เหมาะสำหรับนักศึกษา
ระดับปริญญาตรี
และปริญญาโท คณะรัฐประศาสนศาสตร์
รัฐศาสตร์ นิติศาสตร์ บริหารธุรกิจ
และเศรษฐศาสตร์

สถิติขั้นสูงเพื่อการวิจัยเชิงปริมาณทางสังคมศาสตร์

(Advanced Statistics for Quantitative Research in Social Science)

การวิจัยในระดับเชิงลึกจำเป็นต้องใช้หลักสถิติขั้นสูงเพื่อสืบหาปัญหาการวิจัยที่ถูกต้องและชัดเจน ซึ่งมีความแตกต่างจากการวิจัยในระดับต้นที่อาจจะใช้สถิติพื้นฐาน ทั้งนี้จะพบความแตกต่างอย่างชัดเจนจากคำตอบของการวิจัยนั้นๆ ดังนั้นสถิติขั้นสูงจะช่วยให้การค้นพบคำตอบการวิจัยได้อย่างมีประสิทธิภาพ และเป็นแนวทางสำหรับการพัฒนาต่อไป

รองศาสตราจารย์ ดร. พฤทธ์สรรค์ สุทธิไชยเมธี

ซีเอ็ด

สถิติขั้นสูงเพื่อการวิจัยเชิงปริมาณทางสังคมศาสตร์

(Advanced Statistics for Quantitative Research in Social Science)

รองศาสตราจารย์ ดร. พฤทธ์สรรค์ สุทธิไชยเมธี

บริษัท ซีเอ็ดยูเคชั่น จำกัด (มหาชน)
SE-EDUCATION PUBLIC COMPANY LIMITED

สถิติขั้นสูงเพื่อการวิจัยเชิงปริมาณทางสังคมศาสตร์

(Advanced Statistics for Quantitative Research in Social Science)

โดย รศ.ดร. พฤทธ์สรรค์ สุทธิไชยเมธี

ข้อมูลทางบรรณานุกรมของสำนักหอสมุดแห่งชาติ

พฤทธ์สรรค์ สุทธิไชยเมธี.
สถิติขั้นสูงเพื่อการวิจัยเชิงปริมาณทางสังคมศาสตร์. -- กรุงเทพฯ : ซีเอ็ดยูเคชั่น, 2563.
340 หน้า
1. สถิติ. 2. สถิติวิเคราะห์.
I. ชื่อเรื่อง.
310

Barcode (e-book) 9786160840182

ผลิตและจัดจำหน่ายโดย

บริษัท ซีเอ็ดยูเคชั่น จำกัด (มหาชน)
SE-EDUCATION PUBLIC COMPANY LIMITED

เลขที่ 1858/87-90 ถนนเทพรัตน แขวงบางนาใต้ เขตบางนา กรุงเทพฯ 10260
โทรศัพท์ 0-2826-8000

หากมีคำแนะนำหรือติชม สามารถติดต่อได้ที่ comment@se-ed.com

คำนำ

หนังสือ *สถิติขั้นสูงเพื่อการวิจัยเชิงปริมาณทางสังคมศาสตร์ (Advanced Statistics for Quantitative Research in Social Science)* เป็นหนังสือที่สร้างสรรค์ขึ้นเพื่อนำความรู้นี้ไปใช้ประกอบการทำวิจัยเชิงปริมาณทางสังคมศาสตร์ในสาขาต่างๆ ได้แก่ รัฐประศาสนศาสตร์ รัฐศาสตร์ นิติศาสตร์ บริหารธุรกิจ และเศรษฐศาสตร์ เป็นต้น ซึ่งจุดประสงค์สำคัญของผู้เขียนคือการนำความรู้ในเชิงลึกหรือสถิติขั้นสูงไปประกอบการสร้างสรรค์งานวิจัยให้เกิดประสิทธิภาพและประสิทธิผลสูงสุดต่อไป เนื่องจากงานวิจัยหลายๆ งานที่ผ่านมายังขาดองค์ความรู้ในขั้นสูง ซึ่งส่งผลให้ประสิทธิภาพของงานวิจัยน้อยลง ดังนั้นหนังสือเล่มนี้จึงมุ่งหวังว่าผู้นำไปใช้จะได้ประโยชน์สูงสุดและเป็นแนวทางสำหรับการวิจัยในอนาคตต่อไป

ผู้เขียนขอขอบคุณครูบาอาจารย์ที่สั่งสอนและให้ความรู้แก่ผู้เขียนตลอดมา และขอขอบคุณผู้เขียนหนังสืองานวิจัยทุกท่านที่เกี่ยวข้องต่างๆ ตลอดจนเอกสารต่างๆ ที่เป็นแนวทางให้ผู้เขียนได้สร้างผลงานนี้ขึ้นมา และผู้เขียนหวังเป็นอย่างยิ่งว่าหนังสือนี้จะได้เป็นแนวทางสำคัญให้ผู้ศึกษาและเรียนรู้ได้นำความรู้ไปใช้ให้เกิดประโยชน์สูงสุดต่อไปอย่างยั่งยืน

ขอแสดงความนับถือ

รศ.ดร. พฤทธ์สรรค์ สุทธิไชยเมธี

สารบัญ

การวิเคราะห์สหสัมพันธ์ (Correlation Analysis)

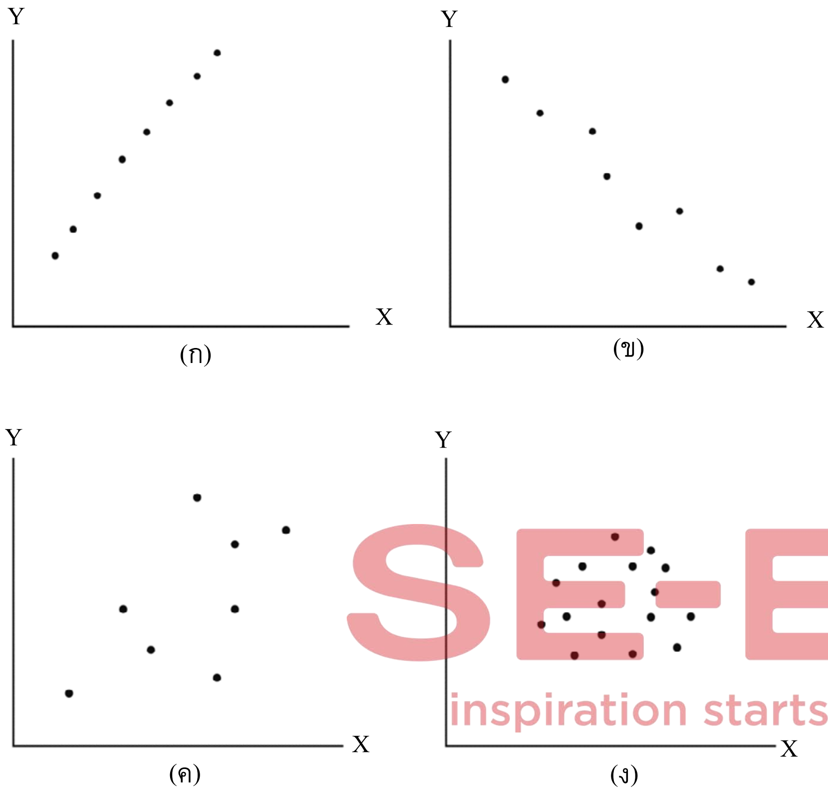
1

การวิเคราะห์สหสัมพันธ์คือ การวิเคราะห์ความสัมพันธ์ที่เกิดขึ้นระหว่างตัวแปรที่ต้องการศึกษา โดยความสัมพันธ์ดังกล่าวอาจจะเป็นความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระ (Dependent Variables and Independent Variables) หรือตัวแปรอิสระกับตัวแปรอิสระ (Independent Variables and Independent Variables) ก็ได้ ซึ่งส่วนใหญ่การวิเคราะห์ความสัมพันธ์เพื่อต้องการวิเคราะห์ผลกระทบที่อาจจะเกิดขึ้นหรือที่กำลังเกิดขึ้นในปัจจุบันและในอนาคต นอกจากนี้การวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรเพื่อต้องการวิเคราะห์ปัญหาที่เกิดขึ้นด้วย ทั้งนี้ขึ้นอยู่กับวัตถุประสงค์ของผู้ศึกษา การวิเคราะห์สหสัมพันธ์แบ่งเป็น 3 ประเภทคือ การวิเคราะห์สหสัมพันธ์อย่างง่าย การวิเคราะห์สหสัมพันธ์พหุคูณ และการวิเคราะห์สหสัมพันธ์บางส่วน โดยมีรายละเอียดดังต่อไปนี้

