บทที่ ๒

วิธีการวิเคราะห์การถดถอยและสหสัมพันธ์

สมมติว่าเรามีตัวแปร ๒ ตัว ถ้าค่าของตัวแปรแต่ละตัวที่เกิดขึ้นต่างเกิดขึ้นโดยไม่เกี่ยวข้องกัน กล่าวคือ การเกิดค่าของตัวแปรตัวหนึ่งไม่มีผลเกี่ยวข้องต่อการเกิดค่าของตัวแปรอีกตัวหนึ่งเลย เราเรียกว่าตัวแปรทั้ง ๒ ตัวไม่มีความสัมพันธ์กัน แต่ถ้าค่าของตัวแปรทั้ง ๒ ตัวต่างเกิดขึ้นอย่างไม่เป็นอิสระ กล่าวคือถ้าค่าของตัวแปรตัวหนึ่งเปลี่ยนไป ค่าของตัวแปรอีกตัวหนึ่งก็เปลี่ยนตามไปด้วย ทำให้เกิดปัญหาที่น่าสนใจติดตามคือ ค่าของตัวแปรแต่ละตัวที่เกิดขึ้นนั้น มีความสัมพันธ์ต่อกันหรือไม่อย่างไร ซึ่งในทางสถิติมีเทคนิคอย่างหนึ่งที่สามารถนำมาใช้ในการหาความสัมพันธ์ระหว่างตัวแปรตัวหนึ่ง กับตัวแปรอีกตัวหนึ่งหรือหลายตัว เทคนิคนี้เรียกว่าการวิเคราะห์การถดถอย (Regression Analysis)

จากการวิเคราะห์การถดถอยจะทำให้ทราบถึงความสัมพันธ์ระหว่างตัวแปรตัวหนึ่งหรือหลายตัว ซึ่งต่อไปจะเรียกว่า ตัวแปรอิสระ (Independent Variables) ว่ามีอิทธิพลต่อการเกิดค่าของตัวแปรอีกตัวหนึ่ง ซึ่งต่อไปจะเรียกว่าตัวแปรตาม (Dependent Variable) หรือไม่ หากผลของการวิเคราะห์การถดถอยพบว่า มีความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตามและเราต้องการที่จะนำผลของการวิเคราะห์ไปใช้ในการอธิบายเกี่ยวกับการคาดคะเนค่าของตัวแปรตามว่าจะใช้คาดคะเนได้ดีเพียงใด เราเรียกว่าการวิเคราะห์สหสัมพันธ์ (Correlation Analysis) มาตรการที่ใช้วัดความใช้ได้ของความสัมพันธ์เรียกว่าสัมประสิทธิ์สหสัมพันธ์[๑] (Coefficient of Correlation)

---

[๑] เอกชัย ชัยประเสริฐสิทธิ์, การวิเคราะห์สหสัมพันธ์และการถดถอย (พระนคร : โรงพิมพ์มหาวิทยาลัยธรรมศาสตร์, ๒๕๒๐), หน้า ๒.

<u>การวิเคราะห์สหสัมพันธ์และการถดถอยอย่างง่าย</u> (Simple Correlation And Regression Analysis)

เป็นการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรตาม ๑ ตัว กับตัวแปรอิสระเพียง ๑ ตัว

<u>การวิเคราะห์สหสัมพันธ์และการถดถอยเชิงซ้อน</u> (Multiple Correlation And Regression Analysis)

เป็นการวิเคราะห์หาความสัมพันธ์ระหว่างตัวแปรตาม ๑ ตัว กับตัวแปรอิสระมากกว่า ๑ ตัว

วิทยานิพนธ์นี้เป็นการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรตาม กับตัวแปรอิสระใหญ่ ๆ ๔ ตัว ดังนั้นในที่นี้จะกล่าวถึงเฉพาะการวิเคราะห์แบบเชิงซ้อนเท่านั้น

<u>รูปแบบการถดถอยเชิงซ้อน</u> (Multiple Regression Model)

