

## การถดถอย

### ความหมาย

การถดถอย (Regression) หมายถึง กระบวนการของการพยากรณ์ (prediction) ของตัวแปรตัวหนึ่ง จากตัวแปรตัวอื่นอย่างน้อย ๑ ตัว ตัวแปรที่ถูกพยากรณ์ เรียกว่า “ตัวแปรตาม (dependent variable)” หรือ “ตัวเกณฑ์ (criterion)” ส่วนตัวแปรที่ใช้ในการพยากรณ์ ตัวแปรตามหรือตัวเกณฑ์ เรียกว่า “ตัวแปรอิสระ (independent variable)” หรือ “ตัวพยากรณ์ (predictor)”

การพยากรณ์ตัวแปรตามโดยใช้ตัวพยากรณ์จะต้องสร้างสมการการถดถอย (regression equation) สมการการถดถอยนี้จะเป็นความสัมพันธ์เชิงคณิตศาสตร์ระหว่างตัวแปรตามและตัวแปรอิสระ การสร้างสมการการถดถอยจะต้องสร้างให้ได้สมการการถดถอยที่เหมาะสมที่สุดของข้อมูล การที่จะให้ได้สมการการถดถอยที่เหมาะสมที่สุดจะต้องมีการวิเคราะห์ข้อมูล ซึ่งจะเรียกว่า “การวิเคราะห์การถดถอย (regression analysis)” ตัวแปรที่จะนำมาวิเคราะห์การถดถอยจะต้องมีลักษณะการแจกแจงปกติ และเป็นอิสระต่อกัน

### ความเป็นมา

การถดถอยได้ถูกนำมาใช้ในวิชาสถิติครั้งแรกในปี ค.ศ. ๑๘๖๗ โดย เซอร์ฟรานซิส แกลตัน (Sir Francis Galton) เซอร์ ฟรานซิส แกลตัน ได้ศึกษาเรื่องเกี่ยวกับพันธุกรรมเพื่อแสดงว่าเด็กซึ่งเกิดจากพ่อแม่ที่สูงมีแนวโน้มจะมีความสูงเท่ากับความสูงเฉลี่ยของพ่อแม่ และได้นำคำว่า “การถดถอย” มาใช้โดยให้ความหมาย การถดถอยว่า หมายถึงกระบวนการของการพยากรณ์ตัวแปรตัวหนึ่ง (ความสูงของเด็ก) จากตัวแปรอีกตัวหนึ่ง (ความสูงของพ่อแม่) ต่อมา

นักสถิติได้นำคำว่า “การถดถอยพหุคูณ (multiple regression) มาใช้และได้เรียกรถดถอยของเซอร์ฟรานซิส แกลตัน ว่า การถดถอยอย่างง่ายหรือการถดถอยเชิงเดียว (simple regression)

### ชนิดของการถดถอย

การถดถอยมี ๒ ชนิด คือ

๑. การถดถอยเชิงเดียว เป็นการถดถอยที่มีสมการการถดถอยประกอบด้วยตัวแปรตาม ๑ ตัว และตัวแปรอิสระ ๑ ตัว เช่น ถ้าให้  $Y$  แทนตัวแปรตาม ให้  $X$  แทนตัวแปรอิสระ และให้  $b_0$ ,  $b_1$  และ  $b_2$  แทนค่าคงที่จะได้ว่า

