสดงว่า $\hat{\beta}_1$ เป็นตัวประมาณที่ไม่เอนเอียงของ β_1

การพิสูจน์ว่า $E(\hat{\beta}_o) = \beta_o$ แสดงได้ดังนี้

จาก
$$\hat{\beta}_o = \bar{y} - \hat{\beta}_1\bar{x}$$

$$= \sum_{i=1}^{n} \{ \frac{1}{n} - \bar{x} w_i\} y_i$$

$$= \sum_{i=1}^{n} \{ \frac{1}{n} - \bar{x}w_i\} (\beta_o + \beta_1 x_i + \varepsilon_i)$$

$$= \beta_o + \sum_{i=1}^{n} (\frac{1}{n} - \bar{x}w_i) \varepsilon_i \text{ ------------(4)}$$

$$E(\hat{\beta}_o) = E \{\beta_o + \sum_{i=1}^{n} (\frac{1}{n} - \bar{x}w_i) \varepsilon_i\}$$

$$= E(\beta_o) + \sum_{i=1}^{n} (\frac{1}{n} - \bar{x}w_i) E(\varepsilon_i)$$

ดังนั้น $E(\hat{\beta}_o) = \beta_o$ แสดงว่า $\hat{\beta}_o$ เป็นตัวประมาณที่ไม่เอนเอียงของ β_o

2.1.1.2.2 <u>ความแปรปรวนของ $\hat{\beta}_o$ และ $\hat{\beta}_1$</u> เมื่อ

$y_i = \beta_o + \beta_1 x_i + \varepsilon_i$, $E(\varepsilon_i) = 0$, $E(\varepsilon_i^2) = \sigma^2$ และ $E(\varepsilon_i\varepsilon_j) = 0$

จะได้ความแปรปรวนของ $\hat{\beta}_o$ และ $\hat{\beta}_1$ คือ $V(\hat{\beta}_o) = \dfrac{\sigma^2 \sum_{i=1}^{n} x_i^2}{n \sum_{i=1}^{n} (x_i-\bar{x})^2}$ และ $V(\hat{\beta}_1) = \dfrac{\sigma^2}{\sum_{i=1}^{n} (x_i-\bar{x})^2}$

การพิสูจน์ว่า $V(\hat{\beta}_o) = \dfrac{\sigma^2 \sum_{i=1}^{n} x_i^2}{n \sum_{i=1}^{n} (x_i-\bar{x})^2}$ แสดงได้ดังนี้

จาก $$V(\hat{\beta}_o) = E(\hat{\beta}_o - E(\hat{\beta}_o))^2$$

$$= E(\hat{\beta}_o - \beta_o)^2$$

จาก(4)จะได้ $$V(\hat{\beta}_o) = E\left\{\sum_{i=1}^{n} \left(\frac{1}{n} - \bar{x}w_i\right) \varepsilon_i\right\}^2$$

$$= E\left\{\sum_{i=1}^{n} \left(\frac{1}{n} - \bar{x}w_i\right)^2 \varepsilon_i^2 + \sum_{i\neq j}^{n} \left(\frac{1}{n} - \bar{x}w_i\right)\left(\frac{1}{n} - \bar{x}w_j\right)\varepsilon_i\varepsilon_j\right\}$$

$$= \sum_{i=1}^{n} \left(\frac{1}{n} - \bar{x}w_i\right)^2 E(\varepsilon_i^2) + \sum_{i\neq j}^{n} \left(\frac{1}{n} - \bar{x}w_i\right)\left(\frac{1}{n} - \bar{x}w_j\right)E(\varepsilon_i\varepsilon_j)$$

$$= \sigma^2 \sum_{i=1}^{n} \left\{\frac{1}{n} - \frac{\bar{x}(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right\}^2$$

$$= \sigma^2 \sum_{i=1}^{n} \left\{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2 - nx_i\bar{x} + n\bar{x}^2}{n\sum_{i=1}^{n}(x_i-\bar{x})^2}\right\}^2$$

$$= \frac{\sigma^2}{n^2\{\sum_{i=1}^{n}(x_i-\bar{x})^2\}^2} \sum_{i=1}^{n}\left(\sum_{i=1}^{n} x_i^2 - nx_i\bar{x}\right)^2$$

$$= \frac{\sigma^2}{n^2\left\{\sum_{i=1}^{n}(x_i-\bar{x})^2\right\}^2} \sum_{i=1}^{n}\left\{\left(\sum_{i=1}^{n}x_i^2\right)^2 - 2nx_i\bar{x}\sum_{i=1}^{n}x_i^2 + n^2x_i^2\bar{x}^2\right\}$$

$$= \frac{n\sigma^2\sum_{i=1}^{n}x_i^2}{n^2\left\{\sum_{i=1}^{n}(x_i-\bar{x})^2\right\}^2} \left\{\sum_{i=1}^{n}x_i^2 - 2\bar{x}\sum_{i=1}^{n}x_i + n\bar{x}^2\right\}$$

ดังนั้น $$V(\hat{\beta}_o) = \frac{\sigma^2\sum_{i=1}^{n}x_i^2}{n\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

การพิสูจน์ว่า $V(\hat{\beta}_1) = \dfrac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$ แสดงได้ดังนี้

จาก $$V(\hat{\beta}_1) = E(\hat{\beta}_1 - E(\hat{\beta}_1))^2$$

$$= E(\hat{\beta}_1 - \beta_1)^2$$

จาก(3)จะได้ $$V(\hat{\beta}_1) = E\left(\sum_{i=1}^{n} w_i\varepsilon_i\right)^2$$

$$= E\left(\sum_{i=1}^{n} w_i^2\varepsilon_i^2 + \sum_{i\neq j}^{n} w_iw_j\varepsilon_i\varepsilon_j\right)$$

$$= \sum_{i=1}^{n} w_i^2 E(\varepsilon_i^2) + \sum_{i\neq j}^{n} w_iw_j E(\varepsilon_i\varepsilon_j)$$

$$= \sigma^2\sum_{i=1}^{n}\left\{\frac{x_i-\bar{x}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right\}^2$$

ดังนั้น $V(\hat{\beta}_1) = \dfrac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$

2.1.1.2.3 <u>การแจกแจงของ $\hat{\beta}_o$ และ $\hat{\beta}_1$</u> เมื่อ

$Y_i = \beta_o + \beta_1 x_i + \varepsilon_i$, $\varepsilon_i \sim NID\ (0, \sigma^2)$ จะได้ว่า

$$\hat{\beta}_o \sim N\left(\beta_o\ , \frac{\sigma^2 \sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n}(x_i-\bar{x})^2}\right) \quad \text{และ} \quad \hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)$$

การพิสูจน์จะทำโดยอาศัยทฤษฎีบท 2.1 ดังนี้

<u>ทฤษฎีบท 2.1</u>[1] ถ้า X เป็นตัวแปรสุ่มที่มีการแจกแจงแบบปกติ ที่มีค่าเฉลี่ยเป็น μ ค่าความแปรปรวนเป็น σ^2 นั่นคือ $X \sim N\ (\mu\ , \sigma^2)$ และ ถ้า $Y = aX + b$ เมื่อ a,b เป็นค่าคงที่ใด ๆ จะได้ว่า Y มีการแจกแจงแบบปกติ ที่มีค่าเฉลี่ยเป็น $a\mu + b$ ค่าความแปรปรวนเป็น $a^2\sigma^2$ นั่นคือ $Y \sim N\ (a\mu + b\ , a^2\sigma^2)$

