บทที่ 3

การใช้สมการการถดถอยและค่าสหสัมพันธ์ในการวิเคราะห์

การวิเคราะห์การถดถอย (Regreesion Analysis) เป็นการใช้เทคนิคทางสถิติเข้ามาช่วยหาความสัมพันธ์ระหว่างตัวแปรตัวหนึ่ง หรือหลายตัว ซึ่งเราเรียก ตัวแปรอิสระ (Independent Variables) ว่า มีอิทธิพลต่อตัวแปรอีกตัวหนึ่ง ซึ่งเรียกว่า ตัวแปรตาม (Dependent Variable) อย่างไร สัมพันธ์ในรูปใด โดยแสดงลักษณะสัมพันธ์นั้นในรูปสมการการถดถอย (Regression Equation)

การวิเคราะห์ค่าสหสัมพันธ์ (Correlation Analysis) เป็นเรื่องที่เกี่ยวกับปัญหาโดยทั่ว ๆ ไปในการวัดระดับความสัมพันธ์ระหว่างตัวแปรต่าง ๆ โดยวัดออกมาเป็นสัมประสิทธิ์สหสัมพันธ์ (Coefficient of Correlation) และสัมประสิทธิ์ของการตัดสินใจ (Coefficient of Determination) ซึ่งรวมไปถึงการทดสอบความสัมพันธ์นี้ โดยใช้ F - test หรือ t - test

การวิเคราะห์ที่เป็นการนำเอาตัวแปรตามตัวหนึ่งกับตัวแปรอิสระเพียงตัวเดียว มาทำการวิเคราะห์ เรียกว่า การถดถอยหรือสหสัมพันธ์อย่างง่าย (Simple Regression or Correlation) แต่ถ้าตัวแปรอิสระมีมากกว่า 1 ตัว เรียกว่า การถดถอย หรือสหสัมพันธ์เชิงซ้อน (Multiple Regression or Correlation)

ในวิทยานิพนธ์ได้กำหนดให้มีตัวแปรอิสระ 5 ตัว จึงเป็นการวิเคราะห์แบบเชิงซ้อน ซึ่งจะได้กล่าวถึงเฉพาะค่าที่จำเป็นต้องใช้ในการวิเคราะห์เท่านั้น จากนั้นจะได้กล่าวถึงวิธีการที่เรียกว่า Step-Wise ซึ่งเป็นเทคนิคที่ใช้ในการนำเอาตัวแปรอิสระเข้ามาพิจารณาทีละตัว

<u>รูปแบบถดถอยเชิงซ้อน</u> (Multiple Regression)

สมการถดถอยเชิงซ้อน เป็นสมการแสดงให้เห็นอิทธิพลของตัวแปรกลุ่มหนึ่งที่มีต่อตัวแปรตามร่วมกัน โดยปกติแบบจำลอง (Model) ที่ใช้ในการวิเคราะห์มี 2 แบบ คือ

1. ตัวแบบเส้นตรง (Linear Regression Model)
2. ตัวแบบที่ไม่ใช่เส้นตรง (Non - Linear Regression Model)

ในที่นี้จะใช้ตัวแบบเส้นตรงในการวิเคราะห์ เพราะเป็นการสะดวกกว่าการใช้แบบที่ไม่ใช่เส้นตรง มีรูปสมการเป็นดังนี้

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 \cdots\cdots + \beta_n X_n + \epsilon$$

โดยที่ Y = ตัวแปรตาม

$X_1, X_2 \cdots\cdots X_n$ = ตัวแปรอิสระแต่ละตัว

α = ค่าคงที่ ซึ่งเป็นค่าของ Y เมื่อ X ทั้งหมดมีค่าเป็นศูนย์

$\beta_1, \beta_2, \cdots\cdots \beta_n$ = เป็นสัมประสิทธิ์ของตัวแปรอิสระแต่ละตัว ซึ่งเป็นค่าที่แสดงว่าเมื่อ X แต่ละตัวเปลี่ยนแปลงไป 1 หน่วย Y จะเปลี่ยนแปลงไปเท่าใด ซึ่งเรียกว่าพารามิเตอร์ถดถอย นั่นคือ ความชันของเส้นถดถอยนั่นเอง

