

วิธีการประมาณค่าพารามิเตอร์ในการถดถอย ด้วยวิธีกำลังสองน้อยที่สุด (Method of Least Squares : MLS)

ศรัยวัฒน์ ลำพุทธา *

การศึกษารูปแบบความสัมพันธ์ของตัวแปรตั้งแต่สองตัวขึ้นไป โดยที่ตัวแปรตัวแรกจะเรียกว่า ตัวแปรอิสระ (X) ซึ่งเป็นตัวที่ส่งผลกระทบ ต่อตัวแปรตัวอื่น ส่วนตัวแปรตัวที่สองจะเรียกว่า ตัวแปรตาม (Y) ซึ่งเป็นตัวแปรที่เราสนใจศึกษาหลังจากได้รับผลจากตัวแปรอิสระ ในการวิเคราะห์เกี่ยวกับตัวแปรที่ศึกษาควรนำตัวแปรที่เกี่ยวข้องทั้งหมดมาพิจารณาด้วย เนื่องจากตัวแปรต่าง ๆ มีความสัมพันธ์กันแล้ว ตัวแปรเหล่านั้นอาจจะมีอิทธิพลต่อการเปลี่ยนแปลงของตัวแปรที่ต้องการศึกษาด้วย ซึ่งเราเรียกว่าการวิเคราะห์การถดถอยและสหสัมพันธ์การวิเคราะห์การถดถอยและสหสัมพันธ์ นั่นเอง ซึ่งเป็นเครื่องมือทางสถิติอย่างหนึ่งที่ใช้ในการวิเคราะห์ ว่าตัวแปรต่าง ๆ มีความสัมพันธ์กันในรูปแบบใดและความสัมพันธ์นั้นมีมากน้อยเพียงใด สำหรับการหาความสัมพันธ์ระหว่างตัวแปร X และ Y ในขั้นแรกจะนำเอาข้อมูลของทั้งสองตัวแปรมาเขียนกราฟแสดงความสัมพันธ์ ซึ่งเรียกกราฟนี้ว่า แผนภาพการกระจาย (Scatter Diagram) ผู้ศึกษาจะต้องพิจารณาจากแผนภาพการกระจายว่า ตัวแปรทั้งสองมีความสัมพันธ์กันอยู่ในระดับใด ซึ่งต้องนำไปเขียนความสัมพันธ์ให้อยู่ในรูปแบบทางคณิตศาสตร์ซึ่งจะมีหลายรูปแบบ แต่ในที่นี้จะศึกษาเฉพาะความสัมพันธ์ของตัวแปร X และ Y ในรูปเส้นตรงหรือเชิงเส้น ซึ่งเราจะเรียกการวิเคราะห์การถดถอยอย่างง่ายที่ความสัมพันธ์ของตัวแปรอยู่ในรูปเชิงเส้นว่า การวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression Analysis) เมื่อพิจารณาแผนภาพการกระจายแสดงความสัมพันธ์ระหว่าง X และ Y แล้วพบว่า X และ Y สัมพันธ์กันในรูปเส้นตรง จะต้องคำนวณหาค่า β_0 และ β_1 ซึ่งจะทำให้ทราบถึงความสัมพันธ์ระหว่าง X และ Y ว่า มีความสัมพันธ์ตามกันหรือตรงข้ามกันและความสัมพันธ์นั้นมีมากหรือน้อยเพียงใด ถ้า β_1 มีค่ามากแสดงว่า Y มีความสัมพันธ์กับ X มากด้วย การที่จะหาค่า β_0 และ β_1 จำเป็นต้องทราบค่า X และ Y ทุกค่าที่เกิดขึ้น ซึ่งเป็นไปได้ยากในทางปฏิบัติเราจึงใช้ข้อมูลตัวอย่างขนาด n ในการประมาณค่า β_0 และ β_1 ซึ่งค่าประมาณของ Y คือ