ดังนั้นจาก (4) $\hat{\beta}_o = \beta_o + \sum_{i=1}^{n}\left(\frac{1}{n} - \bar{x}w_i\right)\varepsilon_i$, คุณสมบัติข้อ 1 และ 2 และทฤษฎีบท 2.1

จะได้ว่า $$\hat{\beta}_o \sim N\left(\beta_o, \frac{\sigma^2 \sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)$$

ในทำนองเดียวกัน

จาก (3) $\hat{\beta}_1 = \beta_1 + \sum_{i=1}^{n} w_i\varepsilon_i$, คุณสมบัติข้อ 1 และ 2 และทฤษฎีบท 2.1

จะได้ว่า $$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)$$

---

[1] ดูพิสูจน์จาก Woodroofe, M., <u>Probability with Applications</u> (Tokyo : McGraw-Hill, Inc, 1975), p.174.

2.1.2 วิธีของบราวน์และมู้ด (Brown and Mood's Method)

วิธีการหาตัวประมาณของพารามิเตอร์วิธีนี้ เป็นวิธีที่เสนอขึ้นโดย บราวน์และมู้ด ในปี ค.ศ. 1951 โดยมีหลักเกณฑ์ดังนี้คือ แบ่งข้อมูลออกเป็น 2 กลุ่ม โดยใช้ค่ามัธยฐาน (median) ของ x เป็นเกณฑ์ในการแบ่ง ซึ่งจะได้ข้อมูลกลุ่มหนึ่งเกิดจาก x ที่มีค่าน้อยกว่าหรือเท่ากับค่ามัธยฐานของ x และอีกกลุ่มหนึ่งจะเกิดจาก x ที่มีค่ามากกว่าค่ามัธยฐานของ x จากนั้นจะหาตัวประมาณค่าพารามิเตอร์ที่จะให้เส้นตรงที่มีมัธยฐานของการเบี่ยงเบน (deviation) ที่ y ห่างจากเส้นตรงนั้นเป็น 0 ในแต่ละกลุ่มของข้อมูลทั้ง 2 กลุ่มนั้น นั่นคือ ต้องการทำให้

$$\text{Median } (y_i - \beta_o - \beta_1 x_i) = 0 \quad , \quad x_i \leq x_m \; , \; i = 1,2,\ldots,n$$

และ $$\text{Median } (y_i - \beta_o - \beta_1 x_i) = 0 \qquad x_i > x_m \; , \; x_m = \text{มัธยฐานของ } x_i$$

สำหรับการหาตัวประมาณโดยวิธีนี้ สามารถทำได้ 2 แบบคือ โดยการลองผิดลองถูก (trial and error) และการพิจารณาจากรูปกราฟ ซึ่งในการวิจัยครั้งนี้จะใช้การพิจารณาจากรูปกราฟ เนื่องจากเป็นวิธีที่สะดวกและมีหลักเกณฑ์ที่แน่นอนกว่า โดยมีขั้นตอนการพิจารณาเป็นดังนี้

1. สร้างแผนภาพการกระจายของข้อมูลตัวอย่าง

2. ลากเส้นในแนวดิ่งผ่านมัธยฐานของ x เพื่อแบ่งข้อมูลออกเป็น 2 กลุ่ม โดยกลุ่มแรกเกิดจาก x ที่มีค่าน้อยกว่าหรือเท่ากับมัธยฐานของ x และกลุ่มที่สองเกิดจาก x ที่มีค่ามากกว่ามัธยฐานของ x (ถ้ามีจุดหนึ่งหรือหลายจุดอยู่บนเส้นมัธยฐานของ x ให้ย้ายเส้นไปทางขวาหรือซ้าย เพื่อให้จำนวนจุดในแต่ละกลุ่มของข้อมูลมีจำนวนเท่าๆกัน)

3. หามัธยฐานของ x และมัธยฐานของ y ของข้อมูลทั้ง 2 กลุ่มนั้น ซึ่งจะได้มัธยฐานทั้งหมด 4 ค่า โดยจะสมมติให้ x_{1m}, y_{1m} เป็นมัธยฐานของ x,y ในกลุ่มที่ 1 และ x_{2m}, y_{2m} เป็นมัธยฐานของ x,y ในกลุ่มที่ 2 ตามลำดับ

4. ลากเส้นตรงเชื่อมต่อระหว่างจุด (x_{1m}, y_{1m}) และ (x_{2m}, y_{2m}) ซึ่งจะได้เส้นตรงเส้นนี้เป็นเส้นประมาณของเส้นถดถอยแรก และจะได้

$$\hat{\beta}_1' = \frac{y_{2m} - y_{1m}}{x_{2m} - x_{1m}} \qquad \text{และ}$$

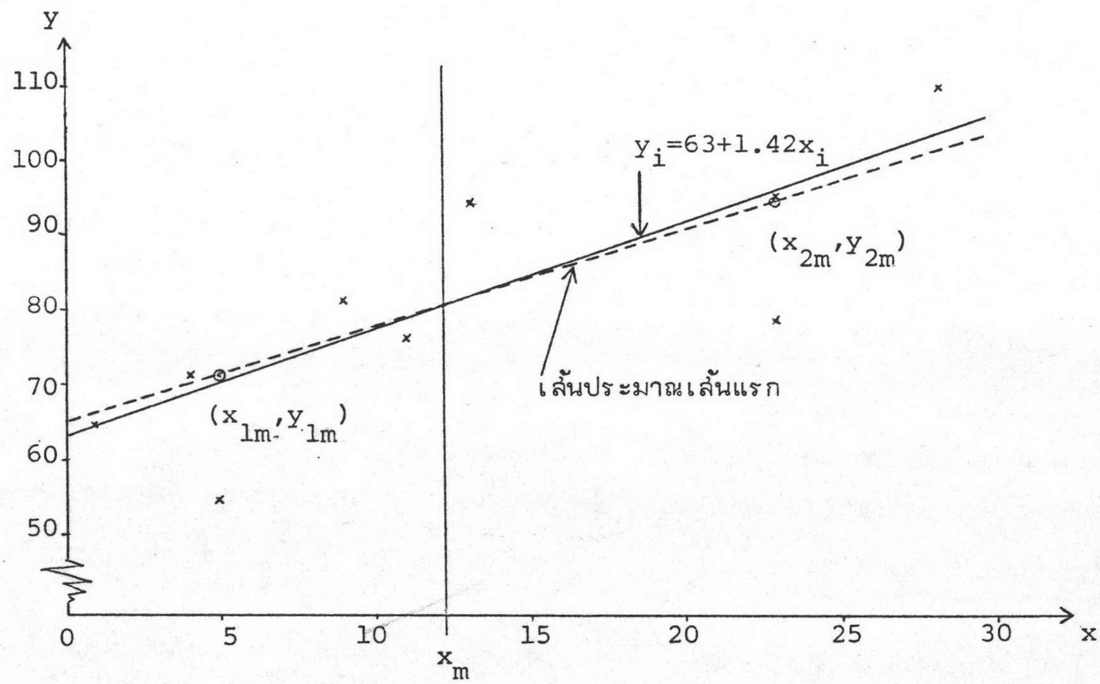
$$\hat{\beta}_o' = y_{1m} - \hat{\beta}_1' x_{1m} \quad \text{หรือ} \quad \hat{\beta}_o' = y_{2m} - \hat{\beta}_1' x_{2m}$$