ϵ = ความคลาดเคลื่อนเชิงสุ่ม ซึ่งแสดงว่าค่า Y ที่ได้แตกต่างไปจากเส้นถดถอยที่แท้จริงเท่าไร การแจกแจงปกติมีค่าเฉลี่ยเป็นศูนย์ และความแปรปรวนเท่ากันหมด

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ข้อกำหนดของตัวแบบถดถอยเชิงซ้อนชนิดเส้น [1]

1. ε เป็นอิสระซึ่งกันและกัน และแต่ละตัวมีการแจกแจงปกติที่มีค่าเฉลี่ย ๐ ความแปรปรวน σ^2

2. ตัวแปรอิสระแต่ละตัวสามารถควบคุมได้ และสำหรับตัวอย่างขนาด n จะมีค่า $\Sigma (X_i - \bar{X})^2 / n$ (i = 1, 2, ... k) มากกว่า ๐ และเป็นจำนวนจำกัด

3. จำนวนค่าสังเกต (Observations) ทั้งหมดต้องมากกว่าจำนวนพารามิเตอร์ถดถอยที่ต้องการประมาณ

4. ไม่มีความสัมพันธ์ระหว่างตัวแปรอิสระ

สมการถดถอยดังกล่าวข้างต้น เป็นสมการของประชากรทั้งหมด ซึ่งเป็นจำนวนที่ใหญ่มาก และไม่ทราบจำนวนแน่นอน ในการวิเคราะห์จึงต้องทำการสุ่มตัวอย่าง (Sampling) ประชากรขึ้นมา และจะได้สมการการถดถอยใหม่ ดังนี้

$$\hat{Y} = a + b_1X_1 + b_2X_2 ..... + b_nX_n + e$$

โดยที่ $\hat{Y}$ เป็น ค่าประมาณของ Y

a เป็น ค่าประมาณของ α

b เป็น ค่าประมาณของ β

e เป็น ค่าประมาณของ ε นั่นคือผลต่างของ $Y - \hat{Y}$

---

[1] วิทยาลัยการค้า, คณะวิชาเศรษฐศาสตร์, "การวิเคราะห์การถดถอย", ในเอกสารประกอบคำบรรยายวิชา หลักสถิติ 2, เลขที่ 785/2518.

การประมาณค่าของพารามิเตอร์ (The Estimation of Parameters)

การหาค่าของ a และ b ในสมการถดถอยเชิงซ้อนเส้นตรง วิธีที่สะดวกที่สุดคือใช้วิธีกำลังสองน้อยที่สุด (Least Square Method) ตามวิธีนี้เป็นการแก้สมการเส้นตรงเพื่อหาตัวไม่ทราบค่า คือ ค่า a 1 ค่า และค่า b ตามจำนวนตัวแปรตาม ซึ่งในการวิจัยนี้มีตัวแปรอิสระ 5 ตัว ดังสมการจะเป็นดังนี้

$$\hat{Y} = a + b_1X_1 + b_2X_2 \ldots\ldots\ldots + b_5X_5$$

สมการที่จะต้องแก้คือ [1]

$$\Sigma Y = na + b_1\Sigma X_1 + b_2\Sigma X_2 + \ldots\ldots + b_5\Sigma X_5 \quad \ldots\ldots(1)$$

$$\Sigma X_1Y = a\Sigma X_1 + b_1\Sigma X_1^2 + b_2\Sigma X_1X_2 + \ldots\ldots + b_5\Sigma X_1X_5 \quad \ldots\ldots(2)$$

$$\Sigma X_2Y = a\Sigma X_2 + b_1\Sigma X_1X_2 + b_2\Sigma X_2^2 + \ldots\ldots + b_n\Sigma X_2X_5 \ldots(3)$$

$$\Sigma X_5Y = a\Sigma X_5 + b_1\Sigma X_1X_5 + b_2\Sigma X_2X_5 + \ldots\ldots + b_5\Sigma X_5^2$$