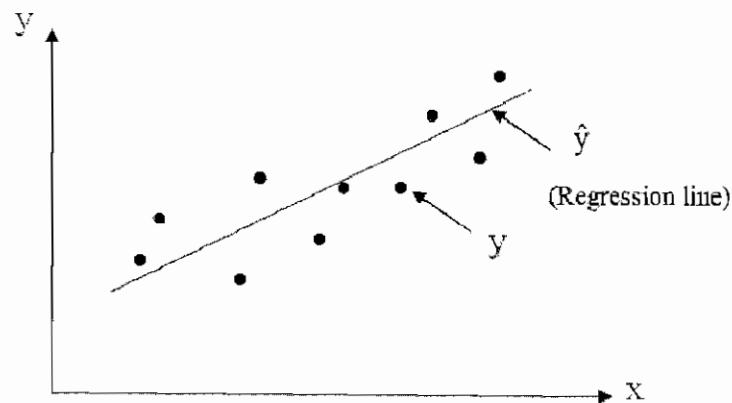
$$Y_i = \beta_0 + \beta_1 X_i \quad \text{หรือ} \quad \hat{Y} = a + b X_i \quad \dots\dots\dots(1)$$

โดยที่ $\beta_0 = a$ (ระยะตัดแกน Y (Y - Intercept) หรือ ค่าของ Y เมื่อ $X = 0$) และ

$\beta_1 = b$ (ความชัน (Slope) เป็นค่าที่แสดงให้ทราบว่าเมื่อ X มีค่าเปลี่ยนไป 1 หน่วย แล้ว Y จะเปลี่ยนไปโดยเฉลี่ยแล้วเท่าไร)

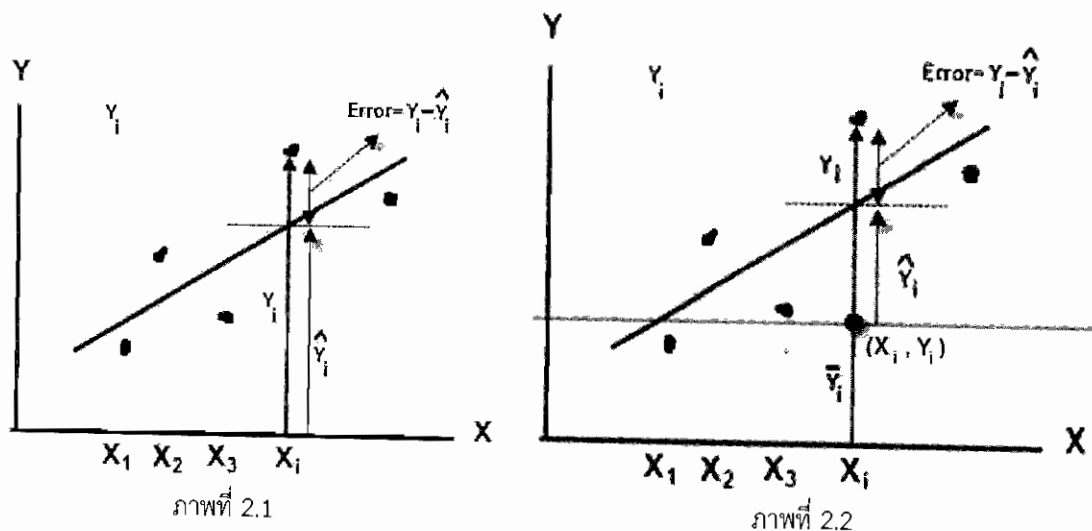
ในการประมาณค่า β_0 และ β_1 ด้วย a และ b นั้น มีเป้าหมายเพื่อให้ความคลาดเคลื่อนในการประมาณค่า Y_i ด้วยค่าต่ำสุด โดยวิธีที่เรียกว่า วิธีกำลังสองน้อยที่สุด (Method of Least Squares : MLS) ซึ่งการวิเคราะห์ข้อมูลเพื่อย้อนกลับหรือถดถอย (Regress) กลับไปหาความสัมพันธ์ที่แท้จริง ถ้าวิเคราะห์ข้อมูลระหว่างสองตัวแปร เรียกว่า Simple Regression แต่ถ้าในกรณีมากกว่าสองตัวแปรขึ้นไป เราเรียกว่า Multiple Regression

ในกรณีสองตัวแปร ให้พิจารณาประเด็นในการลากเส้นตรงเพื่อแทนความสัมพันธ์ระหว่างสองตัวแปร สมมติคือ X และ Y ดังภาพที่ 1