เป็นค่าประมาณค่าแรกของ β_1 และ β_o

5. ถ้ามัธยฐานของค่าเบี่ยงเบนในแนวตั้งของข้อมูลจากเส้นตรงในข้อ 4 ไม่เป็น 0 ทั้ง 2 กลุ่ม ก็ให้เลื่อนเส้นไปสู่ตำแหน่งใหม่ จนเห็นว่าส่วนเบี่ยงเบนในแต่ละกลุ่มมีมัธยฐานเป็น 0 แล้วจึงทำการวัดค่า $\hat{\beta}_o$, $\hat{\beta}_1$ เป็นค่าประมาณของพารามิเตอร์ β_o , β_1

ทั้งนี้เพื่อให้ง่ายต่อการทำความเข้าใจ จึงขอยกตัวอย่างวิธีการประมาณค่าโดยใช้ข้อมูลตัวอย่าง (Snedecor and Cochran 1967 : 139) ดังนี้คือ

x	1	4	5	9	11	13	23	23	28
y	64	71	54	81	76	93	77	95	109



<u>หมายเหตุ</u> ในการวิจัยครั้งนี้ จะใช้ค่าประมาณค่าแรก ($\hat{\beta}'_o$, $\hat{\beta}'_1$) เท่านั้น คือทำเพียงขั้นตอนที่ 1 ถึง 4 เนื่องจากขั้นตอนที่ 5 ยุ่งยาก และใช้เวลามากในการพิจารณาด้วยคอมพิวเตอร์ ซึ่งจากการทดลองกระทำที่ขนาดตัวอย่าง 10 พบว่า ให้ผลสรุปที่ไม่แตกต่างกันมากนัก จึงใช้ค่าประมาณดังกล่าวแทน

2.1.3 <u>วิธีของเซ็นและทิลล์</u> (Sen and Theil's Method)

วิธีการหาตัวประมาณของพารามิเตอร์วิธีนี้ เป็นวิธีที่เสนอขึ้นโดย ทิลล์ ในปี ค.ศ. 1950 และเซ็นในปี ค.ศ. 1968 โดยมีหลักเกณฑ์ดังนี้คือ หาตัวประมาณค่าพารามิเตอร์โดยพิจารณาจากตัวสถิติทาวน์ของแคนดาลล์ (Kendall's tau) ระหว่าง x_i กับ z_i เมื่อ $z_i = y_i - \beta_1 x_i$ คือ

$$U_n(\beta_1) = \{N\binom{n}{2}\}^{-\frac{1}{2}} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} c(x_j-x_i)\ c(z_j-z_i) \text{-----}(1)$$

เมื่อ $N = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} c(x_j - x_i)$, $i = 1,2,\ldots,n$

และ $$c(x_j - x_i) = \begin{cases} -1 & \text{ถ้า } x_j - x_i < 0 \\ 0 & \text{ถ้า } x_j - x_i = 0 \\ 1 & \text{ถ้า } x_j - x_i > 0 \end{cases}$$

เนื่องจาก $U_n(\beta_1)$ เป็น Strictly distribution-free statistic[1] ที่มีการแจกแจงแบบสมมาตรที่ 0 ดังนั้นการประมาณค่าของ β_1 จะทำโดยเลือกตัวประมาณที่ทำให้ $U_n(\beta_1)$ เข้าใกล้ 0 มากที่สุดเท่าที่จะทำได้

ถ้าให้ $\beta_1^* = \text{Sup}\ \{\beta_1 : U_n(\beta_1) > 0\}$

$\beta_2^* = \text{Inf}\ \{\beta_1 : U_n(\beta_1) < 0\}$

ดังนั้นจะได้ $\hat{\beta}_1 = \frac{1}{2}(\beta_1^* + \beta_2^*)$ เป็นตัวประมาณของพารามิเตอร์ β_1

จากหลักเกณฑ์ดังกล่าว สามารถหาตัวประมาณที่แน่นอน (exact estimator) ได้ดังจะพิจารณาต่อไป

2.1.3.1 <u>การหาตัวประมาณที่แน่นอน</u> มีขั้นตอนในการหาดังนี้

1. จากค่าสังเกต n คู่ คือ (x_1,y_1) , $(x_2,y_2),\ldots,$ (x_n,y_n) เมื่อ $x_1 \leq x_2 \leq \ldots \leq x_n$ หาความชัน (slope) ที่เป็นไปได้ทั้งหมดคือ หา

$S_{ij} = \frac{(y_j - y_i)}{(x_j - x_i)}$, $i < j$ ซึ่งมี $N \leq \binom{n}{2}$ ค่า โดยจะพิจารณาเฉพาะ $x_i \neq x_j$ เท่านั้น

---

[1] ดูพิสูจน์จาก Kendall, M.G ., <u>Rank Correlation Methods</u> (London : Charles Griffin and Company, 1955).

2. เรียงค่า S_{ij} จากน้อยไปมาก ซึ่งจะแทนสัญลักษณ์เป็น $S_{(r)}$, $r = 1,2,\ldots,N$ เมื่อ $S_{(1)} < S_{(2)} < \ldots < S_{(N)}$ แล้วพิจารณาตัวสถิติทาวน์ของเคนดาลล์ ดังนี้คือ จาก (1) ถ้าคำนวณ $U_n(S_{(r)})$ จะมี $z_j - z_i$ เป็นลบ $(r-1)$ ค่า เป็นบวก $(N-r)$ ค่า และเป็น 0 1 ค่า คือจะได้ $U_n(S_{(r)}) = (N-2r+1)/\{N\binom{n}{2}\}^{\frac{1}{2}}$ และในทำนองเดียวกันจะได้ $\{N\binom{n}{2}\}^{\frac{1}{2}} U_n(S^+_{(r)}) = (N-2r)$ เมื่อ $S^+_{(r)}$ (หรือ $S^-_{(r)}$) เป็นค่าที่มากกว่า (หรือน้อยกว่า) $S_{(r)}$

ถ้าให้
$$N = \begin{cases} 2M & \text{ถ้า } N \text{ เป็นเลขคู่} \quad , M \in I^+ \\ 2M+1 & \text{ถ้า } N \text{ เป็นเลขคี่} \end{cases}$$