๑.๑ สมการการถดถอยที่มีกราฟเป็นเส้นตรง (Straight line) แทนด้วย

$$Y = b_0 + b_1 X$$

๑.๒ สมการการถดถอยที่มีกราฟเป็นเส้นโค้งพาราโบลา (parabola curve) แทนด้วย

$$Y = b_0 + b_1 X + b_2 X^2$$

๑.๓ สมการการถดถอยที่มีกราฟเป็นเส้นโค้งไฮเพอร์โบลา (hyperbola curve) แทนด้วย

$$Y = \frac{1}{b_0 + b_1 X}$$

๑.๔ สมการการถดถอยที่มีกราฟเป็นเส้นโค้งชี้กำลัง (exponential curve) แทนด้วย

$$Y = b_0 b_1^X$$

๑.๕ สมการการถดถอยที่มีกราฟเป็นเส้นโค้งเรขาคณิต (geometric curve) แทนด้วย

$$Y = b_0 X^{b_1}$$

เป็นต้น

ข้อ ๑.๑ เรียกว่าการถดถอยเชิงเดียวแบบเชิงเส้น (simple linear regression) ส่วนข้อ ๑.๒ ถึง ๑.๕ เรียกว่าการถดถอยเชิงเดียวแบบไม่ใช่เชิงเส้น (simple non-linear regression)

๒. การถดถอยพหุคูณ เป็นการถดถอยที่มีสมการการถดถอยด้วยตัวแปรตาม ๑ ตัวและตัวแปรอิสระมากกว่า ๑ ตัว เช่น ถ้าให้ Y แทนตัวแปรตาม ให้  $X_1, X_2, \dots$  และ  $X_k$  แทนตัวแปรอิสระ k ตัว และให้  $b_0, b_1, \dots$  และ  $b_k$  แทนค่าคงที่ที่จะได้ว่า

๒.๑  $Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k$  เรียกว่าการถดถอยพหุคูณแบบเชิงเส้น (multiple linear regression)

๒.๒  $Y = b_0 + b_1X_1 + b_2X_2 + b_3X_1X_2 + b_4X_1^2 + b_5X_2^2$

เรียกว่า การถดถอยพหุคูณแบบไม่ใช่เชิงเส้น (multiple non-linear regression) เป็นต้น **ตัวแบบและสมการการพยากรณ์ของการถดถอย**

ให้  $Y_i$  แทนค่าที่ i ของตัวแปรตาม  $\hat{Y}$  แทนค่าพยากรณ์ (predicted value) ของตัวแปรตาม  $Y$   $X_i$  แทนค่าที่ i ของตัวแปรอิสระ  $X$   $X_{i,1}, X_{i,2}, \dots$  และ  $X_{i,k}$  แทนค่าที่ i ของตัวแปรอิสระที่ ๑, ๒, ..., และ k ตามลำดับ  $e_i$  แทนค่าที่ i ของค่าผิดพลาดแบบสุ่ม (random error) หรือ เศษตกค้าง (residual) n แทนจำนวนค่าสังเกตของ Y หรือ X  $\beta_0, \beta_1, \dots$  และ  $\beta_k$  แทนพารามิเตอร์ (parameter) k+1 ตัว และให้  $b_0, b_1, \dots$  และ  $b_k$  แทนค่าประมาณของ  $\beta_0, \beta_1, \dots$  และ  $\beta_k$  ตามลำดับ จะได้ตัวแบบ (model) ของการถดถอย สมการการพยากรณ์ (prediction equation) ของการถดถอยและชนิดของการถดถอยซึ่งสอดคล้องกันดังนี้ คือ

1. ตัวแบบของ  $Y = b_0 + b_1X$  คือ

$$Y_i = \beta_0 + \beta_1 X_i + e_i,$$

$$i = 1, 2, \dots, n$$

จะมีสมการการพยากรณ์ของการถดถอยคือ

$$\hat{Y} = b_0 + b_1X$$

2. ตัวแบบของ  $Y = b_0 + b_1X + b_2X^2$  คือ

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + e_i, \quad i = 1, 2, \dots, n$$