ในการแก้สมการ ถ้าใช้ค่าเบี่ยงเบนจากค่าเฉลี่ยของมันในการคำนวณ จะง่ายกว่าแก้จากสมการข้างต้นโดยตรง โดยใช้ $x_1 = X_1 - \bar{X}$, $x_2 = X_2 - \bar{X}$ ..... $x_n = X_n - \bar{X}$ และ $y = Y - \bar{Y}$ ซึ่งวิธีที่จะให้ได้ค่าพวกนี้ง่ายที่สุดคือ รวมกำลังสองและผลคูณของตัวแปรเดิม คือ X และ Y แล้วหักด้วยผลคูณของค่าเฉลี่ย กับผลรวมของตัวแปรนั้น ก็จะได้ผลรวมของ x และ y ดังนี้

---

[1] เอกชัย ชัยประเสริฐสิทธิ์, การวิเคราะห์สหสัมพันธ์และการถดถอย (พระนคร : โรงพิมพ์มหาวิทยาลัยธรรมศาสตร์, 2520), หน้า 69 – 70 .

ผลรวมของตัวแปร	ปรับปรุงด้วย (ค่าเฉลี่ยคูณผลรวม)	ยอดรวมที่ปรับปรุงแล้ว
$\sum X_1^2$	$-\bar{X}_1 \sum X_1$	$\sum x_1^2$
$\sum X_2^2$	$-\bar{X}_2 \sum X_2$	$\sum x_2^2$
$\vdots$	$\vdots$	$\vdots$
$\sum X_n^2$	$-\bar{X}_n \sum X_n$	$\sum x_n^2$
$\sum Y^2$	$-\bar{Y} \sum Y$	$\sum y^2$
$\sum X_1 Y$	$-\bar{X}_1 \sum Y$	$\sum x_1 y$
$\sum X_2 Y$	$-\bar{X}_2 \sum Y$	$\sum x_2 y$
$\vdots$	$\vdots$	$\vdots$
$\sum X_5 Y$	$-\bar{X}_5 \sum Y$	$\sum x_5 y$
$\sum X_1 X_2$	$-\bar{X}_1 \sum X_2$	$\sum x_1 x_2$
$\vdots$	$\vdots$	$\vdots$
$\sum X_4 X_5$	$-X_4 \sum X_5$	$\sum x_4 x_5$

ถ้านำสมการดังกล่าวข้างต้นมาเขียนในรูปของ x และ y สมการที่ (1) จะหายไปเพราะ $\sum y$, $\sum x_1$ และ $\sum x_2$ ต่างเท่ากับศูนย์ ส่วนสมการที่เหลือจะเป็นดังนี้

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

$$\Sigma x_1 y = b_1 \Sigma x_1^2 + b_2 \Sigma x_1 x_2 + \cdots\cdots + b_5 \Sigma x_1 x_5$$

$$\Sigma x_2 y = b_1 \Sigma x_1 x_2 + b_2 \Sigma x_2^2 + \cdots\cdots + b_5 \Sigma x_2 x_5$$

$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots$$

$$\Sigma x_5 y = b_1 \Sigma x_1 x_5 + b_2 \Sigma x_2 x_5 + \cdots\cdots\cdots b_5 \Sigma x_5^2$$

เมื่อมีข้อมูลจริงก็จะได้ค่าต่าง ๆ ยังคงติดค่า $b_1, b_2 \cdots\cdots\cdots b_5$ แล้วทำการแก้สมการ 5 สมการ ซึ่งมีตัวไม่ทราบค่า (Unknown) สมการละ 5 ตัว ก็จะได้ค่า $b_1, b_2 \cdots\cdots\cdots b_5$ ตามลำดับ นำไปแทนค่าในสมการ

$$\hat{Y} = a + b_1 X_1 - b_2 X_2 \cdots\cdots\cdots + b_5 X_5$$

ก็จะได้ค่าของ a และได้สมการการถดถอย

<u>การตีความหมายสมการถดถอย</u>

สมมติว่า จากการคำนวณได้ $\hat{Y} = 0.2 + 1.3 X_1 + \cdots\cdots + .5 X_5$ Y คือ ราคาหุ้น, X_1 = อัตราเงินปันผลต่อราคาตลาด ค่า a เป็นค่า Y - intercept คือ เป็นระยะตัดแกน Y เมื่อ X = 0 หมายความว่า ถ้าไม่มีอัตราเงินปันผลต่อราคาตลาด และตัวแปรอิสระอื่น ๆ เลย ราคาหุ้นจะเท่ากับ 20 บาท