สำหรับ $N = 2M+1$ จะได้ $U_n(S_{(M+1)}) = 0$ เมื่อ $U_n(S^-_{(M+1)}) > 0$ และ $U_n(S^+_{(M+1)}) < 0$ และในทำนองเดียวกัน สำหรับ $N = 2M$ จะได้ว่า สำหรับ β_1 ใด ๆ ในช่วงเปิด $(S_{(M)}, S_{(M+1)})$ ค่า $U_n(\beta_1) = 0$ เมื่อค่า + หรือ − ที่สอดคล้องกับ β_1 คือ $\leqslant S_{(M)}$ หรือ $\geqslant S_{(M+1)}$ ดังนั้นตัวประมาณที่จะใช้ในการประมาณค่าพารามิเตอร์ β_1 สำหรับวิธีนี้ก็คือ

$$\hat{\beta}_1 = \begin{cases} S_{(M+1)} & \text{ถ้า} \quad N = 2M+1 \\ \frac{1}{2}(S_{(M)} + S_{(M+1)}) & \text{ถ้า} \quad N = 2M \end{cases}$$

<u>หมายเหตุ</u> ตัวประมาณที่ใช้ในการประมาณค่าพารามิเตอร์ β_o สำหรับวิธีนี้นั้น ยังไม่มีการศึกษาไว้ ดังนั้นในการวิจัยครั้งนี้ จึงขอเสนอตัวประมาณนี้ขึ้นมา โดยอาศัยคุณสมบัติของ ε_i คือ ถ้า ε_i มีการแจกแจงแบบสมมาตรแล้ว ตัวประมาณที่จะใช้ในการประมาณค่าพารามิเตอร์ β_o สำหรับวิธีนี้ก็คือ

$$\hat{\beta}_o = \text{Median } (y_i - \hat{\beta}_1 x_i)$$

2.1.3.2 <u>คุณสมบัติของ $\hat{\beta}_1$ กรณีมีขนาดตัวอย่างใหญ่</u>

กำหนดให้
$$T_n^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$A_n^2 = \frac{1}{12}\{n(n^2-1) - \sum_{j=1}^{a_n} u_j\,(u_j^2-1)\}$$ เมื่อ u_j เป็นจำนวน x_j ที่ซ้ำ และ a_n เป็นจำนวนกลุ่มที่ซ้ำ

$$\rho_n = \sum_{i=1}^{n} \frac{(i-\frac{1}{2}(n+1))(x_i-\bar{x})}{T_n A_n}$$

$$B(F) = \int_{-\infty}^{\infty} f^2(y)\,dy < \infty$$

<u>ทฤษฎีบท 2.2</u>[1] ถ้า $T_n \to \infty$ เมื่อ $n \to \infty$ แล้ว $\rho_n T_n(\hat{\beta}_1-\beta_1)$ จะมีการแจกแจงแบบปกติ ที่มีค่าเฉลี่ยเป็น 0 และค่าความแปรปรวนเป็น $\frac{1}{12B^2(F)}$

จากทฤษฎีบท 2.2 จะได้ $\hat{\beta}_1 \sim N(\beta_1, \frac{1}{12\rho_n^2 T_n^2 B^2(F)})$

<u>หมายเหตุ</u> ρ_n เป็นสหสัมพันธ์ระหว่าง x_i กับ i

2.1.4 <u>วิธีของซีเวอร์</u> (Sievers' Method)

วิธีการหาตัวประมาณของพารามิเตอร์วิธีนี้ เป็นวิธีที่เสนอขึ้นโดย ซีเวอร์ ในปี ค.ศ. 1978 โดยมีหลักเกณฑ์ดังนี้คือ หาตัวประมาณค่าพารามิเตอร์ด้วยวิธีการคล้าย ๆ กับวิธีของเซ็นและทิลล์ เพียงแต่วิธีนี้จะมีค่าถ่วงน้ำหนักมาเกี่ยวข้องด้วย โดยพิจารณาตัวสถิติอันดับถ่วงน้ำหนัก (Weighted Rank Statistic)

---

[1] ดูพิสูจน์จาก Sen, P.K., "Estimates of the Regression Coefficient Based on Kendall's Tau," <u>Journal of the American Statistical Association</u> 63(1968) : 1379-1389.

$$T_{(\beta_1)} = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} a_{ij}\, \phi\,(y_i - \beta_o - \beta_1 x_i,\ y_j - \beta_o - \beta_1 x_j)$$

เมื่อ $a_{ij} = x_j - x_i \quad , \; i < j$

และ $\phi(u,v) = \begin{cases} 1 & \text{ถ้า} \quad u \leqslant v \quad \text{โดย } u = y_i - \beta_o - \beta_1 x_i \\ 0 & \text{ถ้า} \quad u > v \quad \quad\; v = y_j - \beta_o - \beta_1 x_j \end{cases}$

ถ้าให้ $\beta_1' = \text{Sup}\,\{\beta_1 : T_{(\beta_1)} \geqslant \frac{a..}{2}\}$, $a.. = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} a_{ij}$

$\beta_2' = \text{Inf}\,\{\beta_1 : T_{(\beta_1)} \leqslant \frac{a..}{2}\}$

ดังนั้นจะได้ $\hat{\beta}_1 = \frac{1}{2}(\beta_1' + \beta_2')$ เป็นตัวประมาณของพารามิเตอร์ β_1

จากหลักเกณฑ์ดังกล่าว สามารถหาตัวประมาณที่แน่นอนได้ดังจะพิจารณาต่อไป

2.1.4.1 <u>การหาตัวประมาณที่แน่นอน</u> มีขั้นตอนในการหาดังนี้

1. จากค่าสังเกต n คู่ คือ (x_1,y_1), (x_2,y_2),...., (x_n,y_n) เมื่อ $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$ หาค่าถ่วงน้ำหนัก และความชันที่เป็นไปได้ทั้งหมด คือหา $a_{ij} = x_j - x_i$ และ $S_{ij} = \frac{y_j - y_i}{x_j - x_i}$, $i < j$ ซึ่งมี $N \leqslant \binom{n}{2}$ ค่า โดยจะพิจารณาเฉพาะ $x_i \neq x_j$ เท่านั้น

2. หาการแจกแจงความน่าจะเป็นแบบไม่ต่อเนื่อง (Discrete probability distribution) ของ S_{ij} คือ หา $P_{ij} = \frac{a_{ij}}{a..}$, $a.. = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} a_{ij}$

3. เรียงค่า (S_{ij}, P_{ij}) โดยเรียงตามค่าของ S_{ij} จากน้อยไปมาก ซึ่งจะแทนสัญลักษณ์เป็น $(S_{(1)},P_{(1)})$, $(S_{(2)},P_{(2)})$,....,$(S_{(N)},P_{(N)})$ เมื่อ $S_{(1)} < S_{(2)} < \ldots < S_{(N)}$