เพื่อเป็นการง่ายและสะดวกในการวิเคราะห์การถดถอย จะถือว่าความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระเป็นไปในลักษณะที่เป็นเส้นตรง ดังนั้น ถ้ากำหนดให้ Y เป็นตัวแปรผันตาม และ X เป็นตัวแปรผันอิสระแล้ว สมการเส้นตรงของการถดถอยเชิงซ้อนจะมีรูปแบบดังนี้

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 \,.....\, + \beta_n X_n + E$$

Y	คือ	ตัวแปรผันตาม
α	คือ	ตัวคงที่และเป็นค่าของ Y เมื่อค่าของ X แต่ละตัวมีค่าเป็นศูนย์
β	คือ	ค่าที่เปลี่ยนแปลงไปของ Y เมื่อ X แต่ละตัวมีค่าเปลี่ยน-แปลงไป ๑ หน่วย ค่าของ B จึงเป็นความชันของเส้นตรง หรือเรียกว่าสัมประสิทธิ์การถดถอย

E คือ ค่าของ Y ที่ต่างไปจากเส้นถดถอยที่แท้จริง

$X_1 \ldots X_n$ คือ ตัวแปรผันอิสระ

สมการของการถดถอยข้างต้นนี้เป็นสมการถดถอยของประชากร ซึ่งมีจำนวนที่ใหญ่มาก และยากที่จะทราบจำนวนที่แน่นอน ดังนั้น จะต้องใช้วิธีการมองในแง่ของการอ้างอิงเชิงสถิติ จากตัวอย่างชุดใดชุดหนึ่งไปสู่ประชากรอันเป็นที่มาของตัวอย่างชุดนั้น เส้นถดถอยของตัวอย่างชุดหนึ่งเป็นเพียงเส้นถดถอยในหลาย ๆ เส้นที่ได้จากตัวอย่างชุดต่าง ๆ ซึ่งทำการสุ่มตัวอย่างมาจากประชากรกลุ่มเดียวกัน

จากการสุ่มตัวอย่างประชากร จะได้สมการถดถอยใหม่ดังนี้

$$\hat{y} = a + b_1x_1 + b_2x_2 \ldots\ldots + b_nx_n + e$$

โดยกำหนดให้ $\hat{y}$ คือ ค่ากะประมาณของ Y

a คือ ค่ากะประมาณของ α

b คือ ค่ากะประมาณของ β

e คือ ค่ากะประมาณของ E

เพื่อให้การอ้างอิงเกี่ยวกับความสัมพันธ์ของประชากรจากข้อมูลตัวอย่างเป็นการอ้างอิงที่ใช้ได้จึงจำเป็นจะต้องกำหนดข้อสมมติบางประการต่อไปนี้

๑. ความสัมพันธ์ของประชากรจะต้องเป็นไปในลักษณะที่เป็นเส้นตรง

๒. ค่า E แต่ละค่าเป็นอิสระจากกัน และมีการแจกแจงเป็นปกติ

๓. ตัวแปรอิสระแต่ละตัวเป็นอิสระจากกัน

๔. จำนวนค่าสังเกต (Observations) ทั้งหมดต้องมากกว่าจำนวนพารามิเตอร์ที่ต้องการกะประมาณ

การหาสมการการถดถอยโดยใช้วิธีกำลังสองน้อยที่สุด (Least Square Method)

วิธีการเขียนเส้นตรงที่ถูกต้องและอาจนำไปใช้กับข้อมูลที่มีจำนวนมากได้สะดวก คือ ต้องอาศัยวิธีกำลังสองน้อยที่สุด สำหรับเส้นกำลังสองน้อยที่สุดจะหาค่ากะประมาณของพารามิเตอร์โดยแก้สมการที่เรียกว่า สมการปกติ (Normal Equations)

ความคลาดเคลื่อนมาตรฐานของการกะประมาณ (Standard Error of Estimates)