จะมีสมการการพยากรณ์ของการถดถอยคือ

$$\hat{Y} = b_0 + b_1X + b_2X^2$$

3. ตัวแบบของ  $Y = \frac{1}{b_0 + b_1X}$  คือ

$$Y_i = \frac{1}{\beta_0 + \beta_1 X_i} + e_i, \quad i = 1, 2, \dots, n$$

จะมีสมการการพยากรณ์ของการถดถอยคือ

$$\hat{Y} = \frac{1}{b_0 + b_1X}$$

4. ตัวแบบของ  $Y = b_0b_1^X$  คือ

$$Y_i = \beta_0\beta_1^{X_i} + e_i, \quad i = 1, 2, \dots, n$$

จะมีสมการการพยากรณ์ของการถดถอยคือ

$$\hat{Y} = b_0 b_1^X$$

5. ตัวแบบของ  $Y = b_0X^{b_1}$  คือ

$$Y_i = \beta_0 X_i^{\beta_1} + e_i, \quad i = 1, 2, \dots, n$$

จะมีสมการการพยากรณ์ของการถดถอยคือ

$$\hat{Y} = b_0X^{b_1}$$

6. ตัวแบบของ  $Y = b_0 + b_1X_1 +$

$$b_2X_2 + \dots + b_kX_k$$
 คือ

$$\dots + \beta_k X_k^2$$

จะมีสมการ  $\hat{Y} = b_0$

7. ตัวแบบ

$$b_3X_1X_2 +$$

$$Y$$

$$\beta_3X_{i,1}X_{i,2}$$

จะมีสมการ

$$\hat{Y} = b_0$$

$$b_4X_1^2 + b$$

เราจะเ

predicted value) ของตัวเชื่อมั่น (confidence) จะเรียกค่าพยา (prediction i

เราจะเ

ค่าของ  $b_0, b_1$

และ  $b_k$  จะ

(method of l

หลักการ

ต้องหา  $b_0, b_1$

$\sum_{i=1}^n$  (ค

$i=1$

มีค่าน้อยที่สุด

เช่น ถ้าสมการ

$$Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \dots + \beta_k X_{i,k} + e_i, \quad i = 1, 2, \dots, n$$

จะมีสมการการพยากรณ์ของการถดถอยคือ

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_k X_k$$

7. รูปแบบของ  $Y = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_1 X_2 + b_4 X_1^2 + b_5 X_2^2$  คือ

$$Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \beta_3 X_{i,1} X_{i,2} + \beta_4 X_{i,1}^2 + \beta_5 X_{i,2}^2 + e_i, \quad i = 1, 2, \dots, n$$

จะมีสมการการพยากรณ์ของการถดถอยคือ

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_1 X_2 + b_4 X_1^2 + b_5 X_2^2$$

เราจะเรียก  $\hat{Y}$  ว่าค่าพยากรณ์แบบจุด (point predicted value) ของตัวแปรตาม  $Y$  นอกจากนี้เราสามารถจะหาค่าพยากรณ์แบบช่วง (interval predicted value) ของตัวแปรตาม  $Y$  ได้โดยวิธีการหาช่วงความเชื่อมั่น (confidence interval) ของพารามิเตอร์ และจะเรียกค่าพยากรณ์แบบช่วงของ  $Y$  นี้ว่า ช่วงพยากรณ์ (prediction interval)

เราจะเห็นว่าการหาค่าของ  $\hat{Y}$  นั้น เราต้องทราบค่าของ  $b_0, b_1, \dots$  และ  $b_k$  การหาค่าของ  $b_0, b_1, \dots$  และ  $b_k$  จะทำได้โดยระเบียบวิธีกำลังสองน้อยที่สุด (method of least squares)

หลักการของระเบียบวิธีกำลังสองน้อยที่สุดคือ ต้องการ  $b_0, b_1, \dots$  และ  $b_k$  ซึ่งทำให้

$$\sum_{i=1}^n (\text{ค่าสังเกต } Y_i - \text{ค่าทำนาย } \hat{Y}_i)^2$$

มีค่าน้อยที่สุด

เช่น ถ้าสมการการพยากรณ์ของการถดถอย คือ

$$Y = b_0 + b_1 X$$

จะต้องหา  $b_0$  และ  $b_1$  ที่ทำให้

$$\sum_{i=1}^n [Y_i - (b_0 + b_1 X_i)]^2 \text{ มีค่าน้อยที่สุด}$$

โดยวิธีการทางคณิตศาสตร์เราสามารถหา  $b_0$  และ  $b_1$  ได้ดังนี้

$$b_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} =$$

$$\frac{\sum_{i=1}^n X_i Y_i - [(\sum X_i)(\sum Y_i)]/n}{\sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2/n}$$