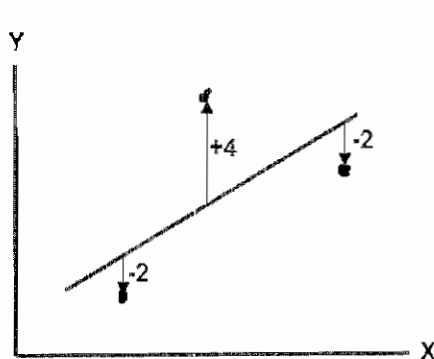
ภาพที่ 1 การลากเส้นตรงเพื่อแทนความสัมพันธ์ระหว่างสองตัวแปรให้ได้ใกล้เคียง

ในการลากเส้นตรงเพื่อใช้แทนข้อมูลทั้งหมดให้ได้ใกล้เคียงและดีที่สุด อาจทำได้หลายวิธี เช่น การกะด้วยสายตา หรือวิธีอื่นๆ แล้วแต่แนวความคิดของแต่ละคน อย่างไรก็ตาม ในการจะลากเส้นเพื่อใช้แทนข้อมูลทั้งหมดต้องมีหลักเกณฑ์ในการลาก ซึ่งหลักเกณฑ์ง่าย ๆ คือ ให้เลือกเส้นที่มีความคลาดเคลื่อนรวม มีค่าน้อยที่สุด นั่นก็คือ เลือกเส้นที่ทำให้ $\sum(Y_i - \hat{Y}_i)$ มีค่าน้อยที่สุด ดังภาพที่ 2

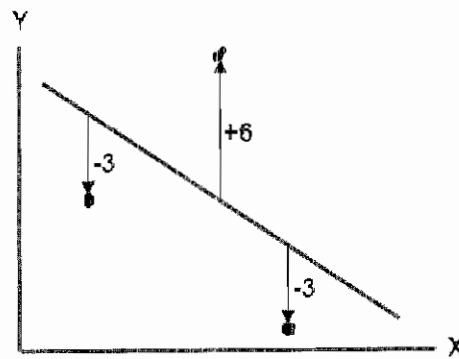


ภาพที่ 2.1 และ 2.2 แสดงความคลาดเคลื่อนเมื่อประมาณ Y_i ด้วย $\hat{Y}_i$

จากภาพที่ 2.1 และ 2.2 ให้พิจารณาภาพที่ 2.3 และ 2.4 ซึ่งเป็นข้อมูลชุดเดียวกัน จะได้เส้นตรงแทนข้อมูล 2 เส้น ปัญหาที่เกิดขึ้นคือ ตัดสินไม่ได้ว่าเส้นใดดีกว่ากัน จากหลักเกณฑ์ที่กำหนดคือ $\sum (y_i - \hat{y}_i) = 0$ ซึ่งทั้งสองภาพเท่ากับศูนย์ทั้งคู่ นั่นคือเกิดปัญหาในเรื่องของทิศทางหรือเครื่องหมายบวกและลบ ของความคลาดเคลื่อนที่หักล้างกัน