1.1 การวิเคราะห์สหสัมพันธ์อย่างง่าย (Simple Correlation Analysis)

การวิเคราะห์สหสัมพันธ์อย่างง่ายหมายถึง การวิเคราะห์ความสัมพันธ์ของตัวแปร 2 ตัวแปร ซึ่งอาจจะเป็นความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระ หรือตัวแปรอิสระด้วยกันเองเพียงแค่ 2 ตัวแปรหรือเพียงแค่คู่เดียวเท่านั้น โดยมีทิศทางของความสัมพันธ์ที่สามารถเกิดขึ้นได้ทั้งในทิศทางเดียวกัน (แนวบวก) หรือทิศทางตรงข้ามกัน (แนวลบ) และสามารถทราบขนาดของความสัมพันธ์ได้จากค่าสัมประสิทธิ์ของความสัมพันธ์ (Coefficient of Correlation) โดยค่าสัมประสิทธิ์ของความสัมพันธ์จะมีค่าสูงสุดได้เท่ากับ 1 เท่านั้น กล่าวคือ ถ้าค่าสัมประสิทธิ์ของความสัมพันธ์เป็นบวก แสดงว่ามีความสัมพันธ์กันในทิศทางเดียวกัน และถ้า

เป็นลบแสดงว่ามีความสัมพันธ์กันในทิศทางตรงข้ามกัน และถ้าเป็นศูนย์แสดงว่าไม่มีความสัมพันธ์ใดๆ เกิดขึ้นเลย แสดงได้ดังรูปที่ 1.1



รูปที่ 1.1 ลักษณะความสัมพันธ์ระหว่างตัวแปร X และ Y

จากรูปที่ 1.1 พบว่า รูป (ก) เป็นการแสดงความสัมพันธ์ระหว่าง X และ Y ในลักษณะทิศทางเดียวกัน และมีความสัมพันธ์กันค่อนข้างสูง รูป (ข) เป็นการแสดงความสัมพันธ์ระหว่าง X และ Y ในทิศทางตรงข้ามกันและมีความสัมพันธ์กันค่อนข้างสูง รูป (ค) เป็นความสัมพันธ์กันในทิศทางเดียวกัน แต่ความสัมพันธ์น้อยกว่ารูป (ก) เนื่องจากลักษณะการกระจายตัวของข้อมูลสูงกว่า และสำหรับรูป (ง) แสดงให้เห็นว่า X และ Y ไม่มีความสัมพันธ์กันในลักษณะเส้นตรงเลย เนื่องจากไม่สามารถบอกทิศทางของความสัมพันธ์และขนาดได้เลย กล่าวคือมีค่าสัมประสิทธิ์ของความสัมพันธ์เท่ากับศูนย์

ดังนั้นทุกครั้งที่จะนำตัวแปรใดๆ มาใช้ในแบบจำลอง (Model) จำเป็นอย่างยิ่งที่ผู้วิจัยหรือผู้ศึกษาจะต้องให้ความสัมพันธ์กับการวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์ และควรจะต้องมีการวิเคราะห์ก่อนเสมอ เนื่องจากหากตัวแปรที่นำมาใช้ขาดความสัมพันธ์กันแล้วจะไม่สามารถบอกได้เลยว่าแบบจำลองนั้นมีความน่าเชื่อถือ และจะส่งผลต่อการพยากรณ์ต่อไปในอนาคตด้วย นอกจากนี้การวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์ ทำให้ทราบขนาดของความสัมพันธ์เพื่อช่วยให้การคัดเลือกตัวแปรทำได้ง่ายยิ่งขึ้น ทั้งนี้ในกรณีที่มีตัวแปรอิสระตั้งแต่ 2 ตัวแปรขึ้นไป มีข้อควรระวังสำหรับผู้ที่จะนำไปใช้คือ ต้องระวังเกี่ยวกับทิศทางของความสัมพันธ์ด้วยเสมอ กล่าวคือ ถ้าทิศทางที่แสดงโดยค่าสัมประสิทธิ์สหสัมพันธ์เป็นบวก ในแบบจำลองก็ควรจะต้องเป็นบวกเช่นเดียวกัน และถ้าสัมประสิทธิ์สหสัมพันธ์เป็นลบ ในแบบจำลองก็จะต้องเป็นลบด้วยเสมอ เนื่องจากบางแบบจำลองจะต้องมีการพิจารณาค่าความล่าช้า (Lagged) ซึ่งจะส่งผลให้ทิศทางของสัมประสิทธิ์สหสัมพันธ์มีการเปลี่ยนแปลงได้ ดังนั้นผู้นำความรู้เรื่องดังกล่าวไปใช้ประกอบการวิจัยหรือประกอบการพิจารณาต่างๆ จะต้องให้ความสำคัญกับทิศทางเป็นสำคัญเสมอ

สำหรับการคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์กรณีที่วิเคราะห์ความสัมพันธ์แบบอย่างง่าย หรือวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรเพียงสองตัวแปรนั้น สามารถคำนวณได้จากสูตรต่อไปนี้

$$r_{xy} = \frac{n\sum xY - \sum x \sum Y}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum Y^2 - (\sum Y)^2]}} \qquad ...(1.1)$$

โดยกำหนดให้

r_{xy} คือสัมประสิทธิ์สหสัมพันธ์ ระหว่างตัวแปร **X** และ **Y**

จากสมการที่ (1.1) พบว่า ผลการคำนวณจะทำให้ทราบค่าสัมประสิทธิ์สหสัมพันธ์ โดยค่าสัมประสิทธิ์สหสัมพันธ์จะมีค่าได้สูงสุดคือ 1 ซึ่งอาจจะมีค่าเป็นบวก เป็นลบ หรือเป็นศูนย์ก็ได้ แสดงได้ดังสมการที่ (1.2)

$$-1 \le r_{xy} \le 1 \qquad ...(1.2)$$

จากสมการที่ (1.2) พบว่า ถ้า r_{xy} มีค่าเข้าใกล้ 1 หรือเข้าใกล้ -1 เท่าไร จะแสดงให้เห็นว่าตัวแปรทั้งสองมีความสัมพันธ์กันมากเท่านั้น และทิศทางความสัมพันธ์จะทราบได้จาก

เครื่องหมายของสัมประสิทธิ์สหสัมพันธ์นั้นๆ และถ้า $r_{xy} = 1$ หรือ -1 แสดงให้เห็นความสัมพันธ์แบบสมบูรณ์

ตัวอย่างที่ 1.1 ข้อมูลจากงานวิจัยโดยพฤทธ์สรรค์ สุทธิไชยเมธี แสดงความสัมพันธ์ระหว่างตัวแปรตาม Y กับตัวแปรอิสระ X ดังนี้

วิธีทำ

จากสูตร $r_{xy} = \dfrac{n\sum xY - \sum x \sum Y}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum Y^2 - (\sum Y)^2]}}$ สามารถคำนวณได้จากตารางต่อไปนี้