และ  $b_0 = \bar{Y} - b_1 \bar{X}$

เมื่อ  $\bar{Y} = \sum_{i=1}^n Y_i / n$

และ  $\bar{X} = \sum_{i=1}^n X_i / n$

ในการทำงานเดียวกัน ถ้าสมการการพยากรณ์ของการถดถอย คือ  $\hat{Y} = b_0 + b_1 X + b_2 X^2$  จะต้องการ  $b_0, b_1$  และ  $b_2$  ที่ทำให้

$$\sum_{i=1}^n [Y_i - (b_0 + b_1 X_i + b_2 X_i^2)]^2$$

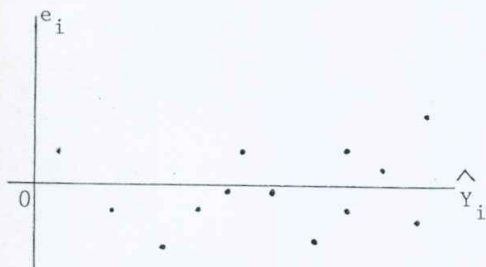
มีค่าน้อยที่สุด เป็นต้น

**ข้อตกลงเบื้องต้นของการถดถอย**

๑. ค่าเฉลี่ยของแต่ละ  $e_i$  เท่ากับศูนย์ เมื่อ  $i=1, 2, \dots, n$
๒. ความแปรปรวนของแต่ละ  $e_i$  เท่ากัน เมื่อ  $i=1, 2, \dots, n$
๓. แต่ละ  $e_i$  เป็นอิสระต่อกัน เมื่อ  $i=1, 2, \dots, n$
๔.  $e_i$  เมื่อ  $i = 1, 2, \dots, n$  มีการแจกแจงปกติ (normal distribution)

การตรวจสอบข้อตกลงเบื้องต้นของการถดถอย มีหลายวิธี แต่ในที่นี้จะกล่าวถึงเพียงวิธีเดียว คือ การ

ทำแผนภาพกระจาย (scatter - diagram) ระหว่างค่าผิดพลาดแบบสุ่ม( $e_i$ )และค่าพยากรณ์ ( $\hat{Y}_i$ ) ซึ่ง  $e_i = Y_i - \hat{Y}_i$  ถ้าแผนภาพกระจายมีรูปร่างเป็นแถบแนวนอน(horizontal band) ดังแสดงในรูป ๑ แสดงว่ามีข้อดกลงเบื้องต้นทุก ๆ ข้อของการถดถอย



รูป 1 แผนภาพกระจายระหว่าง  $e_i$  และ  $\hat{Y}_i$

**ขั้นตอนของการวิเคราะห์การถดถอย**

๑. เลือกตัวแบบของการถดถอย (หรือชนิดของการถดถอย) ที่คาดว่าจะเหมาะสมกับข้อมูล

๒. ตรวจสอบข้อดกลงเบื้องต้นของการถดถอย

๒.๑ ถ้าข้อมูลมีข้อดกลงเบื้องต้นของการถดถอยทุก ๆ ข้อแล้ว ข้อมูลอาจจะไม่มีตัวแบบของการถดถอยที่คาดว่าจะเหมาะสมกับข้อมูลตามข้อ ๑ หรือไม่ก็ได้ จะทราบว่าข้อมูลมีตัวแบบของการถดถอยตามที่คาดไว้หรือไม่นั้น จะต้องทดสอบนัยสำคัญของการถดถอย (significance of regression) เช่น การวิเคราะห์ความแปรปรวนของการถดถอย (analysis of variance of regression) เป็นต้น ถ้าข้อมูลมีตัวแบบของการถดถอยตามที่ได้คาดไว้แล้ว จะได้สมการการพยากรณ์ของการถดถอยและหาค่าพยากรณ์ของตัวแปรได้ แต่