การใช้ประโยชน์จากเส้นถดถอยเพื่อวัตถุประสงค์ในการคาดคะเนล่วงหน้าขึ้นอยู่กับว่าข้อมูลที่รวบรวมได้ กระจัดกระจายไปจากเส้นถดถอยที่คำนวณได้มากน้อยเพียงใด ถ้าค่าของ Y ที่สังเกตได้แตกต่างไปจากเส้นถดถอยมาก ค่า Y ที่กะประมาณจากเส้นนี้ก็มีโอกาสที่จะแตกต่างไปจากค่าที่เกิดขึ้นจริงได้ง่าย และไม่อาจนำไปใช้ประโยชน์ในการคาดคะเนล่วงหน้าได้ดีเท่าที่ควร ในทางตรงกันข้าม ถ้าค่าของ Y ที่สังเกตอยู่ใกล้กับเส้นถดถอยแล้ว การกะประมาณค่าของ Y จากเส้นถดถอยที่คำนวณได้ก็จะมีความเชื่อถือได้มากขึ้น มาตรการที่ใช้วัดการกระจัดกระจายของข้อมูลที่สังเกตได้จากเส้นถดถอยเรียกว่า ความคลาดเคลื่อนมาตรฐานของการกะประมาณ[๑]

สัมประสิทธิ์สหสัมพันธ์ (Coefficient of Correlation)

สัมประสิทธิ์สหสัมพันธ์ เป็นมาตรการที่ใช้วัดความสัมพันธ์ระหว่างตัวแปรผันหรือถ้าจะกล่าวให้เฉพาะเจาะจง อาจจะนิยามสัมประสิทธิ์สหสัมพันธ์ได้ว่า เป็นมาตรการที่ชี้ให้เห็นว่าตัวแปรอิสระมีส่วนทำให้เกิดการเปลี่ยนแปลงขึ้นในตัวแปรตามมากน้อยเพียงใด สัญญลักษณ์ที่ใช้แทนสัมประสิทธิ์สหสัมพันธ์ในที่นี้คือ R ค่าของ R จะอยู่ระหว่าง +1 กับ -1 กล่าวคือ ถ้าค่า R เป็นศูนย์จะชี้ให้เห็นว่าตัวแปรทั้งสองไม่มีสหสัมพันธ์กันเลย แต่ถ้าค่า

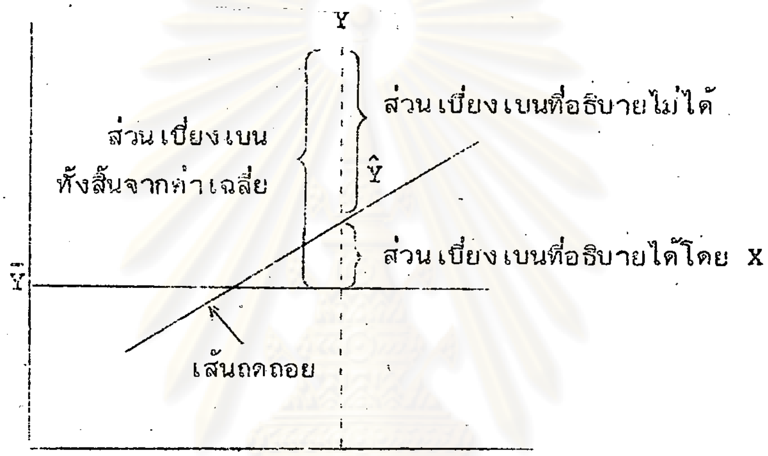
---

[๑] เรื่องเดียวกัน, หน้า ๑๕.

ของ R เข้าใกล้ +1 หรือ -1 มากเพียงใด ยิ่งแสดงให้เห็นถึงการมีสหสัมพันธ์ต่อกันมากเท่านั้น แนวความคิดเกี่ยวกับเรื่องนี้แสดงให้เห็นได้จากรูปที่ ๒-๑ ต่อไปนี้