Y	X	XY	X^2	Y^2
5	2	10	4	25.00
15	15	225	225	225.00
12	8	96	64	144.00
12	10	120	100	144.00
17	10	170	100	289.00
16	8	128	64	256.00
21	12	252	144	441.00
20	15	300	225	400.00
20	17	340	289	400.00
25	20	500	400	625.00
24	15	360	225	576.00
30	14	420	196	900.00
15	18	270	324	225.00
30	15	450	225	900.00
27	17	459	289	729.00
$\sum Y = 289$	$\sum X = 196$	$\sum XY = 4,100$	$\sum x^2 = 2,874$	$\sum Y^2 = 6,279$

$$\text{สูตร } r_{xy} = \frac{n\sum xY - \sum x \sum Y}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum Y^2 - (\sum Y)^2]}}$$

$$= \frac{15(4,100) - (196)(289)}{\sqrt{\left[15(2,874) - (38,416)\right]\left[15(6,279) - (83,521)\right]}}$$

$$= \frac{61,500 - 56,644}{\sqrt{\left[43,110 - 38,416\right]\left[94,185 - 83,521\right]}}$$

$$= \frac{4,856}{\sqrt{4,694 \times 10,664}}$$

$$r_{xy} = 0.68$$

ดังนั้น ถ้า X เปลี่ยนแปลงส่งผลให้ Y เกิดการเปลี่ยนแปลงด้วยในทิศทางเดียวกัน ซึ่งขนาดของความสัมพันธ์เท่ากับ 0.68 และการเปลี่ยนแปลงในทิศทางเดียวกัน กล่าวคือ ถ้า X เพิ่มขึ้น Y ก็จะเพิ่มขึ้น แต่ถ้า X ลดลง Y ก็จะต้องลดลงด้วยเสมอ

อย่างไรก็ตาม เมื่อผู้วิจัยหรือผู้ศึกษาสามารถคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ได้แล้วจะมี ผลดีกับแบบจำลอง กล่าวคือ สามารถทราบค่าสัมประสิทธิ์การตัดสินใจ (Coefficient of Determination) หรือเป็นค่าที่บอกให้ทราบว่าตัวแปรอิสระสามารถทำนายตัวแปรตาม โดยได้จากค่าสัมประสิทธิ์สหสัมพันธ์ยกกำลังสอง (r^2) ดังนั้นในตัวอย่างที่ 1 พบว่าค่าสัมประสิทธิ์การตัดสินใจจึงมีค่าเท่ากับ $(0.68)^2 = 0.46$ แสดงให้เห็นว่าตัวแปรอิสระ X สามารถอธิบายตัวแปรตาม Y ได้ 46 เปอร์เซ็นต์

1.2 การวิเคราะห์สหสัมพันธ์พหุคูณ (Multiple Correlation Analysis)

การวิเคราะห์สหสัมพันธ์พหุคูณหมายถึง การวัดความสัมพันธ์ของตัวแปรตั้งแต่ 3 ตัวขึ้นไป จะต้องเป็นความสัมพันธ์ระหว่างตัวแปรตามกับกลุ่มตัวแปรอิสระ โดยค่าที่ได้จากการคำนวณจะเป็นค่าสัมประสิทธิ์สหสัมพันธ์ (Coefficient of Multiple Correlation) ซึ่งบอกถึงขนาดความสัมพันธ์และทิศทางของความสัมพันธ์ดังกล่าว โดยมีลักษณะเหมือนกรณีสหสัมพันธ์

เชิงเดี่ยว กล่าวคือ ทิศทางความสัมพันธ์ในทิศทางเดียวกัน ทิศทางตรงข้ามกัน และไม่มีความสัมพันธ์กันเลย

$$r = \sqrt{\frac{\sum(\hat{Y}-\overline{Y})^2}{\sum(Y-\overline{Y})^2}} \qquad ...(1.3)$$

โดยกำหนดให้

Y คือค่าของตัวแปรตาม

$\overline{Y}$ คือค่าเฉลี่ยของตัวแปรตาม

$\hat{Y}$ คือค่าพยากรณ์ของตัวแปรตาม

สำหรับในกรณีที่มีตัวแปรอิสระตั้งแต่ 2 ตัวขึ้นไปนั้น จะสามารถใช้สูตรต่อไปนี้ในการคำนวณก็ได้เพื่อให้ง่ายต่อการนำไปใช้มากยิ่งขึ้น ดังนี้

$$r = \sqrt{\frac{n(\beta_0\sum Y + \beta_1\sum X_1Y + \beta_2\sum X_2Y) - (\sum Y)^2}{n\sum Y^2 - (\sum Y)^2}} \qquad ...(1.4)$$

โดยกำหนดให้

Y คือตัวแปรตาม

X_1 คือตัวแปรอิสระตัวที่หนึ่ง

X_2 คือตัวแปรอิสระตัวที่สอง

β_0 คือค่าคงที่

β_1 คือค่าสัมประสิทธิ์การถดถอยพหุคูณของ X_1

β_2 คือค่าสัมประสิทธิ์การถดถอยพหุคูณของ X_2

จากสมการที่ (1.4) พบว่า สามารถคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณซึ่งจะมีขนาดอยู่ระหว่าง 0 ถึง 1 เท่านั้น สำหรับเครื่องหมายบวกหรือลบคือ ทิศทางของความสัมพันธ์ แต่จะไม่เกี่ยวกับขนาดของความสัมพันธ์ โดยแสดงวิธีการคำนวณได้ดังตัวอย่างต่อไปนี้

ตัวอย่างที่ 1.2 ข้อมูลจากงานวิจัยโดยพฤทธ์สรรค์ สุทธิไชยเมธี แสดงการวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณ

Y	X_1	X_2	X_1Y	X_2Y	X_1X_2	X_1^2	X_2^2	Y^2
5	2	10	10	50	50	20	4	100
15	15	5	225	75	75	75	225	25
12	8	20	96	240	240	160	64	400
12	10	25	120	300	300	250	100	625
17	10	12	170	204	204	120	100	144
16	8	15	128	240	240	120	64	225
21	12	5	252	105	105	60	144	25
20	15	10	300	200	200	150	225	100
20	17	15	340	300	300	255	289	225
25	20	30	500	750	750	600	400	900
24	15	35	360	840	840	525	225	1225
30	14	30	420	900	900	420	196	900
15	18	12	270	180	180	216	324	144
30	15	10	450	300	300	150	225	100
27	17	15	459	405	405	255	289	225
289	**196**	**249**	**4,100**	**5,089**	**3,376**	**2,874**	**5,363**	**6,279**

$n = 15$ $\beta_0 = 4.14$ $\beta_1 = 0.97$ $\beta_2 = 0.14$

$$r = \sqrt{\frac{n(\beta_0\sum Y + \beta_1\sum X_1Y + \beta_2\sum X_2Y) - (\sum Y)^2}{n\sum Y^2 - (\sum Y)^2}}$$

$$= \sqrt{\frac{15[(4.14)(289) + (0.97)(4,100) + (0.14)(5,089)] - [289]^2}{15(6,279) - (289)^2}}$$

$$= 0.70$$

ดังนั้นการวิเคราะห์สหสัมพันธ์พหุคูณในตัวอย่างที่ 1.2 ตัวแปรตามกับตัวแปรอิสระทั้งสองตัว มีความสัมพันธ์กันในทิศทางเดียวกันเท่ากับ 0.70

อย่างไรก็ตาม จากตัวอย่างดังกล่าวทำให้ผู้วิจัยสามารถนำค่าสหสัมพันธ์พหุคูณไปวิเคราะห์ค่าสัมประสิทธิ์การตัดสินใจได้ โดยการนำไปยกกำลังสอง ดังนี้

$$R^2 = \sqrt{r} \qquad \text{...(1.5)}$$

จากสมการที่ (1.5) พบว่า ค่าสัมประสิทธิ์แห่งการตัดสินใจมีค่าเท่ากับ 0.50 กล่าวคือ ตัวแปรอิสระทั้งสองตัวแปรสามารถอธิบายการตัดสินใจของตัวแปรตามได้ 50 เปอร์เซ็นต์