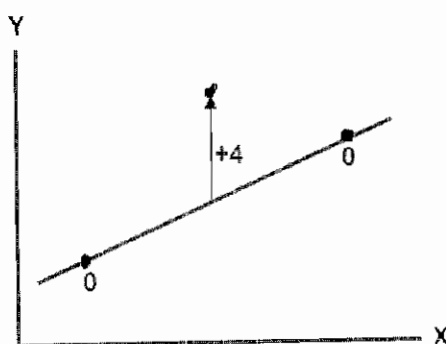
ภาพที่ 2.3



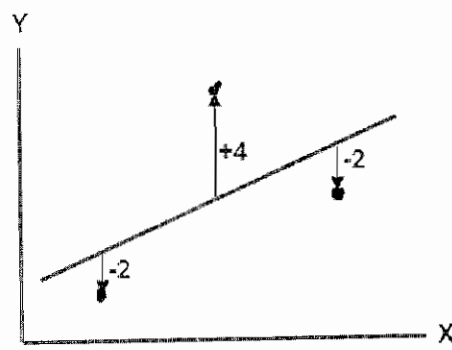
ภาพที่ 2.4

ภาพที่ 2.3 และ 2.4 ปัญหาของการใช้เกณฑ์ $\sum (Y_i - \hat{Y}_i)$ น้อยที่สุด

เพื่อแก้ไขปัญหาดังกล่าว ลองพิจารณาวิธีใหม่คือ ใช้เกณฑ์ $\sum (|Y_i - \hat{Y}_i|)$ มีค่าน้อยที่สุด ซึ่งกรณีนี้ก็คือการวัดเฉพาะขนาดไม่คำนึงถึงทิศทางคือบวกและลบ สามารถสร้างเส้นได้ดังภาพที่ 2.5 และ 2.6



ภาพที่ 2.5



ภาพที่ 2.6

ภาพที่ 2.5 และ 2.6 ปัญหาของการใช้เกณฑ์ $\sum (|Y_i - \hat{Y}_i|)$ น้อยที่สุด

จากภาพที่ 2.5 และ 2.6 ถ้าใช้เกณฑ์ $\sum (|Y_i - \hat{Y}_i|)$ มีค่าน้อยที่สุด เราควรจะเลือกกรณีดังภาพที่ 2.5 เพราะมีความคลาดเคลื่อนรวมเท่ากับ $|0| + |4| + |0| = 4$ หน่วย แต่เมื่อพิจารณาภาพที่ 2.6 มีความคลาดเคลื่อนรวมเท่ากับ $|-2| + |4| + |-2| = 8$ หน่วย แต่ถ้าพิจารณาตามหลักสถิติแล้ว กรณีภาพที่ 2.5 ถือว่าไม่ดี เพราะไม่ได้

ใช้ข้อมูลที่มีอยู่ทั้งหมดอย่างเต็มที่ใช้เพียงแค่สองจุดเท่านั้นคือ ลากเส้นจากจุดต้นทาง คือ 0 ไปเชื่อมกับจุดปลายทาง คือ 0 โดยที่ไม่ได้ใช้ประโยชน์จากข้อมูลในจุดกลาง ในที่นี้ คือ 4 เลย ซึ่งต่างจากภาพที่ 2.6 ที่มีการนำข้อมูลทุกจุดไปใช้อย่างคุ้มค่า

ดังนั้นถึงแม้ว่าจะสามารถแก้ไขปัญหาในเรื่องของทิศทางได้ แต่ก็ยังไม่ใช่วิธีที่ดีที่สุด เพราะยังเกิดปัญหาในเรื่องการใช้ประโยชน์ของข้อมูล พิจารณาอีกวิธีหนึ่งคือการใช้เกณฑ์โดยเลือกเส้นที่ทำให้ผลบวกของความคลาดเคลื่อนใน

รูปกำลังสองมีค่าน้อยที่สุด ซึ่งก็คือ $\sum (Y_i - \hat{Y}_i)^2$ น้อยที่สุดนั่นคือ คือ $\sum (Y_i - \hat{Y}_i)^2 = 0$ ซึ่งหลักเกณฑ์นี้เรียกว่า **วิธีกำลังสองน้อยที่สุด** (Ordinary Least Squares : OLS หรือ Method of Least Squares : MLS) การนำค่าความแตกต่างของตัวอย่างที่ตอบแต่ละคนยกกำลังสอง แล้วนำมารวมกันจะต้องมีค่าเท่ากับศูนย์ แล้วหาวิธีการทางคณิตศาสตร์ แก่สมการกำลังสองน้อยที่สุด จนได้สมการชุดหนึ่งเรียกว่า **สมการมาตรฐาน** (Normal Equations) เป็นสมการระบบเส้นตรง จากสมการนี้ใช้ความรู้ทางคณิตศาสตร์พีชคณิตหรือเรื่องเมทริกซ์ (Matrix) เพื่อแก้สมการหาค่าสัมประสิทธิ์ของสมการและค่าคงที่ได้ วิธีนี้ถือว่าเป็นหลักเกณฑ์ในการวิเคราะห์ข้อมูลเพื่อย้อนกลับหรือถดถอย (Regress) ที่นิยมใช้กันอย่างแพร่หลาย วิธีนี้คิดโดยนักคณิตศาสตร์ชาวเยอรมัน ชื่อ Carl Friederich Gauss ช่วงปี ค.ศ. 1821-1823 และ A.A.Markov ในปี ค.ศ. 1912 ทฤษฎีเกี่ยวกับการประมาณค่ากำลังสองน้อยที่สุดเป็นที่รู้จักกันดีในชื่อ The Gauss-Markov Theorem เป็นที่รู้จักกันว่าเป็นโมเดลการถดถอยเชิงเส้นแบบดั้งเดิมหรือแบบทั่วไป