รูปที่ ๒-๑

การแยกส่วนเบี่ยงเบนของตัวแปรตาม



จากรูปที่ ๒-๑ นี้ จะสังเกตเห็นว่าส่วนเบี่ยงเบนทั้งสิ้นของตัวแปรตาม Y จากค่าเฉลี่ยเลขคณิต Y อาจแยกออกได้เป็น ๒ ส่วน คือ ส่วนแรกได้แก่ ส่วนเบี่ยงเบนของค่าที่อยู่บนเส้นถดถอย จากค่าเฉลี่ยเลขคณิต $(\hat{Y} - \bar{Y})$ ซึ่งอาจอธิบายได้โดยค่า X ที่กำหนดให้ ส่วนที่สองได้แก่ส่วนเบี่ยงเบนของ Y จากค่าที่อยู่บนเส้นถดถอย $(Y - \hat{Y})$ ซึ่งไม่อาจอธิบายได้โดยค่าของ X ดังนั้น จะได้ว่า

ส่วนเบี่ยงเบนทั้งสิ้น = ส่วนเบี่ยงเบนที่อธิบายได้ + ส่วนเบี่ยงเบนที่อธิบายไม่ได้

(Total Variation) = (Explained Variation)+(Unexplained Variation)

หรือ $S^2_Y = S^2_{\hat{Y}.\bar{Y}} + S^2_{Y.X}$

ถ้าเอา S^2_Y หารตลอด: $1 = \dfrac{S^2_{\hat{Y}.\bar{Y}}}{S^2_Y} + \dfrac{S^2_{Y.X}}{S^2_Y}$

S^2_Y คือ ค่าเบี่ยงเบนทั้งหมดของ Y ที่ต่างไปจากค่า $\bar{Y}$

$S^2_{\hat{Y}.\bar{Y}}$ คือ ค่าของ $\hat{Y}$ ที่เบี่ยงเบนไปจากค่า $\bar{Y}$ ซึ่งสามารถจะทราบ หรืออธิบายส่วนเบี่ยงเบนนี้ได้ด้วยค่าของตัวแปรอิสระ X

$S^2_{Y.X}$ คือ ค่าของ Y ที่เบี่ยงเบนไปจากค่า $\hat{Y}$ ซึ่งไม่สามารถจะทราบ หรืออธิบายส่วนเบี่ยงเบนนี้ได้ด้วยค่าของตัวแปรอิสระ X

ดังนั้น $\dfrac{S^2_{\hat{Y}.\bar{Y}}}{S^2_Y}$ ก็คือ สัดส่วนของส่วนเบี่ยงเบนที่สามารถอธิบายได้ด้วยตัวแปรอิสระ X จากส่วนเบี่ยงเบนที่เกิดขึ้นทั้งหมด

ส่วน $\dfrac{S^2_{Y.X}}{S^2_Y}$ ก็คือ สัดส่วนของส่วนเบี่ยงเบนที่ไม่สามารถอธิบายได้ด้วยตัวแปรอิสระ X จากส่วนเบี่ยงเบนทั้งหมดที่เกิดขึ้น

ถ้ากำหนดให้ $R^2 = S^2_{\hat{Y}.\bar{Y}}/S^2_Y$ ค่าของ R^2 จะชี้ให้เห็นว่าการเปลี่ยนแปลงไปของค่า Y ที่ต่างไปจากค่า $\bar{Y}$ นั้น ส่วนหนึ่งสามารถจะทราบหรืออธิบายการเปลี่ยนแปลงไปของ Y ได้ด้วยตัวแปรอิสระ X นั่นคือ ค่า R^2 จะชี้ให้เห็นว่าการเปลี่ยนแปลงไปของค่า Y ส่วนหนึ่งนั้น เกิดจากการเปลี่ยนแปลงไปของค่า X เป็นจำนวนมากน้อยเพียงใด