1.3 สหสัมพันธ์บางส่วน (Partial Correlation)

สหสัมพันธ์บางส่วนหมายถึง การวิเคราะห์ความสัมพันธ์ของตัวแปรอิสระตั้งแต่ 2 ตัวขึ้นไปโดยการพิจารณาได้แต่ละคู่ความสัมพันธ์ ถ้ากรณีที่ตัวแปรอิสระมีมากกว่า 2 ตัวแปร จะต้องสมมติให้ตัวแปรอิสระอื่นๆ คงที่ โดยปล่อยให้เป็นอิสระเฉพาะตัวแปรคู่ที่ศึกษาเท่านั้น โดยกำหนดให้ตัวแปรอิสระคู่ที่ศึกษาได้แก่ ตัวแปร i และตัวแปร j และสำหรับตัวแปรอิสระอื่นๆ คงที่ คือ C โดยแสดงได้ดังนี้

$$r_{ij,C} = \frac{r_{ij} - (r_{iC})(r_{jC})}{\sqrt{(1-r_{iC}{}^2)(1-r_{jC}{}^2)}} \qquad \text{...(1.6)}$$

สำหรับขนาดของสัมประสิทธิ์สหสัมพันธ์บางส่วนนั้นจะมีค่าอยู่ระหว่าง 0 ถึง 1 เท่านั้น จะไม่พิจารณาเครื่องหมายลบ เพราะเครื่องหมายเป็นแค่ทิศทางของความสัมพันธ์เท่านั้นเอง

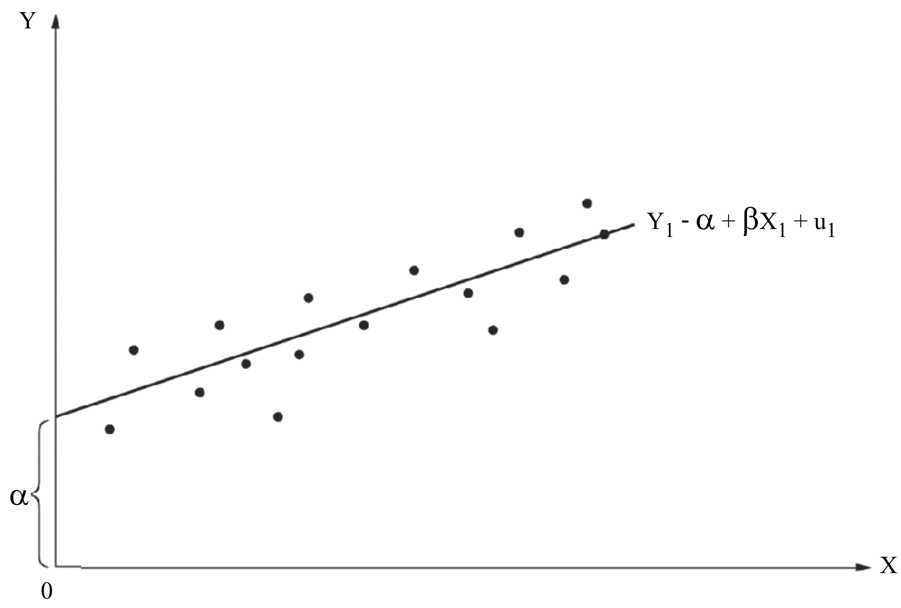
ดังนั้นจะเห็นได้ว่าการคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์แต่ละกรณีนั้นมีความสำคัญยิ่งกับผู้ที่ศึกษาและผู้วิจัย เนื่องจากถ้าตัวแปรที่นำมาศึกษาในแบบจำลองไม่มีความสัมพันธ์กันแล้วแบบจำลองที่สร้างขึ้นก็ขาดความน่าเชื่อถือและไม่สามารถนำไปใช้ประโยชน์ต่อได้ เช่น การนำไปใช้สำหรับการพยากรณ์ก็จะเกิดความคลาดเคลื่อนสูง เนื่องจากตัวแปรที่ศึกษาไม่ได้มีความสัมพันธ์มาตั้งแต่ต้น นอกจากนี้ผู้วิจัยจะต้องให้ความสำคัญกับการพิจารณาค่าตัวแปรที่นำมาใส่ในแบบจำลองแต่ละแบบจำลอง โดยจะต้องไม่ให้ตัวแปรอิสระมีความสัมพันธ์กันเองด้วย เนื่องจากจะทำให้เกิดปัญหา Multicollinearity ซึ่งจะมีผลทำให้เกิด Spurious ในแบบจำลอง

การวิเคราะห์สมการถดถอย (Linear Regression Analysis)

2

2.1 การวิเคราะห์ถดถอยอย่างง่าย (Simple Linear Regression Analysis)

การวิเคราะห์ถดถอยอย่างง่าย หรือการวิเคราะห์ถดถอยเชิงเดี่ยว เป็นการศึกษาถึงความสัมพันธ์ของตัวแปร 2 ตัวแปรคือ ตัวแปรตาม (Dependent Variable) กับตัวแปรอิสระ (Independent Variable) ซึ่งความสัมพันธ์ดังกล่าวจะอยู่ในรูปเส้นตรง ดังแสดงในรูปที่ 2.1



รูปที่ 2.1 แสดงความสัมพันธ์ของตัวแปรตามกับตัวแปรอิสระในรูปแบบเชิงเส้น

จากรูปที่ 2.1 เมื่อนำตัวแปรตามและตัวแปรอิสระมาเขียนกราฟแสดงความสัมพันธ์ เรียกว่า Scatter Diagram รูปแบบสมการเชิงเส้นอย่างง่าย (Simple Regression) มีเป้าหมายที่สำคัญคือ การวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรทั้งสอง กล่าวคือ เมื่อมีการเปลี่ยนแปลงของ X จะมีอิทธิพลทำให้ Y เปลี่ยนแปลงตาม โดยจะเปลี่ยนแปลงมากน้อยเพียงใดขึ้นอยู่กับขนาดความสัมพันธ์ของตัวแปรทั้งสอง ซึ่งเมื่อได้ความสัมพันธ์แล้วสามารถนำมาวิเคราะห์การพยากรณ์ต่อไปในอนาคตได้ รูปแบบสมการเชิงเส้นอย่างง่าย (Simple Regression) แสดงดังนี้

$$Y_i = \alpha + \beta_1 X_i + u_i \qquad \text{...(2.1)}$$

โดยกำหนดให้

Y_i คือตัวแปรตาม (Dependent Variable)

X_i คือตัวแปรอิสระ (Independent Variable)

α คือส่วนตัดแกน Y (Intercept)

β_1 คือ Slope

u_i คือความคลาดเคลื่อนอย่างสุ่ม (Random Error)

k คือจำนวนตัวแปรอิสระ + 1 (ใน Simple Regression k = 2)

i คือ 1, 2, 3, 4, …, n

จากสมการที่ (2.1) พบว่า ความสัมพันธ์ระหว่าง Y กับ X เป็นความสัมพันธ์ที่แท้จริงที่แสดงการเปลี่ยนแปลงของ Y อันเนื่องมาจากการเปลี่ยนแปลงของ X แต่เราไม่สามารถทราบได้ว่าความสัมพันธ์ดังกล่าววางอยู่ ณ ตำแหน่งใด จึงจำเป็นจะต้องหาตำแหน่งที่ตั้งที่ใกล้เคียงมากที่สุดจากข้อมูลที่มีอยู่ เรียกว่า วิธีการประมาณค่า (Estimate) นอกจากปัจจัย X ที่ทำให้ Y เปลี่ยนแปลงแล้ว ยังพบว่าตัวแปรภายนอกอื่นๆ ก็ย่อมส่งผลกระทบให้เกิดการเปลี่ยนแปลงของ Y ได้เช่นเดียวกัน เรียกตัวแปรนั้นว่าตัวแปรที่ไม่อาจวัดค่าหรือสังเกตได้คือ u_i (Unobservable Variable) หรือแม้ว่าค่าสังเกต u_i จะมีอยู่ตามธรรมชาติหรือสามารถวัดได้ก็ตาม แต่บางครั้งเป็นการยากที่จะควบคุมได้ทั้งขนาดเครื่องหมายและทิศทาง ดังนั้นการประมาณค่าให้ได้ตัวประมาณค่าที่ดี (Estimator) จะต้องมีข้อสมมติ (Assumption) ดังต่อไปนี้