ค่า b ซึ่งเป็นความชัน (Slope) ของเส้นถดถอย (Regression Line) เป็นค่าที่แสดงว่า เมื่อ X เปลี่ยนไป 1 หน่วย Y จะเปลี่ยนไปเท่าไร จากสมการ b_1 เท่ากับ 1.3 หมายความว่า เมื่ออัตราเงินปันผลต่อราคาตลาดเพิ่มขึ้น 1 % จะทำให้ราคาหุ้นเพิ่มขึ้นเป็น 1.30 บาท ค่า b นี้เรียกว่าสัมประสิทธิ์ของการถดถอย (Regression Coefficient)

ความคลาดเคลื่อนมาตรฐานของการกะประมาณ (Standard Error of Estimates)

คือส่วนเบี่ยงเบนมาตราฐานของผลแตกต่างระหว่างค่าของ Y กับ $\hat{Y}$ $(Y - \hat{Y})$ ความคลาดเคลื่อนมาตราฐานนี้ใช้วัดการกระจายของค่า Y รอบ ๆ เส้นถดถอยทั้งหมด คำนวณได้จากสูตร

$$S_{Y.12345} = \sqrt{\frac{\sum(Y - \hat{Y})^2}{n - k}}$$

ในที่นี้ n คือจำนวนของข้อมูลที่สังเกต (Observations)

k คือจำนวนตัวแปรในสมการถดถอย (6 ตัว)

$S_{Y.12345}$ คือความคลาดเคลื่อนมาตราฐานของการกะประมาณตัวแปรตาม Y กับตัวแปรอิสระ X_1, X_2, X_3, X_4, X_5

การคำนวณ $\sum (Y - \hat{Y})^2$ โดยตรงเป็นเรื่องยุ่งยากมาก จึงใช้สูตรต่อไปนี้แทน

$$S_{Y\ 1,\ \ldots 5} = \sqrt{\frac{\sum y^2 - b_1\sum x_1y_1 \ \ldots\ldots\ b_5\sum x_5y_5}{n - k}}$$

การตีความหมายของความคลาดเคลื่อนมาตราฐานของการกะประมาณ

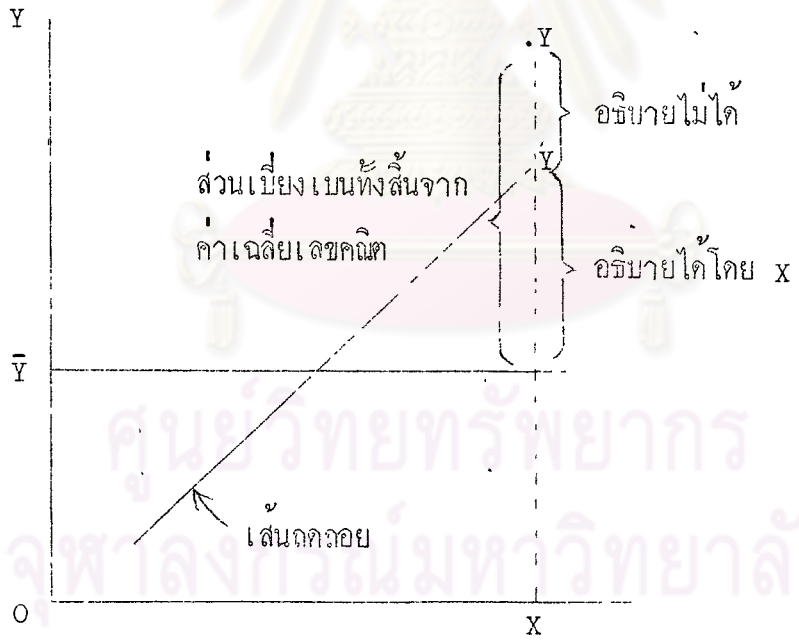
สมมติว่าได้ $S_{Y\ 1,\ldots 5} = 0.01$ หมายความว่า ถ้าราคาหุ้นกระจายอยู่รอบ ๆ เส้นถดถอยในลักษณะปกติแล้ว ราคาหุ้นนี้อาจจะสูงขึ้นหรือต่ำลงได้อีก 1 บาท