ตาราง 1 คะแนนของวิชาวิธีทางสถิติและวิชาสถิติคณิตศาสตร์ของนักเรียน

วิชา	1	2	3	4	5	6	7	8	9	10	11	12
สถิติคณิตศาสตร์ (X)	65	63	67	64	68	62	70	66	68	67	69	71
วิธีทางสถิติ (Y)	68	66	68	65	69	66	68	65	71	67	68	70

ถ้าข้อมูลไม่มีตัวแบบของการถดถอยตามที่คาดไว้แล้ว จะต้องเลือกตัวแบบของการถดถอยใหม่ต่อไป

๒.๒ ถ้าข้อมูลไม่มีข้อดกลงเบื้องต้นทุก ๆ

ข้อของการถดถอยแล้ว อาจจะต้องเลือกตัวแบบของการถดถอยใหม่ต่อไป

**เทคนิคของการเลือกการถดถอยที่เหมาะสมที่สุด**

เทคนิคของการเลือกการถดถอยที่เหมาะสมที่สุดสำหรับข้อมูลมีหลายวิธี คือ

๑. กระบวนการการกำจัดย้อนหลัง (backward elimination procedure)

๒. กระบวนการการเลือกไปข้างหน้า (forward selection procedure)

๓. กระบวนการการถดถอยสเต็ปไวส์ (stepwise regression procedure)

๔. กระบวนการการถดถอยสเตจไวส์ (stagewise regression procedure)

**ตัวอย่าง**

การศึกษาเพื่อหาสมการการพยากรณ์ของการถดถอยของคะแนนวิชาวิธีทางสถิติ (statistical method) ของนักศึกษาโดยจะทำนายคะแนนของวิชานี้จากคะแนนของวิชาสถิติคณิตศาสตร์ (mathematical statistics) สุ่มตัวอย่างนักศึกษาที่เรียนทั้งวิชาวิธีทางสถิติและวิชาสถิติ คณิตศาสตร์ จำนวน ๑๒ คน ซึ่งคะแนนที่ได้จากการสอบวิชาทั้งสองแสดงในตาราง ๑

**นักเรียนคนที่**

วิธีทำ  
คะแนน  
คะแนน  
ตั้ง  
ขั้นตอน  
๑.  
แบบของ

๒.  
สมมติว่า

๓.  
การศึกษา  
ตามที่คาด  
การถดถ

แล  
ระเบียบวิ

$b_1 = d_0$

เป็  
คะแนนวิ  
การตีค

$b_0$

วิธีทำ จากตาราง ๑ จะเห็นว่า ตัวแปรอิสระ X คือ คะแนนของวิชาสถิติคณิตศาสตร์ ตัวแปรตาม Y คือ คะแนนของวิชาวิธีทางสถิติ

ดังนั้น การศึกษาครั้งนี้จะเป็นการถดถอยเชิงเดียว ขั้นตอนของการวิเคราะห์การถดถอยมีดังนี้ คือ

๑. ในกรณีที่เป็นการถดถอยเชิงเดียว จะเลือกตัวแบบของการถดถอยที่คาดว่าจะเหมาะสม คือ

$$Y_i = \beta_0 + \beta_1 X_i + e_i, \\ i = 1, 2, \dots, 12$$

๒. ตรวจสอบข้อตกลงเบื้องต้นของการถดถอย สมมติว่ามีข้อตกลงเบื้องต้นของการถดถอยทุก ๆ ข้อ