2.1.1 ข้อสมมติเบื้องต้น (Basic Assumption) ของการวิเคราะห์ถดถอยอย่างง่าย

1. กำหนดให้ u_i เป็นตัวแปรเชิงสุ่ม

จากสมการที่ (2.1)

$$Y_i = \alpha + \beta_1 X_i + u_i \qquad ...(2.1)$$

พบว่า Y_i เป็นฟังก์ชันเชิงเส้น (Linear Function) ของ u_i กล่าวคือ ถ้ากำหนดให้ u_i เป็นตัวแปรสุ่ม

การเปลี่ยนแปลงหรือการเคลื่อนไหวของ u_i จะถูกควบคุมด้วยทฤษฎีความน่าจะเป็น ทำให้สามารถควบคุมการเคลื่อนไหวต่างๆ ได้ ซึ่งจะส่งผลให้ Y_i เป็นตัวแปรสุ่มด้วย

2. ค่าเฉลี่ย (Mean) ของ u_i มีค่าเท่ากับ 0

$$E(u_i) = 0 \quad ; i = 1, 2, 3, 4, 5 ..., n$$

เมื่อทำการทดลองซ้ำๆ ณ $X = X_i$ พบว่าจะมีค่า u_i หลายค่าซึ่งมีทั้งบวกและลบ แต่เมื่อนำมา ถัวเฉลี่ย (Mean) แล้วจะมีค่าเท่ากับ 0 ($E(u_i) = 0$) ดังนั้น

$$\mu_{y/x} = E(Y_i)$$

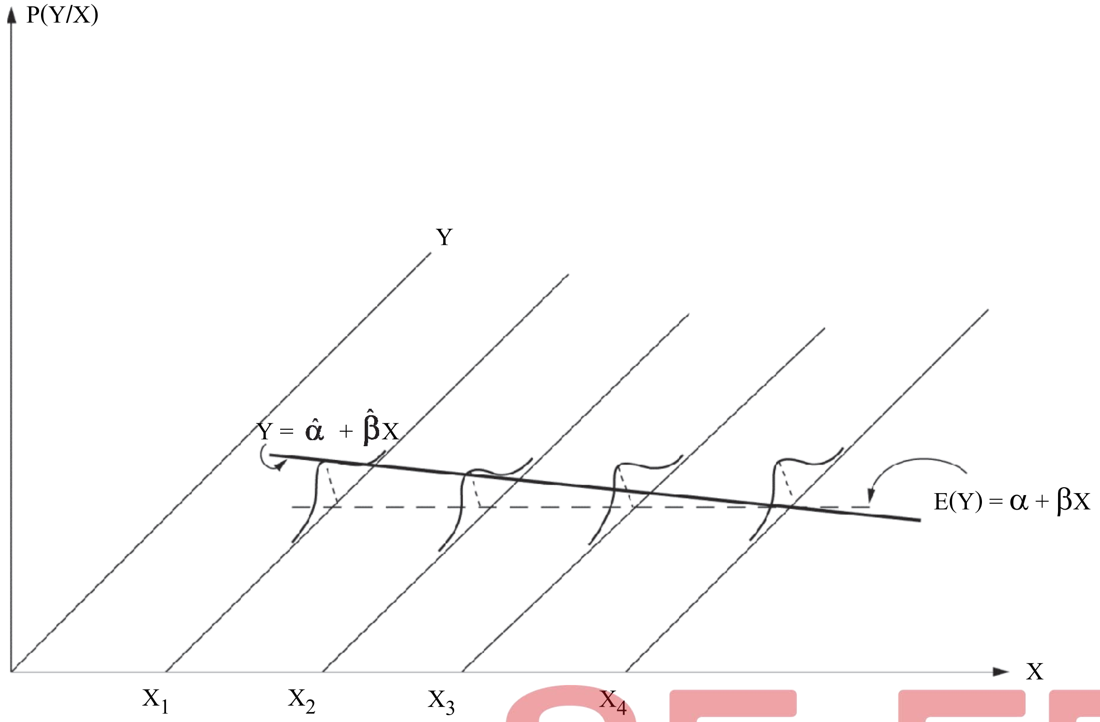
$$= E(\alpha + \beta_1 X_i + u_i)$$

$$\mu_{y/x} = \alpha + \beta_1 X_i \qquad ...(2.2)$$

3. ณ ความคลาดเคลื่อนต่างกัน

u_i และ u_j ค่า $E(u_i u_j) = 0$

กล่าวคือ ค่าความคลาดเคลื่อน u_i และ u_j จะต้องเป็นอิสระจากกัน เนื่องจากเมื่อ $E(u_i u_j) \neq 0$ หรือ $E(u_i u_j) = 0$; $i \neq j$ แสดงว่ารูปแบบสมการนั้นจะเกิดปัญหา Autocorrelation ขึ้น และเมื่อตัวคลาดเคลื่อนเป็นอิสระจากกัน สามารถแสดงความสัมพันธ์ได้ดังรูปที่ 2.2



รูปที่ 2.2 แสดงความสัมพันธ์ของ u_i และ u_j ที่เป็นอิสระกัน

4. ความแปรปรวน (Variance) ของ u_i มีค่าคงที่

$$Var(u_i) = E(u_i)^2 = \sigma_u^2$$

ในการประมาณค่านั้น จำเป็นจะต้องกำหนดให้ค่า Variance หรือการกำหนดให้ฐานโค้งของตัวแปรสุ่ม u_i มีความกว้างคงที่เท่ากับ σ^2 แสดงได้ ดังนี้

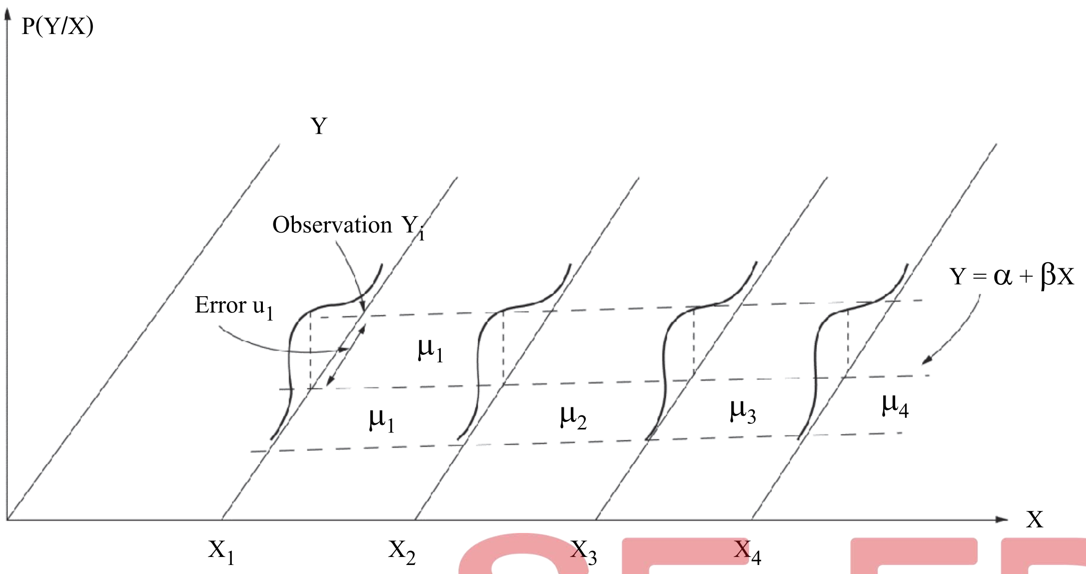
$$Var(u_i) = \sigma_u^2 \qquad ...(2.3)$$