4. หาการแจกแจงสะสม (Cumulative distribution) ของ $S_{(r)}$, $r=1,2,\ldots,N$ คือ หา $H_{(r)} = \sum_{i=1}^{r} P_{(r)}$ และจะได้ตัวประมาณที่แน่นอนในการประมาณค่าพารามิเตอร์ β_1 สำหรับวิธีนี้คือ $S_{(r)}$ ที่ทำให้ $H_{(r)} = .5$ นั่นคือ

$$\hat{\beta}_1 = \begin{cases} S_{(r)} & \text{ถ้า } H_{(r)} = .5 \\ \frac{1}{2}\left(S_{(r)}+S_{(r+1)}\right) & \text{ถ้า } H_{(r)} < .5 \text{ และ } H_{(r+1)} > .5 \end{cases}$$

<u>หมายเหตุ</u> ตัวประมาณที่ใช้ในการประมาณค่าพารามิเตอร์ β_o สำหรับวิธีนี้นั้น ยังไม่มีการศึกษาไว้ ดังนั้นในการวิจัยครั้งนี้ จึงขอเสนอตัวประมาณนี้ขึ้นมา โดยอาศัยคุณสมบัติของ ε_i คือ ถ้า ε_i มีการแจกแจงแบบสมมาตรแล้ว ตัวประมาณที่จะใช้ในการประมาณค่าพารามิเตอร์ β_o สำหรับวิธีนี้ ก็คือ

$$\hat{\beta}_o = \text{Median}\ (y_i-\hat{\beta}_1 x_i)$$

2.1.4.2 <u>คุณสมบัติของ $\hat{\beta}_1$ กรณีมีขนาดตัวอย่างใหญ่</u>

กำหนดให้ $$\sigma_n^2 = \sum_{i=1}^{n} \frac{(x_i-\bar{x})^2}{n}$$

$$\rho_n = \frac{\sum_{i=1}^{n} A_i(x_i-\bar{x})}{\left(\sum_{i=1}^{n}(x_i-\bar{x})^2 \sum_{i=1}^{n} A_i^2\right)^{\frac{1}{2}}} \quad \text{เมื่อ} \quad A_i = n\ (x_i-\bar{x})$$

และ $$B_{(F)} = \int_{-\infty}^{\infty} f^2(y)\ dy < \infty$$

ทฤษฎีบท 2.3[1] ถ้า $\sigma_n^2 \to \sigma^2 > 0$ และ $\rho_n \to \rho \neq 0$ เมื่อ $n \to \infty$ แล้ว $n^{\frac{1}{2}}(\hat{\beta}_1 - \beta_1)$ จะมีการแจกแจงแบบปกติ ที่มีค่าเฉลี่ยเป็น 0 และค่าความแปรปรวนเป็น

$$\frac{1}{12\ \rho^2\sigma^2B^2(F)}$$

จากทฤษฎีบท 2.3 จะได้ $\hat{\beta}_1 \sim N\left(\beta_1, \frac{1}{12n\rho^2\sigma^2B^2(F)}\right)$

หมายเหตุ ρ_n เป็นสหสัมพันธ์ระหว่าง x_i กับ A_i

2.2 การทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์ β_1

สมมติฐานที่ทดสอบ

A : การทดสอบแบบสองทาง (Two-sided test)

H_o : $\beta_1 = \beta_{10}$

H_A : $\beta_1 \neq \beta_{10}$, $\beta_1 < \beta_{10}$ หรือ $\beta_1 > \beta_{10}$

B : การทดสอบแบบทางเดียว (One-sided test)

H_o : $\beta_1 = \beta_{10}$

H_A : $\beta_1 < \beta_{10}$

---

[1] ดูพิสูจน์จาก Sievers, G.L., "Weighted Rank Statistics for Simple Linear Regression," Journal of the American Statistical Association 73 (1978) : 628-631.

C : การทดสอบแบบทางเดียว (One-sided test)

H_o : $\beta_1 = \beta_{10}$

H_A : $\beta_1 > \beta_{10}$

ในกรณีที่ผู้วิจัยต้องการทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์ β_1 นั้น ตัวสถิติแบบพาราเมตริกที่ใช้คือ ตัวสถิติทดสอบที (t test) ซึ่งได้จากวิธีกำลังสองต่ำสุด สำหรับตัวสถิติแบบนอนพาราเมตริก ที่ใช้ในการทดสอบนี้ หาได้จากหลายวิธีเช่น วิธีของบราวน์และมู้ด วิธีของเซ็นและทิลล์ และวิธีของซีเวอร์ ซึ่งรายละเอียดของแต่ละวิธีเป็นดังนี้

2.2.1 <u>วิธีกำลังสองต่ำสุด</u> (Least Squares Method)

ในการทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์ β_1 แบบพาราเมตริก ถ้าทราบ σ^2 จะใช้การทดสอบซี (z test) แต่โดยปกติแล้วจะไม่ทราบ σ^2 ดังนั้นจะใช้การทดสอบที (t test) แทน ซึ่งมีรายละเอียดของตัวสถิติทดสอบแสดงได้ดังนี้

2.2.1.1 สถิติทดสอบ[1]

$$t = \frac{\hat{\beta}_1 - \beta_{10}}{S_{\hat{\beta}_1}}$$

โดย $\hat{\beta}_1$ เป็นตัวประมาณกำลังสองต่ำสุดของ β_1

$$S_{\hat{\beta}_1} = \sqrt{\frac{S^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}}$$

$$S^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{n-2} \quad \text{เมื่อ} \quad \hat{y}_i = \hat{\beta}_o - \hat{\beta}_1 x_i$$

---

[1] ดูรายละเอียดจาก Johnston, J., <u>Econometric Methods</u> (Auckland : McGraw-Hill, Inc, 1984), p. 34.

2.2.1.2 เกณฑ์การตัดสินใจ

สมมติฐานแบบ A จะปฏิเสธ H_o เมื่อ $t \leq -t(\alpha/2, n-2)$ หรือ $t \geq t(\alpha/2, n-2)$

สมมติฐานแบบ B จะปฏิเสธ H_o เมื่อ $t \leq -t(\alpha, n-2)$

สมมติฐานแบบ C จะปฏิเสธ H_o เมื่อ $t \geq t(\alpha, n-2)$

เมื่อ $t(\alpha/2, n-2)$ หรือ $t(\alpha, n-2)$ เป็นค่าวิกฤต (critical value) ที่ได้จากตาราง t ที่องศาอิสระ (degree of freedom (df)) n-2 ณ ระดับนัยสำคัญ α

2.2.2 วิธีของบราวน์และมู้ด (Brown and Mood's Method)

การทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์ β_1 วิธีนี้ เป็นวิธีที่เสนอขึ้นโดยบราวน์และมู้ด ในปี ค.ศ. 1951 เป็นการพิจารณาจากรูปกราฟ ซึ่งมีหลักเกณฑ์ ดังนี้คือ สร้างแผนภาพการกระจายของข้อมูลตัวอย่าง แล้วลากเส้นในแนวดิ่งผ่านมัธยฐานของ x จากนั้นลากเส้น $y_i = a + \beta_{10} x_i$, $a =$ มัธยฐานของ $(y_i - \beta_{10} x_i)$ บนแผนภาพการกระจาย แล้วจึงหาตัวสถิติทดสอบ ซึ่งมีรายละเอียดแสดงได้ดังนี้