<u>สัมประสิทธิ์ของการตัดสินใจ</u> (Coefficient of Determination)

ตัวแปรตามที่กระจายอยู่รอบ ๆ เส้นถดถอยนี้ สามารถแยกออกได้เป็น 2 ส่วน คือ ส่วนแรกเป็นส่วนเบี่ยงเบนของค่าที่อยู่บนเส้นถดถอยจากค่าเฉลี่ยของมัน นั่นคือ $\hat{Y} - \bar{Y}$ ซึ่งอธิบายได้ (Explained) โดยค่าของ X ค่าใดค่าหนึ่งที่กำหนดให้ ส่วนที่สองเป็นส่วนเบี่ยงเบนของค่าของ Y จากเส้นถดถอย นั่นคือ $Y - \hat{Y}$ ซึ่งเป็นส่วนที่ไม่สามารถอธิบายได้ (Unexplained) โดยค่าของ X ดังนั้น $(Y - \bar{Y}) = (\hat{Y} - \bar{Y}) + (Y - \hat{Y})$

รูปที่ 3 – 1

การแยกความผันแปรของตัวแปรตาม



จากรูป ส่วนเบี่ยงเบนทั้งสองส่วนเป็นอิสระซึ่งกันและกัน ค่าแปรปรวนทั้งสิ้น (Total Variance) ของ Y จึงเท่ากับผลรวมของค่าแปรปรวนทั้งสองส่วน ดังนี้

Total Variance = Explained Variance + Unexplained Variance

$$S_Y^2 = S_{\hat{Y} \cdot \bar{Y}}^2 + S_{Y \cdot X}^2$$

เมื่อนำเอา Explained Variance หารด้วย Total Variance ของ Y จะได้สัมประสิทธิ์ของการตัดสินใจ ซึ่งใช้สัญลักษณ์ R^2

$$R^2 = \frac{\text{Explained Variance}}{\text{Total Variance}}$$

หรือนั่นคือ

$$R^2 = 1 - \frac{\text{Unexplained Variance}}{\text{Total Variance}}$$

$$= 1 - \frac{S^2_{Y \cdot X}}{S^2_Y}$$

$S^2_{Y \cdot X}$ คือตัวเดียวกันกับการหาความคลาดเคลื่อนจากการกะประมาณ ซึ่งได้อธิบายข้างต้นแล้ว สำหรับ S^2_Y คำนวณได้จากสูตร

$$S^2_Y = \frac{\sum (Y - \bar{Y})^2}{n - 1} \quad \text{หรือ} \quad \frac{\sum y^2}{n - 1}$$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

<u>การตีความหมายของ R^2</u>

R^2 เป็นมาตรวัดการปรับที่ดีที่สุด (Goodness of Fit) ซึ่งเป็นสัดส่วนของตัวแปร Y ที่เนื่องมาจากความผันแปรของตัวแปร X นั่นเอง นั่นคือ เป็นตัวที่ใช้บรรยายว่าเส้นถดถอยนั้นปรับเข้าได้กับข้อมูลหรือค่าสังเกตที่นำมาวิจัยได้ดีเพียงไร ค่าของ R^2 จะเป็น ดังนี้ $0 \leqslant R^2 \leqslant 1$ ถ้าตัวอย่างขนาดค่อนข้างเล็ก R^2 มีแนวโน้มที่จะเอนไปในทางบวก (Positively Biased) จึงต้องแก้ความโน้มเอียงนี้โดยนำเอา Degree of Freedom เข้ามาปรับ ดังนี้

$$R^2 = 1 - (1-R)^2 \frac{(n - 1)}{(n - 2)}$$

สมมติว่าได้ค่า R^2 = .889 หมายความว่า 88.90% ของความผันแปรใน Y เนื่องมาจากความผันแปรของค่าที่ปรับ (Fitted Value) ของ Y โดยใช้เส้นถดถอยเป็นตัวเทียบ หรือกล่าวง่าย ๆ คือ ราคาหุ้นที่เปลี่ยนแปลงนั้นอาจอธิบายได้ หรือเป็นผลสืบเนื่องมาจากตัวแปรที่กำหนดขึ้นอยู่ 88.90% อีก (1 – 88.90) = 11.10% แสดงว่าราคาหุ้นไม่ได้เป็นผลมาจากตัวแปรต่าง ๆ ที่กำหนดขึ้น