๓. ทดสอบนัยสำคัญของการถดถอย สมมติว่า การศึกษาครั้งนี้มีตัวแบบของการถดถอยที่เหมาะสมตามที่กล่าวไว้ในข้อ ๑ ดังนั้น สมการการพยากรณ์ของการถดถอย คือ

$$y = b_0 + b_1 X$$

และสามารถหา  $b_0$  และ  $b_1$  จากสูตรซึ่งได้จากระเบียบวิธีกำลังสองน้อยที่สุดได้  $b_0 = 35.82$  และ

$b_1 = 0.476$  และจะได้ว่า

$$\hat{Y} = 35.82 + 0.476X$$

เป็นสมการการพยากรณ์ของการถดถอยของคะแนนวิชาวิธีทางสถิติของนักศึกษา

**การตีความ  $b_0$  และ  $b_1$**

$b_0 = 35.82$  หมายความว่า ถ้านักศึกษาได้คะแนน

วิชาสถิติคณิตศาสตร์เท่ากับศูนย์แล้ว นักศึกษาจะได้คะแนนวิชาวิธีทางสถิติเท่ากับ ๓๕.๘๒ คะแนน

$b_1 = 0.476$  หมายความว่า ถ้านักศึกษาได้คะแนนวิชาสถิติคณิตศาสตร์เพิ่มขึ้น ๑ คะแนนแล้ว นักศึกษาจะได้คะแนนวิชาวิธีทางสถิติเพิ่มขึ้นเท่ากับ ๐.๔๗๖ คะแนน

## ประโยชน์ของการถดถอย

การถดถอยนับว่ามีประโยชน์ในการพยากรณ์เหตุการณ์ในอนาคตเกี่ยวกับการศึกษา การวิจัย การวัดผล ประเมินผล และธุรกิจต่าง ๆ เป็นอันมาก กล่าวคือ ใช้สำหรับหาสมการการพยากรณ์ของการถดถอยของตัวแปร เพื่อพยากรณ์ค่าของตัวแปรตัวหนึ่งจากตัวแปรตัวหนึ่ง หรือจากตัวแปรอีกหลายตัว เช่น

๑. ต้องการพยากรณ์คะแนนวิชาภาษาอังกฤษ (ตัวแปรตาม) จากจำนวนปีที่นักเรียนได้เรียนภาษาอังกฤษ (ตัวแปรอิสระ)

๒. ต้องการพยากรณ์เกรดเฉลี่ยของนักเรียน (ตัวแปรตาม) จาก I.Q. ของนักเรียน (ตัวแปรอิสระ)

๓. ต้องการพยากรณ์ความสูงของลูก (ตัวแปรตาม) จากความสูงของพ่อและแม่ (ตัวแปรอิสระ ๒ ตัว)

๔. ต้องการพยากรณ์ปริมาณสินค้าที่จะขายได้ (ตัวแปรตาม) จากราคาของสินค้า (ตัวแปรอิสระ)

อรพินท์ เจียรระพงษ์

บรรณานุกรม

- Conover, W.J. and Ronald L. Iman. Introduction to Modern Business Statistics. New York: John Wiley & Sons, 1983.
- Ferguson, George A. Statistical Analysis in Psychology and Education. Tokyo: McGraw-Hill International Book Company, 1981.
- Freund, John E. Modern Elementary Statistics. London: Prentice-Hall International Inc., 1979.
- Levin, Richard I. Statistics for Management. New Jersey: Prentice-Hall, Inc., 1981
- Pfaffenberger, Roger C. and James H. Patterson. Statistical Methods for Business and Economics. Illinois: Richard D. Irwin, Inc., 1977.

กว

แสด  
ที่เก

ชน

และ

คือ

rela

แป

เปรี

ถ้า

ฐาน

ช่วง

จะ

คชันร

ind

วค

กำ

คือ

agg

นำ

ฐาน

ที่

Q<sub>1</sub>

ขอ