หลักการของการประมาณค่าด้วยวิธี MLS คือ การหาค่า a และ b ในสมการถดถอยเพื่อให้ได้ค่าการพยากรณ์ค่า Y เมื่อทราบค่า X มีความถูกต้องมากที่สุด หรือให้ความคลาดเคลื่อนในการพยากรณ์มีค่าน้อยที่สุด ซึ่งก็คือค่าความคลาดเคลื่อน มีได้ทั้งบวกและลบ ค่าประมาณพารามิเตอร์ที่หาตามวิธี MLS เป็นค่าประมาณค่าเชิงเส้นที่ไม่คลาดเคลื่อนและเป็นค่าประมาณที่ดีที่สุด (Best Linear Unbiased Estimator : BLUE) ค่าประมาณจะไม่คลาดเคลื่อนเมื่อค่าเฉลี่ยของตัวประมาณค่าที่ได้จากกลุ่มตัวอย่างหลายกลุ่มมีค่าเท่ากับพารามิเตอร์ของประชากร ค่าประมาณจะเป็นค่าที่ดีที่สุด เมื่อความแปรปรวนของการประมาณค่านั้นมีค่าน้อยที่สุดเมื่อเทียบกับค่าความแปรปรวนของตัวประมาณค่าตัวอื่นๆ จากการประมาณค่าด้วยวิธี MLS จะได้ค่าของสัมประสิทธิ์ a และ b ซึ่งเมื่อนำไปแทนค่าในสมการถดถอยจะได้สมการพยากรณ์ เพื่อใช้ในการประมาณค่าของตัวแปร Y เมื่อทราบค่าตัวแปร X ค่า Y ที่ได้จากการพยากรณ์ $\hat{Y}$ ต่างจากค่าสังเกตจากตัวแปร Y ถ้านำมาลบกันจะกลายเป็นความคลาดเคลื่อน ผลรวมกำลังสองของความคลาดเคลื่อนเมื่อหารด้วยองศาอิสระ จะได้ค่าความแปรปรวนในการพยากรณ์หรือความแปรปรวนของความคลาดเคลื่อน ประโยชน์ของค่าสถิติที่หาได้คือ ช่วยในการทดสอบสมมติฐานเกี่ยวกับสัมประสิทธิ์การถดถอยและสมมติฐานเกี่ยวกับสัมประสิทธิ์การพยากรณ์

ข้อตกลงเบื้องต้นในการประมาณค่าพารามิเตอร์ด้วยวิธี MLS ตามวิธีของ Gauss มี 4 ข้อคือ

1. เทอมความคลาดเคลื่อนทุกตำแหน่งของค่าตัวแปร X มีการแจกแจงปกติ และมีค่าเฉลี่ยเท่ากับศูนย์ ($\sum e_i = 0$)
2. เทอมความคลาดเคลื่อนของค่าสังเกตแต่ละค่าต้องเป็นอิสระต่อกัน
3. การกระจายต้องมีค่าความแปรปรวนเท่ากันทุกตำแหน่งของค่าตัวแปร X
4. เทอมความคลาดเคลื่อนและตัวแปรต้องเป็นอิสระต่อกัน

การหาค่าสัมประสิทธิ์ของสมการและค่าคงที่ตามวิธีกำลังสองน้อยที่สุด

การเลือกเส้นตรงหนึ่งเส้นก็เท่ากับการหาสัมประสิทธิ์ของเส้นตรงนั้นคือ จุดตัดแกน Y (Intercept : a) และ

ความชัน (slope : b) จากภาพที่ 2 สมมติเส้นตรงที่ลากขึ้นมาคือ $\hat{Y} = a + b X_i$