<u>สัมประสิทธิ์สหสัมพันธ์</u> (Coefficient of Correlation)

เป็นมาตราการที่ใช้วัดความสัมพันธ์ระหว่างตัวแปร 2 ตัว ค่าที่ได้จะเช่นเดียวกับ R^2 คือ อยู่ระหว่าง – 1 กับ +1 ใช้สัญญลักษณ์ R แทนในการกล่าวถึง เราคำนวณหาค่าของ R ได้โดยถอดรากกำลังที่ 2 ของ R^2

<u>การตีความหมายของ R</u>

สมมุติว่าได้ค่า R = .992 ชี้ให้เห็นระดับความสัมพันธ์ (Degree of Relationships) ระหว่างราคาหุ้นกับตัวแปรที่กำหนดขึ้น จากค่าตัวอย่างนี้แสดงให้เห็นว่า ราคาหุ้นมีความสัมพันธ์กับตัวแปรที่กำหนดให้ไปในทางเดียวกัน

<u>สหสัมพันธ์บางส่วน</u> (Partial Correlation)

เป็นมาตราการวัดสหสัมพันธ์สุทธิระหว่างตัวแปรอิสระแต่ละตัวกับตัวแปรตามค่าที่คำนวณได้ แสดงถึงความแตกต่างที่เกิดขึ้นกับตัวแปรตาม ซึ่งมิอาจอธิบายได้โดยตัวแปรอิสระอื่น ๆ แต่อาจอธิบายได้โดยการนำเอาตัวแปรอิสระตัวอื่น ๆ เข้ามา ในการวิเคราะห์สหสัมพันธ์บางส่วนถือว่าอิทธิพลของตัวแปรอิสระอื่น ๆ อยู่คงที่ ซึ่งแตกต่างไปจากการวิเคราะห์สหสัมพันธ์ที่กล่าวมาแล้ว เพราะที่กล่าวมาแล้วเป็นการพิจารณาความสัมพันธ์ระหว่างตัวแปร 2 ตัวเท่านั้น

สูตรในการคำนวณหาสัมประสิทธิ์การตัดสินใจบางส่วน (Coefficient of Partial Correlation) จะเป็นดังนี้ สมมติว่า เดิมมีสมการถดถอย

$$\hat{Y} = a + b_1 X_1$$

[1] เมื่อตัวแปรอิสระเพิ่มมา 1 ตัว คือ X_2 และจะนำเข้ามาพิจารณาในสมการการถดถอยด้วย ก็จะให้ค่า X_1 คงที่ สูตรที่ใช้ในการคำนวณหาสัมประสิทธิ์ การตัดสินใจบางส่วนจะเป็นดังนี้

$$R^2_{Y2.1} = 1 - \frac{S^2_{Y.12}}{S^2_{Y.1}}$$

$$= \frac{S^2_{Y.1} - S^2_{Y.12}}{S^2_{Y.1}}$$

ในที่นี้ $S^2_{Y.12}$ คือ ค่าแปรปรวนของตัวแปร Y ที่มิอาจอธิบายได้โดย X_1 และ X_2 ในการวิเคราะห์สหสัมพันธ์เชิงซ้อน

$S^2_{Y.1}$ คือ ค่าความแปรปรวนของ Y ที่มิอาจอธิบายได้โดย X_1 ในการวิเคราะห์สหสัมพันธ์อย่างง่าย