2.2.2.1 สถิติทดสอบ

$$\chi_b^2 = \frac{16}{n}\left(n_1 - \frac{n}{4}\right)^2$$

โดย n_1 = จำนวนจุดที่อยู่เหนือเส้น $y_i = a + \beta_{10} x_i$ และอยู่ทางซ้ายของมัธยฐานของค่า x

และ a = มัธยฐานของ $(y_i - \beta_{10} x_i)$

2.2.2.2 เกณฑ์การตัดสินใจ

สมมติฐานทั้ง 3 แบบ จะปฏิเสธ H_o เมื่อ $\chi_b^2 \geq \chi^2(\alpha, 1)$

เมื่อ $\chi^2(\alpha, 1)$ เป็นค่าวิกฤตที่ได้จากตารางไคสแควร์ (Chi-Square (χ^2)) ที่องศาอิสระ 1 ณ ระดับนัยสำคัญ α

2.2.3 วิธีของเซ็นและทิลล์ (Sen and Theil's Method)

การทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์ β_1 วิธีนี้ เป็นวิธีที่เสนอขึ้นโดยทิลล์ ในปี ค.ศ. 1950 และเซ็นในปี ค.ศ. 1968 โดยมีหลักเกณฑ์ดังนี้คือ พิจารณาจากตัวสถิติทาวน์-ของเคนดาลล์ระหว่าง x_i กับ z_i เมื่อ $z_i = y_i - \beta_{10} x_i$ ซึ่งมีรายละเอียดของตัวสถิติทดสอบแสดงได้ดังนี้คือ

2.2.3.1 สถิติทดสอบ

$$\tau = \frac{S}{\sqrt{N\binom{n}{2}}} \quad \text{(หรือ} \quad \tau = \frac{S}{n(n-1)/2} \quad \text{กรณีไม่มีค่า } x \text{ ซ้ำ)}$$

โดย
$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} c(x_j - x_i)\, c(z_j - z_i) \quad , \quad i = 1,2,\ldots,n$$

$$N = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} c(x_j - x_i)$$

และ
$$c(x_j - x_i) = \begin{cases} -1 & \text{ถ้า } x_j - x_i < 0 \\ 0 & \text{ถ้า } x_j - x_i = 0 \\ 1 & \text{ถ้า } x_j - x_i > 0 \end{cases}$$

2.2.3.2 เกณฑ์การตัดสินใจ

สมมติฐานแบบ A จะปฏิเสธ H_o เมื่อ $\tau \leqslant -\tau\,(\alpha/2,n)$ หรือ $\tau \geqslant \tau(\alpha/2,n)$

สมมติฐานแบบ B จะปฏิเสธ H_o เมื่อ $\tau \leqslant -\tau\,(\alpha,n)$

สมมติฐานแบบ C จะปฏิเสธ H_o เมื่อ $\tau \geqslant \tau\,(\alpha,n)$

เมื่อ $\tau(\alpha/2,n)$ หรือ $\tau(\alpha\,,\,n)$ เป็นค่าวิกฤตที่ได้จากตารางของเคนดาลล์ที่ขนาดตัวอย่าง n ณ ระดับนัยสำคัญ α

2.2.2.3 <u>กรณีที่มีขนาดตัวอย่างใหญ่</u>

ตัวสถิติทดสอบจะประมาณด้วยการแจกแจงแบบปกติ ดังนี้

$$Z = \frac{\tau - E(\tau)}{v(\tau)}$$

เมื่อ $E(\tau) = 0$

$$v(\tau) = \frac{2(2n+5)}{9n(n-1)}$$

2.2.4 <u>วิธีของซีเวอร์</u> (Sievers' Method)

การทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์ β_1 วิธีนี้ เป็นวิธีที่เสนอขึ้นโดยซีเวอร์ ในปี ค.ศ. 1978 โดยมีหลักเกณฑ์ดังนี้คือ พิจารณาจากตัวสถิติอันดับถ่วงน้ำหนัก ซึ่งคล้ายกับตัวสถิติทาวน์ของเคนดาลล์ แต่จะมีค่าถ่วงน้ำหนักมาเกี่ยวข้องด้วย ซึ่งมีรายละเอียดของสถิติทดสอบแสดงได้ดังนี้

2.2.4.1 <u>สถิติทดสอบ</u>

$$T_o = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} a_{ij}\, \phi\,(y_i-\beta_{10}x_i\ ,\ y_j-\beta_{10}x_j)$$

เมื่อ $a_{ij} = x_j - x_i$

และ
$$\phi(u,v) = \begin{cases} 1 & \text{ถ้า } u \leq v \quad \text{โดย } u = y_i - \beta_{10}x_i \\ 0 & \text{ถ้า } u > v \quad\quad\;\; v = y_j - \beta_{10}x_j \end{cases}$$

2.2.4.2 <u>เกณฑ์การตัดสินใจ</u>

สมมติฐานแบบ A จะปฏิเสธ H_o เมื่อ $T_o \leq \frac{a..}{2} - z_{\alpha/2}\left(\sum_{i=1}^{n} \frac{A_i^2}{12}\right)^{\frac{1}{2}}$ หรือ

$$T_o \geq \frac{a..}{2} + z_{\alpha/2}\left(\sum_{i=1}^{n} \frac{A_i^2}{12}\right)^{\frac{1}{2}}$$

สมมติฐานแบบ B จะปฏิเสธ H_o เมื่อ $T_o \leqslant \frac{a..}{2} - z_\alpha \left(\sum_{i=1}^{n} \frac{A_i^2}{12}\right)^{\frac{1}{2}}$

สมมติฐานแบบ C จะปฏิเสธ H_o เมื่อ $T_o \geqslant \frac{a..}{2} + z_\alpha \left(\sum_{i=1}^{n} \frac{A_i^2}{12}\right)^{\frac{1}{2}}$

เมื่อ $a.. = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} a_{ij}$, $A_i = n(x_i - \bar{x})$ ส่วน

$z_{\alpha/2}$ หรือ z_α เป็นค่าวิกฤตที่ได้จากตาราง z ที่ระดับนัยสำคัญ α

<u>หมายเหตุ</u>[1] ถ้า n มีขนาดเล็ก T_o ไม่มีการแจกแจงที่แน่นอน แต่ถ้า n มีขนาดใหญ่ T_o จะมีการแจกแจงแบบปกติ ที่มีค่าเฉลี่ยเป็น $\frac{a..}{2}$ และค่าความแปรปรวนเป็น $\sum_{i=1}^{n} \frac{A_i^2}{12}$

2.3 การทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์ β_o , β_1

สมมติฐานที่ทดสอบ

H_o : $\beta_o = \beta_{oo}$, $\beta_1 = \beta_{10}$

H_A : $\beta_o \neq \beta_{oo}$, $\beta_1 \neq \beta_{10}$, $\beta_o < \beta_{oo}$ หรือ $\beta_o > \beta_{oo}$

และ $\beta_1 < \beta_{10}$ หรือ $\beta_1 > \beta_{10}$

---

[1] ดูพิสูจน์จาก Sievers, G.L., "Weighted Rank Statistics for Simple Linear Regression," <u>Journal of the American Statistical Association</u> 73(1978) : 628-631.