ดังนั้น ความคลาดเคลื่อน (Error : e_i) คือ

$$e_i = Y_i - \hat{Y}$$

แทนค่า $\hat{Y}$ ในสมการ e_i จะได้

$$e_i = Y_i - a - b X_i$$

ยกกำลังสองจะได้

$$e_i^2 = (Y_i - a - b X_i)^2$$

ผลรวมของทุกจุดจะได้

$$\sum e_i^2 = \sum (Y_i - a - b X_i)^2$$

ตามหลักเกณฑ์ของวิธี MLS เราจะหาค่า a และ b ที่ทำให้ $\sum e_i^2$ มีค่าน้อยที่สุด ใช้หลักแคลคูลัส จะได้

$$\sum e_i^2 = \sum (Y_i - a - b X_i)^2 = M$$

$$\frac{\partial \sum e_i^2}{\partial a} = \frac{\partial M}{\partial a} = 0 = 2 \sum_{i=1}^n [(Y_i - a - b X_i)(-1)] \dots\dots\dots(2)$$

$$\text{และ} \quad \frac{\partial \sum e_i^2}{\partial b} = \frac{\partial M}{\partial b} = 0 = 2 \sum_{i=1}^n [(Y_i - a - b X_i)(-X_i)] \dots\dots\dots(3)$$

แก้สมการ (2) และ (3) จะได้ค่าของ a และ b ดังนี้

$$b = \frac{\sum (X_i Y_i) - \frac{(\sum X_i)(\sum Y_i)}{n}}{\sum X_i^2 - \frac{(\sum X_i)^2}{n}}$$

$$a = \bar{Y} - b \bar{X}$$

จากสมการถดถอยเชิงเส้นตรงตามสมการ (1) จะได้สมการ (4) และ (5) ดังนี้

$$\sum Y_i = na + b \sum X_i \dots\dots\dots(4)$$

$$\sum Y_i X_i = a \sum X_i + b \sum X_i^2 \dots\dots\dots(5)$$

สมการ (4) และ (5) รวมกันเรียกว่า สมการมาตรฐาน (Normal Equations) และเมื่อแก้สมการ (4) และ (5) เพื่อหาค่า a และ b จะได้สมการ (6) และ (7)

$$a = \frac{\sum X^2 \sum Y - \sum x \sum xy}{n \sum X^2 - (\sum X)^2} \dots\dots\dots(6)$$

$$b = \frac{n \sum xy - \sum x \sum Y}{n \sum X^2 - (\sum X)^2} \dots\dots\dots(7)$$

สมการ (6) และ (7) อยู่ในรูป Original forms สามารถดัดแปลงให้อยู่ในรูป deviated forms เพื่อให้สูตรมีความกะทัดรัดเข้าสะดวกในการนำไปใช้ ดังสมการ (8) และ (9)

$$a = \bar{Y} - b\bar{X} \dots\dots\dots(8)$$

$$b = \frac{\sum X_i Y_i}{\sum X_i^2} \dots\dots\dots(9)$$

โดยที่ $X_i = X - \bar{X}$ และ $Y_i = Y - \bar{Y}$

ตัวอย่างการคำนวณ

ในการหาความสัมพันธ์ระหว่างความตั้งใจเรียน(X) กับผลสัมฤทธิ์ทางการเรียน(Y) เพื่อนำมาใช้ในการพยากรณ์ผลสัมฤทธิ์ทางการเรียน(Y) ต่างๆกัน จึงสุ่มนักเรียนมา จำนวน 13 คน แล้วศึกษาเกี่ยวกับความตั้งใจเรียน(X) กับผลสัมฤทธิ์ทางการเรียน(Y) ได้ผลดังนี้