จากสมการที่ (2.3) พบว่า เมื่อกำหนด u_i ก็แสดงว่าเป็นการกำหนด Y_i นั่นเอง แสดงได้ดังนี้

$$\begin{aligned} Var(Y_i) &= Var(\alpha + \beta_1 X_i + u_i) \qquad ...(2.4) \\ &= Var(\alpha) + Var(\beta_1 X_i) + Var(u_i) \\ &= 0 + 0 + Var(u_i) \end{aligned}$$

$$Var(Y_i) = \sigma^2 \;;\; I = 1, 2, 3, 4, 5, \ldots, n \qquad ...(2.5)$$

จากสมการที่ (2.3) และสมการที่ (2.5) พบว่า เมื่อกำหนดให้ความกว้างของ u_i คงที่เท่ากับ σ^2 ก็จะส่งผลให้ความกว้างของ Y_i คงที่เท่ากับ σ^2 เช่นกัน แสดงได้ดังรูปที่ 2.3



รูปที่ 2.3 แสดงความกว้างของ u_i และ Y_i ซึ่งมีค่าคงที่เท่ากับ σ^2

5. ui มีการแจกแจงแบบ N(0, σ^2) หรือเรียกว่า การแจกแจงแบบปกติ (Normal Distribution)

$$u_i \sim N(0, \sigma_u^2) \qquad ...(2.6)$$

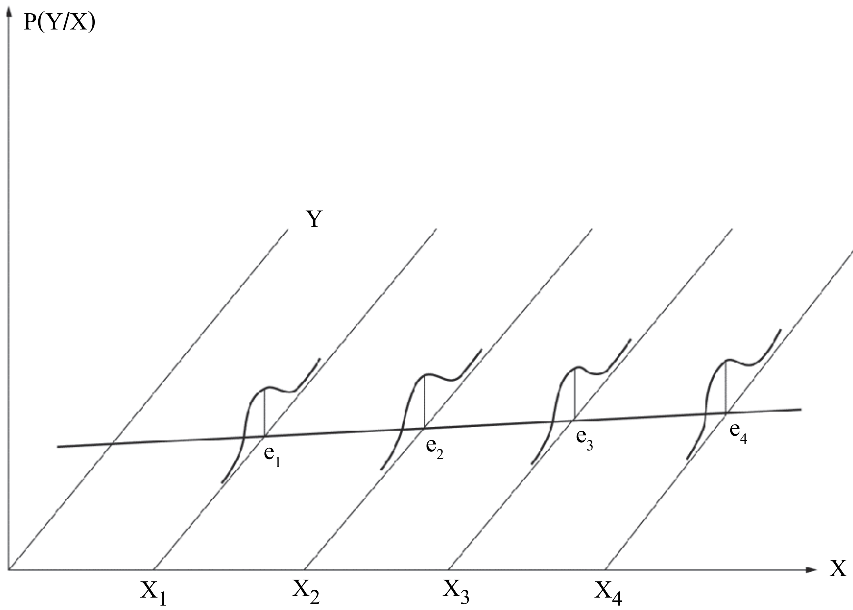
จากสมการที่ (2.6) แสดงให้ทราบได้ว่า

$$u_i \sim N(E(Y_i), \sigma_u^2)$$

$$Y_i \sim N(E(\alpha + \beta X_i), \sigma_u^2)$$

$$Y_i \sim N(0, \sigma_u^2)\ ;\ i = 1, 2, 3, 4, 5, \ldots, n \qquad ...(2.7)$$

จากสมการที่ (2.6) และสมการที่ (2.7) พบว่า เส้น True Line จะผ่านจุด Mean ของ Y_i โค้งปกตินั้น จะมีโค้งกว้างเท่ากัน หรือมีส่วนสูงเท่ากัน ดังแสดงในรูปที่ 2.4



รูปที่ 2.4 แสดง u_i หรือ Y_i มีการแจกแจงแบบ $N(0, \sigma^2)$

อย่างไรก็ตามพบว่า เมื่อ u_i มีการแจกแจงแบบไม่ปกติ (Non Normal Error) (การทดสอบว่า u_i มีการแจกแจงแบบปกติหรือไม่ อาจจะทดสอบได้หลายวิธีด้วยกัน เช่น Kolmogorov Smirnov Test หรือ Shapiro–Wilk Statistics) จะไม่สามารถใช้ตัวทดสอบ t–test, F–test, χ^2 –test และ test ต่างๆ ที่เกี่ยวข้องกับการแจกแจงแบบปกติได้ แต่จะใช้ตัวทดสอบที่เรียกว่า Robust Regression

6. u_i กับ X_i จะต้องเป็นอิสระจากกัน

$$Cov(u_i, X_j) = 0 \; ; i = 1, 2, 3, \ldots, n \; ; j = 1, 2, 3, \ldots, m$$

u_i กับ X_i จะต้องเป็นอิสระจากกัน กล่าวคือ เมื่อเป็นการกำหนดให้สมการเชิงเส้นยังคงเดิมคือ กำหนดให้ตัวแปรตามคือ Y_i ส่วนตัวแปรอิสระคือ X_i แสดงในรูปฟังก์ชันได้ว่า $Y_i = f(X_i)$ หรือเพื่อป้องกันไม่ให้สมการเชิงเส้นเปลี่ยนแปลงไปจากเดิมเป็น $X_i = f(Y_i)$ นั่นเอง ส่วนสาเหตุของการเปลี่ยนแปลงเป็นเพราะตัวแปรบางคู่มีลักษณะเป็น Casual Relationship หรือเป็นเหตุเป็นผลซึ่งกันและกัน

อย่างไรก็ตาม พบว่า ในสมการ Regression Model จะกำหนดให้ตัวแปรตามกับตัวแปรอิสระมีความสัมพันธ์กันในรูปสมการเชิงเส้น $Y_i = f(X_i)$ กล่าวคือ การที่ X_i เปลี่ยนแปลงจะส่งผลให้ Y_i เปลี่ยนแปลงตาม นั่นก็คือ Y_i เป็นตัวแปรภายใน (Endogenous Variable) ส่วน X_i เป็นตัวแปรภายนอก (Exogenous Variable) และยังเป็นตัวแปรที่เรียกว่า Mathematical Variable หรือ Stochastic Variable แต่ถ้าในสมการ Simultaneous Model หรือ Multi Equation Model พบว่า X_i สามารถเป็นได้ทั้งตัวแปรตามและตัวแปรอิสระหรือเป็นได้ทั้งตัวแปรภายในและตัวแปรภายนอก

2.1.2 การประมาณค่าพารามิเตอร์ของสมการถดถอยอย่างง่าย

การประมาณค่าพารามิเตอร์ (Parameter) ของสมการถดถอยอย่างง่าย (Simple Regression) เป็นการวิเคราะห์หาขนาดของความสัมพันธ์และทิศทางของความสัมพันธ์ระหว่างตัวแปรตาม (Dependent Variable) กับตัวแปรอิสระ (Independent Variable) ซึ่งผลลัพธ์ที่ได้จะแสดงออกมาในรูปของ α และ β แสดงรายละเอียดได้ ดังนี้

$$Y_i = f(X_i)$$

จาก True Line ในสมการที่ (2.1)

$$Y_i = \alpha + \beta_1 X_i + u_i$$

จากสมการที่ (2.1) จะไม่สามารถทราบได้ว่าเส้น $Y_i = \alpha + \beta_1 X_i$ วางอยู่แน่นอน ณ ตำแหน่งใด วิธีการที่จะทำให้ทราบได้ว่าอยู่ ณ ตำแหน่งใดหรือให้ใกล้เคียงมากที่สุดคือ การประมาณการ