จากสมการ $$\frac{S^2_{\hat{Y}.\bar{Y}}}{S^2_Y} + \frac{S^2_{Y.X}}{S^2_Y} = 1$$

ดังนั้น $$\frac{S^2_{\hat{Y}.\bar{Y}}}{S^2_Y} = 1 - \frac{S^2_{Y.X}}{S^2_Y}$$

นั่นคือ $$R^2 = 1 - \frac{S^2_{Y.X}}{S^2_Y}$$

ค่าของ R^2 จะเป็น + เสมอ และมีค่าไม่เกิน 1 ถ้า R^2 มีค่าเข้าใกล้ 1 มากเพียงใดก็แสดงว่าการเปลี่ยนแปลงไปของตัวแปรตาม Y สามารถอธิบายได้ด้วยตัวแปรอิสระ X มากเพียงนั้น นั่นคือ R^2 จะชี้ให้เห็นระดับความสัมพันธ์ระหว่างตัวแปรตาม Y กับตัวแปรอิสระ X ว่ามีมากหรือน้อยเพียงใด จึงเรียกค่า R^2 ว่าสัมประสิทธิ์ของการตัดสินใจ (Coefficient of Determination)

สัมประสิทธิ์ของการตัดสินใจบางส่วน (Coefficient of Partial Determination)

สัมประสิทธิ์ของการตัดสินใจตามที่กล่าวมาแล้วเป็นการวัดอิทธิพลของตัวแปร-อิสระหลาย ๆ ตัวที่มีต่อตัวแปรตามร่วมกัน แต่สัมประสิทธิ์ของการตัดสินใจในหัวข้อนี้จะเป็นการวัดอิทธิพลของตัวแปรอิสระทีละตัวกับตัวแปรตามว่า ตัวแปรอิสระตัวใดจะมีอิทธิพลต่อตัวแปรตามมากน้อยกว่ากัน โดยในการวิเคราะห์เพื่อหาสัมประสิทธิ์ของการตัดสินใจบาง-ส่วนของตัวแปรอิสระตัวหนึ่งกับตัวแปรตามนั้น จะถือว่าตัวแปรอิสระตัวอื่นอยู่คงที่

เปรียบเทียบการนำ R^2 และ R ไปใช้ประโยชน์

บางครั้ง R^2 จะเป็นมาตรการที่ดีกว่า R เนื่องจากค่าของ R อาจชี้ให้เห็นสหสัมพันธ์ที่มีค่าสูงกว่าที่ควรจะเป็น เช่น ถ้าค่าเบี่ยงเบนของ Y ร้อยละ ๕๐ อาจอธิบายได้โดยตัวแปรอิสระ X ส่วนอีกร้อยละ ๕๐ ไม่อาจอธิบายได้โดย X

ดังนั้น $R^2 = 0.50$

แต่เมื่อหาค่าของ R จะ $= \sqrt{0.50} = 0.71$ ซึ่งสูงกว่าที่ควรจะเป็น

การทดสอบความมีนัยสำคัญทางด้านความสัมพันธ์

บางครั้งตัวอย่างที่สุ่มมาอาจจะทำให้เห็นไปได้ว่า ตัวแปรตาม Y กับตัวแปรอิสระ X ทุกตัวมีความสัมพันธ์กัน ซึ่งจริง ๆ แล้วอาจเป็นเรื่องที่เกิดขึ้นโดยบังเอิญก็ได้ อย่างไรก็ตาม ถ้าไม่มีความสัมพันธ์ระหว่าง Y กับ X ทุกตัวเลย ค่าของสัมประสิทธิ์

การถดถอย $\beta_1, \beta_2, \beta_3 \ldots\ldots \beta_n$ จะเท่ากับศูนย์ เพราะฉะนั้นจึงสามารถที่จะตั้งสมมติฐานเพื่อใช้ในการทดสอบได้ดังนี้

$$H_o : \beta_1 = \beta_2 = \beta_3 \ldots\ldots\ldots\ldots = \beta_n = 0$$

$$H_a : \beta_1 \neq \beta_2 \neq \beta_3 \ldots\ldots\ldots\ldots \neq \beta_n = 0$$