คนที่	Y	X	X^2	XY	Y_i (Y - $\bar{Y}$)	X_i (X - $\bar{X}$)	$X_i Y_i$	X_i^2
1	45	16	256	720	4.31	0.46	1.99	0.21
2	43	15	225	645	2.31	-0.54	-1.24	0.29
3	46	15	225	690	5.31	-0.54	-2.86	0.29
4	44	16	256	704	3.31	0.46	1.53	0.21
5	47	19	361	893	6.31	3.46	21.83	11.98
6	35	18	324	630	-5.69	2.46	-14.01	6.06
7	36	15	225	540	-4.69	-0.54	2.53	0.29
8	38	13	169	494	-2.69	-2.54	6.83	6.44
9	40	15	225	600	-0.69	-0.54	0.37	0.29
10	42	15	225	630	1.31	-0.54	-0.70	0.29
11	41	16	256	656	0.31	0.46	0.14	0.21
12	33	12	144	396	-7.69	-3.54	27.22	12.52
13	39	17	289	663	-1.69	1.46	-2.47	2.14
Σ	529.00	202.00	3180.00	8261.00	0.00	0.00	41.15	41.23
ค่าเฉลี่ย	40.69	15.54						

ไม่ว่าจะคำนวณในสมการที่ (3) และ (4) หรือ สมการที่ (5) และ (6) จะได้ผลลัพธ์เหมือนกัน แทนค่า ในสมการ (3) และ (4) จะได้

$$a = \frac{\sum x^2 \sum y - \sum x \sum xy}{n \sum x^2 - (\sum x)^2} = \frac{(21,755 \times 529) - (202 \times 8,261)}{(13 \times 21,755) - (202 \times 202)}$$

$$= 25.1828$$

$$b = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2} = \frac{13 \times 8,261 - (202 \times 529)}{13 \times 21,755 - (202 \times 202)}$$

$$= 0.9981$$

เพราะฉะนั้น สมการถดถอยคือ $\hat{Y} = a + bX_i$

แทนค่าจะได้ $\hat{Y} = 25.1828 + 0.9981 X_i$

จากสมการการถดถอยมีค่า b เท่ากับ 0.9981 อธิบายได้ว่าเมื่อความตั้งใจเรียน (X) เพิ่มขึ้น 1 หน่วยผลสัมฤทธิ์ทางการเรียนจะเพิ่มขึ้น 0.9981 หน่วย สำหรับ a เท่ากับ 25.1828 อธิบายว่าเมื่อไม่มีความตั้งใจเรียนผลสัมฤทธิ์ทางการเรียนจะมีค่าเท่ากับ 25.1828 หน่วย อย่างไรก็ตามควรอธิบายสมการถดถอยเฉพาะเมื่อความตั้งใจเรียนอยู่ในช่วงที่นำมาศึกษาเท่านั้น เช่น เมื่อนักเรียนมีความตั้งใจเรียนเท่ากับ 16 คะแนน ($X=16$) จะมีผลสัมฤทธิ์ทางการเรียนประมาณ 41.1524 คะแนน

$$\text{แทนค่า } X=16 \text{ ในสมการ } \hat{Y} = 25.1828 + 0.9981 \times 16 = 41.1524$$