การประมาณการจำเป็นจะต้องมีข้อมูลเชิงสังเกต (Observation) แล้วคำนวณหาค่า α, β_1 เพื่อให้ทราบความสัมพันธ์ระหว่าง Y_i กับ X_i ว่าจะมากน้อยเพียงใด และมีทิศทางความสัมพันธ์อย่างไร แสดงได้ดังนี้

กำหนดให้ฟังก์ชันค่าสังเกตคือ

$$Z_i = (Y_i, X_i) \ ; \ i = 1, 2, 3, 4, 5, \ldots, n \qquad \ldots(2.8)$$

สำหรับ Simple Linear Regression กำหนดให้ $X_i = X_i$ ดังนั้นจากสมการที่ (2.8) สามารถนำมาประมาณการได้ ดังนี้

$$Y_i = \hat{\alpha} + \hat{\beta}_1 X_1 + \hat{e}_i$$

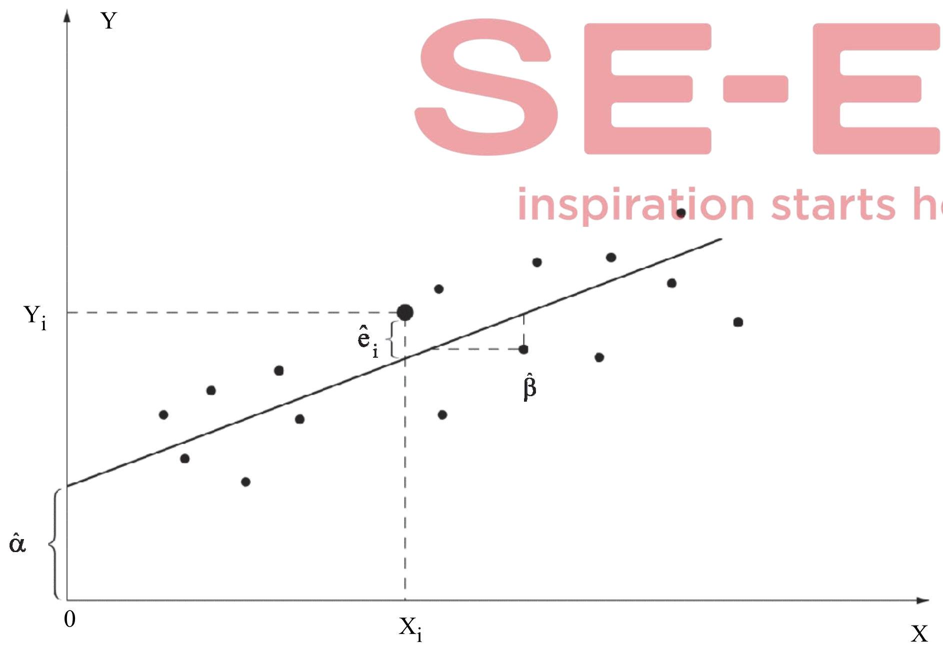
หรือ

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}_1 X_1 \qquad ...(2.9)$$

โดยกำหนดให้

$\hat{Y}_i$	คือค่าประมาณการของ Y_i หรือความสูงของเส้นตรง
$\hat{\alpha}$	คือ Y–intercept หรือตัวประมาณค่า α
$\hat{\beta}_1$	คือ Slope ของเส้น Regression หรือตัวประมาณค่า β_1
$\hat{e}_i$ หรือ e_i	คือส่วนเบี่ยงเบน หรือค่าความคลาดเคลื่อน หรือส่วนเหลือ (Error Term or Residual Term or Disturbances Term)

จากสมการที่ (2.9) แสดงความสัมพันธ์ของสมการประมาณการได้ดังแสดงในรูปที่ 2.5



รูปที่ 2.5 แสดงความสัมพันธ์ของสมการประมาณการ

จากรูปที่ 2.5 พบว่า มีหลายวิธีในการคำนวณหาตัวประมาณค่าที่ดีที่สุด (Best Estimator) แต่วิธีที่นิยมมากและให้ผลการประมาณการใกล้เคียงกับสมการจริงมากที่สุด (True Equation)

คือวิธี Ordinary Least Square (OLS) หรือ Least Square หรือ Classical Least Square (CLS) แสดงรายละเอียดได้ดังนี้

นำสมการที่ (2.1) – สมการที่ (2.9) $\hat{\alpha} + \hat{\beta}_1 X_1$

$$\hat{e}_i = Y_i - \hat{Y}_i$$

$$\hat{e}_i = Y_i - \hat{\alpha} - \hat{\beta}_1 X_1 \qquad ...(2.10)$$

จากสมการที่ (2.9) และสมการที่ (2.10) เมื่อรวมความคลาดเคลื่อนของแต่ละจุดเข้าด้วยกันและไม่ต้องคำนึงถึงเครื่องหมาย แสดงได้ดังนี้

$$\sum_{i=1}^{n} |\hat{e}_i| = \sum_{i=1}^{n} |Y_i - \hat{Y}_i|$$

$$\sum_{i=1}^{n} |\hat{e}_i| = \sum_{i=1}^{n} |Y_i - \hat{\alpha} - \hat{\beta}_1 X_1| \qquad ...(2.11)$$

จากสมการที่ (2.10) และสมการที่ (2.11) ยังพบปัญหาที่อาจจะเกิดขึ้นได้กับสมการประมาณการคือ อาจทำให้ $\sum^{n} \hat{e}_i$ มีค่าเป็นศูนย์หรือน้อยกว่าศูนย์ ซึ่งผิดจากค่าที่ควรจะเป็น ทั้งนี้ หากต้องการตัดปัญหาเรื่องเครื่องหมายดังกล่าว สามารถคำนวณด้วยการยกกำลังสอง แสดงได้ ดังนี้

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

$$\sum_{i=1}^{n} \hat{e}_i^2 = \sum_{i=1}^{n} (Y_i - \hat{\alpha} - \hat{\beta}_1 X_1)^2 \qquad ...(2.12)$$

จากสมการที่ (2.12) พบว่า ผลรวมของส่วนที่เบี่ยงเบนไปจากเส้นตรงนั้น จะเบี่ยงเบนไปไม่ว่าทางด้านบวกหรือทางด้านลบ แต่ผลรวมของการเบี่ยงเบนจะเข้าใกล้ศูนย์มากที่สุด ดังนั้นในการคำนวณเส้นที่ดีที่สุดจะคำนวณโดยใช้วิธี กำลังสองน้อยที่สุด (Ordinary Least Square; OLS)

การหาค่า α, β_1 โดยการ Take Partial Derivative มุ่งตรงต่อ α และ β_1 แล้ว Set ให้เท่ากับศูนย์

$$\frac{\partial}{\partial \hat{\alpha}} \sum_{i=1}^{n} \hat{e}_i^2 = \frac{\partial}{\partial \hat{\alpha}} \sum_{i=1}^{n} (Y_i - \hat{\alpha} - \hat{\beta}_1 X_1)^2 = 0 \qquad ...(2.13)$$

$$2\sum_{i=1}^{n}(Y_i - \hat{\alpha} - \hat{\beta}_1 X_1)(-1) = 0$$

$$-2\sum_{i=1}^{n}(Y_i - \hat{\alpha} - \hat{\beta}_1 X_1) = 0$$

$$-2\sum_{i=1}^{n} Y_i + 2\sum_{i=1}^{n} \hat{\alpha} + 2\sum_{i=1}^{n} \hat{\beta}_1 X_1 = 0$$

$$-2\sum_{i=1}^{n} Y_i + 2n\hat{\alpha} + 2\sum_{i=1}^{n} \hat{\beta}_1 X_1 = 0 \qquad ...(2.14)$$

จากสมการที่ (2.14) หารด้วย 2 ตลอด

$$-\sum_{i=1}^{n} Y_i + n\hat{\alpha} + \sum_{i=1}^{n} \hat{\beta}_1 X_1 = 0 \qquad ...(2.15)$$