ตัวสถิติที่ใช้ในการทดสอบสมมติฐาน คือ

$$F = \frac{\text{Mean Square (Due to Regression)}}{\text{Error Mean Square}}$$

ถ้าค่าของ F ที่คำนวณได้นี้มากกว่าค่า $F_{1-\frac{\alpha}{2}, (k-1)(n-k)}$ ซึ่งได้จากการเปิดตารางโดย n คือ จำนวนตัวอย่าง และ k คือจำนวนพารามิเตอร์ทั้งหมดในการวิเคราะห์แล้ว ก็จะปฏิเสธสมมติฐาน H_o และยอมรับสมมติฐาน H_a นั่นคือ ยอมรับว่าตัวแปรตาม Y มีความสัมพันธ์กับตัวแปรอิสระทุกตัวร่วมกัน

สำหรับการทดสอบว่าเฉพาะตัวแปรอิสระแต่ละตัวจะมีความสัมพันธ์กับตัวแปรตามหรือไม่ ก็สามารถทดสอบโดยใช้ F-test เช่นกัน แต่สมมติฐานที่ใช้ทดสอบจะเป็นดังนี้

$$H_o : \beta_i = 0 \quad (i = 1, 2, 3, \ldots\ldots, n)$$

$$H_a : \beta_i \neq 0$$

ตัวสถิติที่ใช้ในการทดสอบสมมติฐาน คือ

$$F = \frac{\text{Mean Square (Due to } X_i)}{\text{Error Mean Square}}$$

ถ้าค่าของ F ที่คำนวณได้มากกว่าค่า F จากการเปิดตาราง ก็แสดงว่าปฏิเสธสมมติฐาน H_o และยอมรับสมมติฐาน H_a นั่นคือ ยอมรับว่าตัวแปรอิสระ X ที่ทดสอบตัวนั้นมีความสัมพันธ์กับตัวแปรตาม Y

<u>วิธีการเลือกตัวแบบการถดถอย</u>

ถ้าตัวแปรอิสระมีมาก สมมติว่ามี k ตัวปัญหาที่อาจจะต้องพิจารณาก็คือ ควรจะเลือกใช้ตัวแปรอิสระกี่ตัวจึงจะเหมาะสม เนื่องจากถ้าใช้ตัวแปรอิสระทุกตัวก็อาจจะต้องเสียเวลาและค่าใช้จ่ายมาก หากใช้น้อยตัวก็อาจจะให้ข้อเท็จจริงเกี่ยวกับตัวแปรตามไม่เพียงพอ การพิจารณาว่าตัวแปรอิสระใดควรจะเข้าไปร่วมอยู่ในสมการการถดถอยนั้น เรียกว่าวิธีการเลือกตัวแบบการถดถอย ซึ่งมีอยู่หลายวิธี คือ[๑]

๑. การเลือกตัวแบบการถดถอยโดยการพิจารณาทั้งหมด (All Possible Regressions)

๒. การเลือกตัวแบบการถดถอยแบบถอยหลัง (Backward Elimination Regressions)

๓. การเลือกตัวแบบการถดถอยไปข้างหน้า (Forward Selection Procedure)

๔. การเลือกตัวแบบการถดถอยแบบขั้นบันได (Stepwise Regression Procedure)

๕. การเลือกตัวแบบการถดถอยโดยใช้หลักเกณฑ์ผสม

๖. การเลือกตัวแบบการถดถอยเป็นขั้นตอน (Stagewise Regression Procedure)

วิทยานิพนธ์นี้จะใช้วิธีการเลือกตัวแบบการถดถอยแบบขั้นบันได ดังนั้น จึงจะกล่าวถึงการเลือกตัวแบบด้วยวิธีนี้เพียงวิธีเดียวเท่านั้น

---

[๑] สถาบันบัณฑิตพัฒนบริหารศาสตร์, คณะสถิติประยุกต์, "การวิเคราะห์การถดถอย," ใน <u>เอกสารประกอบคำบรรยายวิชาสถิติเศรษฐศาสตร์และธุรกิจ</u>, หน้า ๑.