สรุปแล้วความเข้าใจในสถานการณ์ต่าง ๆ และความสามารถในการคาดการณ์ล่วงหน้าให้ถูกต้องจึงเป็นประโยชน์อย่างมากในการตัดสินใจ เมื่อทราบว่าตัวแปรมีความสัมพันธ์กันสูง ก็ต้องหาสมการแสดงความสัมพันธ์ระหว่างตัวแปรที่บอกว่าค่าของตัวแปรที่สนใจเปลี่ยนแปลงตามค่าของตัวแปรอื่นอย่างไร สิ่งที่ได้จากสมการดังกล่าวก็คือ การประมาณหรือคาดคะเนค่าของตัวแปรนั้นจากค่าของตัวแปรอื่น ตัวแปรที่สนใจทำนายค่าเราเรียกว่า ตัวแปรตาม (dependent variable) ส่วนตัวแปรอื่นเราเรียกว่า ตัวแปรต้นหรือตัวแปรอิสระ (independent variable) ทั้งนี้ตัวแปรอิสระมีอิทธิพลทำให้ตัวแปรตามเปลี่ยนค่าตามไป เช่น ความสูงของบิดาคือตัวแปรอิสระที่มีอิทธิพลต่อความสูงของบุตรชายซึ่งเป็นตัวแปรตามหรือในเรื่องของเงินเดือนและ GPA เงินเดือนเริ่มต้น คือ ตัวแปรตามที่ต้องกาาคาดคะเนค่าจากตัวแปรอิสระ GPA ถ้ามีข้อมูลเงินเดือนเริ่มต้นและ GPA ของบัณฑิตหลายคนที่ผ่านมา ข้อมูลนั้นนำมาศึกษาหาความสัมพันธ์ระหว่างเงินเดือนเริ่มต้นและ GPA ได้ อย่างไรก็ตามในหลาย ๆ กรณีที่เกิดขึ้น ตัวแปรตามมักได้รับอิทธิพลจากตัวแปรอิสระหลายตัว เงินเดือนเริ่มต้นนอกจากจะขึ้นกับ GPA แล้ว ยังอาจขึ้นกับสาขาวิชาและสถาบันที่จบมา เพศ ประเภทของงาน สถานที่ตั้งของบริษัท และตัวแปรอื่น ๆ อีกมาก โดยทั่วไปจะไม่สามารถรวบรวมข้อมูลของปัจจัยทุกอย่างที่เกี่ยวข้องกับตัวแปรตามมาได้หมด การทำนายค่าของตัวแปรตามให้ถูกต้องแน่นอน จึงเป็นไปได้ยาก สำหรับในที่นี้ จะสนใจเฉพาะรูปแบบความสัมพันธ์อย่างง่ายคือ มีตัวแปรอิสระตัวเดียวและลักษณะความสัมพันธ์อยู่ในรูปเส้นตรง การจะทำนายค่าตัวแปรตามให้ใกล้เคียงจึงอยู่ที่สหสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระต้องมีค่าสูง เพื่อให้สอดคล้องกับแผนภาพการกระจาย ให้ X เป็นตัวแปรอิสระ และ Y เป็นตัวแปรตาม ความสัมพันธ์แบบเส้นตรงระหว่างตัวแปร X และ Y เขียนเป็นสมการเส้นตรงที่ลักษณะของเส้นกำหนดโดยค่าคงที่ 2 ค่า ได้แก่ ค่าของ a และค่าของ b คือ $Y = a + bX$

เอกสารอ้างอิง

- กัลยา วานิชย์บัญชา. หลักสถิติ. พิมพ์ครั้งที่ 11 . กรุงเทพฯ : โรงพิมพ์แห่งจุฬาลงกรณ์, 2552.
- _____. การวิเคราะห์สถิติ : สถิติสำหรับการบริหารและวิจัย. พิมพ์ครั้งที่ 12 . กรุงเทพฯ :
โรงพิมพ์แห่งจุฬาลงกรณ์, 2553.
- คงศักดิ์ สันติพิทยวงศ์ . สถิติสำหรับเศรษฐศาสตร์. กรุงเทพฯ : โรงพิมพ์แห่งจุฬาลงกรณ์, 2551.
- จรัญ จันทลักษณ์. สถิติการวิเคราะห์และการวางแผนงานวิจัย. กรุงเทพฯ : สำนักพิมพ์
มหาวิทยาลัยเกษตรศาสตร์, 2549.
- นงลักษณ์ วิรัชชัย . โมเดลลิสม์: สถิติวิเคราะห์สำหรับการวิจัย. พิมพ์ครั้งที่ 3. กรุงเทพฯ :
จุฬาลงกรณ์มหาวิทยาลัย, 2542.
- สมบัติ ท้ายเรือคำ. สถิติขั้นสูงสำหรับการวิจัยทางการศึกษา. กอสนันท์ : ประสานการพิมพ์, 2552.
- D. Moore and G. McCabe. *Introduction to the Practice of Statistics*, W. H. Freeman and Co.,
London, 2003.
- Pedhazur, E. J. *Multiple regression in behavioral research*. New York : CBS College, 1982.
- MICHAEL L. JOHNSON and LINDSAY M. FAONT. *Parameter Estimation by Least-Squares Methods*.
University of Virginia , 1992.
- Woods, C. M., & Edwards, M. C. *Factor analysis and related methods*.
In C. R. Rao, J. P. Miller, & D. C. Rao
(Eds.), *Handbook of statistics 27: Epidemiology and medical statistics* (pp.367–394). New York :
Elsevier, 2008.