ตัวเศษในสูตรคือ $S^2_{Y.1} - S^2_{Y.12}$ แทนส่วนของค่าความแตกต่างที่เกิดขึ้นที่มิอาจอธิบายได้โดย X_1 ที่ลดลงเนื่องจากการนำเอาตัวแปร X_2 เข้ามาอีกตัวหนึ่ง ดังนั้น อัตราส่วน $S^2_{Y.12}/S^2_{Y.1}$ จึงอาจตีความได้ว่าเป็นเครื่องวัดที่ออกมาในรูปค่าสัมพัทธ์ระหว่างความสัมพันธ์เกี่ยวเนื่องของตัวแปรทั้งสาม กับความสัมพัทธ์เกี่ยวเนื่องเฉพาะตัวแปรตาม Y กับ X_1 เท่านั้น

และถ้าเราถอดรากกำลังที่สองของ R^2 ก็จะได้ค่าสหสัมพันธ์บางส่วน คือ R ตัวกำกับ (Subscipt) ที่อยู่ใน $R^2_{Y2.1}$ ตัวหน้าจุด ในที่นี้คือ 2 แทนตัวแปรที่กำลังพิจารณาวัดสหสัมพันธ์บางส่วน ส่วนตัวเลขหลังจุดซึ่งในที่นี้คือ 1 แทนตัวแปรอิสระซึ่งอาจเป็นหนึ่งหรือหลายตัวก็ได้ที่เราสมมติว่าคงที่

---

[1] เรื่องเดียวกัน, หน้า 78 – 80.

การตีความหมายสหสัมพันธ์บางส่วน

ถ้า $R_{Y2.1} = 0$ หมายความว่าการนำตัวแปรอิสระตัวที่ 2 คือ X_2 เข้ามาพิจารณาเพิ่มขึ้นอีกหนึ่งตัว ไม่มีส่วนช่วยให้ในการอธิบายค่าแปรปรวนที่มีอยู่ในตัวแปรตาม Y ในการวิเคราะห์นี้เลย หรืออาจกล่าวได้ว่า ระหว่าง Y กับ X_2 โดยให้ X_1 คงที่ ไม่มีสหสัมพันธ์บางส่วนเลย แต่ถ้า $R_{Y2.1} = 1$ หมายความว่า ค่าแปรปรวนที่มิอาจอธิบายได้โดย X_1 ในการวิเคราะห์ครั้งแรก อธิบายได้โดย X_2 โดยสิ้นเชิง

การทดสอบสมมติฐานเกี่ยวกับการถดถอย

สมมติฐานที่จะทดสอบคือ "ไม่มีความสัมพันธ์ระหว่างตัวแปรอิสระที่อธิบายได้กับตัวแปรตาม" เป็นการทดสอบว่ากลุ่มของตัวแปรอิสระที่ถูกเลือกมาทำการวิจัยนี้มีความสัมพันธ์กับราคาหุ้นหรือไม่ นั่นคือ

$$H_o : \beta_1 = \beta_2 \ldots\ldots\ldots = \beta_5 = 0$$

$$H_a : \beta_1 \neq \beta_2 \ldots\ldots\ldots \neq \beta_5 \neq 0$$

ตัวสถิติที่ใช้ในการทดสอบ สมมติฐานคือ $F = MSR/MSE$ ซึ่งคำนวณได้จากตารางวิเคราะห์ความแปรปรวน (Analysis of Variance)หรือที่เรียกว่า ตาราง ANOVA ซึ่งจะเป็นดังนี้ [1]

---

[1] วิทยาลัยการค้า, คณะวิชาเศรษฐศาสตร์, "การวิเคราะห์การถดถอย", ใน เอกสารประกอบคำบรรยายวิชาหลักสถิติ 2, หน้า 10.

ตารางที่ 3 – 1

ANOVA

Source of Variation : SOV	d.f.	Sum of Squares : SS	Mean of Squares : MS	F
Explained Variables	k - 1	SSR = $\sum(\hat{Y} - \bar{Y})^2$	MSR = $\frac{SSR}{k-1}$	$MSR/_{MSE}$
Unexplained Variables	n - k	SSE = $\sum(Y - \hat{Y})^2$	MSE = $\frac{SSE}{n-k}$	
Total	n - 1	SST = $\sum(Y - \bar{Y})^2$		