ตัวแบบการถดถอยโดยวิธีแบบขั้นบันได[๑] (Stepwise Regression Procedure)

ตัวแบบการถดถอยโดยวิธีนี้ เป็นวิธีที่พยายามเพิ่มจำนวนตัวแปรอิสระเข้าไปในสมการทีละตัวจนกว่าจะได้สมการการถดถอยเป็นที่น่าพอใจ สำหรับในแต่ละขั้นที่มีการเพิ่มจำนวนตัวแปรอิสระเข้าไปในตัวแบบ จะมีการพิจารณาตัวแปรที่รวมอยู่ในตัวแบบไว้แล้วอีกด้วย ดังนั้น จึงมีการทดสอบ $H_o : \beta_i = 0$ ทุก ๆ ค่าของ X_i ในตัวแบบ หากพบว่ายอมรับ H_o นี้เมื่อไรก็จะเอา X_i ตัวนั้นออกจากตัวแบบ การพิจารณาตัวแปรอิสระที่จะเข้าไปร่วมอยู่ในสมการการถดถอยจะเสร็จสิ้นก็ต่อเมื่อไม่มีตัวแปรอิสระตัวใดจะร่วมอยู่ในตัวแบบได้อีก และไม่มีตัวแปรอิสระใดถูกกำจัดออกจากตัวแบบนั้น วิธีการนี้จึงมีขั้นตอนดังต่อไปนี้

๑. หาเมทริกซ์สหสัมพันธ์ (Correlation Matrix) ระหว่างตัวแปรแต่ละคู่ แล้วเลือก X ที่มีสัมประสิทธิ์สหสัมพันธ์กับ Y สูงที่สุด สมมติว่าเป็น X_1 เข้ามาพิจารณาในสมการการถดถอย

๒. คำนวณหาสัมประสิทธิ์สหสัมพันธ์บางส่วนระหว่าง Y กับ X ที่เหลือ โดยถือว่า X_1 อยู่ในตัวแบบแล้วจากนั้นเลือกตัวแปรอิสระที่ให้ค่ากำลังสองของสหสัมพันธ์บางส่วนสูงที่สุด เข้ามาในตัวแบบ สมมติว่าเป็น X_2

๓. พิจารณาสมการการถดถอย $\hat{Y} = a + b_1X_1 + b_2X_2$ โดยทดสอบสมมติฐานที่ว่า $H_o : \beta_1 = 0$ และ $H_o : \beta_2 = 0$ หากยังไม่ยอมรับ H_o ทั้งสอง ก็ถือว่า X_1 และ X_2 ควรจะอยู่ในตัวแบบ

๔. คำนวณหาสัมประสิทธิ์สหสัมพันธ์บางส่วนระหว่าง Y กับ X ที่เหลือ โดยถือว่า X_1 และ X_2 อยู่ในสมการแล้ว จากนั้นเลือก X ที่มีกำลังสองของสัมประสิทธิ์สหสัมพันธ์บางส่วนสูงที่สุดเข้ามาพิจารณาในสมการการถดถอย สมมติว่าเป็น X_3

---

[๑] เรื่องเดียวกัน, หน้า ๓ - ๔.

๕. พิจารณาสมการถดถอย $\hat{Y} = a + b_1X_1 + b_2X_2 + b_3X_3$ โดยทดสอบ $H_0 : \beta_i = 0$ (ถือว่า X_i เข้าไปร่วมอยู่ในตัวแบบ ๓ ตัว, $i = 1, 2, 3$) ผลการพิจารณาอาจจะทิ้ง X บางตัวที่เลือกไว้แล้วไปก็ได้

๖. ดำเนินการต่อไปเช่นเดิม กล่าวคือ คำนวณสัมประสิทธิ์สหสัมพันธ์บางส่วนระหว่าง Y กับ X ที่เหลืออยู่อีก แล้วเลือก X ที่ให้ค่ากำลังสองของสัมประสิทธิ์สหสัมพันธ์บางส่วนสูงที่สุดเข้ามาในสมการ แล้วทดสอบ $H_0 : \beta_i = 0$ ทำเช่นนี้จนกระทั่งไม่มีตัวแปรอิสระใดจะเข้าไปร่วมในตัวแบบได้อีก และไม่มี X ตัวใดถูกตัดออกจึงจะยุติ