ในกรณีที่ผู้วิจัยต้องการทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์ β_o, β_1 นั้น ตัวสถิติแบบพาราเมตริกที่ใช้คือ ตัวสถิติทดสอบเอฟ (F test) ซึ่งได้จากวิธีกำลังสองต่ำสุด สำหรับตัวสถิติแบบนอนพาราเมตริกที่ใช้ในการทดสอบนี้ หาได้จาก 2 วิธีคือ วิธีของบราวน์และมู้ด และวิธีของแลนคาสเตอร์และเควด สำหรับวิธีของแลนคาสเตอร์และเควดเป็นวิธีที่พิจารณาต่อจากวิธีของเซ็นและทิลล์ในหัวข้อ 2.2 ซึ่งรายละเอียดของแต่ละวิธีเป็นดังนี้

2.3.1 <u>วิธีกำลังสองต่ำสุด</u> (Least Squares Method)

ในการทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์ β_o, β_1 แบบพาราเมตริก ถ้าทราบ σ^2 จะใช้การทดสอบไคสแควร์ (X^2 test) แต่โดยปกติแล้ว จะไม่ทราบ σ^2 ดังนั้น จะใช้การทดสอบเอฟ (F test) แทน ซึ่งมีรายละเอียดของตัวสถิติทดสอบแสดงได้ดังนี้

2.3.1.1 <u>สถิติทดสอบ</u>[1]

$$F = \frac{n^2(\hat{\beta}_o - \beta_{oo})^2 + (\hat{\beta}_1 - \beta_{10})^2 \sum_{i=1}^{n} x_i^2 + 2(\hat{\beta}_o - \beta_{oo})(\hat{\beta}_1 - \beta_{10}) \sum_{i=1}^{n} x_i}{2S^2}$$

โดย $\hat{\beta}_o$, $\hat{\beta}_1$ เป็นตัวประมาณกำลังสองต่ำสุดของ β_o, β_1

$$S^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{n-2} \quad , \quad \hat{y}_i = \hat{\beta}_o - \hat{\beta}_1 x_i$$

2.3.1.2 <u>เกณฑ์การตัดสินใจ</u>

จะปฏิเสธสมมติฐาน H_o เมื่อ $F \geq F(\alpha, 2, n-2)$

เมื่อ $F(\alpha, 2, n-2)$ เป็นค่าวิกฤตที่ได้จากตาราง F ที่องศาอิสระ 2 กับ n-2 ณ ระดับนัยสำคัญ α

---

[1] ดูพิสูจน์จาก Jonhston, J., <u>Econometric Methods</u> (Auckland : McGraw-Hill, Inc, 1984), p. 182-185.

2.3.2 <u>วิธีของบราวน์และมู้ด</u> (Brown and Mood's Method)

การทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์ β_o, β_1 วิธีนี้ เป็นวิธีที่เสนอขึ้นโดยบราวน์และมู้ด ในปี ค.ศ. 1951 เป็นการพิจารณาจากรูปกราฟ ซึ่งมีหลักเกณฑ์ดังนี้คือ สร้างแผนภาพการกระจายของข้อมูลตัวอย่าง แล้วลากเส้นในแนวดิ่งผ่านมัธยฐานของ x จากนั้นลากเส้น $Y_i = \beta_{oo} + \beta_{10}X_i$ บนแผนภาพการกระจาย แล้วจึงหาตัวสถิติทดสอบ ซึ่งมีรายละเอียดแสดงได้ดังนี้

2.3.2.1 <u>สถิติทดสอบ</u>

$$\chi^2 = \frac{8}{n}\left[(n_1 - \frac{n}{4})^2 + (n_2 - \frac{n}{4})^2\right]$$

โดย n_1 = จำนวนจุดที่อยู่เหนือเส้น $y_i = \beta_{oo} + \beta_{10}x_i$ และอยู่ทางซ้ายของมัธยฐานของค่า x

n_2 = จำนวนจุดที่อยู่เหนือเส้น $y_i = \beta_{oo} + \beta_{10}x_i$ และอยู่ทางขวาของมัธยฐานของค่า x

2.3.2.2 <u>เกณฑ์การตัดสินใจ</u>

จะปฏิเสธสมมติฐาน H_o เมื่อ $\chi^2 \geqslant \chi^2(\alpha, 2)$

เมื่อ $\chi^2(\alpha, 2)$ เป็นค่าวิกฤตที่ได้จากตาราง χ^2 ที่องศาอิสระ 2 ณ ระดับนัยสำคัญ α

2.3.3 <u>วิธีของแลนคาสเตอร์และเควด</u> (Lancaster and Quade's Method)

การทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์ β_o, β_1 วิธีนี้ เป็นวิธีที่เสนอขึ้นโดยแลนคาสเตอร์และเควด ในปี ค.ศ. 1985 โดยมีหลักเกณฑ์ดังนี้คือ เป็นการรวมตัวสถิติทาวน์ของเคนดาลล์ (Kendall's tau) ระหว่าง x_i กับ R_i, $R_i = y_i - \beta_{oo} - \beta_{10}x_i$ และตัวสถิติเครื่องหมาย (Sign statistic) เข้าด้วยกัน ซึ่งมีรายละเอียดของสถิติทดสอบแสดงได้ดังนี้

2.3.3.1 สถิติทดสอบ

$$C = \frac{9n(n-1)}{2(2n+5)} T^2 + \frac{(2L-n)^2}{n}$$

โดย $$T^2 = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} Sgn(R_i - R_j)(x_i - x_j)$$

$$L = \sum_{i=1}^{n} L_i$$

$$L_i = \frac{(1+Sgn(R_i))}{2} = \begin{cases} 1 & \text{ถ้า } R_i > 0 \\ 0 & \text{ถ้า } R_i < 0 \end{cases}$$

และ $$R_i = y_i - \beta_{00} - \beta_{10} x_i$$

2.3.3.2 เกณฑ์การตัดสินใจ

ถ้า $n \leqslant 15$ จะปฏิเสธสมมติฐาน H_0 เมื่อ $c \geqslant c(\alpha, n)$ และถ้า $n > 15$ จะปฏิเสธสมมติฐาน H_0 เมื่อ $c \geqslant \chi^2(\alpha, 2)$

เมื่อ $c(\alpha, n)$ เป็นค่าวิกฤตที่ได้จากตารางการแจกแจงที่แน่นอนของ c(Exact distribution of c) ที่ขนาดตัวอย่าง n ซึ่งเสนอโดยแสนคาสเตอร์และเควด (1985:393-397) สำหรับ $\chi^2(\alpha, 2)$ เป็นค่าวิกฤตที่ได้จากตาราง χ^2 ที่องศาอิสระ 2 ณ ระดับนัยสำคัญ α