โดยที่ n = จำนวนตัวอย่าง

k = จำนวนพารามิเตอร์ถดถอย ในการวิจัยครั้งนี้มี 6 ตัว

k - 1 = จำนวนพารามิเตอร์ของตัวแปรอิสระ ซึ่งมี 5 ตัว

เมื่อได้ค่า F จากการเอา MSR/MSE แล้ว นำมาเปรียบเทียบกับค่า F ที่ได้จากการเปิดตาราง โดยมี Degree of Freedom เป็น (k - 1), (n - k) จะปฏิเสธสมมติฐานก็เมื่อค่า F ที่คำนวณได้มากกว่าค่า $F_{\alpha/2}$ (k - 1) (n - k)

สำหรับการที่จะทดสอบว่า ตัวแปรอิสระแต่ละตัวมีความสัมพันธ์กับราคาหุ้นหรือไม่นั้น ตั้งสมมติฐานว่า

H_o : $\beta_i = 0$

H_a : $\beta_i \neq 0$

ใช้ตัวสถิติ $t = \frac{b_i - \beta}{S_{b_i}}$ ทำการทดสอบสมมติฐาน ขอบเขตของการปฏิเสธสมมติฐาน H_o ณ ระดับนัยสำคัญ α คือ ค่าที่คำนวณได้ $> t_{\alpha/2}(n - 2)$

<u>สมการถดถอยเชิงซ้อนแบบ Step-wise</u>

เทคนิคที่ใช้ในการคำนวณในการวิจัยนี้ เรียกว่า Step-wise เป็นวิธีการนำเอาตัวแปรอิสระเข้ามาพิจารณาทีละตัว โดยมีขั้นตอนดังนี้

<u>ขั้นที่ 1</u> หาค่าสหสัมพันธ์ (Correlation) ระหว่างตัวแปรแต่ละคู่ เพื่อดูว่าถ้าไม่พิจารณาตัวแปรตัวอื่นแล้ว ระหว่างตัวแปร 2 ตัว จะมีความสัมพันธ์ไปในทางเดียวกัน (ทางบวก) หรือในทางตรงข้าม (ทางลบ) และระดับความสัมพันธ์จะเป็นเท่าใด

<u>ขั้นที่ 2</u> เลือกตัวแปรอิสระที่มีค่าสหสัมพันธ์ (R) กับตัวแปรตามมากที่สุดมาพิจารณาในสมการถดถอย ซึ่งในขั้นนี้สมการถดถอยจะเป็น

$$\hat{Y} = a + b_1X_1$$

ในขั้นนี้จะคำนวณค่าสถิติต่าง ๆ ที่กล่าวมาแล้วข้างต้น เพื่อจะดูว่าเมื่อมีตัวแปรอิสระ 1 ตัวในสมการ ค่าต่าง ๆ เหล่านั้นจะเป็นเท่าใด

<u>ขั้นที่ 3</u> เป็นการเลือกตัวแปรอิสระตัวที่สองเข้ามาในสมการ โดยการหาค่ากำลังที่สองของสหสัมพันธ์ส่วนย่อย ระหว่างค่า Y และค่าแต่ละค่าของตัวแปรอิสระที่เหลืออยู่จำนวน 4 ตัว โดยจะเลือกตัวแปรอิสระที่ให้ค่ากำลังสองของสหสัมพันธ์ส่วนย่อยสูงที่สุด. (ไม่พิจารณาตัวแปรอิสระตัวแรกที่เข้ามาในสมการแรกแล้ว) สำหรับตัวแปรอิสระที่เหลือก็จะใช้วิธีการเดียวกันนี้

หลังจากที่มีการเลือกตัวแปรอิสระเข้ามาในสมการแต่ละครั้ง จะมีการทดสอบนัยสำคัญของตัวแปรอิสระที่อยู่ร่วมกันในสมการ เพื่อจะพิจารณาว่า หลังจากมีตัวแปรอิสระเพิ่มเข้ามาในสมการแต่ละตัวแล้ว ตัวแปรอิสระเดิมที่อยู่ในสมการยังคงอธิบายตัวแปรตามได้หรือไม่ โดยพิจารณาจากค่า t - test ตัวแปรอิสระตัวใดให้ค่า t ต่ำกว่าที่กำหนดไว้ ก็จะถูกออกจากสมการ