2.4 ผลงานวิจัยที่เกี่ยวข้อง

ในการศึกษาวิธีการประมาณค่า และการทดสอบสมมติฐานของความถดถอยเชิงเส้นแบบง่าย ด้วยวิธีการนอนพาราเมตริกนั้น มีผู้ศึกษาไว้ไม่มากนัก ดังนั้นผลงานวิจัยที่เกี่ยวข้องจึงมีอยู่น้อย ซึ่งจากการศึกษาที่ผ่านมา มีนักวิจัยบางท่านได้ศึกษาค่าประสิทธิภาพสัมพัทธ์ใกล้อนันต์ (Asymptotic Relative Efficiency) ระหว่างวิธีพาราเมตริก และวิธีนอนพาราเมตริก หรือระหว่างวิธีนอนพาราเมตริกด้วยกัน ซึ่งมีรายละเอียดที่เสนอไว้ดังนี้

ซีเวอร์ (1978:628-631) ศึกษาเปรียบเทียบประสิทธิภาพของการประมาณค่าพารามิเตอร์ β_1 ในรูปแบบของค่าประสิทธิภาพสัมพัทธ์ใกล้อนันต์ ที่พิจารณาจากค่าความแปรปรวนของตัวประมาณ ระหว่างวิธีกำลังสองต่ำสุด กับวิธีของซีเวอร์ ซึ่งได้ผลสรุปดังนี้คือ

1. ถ้าความคลาดเคลื่อน มีการแจกแจงแบบปกติ จะได้ค่าประสิทธิภาพสัมพัทธ์ใกล้อนันต์เท่ากับ .955

2. ถ้าความคลาดเคลื่อน มีการแจกแจงลักษณะที่มีการกระจายไปทางหางมาก ๆ จะได้ค่าประสิทธิภาพสัมพัทธ์ใกล้อนันต์มากกว่า 1

3. ถ้าความคลาดเคลื่อน มีการแจกแจงแบบต่อเนื่องอื่น ๆ จะได้ค่าประสิทธิภาพสัมพัทธ์ใกล้อนันต์มากกว่าหรือเท่ากับ .864

ซึ่งจากค่าประสิทธิภาพสัมพัทธ์ใกล้อนันต์ดังกล่าว แสดงว่า วิธีของซีเวอร์มีประสิทธิภาพสูงกว่าวิธีกำลังสองต่ำสุด เมื่อความคลาดเคลื่อน มีการแจกแจงลักษณะที่มีการกระจายไปทางหางมาก ๆ ส่วนการแจกแจงแบบอื่น ๆ อาจจะไม่มีประสิทธิภาพสูงกว่า

แลนคาส์เตอร์และเควด (1985:393-397) ศึกษาเปรียบเทียบประสิทธิภาพของการทดสอบระหว่างวิธีกำลังสองต่ำสุด กับวิธีของแลนคาส์เตอร์และเควด และระหว่างวิธีของบราวน์และมู้ด กับวิธีของแลนคาส์เตอร์และเควด โดยใช้ค่าประสิทธิภาพสัมพัทธ์ใกล้อนันต์แบบพิทแมน (Pitman Asymptotic Relative Efficiency) ซึ่งเป็นค่าประสิทธิภาพสัมพัทธ์ใกล้อนันต์ที่พิจารณาจากพารามิเตอร์ที่ไม่มีศูนย์กลาง (Noncentral parameter) ของสถิติทดสอบแต่ละวิธี ในการทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์ β_o, β_1 ซึ่งได้ผลสรุปดังตาราง 2.1

ตารางที่ 2.1[1] ค่าประสิทธิภาพสัมพัทธ์ใกล้อนันต์แบบพิทแมน ภายใต้ลักษณะการแจกแจงแบบต่าง ๆ ของการทดสอบวิธีกำลังสองต่ำสุด (G) วิธีของแสนคาส์เตอร์และเควด (C) และวิธีของบราวน์และมู้ด (B)

การแจกแจงของความคลาดเคลื่อน	ARE(G,C) $\delta_{\beta_o} = 0$	ARE(G,C) $\delta_{\beta_1} = 0$	ARE(B,C) $\delta_{\beta_o} = 0$	ARE(B,C) $\delta_{\beta_1} = 0$
Cauchy	∞	∞	1.00	1.00
Double				
Exponential	1.50	2.00	1.00	1.00
Exponential	3.00	1.00	4.00	1.00
Logistic	1.10($\pi^2/9$)	.82($\pi^2/12$)	1.78(16/9)	1.00
Normal	.95($3/\pi$)	.63($2/\pi$)	2.00	1.00
Triangular				
One-Sided	1.19(32/27)	.44(4/9)	3.56(32/9)	1.00
Symmetric	.89(8/9)	.67(2/3)	1.78(16/9)	1.00
Uniform	1.00	.33(1/3)	4.00	1.00

<u>หมายเหตุ</u> พิจารณาที่ $(\beta_{on}, \beta_{1n}) = (\beta_{oo} + \frac{\delta_{\beta_o}}{\sqrt{n}}, \beta_{10} + \frac{\delta_{\beta_1}}{\sqrt{n}})$

---

[1] Lancaster, J.F. and Quade, D., "A Nonparametric Test for Linear Regression Based on Combining Kendall's Tau with the the Sign Test," <u>Journal of the American Statistical Association</u> 80 (1985): 393-397.

ซึ่งจากค่าประสิทธิภาพสัมพัทธ์ใกล้อนันต์แบบพิทแมน แสดงว่า การทดสอบวิธีของแลนคาสเตอร์และเควดมีประสิทธิภาพสูงที่สุด เทียบกับการทดสอบวิธีกำลังสองต่ำสุด เมื่อความคลาดเคลื่อนมีการแจกแจงแบบ Cauchy แบบ Double Exponential แบบ Exponential แบบ Logistic เฉพาะ $\delta_{\beta_0} = 0$ และแบบ Triangular One-Sided เฉพาะ $\delta_{\beta_0} = 0$ และมีประสิทธิภาพเท่ากัน เมื่อความคลาดเคลื่อนมีการแจกแจงแบบ Exponential เฉพาะ $\delta_{\beta_1} = 0$ และแบบ Uniform เฉพาะ $\delta_{\beta_0} = 0$ นอกนั้นจะมีประสิทธิภาพต่ำกว่า ในขณะเดียวกัน เมื่อเปรียบเทียบประสิทธิภาพระหว่างวิธีของแลนคาสเตอร์และเควด กับวิธีของบราวน์และมู้ด ปรากฏว่า วิธีของแลนคาสเตอร์และเควด มีประสิทธิภาพสูงที่สุดทุกการแจกแจงดังกล่าวเฉพาะ $\delta_{\beta_0} = 0$ ยกเว้นการแจกแจงแบบ Cauchy และแบบ Double Exponential จะมีประสิทธิภาพเท่ากัน และสำหรับ $\delta_{\beta_1} = 0$ จะมีประสิทธิภาพเท่ากันทุกการแจกแจงดังกล่าว ซึ่งสามารถกล่าวโดยสรุปได้ดังนี้คือ เมื่อความคลาดเคลื่อนมีการแจกแจงลักษณะหางยาวหรือกระจายไปทางหางมาก เช่น การแจกแจงแบบ Double Exponential วิธีของแลนคาสเตอร์และเควดจะมีประสิทธิภาพสูงที่สุด