จากสมการที่ (2.15) จัดให้อยู่ในรูปของ Normal Equation

$$\sum_{i=1}^{n} Y_i = n\hat{\alpha} + \sum_{i=1}^{n} \hat{\beta}_1 X_1 \qquad ...(2.16)$$

$$\frac{\partial}{\partial \beta_1}\sum_{i=1}^{n} \hat{e}_i^2 = \frac{\partial}{\partial \hat{\beta}_1}\sum_{i=1}^{n}(Y_i - \hat{\alpha} - \hat{\beta}_1 X_1)^2 = 0 \qquad ...(2.17)$$

$$2\sum_{i=1}^{n}(Y_i - \hat{\alpha} - \hat{\beta}_1 X_1)(-X_1) = 0$$

$$-2\sum_{i=1}^{n} X_i Y_i + 2\sum_{i=1}^{n} \hat{\alpha} X_1 + 2\sum_{i=1}^{n} \hat{\beta}_1 X_1^2 = 0 \qquad ...(2.18)$$

จากสมการที่ (2.18) หารด้วย 2 ตลอด

$$-\sum_{i=1}^{n} X_1 Y_i + \hat{\alpha}\sum_{i=1}^{n} X_1 + \hat{\beta}_1 \sum_{i=1}^{n} X_1^2 = 0 \qquad ...(2.19)$$

จากสมการที่ (2.19) จัดให้อยู่ในรูปของ Normal Equation

$$\sum_{i=1}^{n} X_1 Y_i = \hat{\alpha}\sum_{i=1}^{n} X_i + \hat{\beta}_1 \sum_{i=1}^{n} X_1^2 \qquad ...(2.20)$$

จากสมการที่ (2.16) และสมการที่ (2.20) เรียกว่า สมการ Normal Equations

$$n\hat{\alpha} + \sum_{i=1}^{n} \hat{\beta}_1 X_i = \sum_{i=1}^{n} Y_i$$

$$\hat{\alpha}\sum_{i=1}^{n} X_1 + \hat{\beta}_1 \sum_{i=1}^{n} X_1^2 = \sum_{i=1}^{n} X_1 Y_i$$

จากสมการ Normal Equations สามารถคำนวณหาคำตอบได้หลายวิธีด้วยกัน เช่น การแทนค่า (Substitution) การขจัดตัวแปร (Systematic Elimination) และ Cramer's Rule

การคำนวณหาค่า α จะคำนวณโดยใช้สูตรหรือไม่ก็ได้ กรณีที่ไม่ใช้สูตรสามารถคำนวณได้จากการย้ายข้างสมการ Normal Equation ดังนี้

จากสมการที่ (2.16)

$$\sum_{i=1}^{n} Y_i = n\hat{\alpha} + \sum_{i=1}^{n} \hat{\beta}_1 X_i$$

$$\sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} \hat{\beta}_1 X_1 = n\hat{\alpha} \qquad \text{...(2.21)}$$

จากสมการที่ (2.21) หารด้วย n ตลอด

$$\frac{n\hat{\alpha}}{n} = \frac{\sum_{i=1}^{n} Y_i}{n} - \frac{\sum_{i=1}^{n} \beta_1 Y_i}{n}$$

$$\hat{\alpha} = \overline{Y} - \hat{\beta}_1 \overline{Y} \qquad \text{...(2.22)}$$

สำหรับการคำนวณ $\hat{\beta}$ สามารถคำนวณได้จากหลายวิธีด้วยกัน เช่น การคำนวณด้วยวิธี Cramer's Rule แสดงได้ดังนี้

จากสมการ Normal Equation จัดให้อยู่ในรูปของแมทริกซ์ ดังนี้

$$\begin{bmatrix} n & \sum_{i=1}^{n} X_1 \\ \sum_{i=1}^{n} X_1 & \sum_{i=1}^{n} X_1^2 \end{bmatrix} \begin{bmatrix} \hat{\alpha} \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} X_1 Y_i \end{bmatrix}$$

$$\hat{\beta}_1 = \frac{\begin{bmatrix} n & \sum_{i=1}^{n} X_1 \\ \sum_{i=1}^{n} X_1 & \sum_{i=1}^{n} X_1 Y_i \end{bmatrix}}{\begin{bmatrix} n & \sum_{i=1}^{n} X_1 \\ \sum_{i=1}^{n} X_1 & \sum_{i=1}^{n} X_1^2 \end{bmatrix}}$$

$$\hat{\beta}_1 = \frac{n\sum_{i=1}^{n} X_1 Y_i - (\sum_{i=1}^{n} X_1)(\sum_{i=1}^{n} Y_i)}{n\sum_{i=1}^{n} X_1^2 - (\sum_{i=1}^{n} X_1)^2} \qquad \text{...(2.23)}$$

จากสมการที่ (2.23) หารสมการทางขวามือด้วย n

$$\hat{\beta}_1 = \frac{\dfrac{n\sum_{i=1}^{n} X_1 Y_i}{n} - \dfrac{(\sum_{i=1}^{n} X_1)(\sum_{i=1}^{n} Y_i)}{n}}{\dfrac{n\sum_{i=1}^{n} X_1^2}{n} - \dfrac{(\sum_{i=1}^{n} X_1)^2}{n}}$$

$$= \frac{\sum_{i=1}^{n} X_1 Y_i - \dfrac{(\sum_{i=1}^{n} X_1)(\sum_{i=1}^{n} Y_i)}{n}}{\sum_{i=1}^{n} X_1^2 - \dfrac{(\sum_{i=1}^{n} X_1)^2}{n}}$$

$$= \frac{\sum_{i=1}^{n} X_1 Y_i - \overline{X} \cdot \overline{Y}}{\sum_{i=1}^{n} X_1^2 - \overline{X}^2}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (X_1 - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n} (X_1 - \overline{X})^2} \qquad \text{...(2.24)}$$

จากสมการที่ (2.24) ถ้ากำหนดให้ $(X_1 - \overline{X}) = x_i$ หรือ x_1 (กรณี Simple Linear Regression) และ $(Y_1 - \overline{Y}) = y_i$ จะได้สมการ $\hat{\beta}_1$ ดังนี้

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

$$\hat{\beta}_1 = \frac{\mathrm{Cov}(X_1 Y)}{\mathrm{Var}(X_1)} \qquad \text{...(2.25)}$$

อย่างไรก็ตาม พบว่าในการประมาณค่าสมการถดถอย หรือคำนวณค่า $\hat{\alpha}$ และ $\hat{\beta}_1$ เพื่อจะได้ทราบขนาดของการเปลี่ยนแปลงและทิศทางของการเปลี่ยนแปลงของ X_i ที่มีผลต่อ Y_i แต่ผลลัพธ์ที่ได้จะไม่สามารถอธิบายหรือทราบคำตอบได้เลยถ้าหากว่า Regression Line นั้น

ขาดการทดสอบคุณสมบัติของตัวประมาณค่า หรือคุณสมบัติของ OLS–Estimator หรือคุณสมบัติความเป็น **Best Linear Unbiased Estimator (BLUE)** ตามทฤษฎีของ **Gauss–Markov**

2.1.3 คุณสมบัติความเป็น Best Linear Unbiased Estimator (BLUE)

1. $\hat{\alpha}$ **และ** $\hat{\beta}_1$ **เป็นตัวประมาณค่าที่ไม่เอนเอียง (Unbiased)**

คุณสมบัติความไม่เอนเอียง (Unbiased) ของ $\hat{\alpha}$ และ $\hat{\beta}_1$ แบ่งพิจารณา 2 กรณีดังนี้

$$-E(\hat{\beta}) = \beta$$

จากสมการที่ (2.25)

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2}$$

โดยกำหนดให้

F_i คือค่าคงที่ใดๆ

$$F_i \text{ คือ } \frac{x_i}{\sum_{i=1}^{n} X_i^2} \qquad ...(2.26)$$

แทนสมการที่ (2.26) ลงในสมการที่ (2.25)

$$\hat{\beta}_1 = \sum_{i=1}^{n} F_i y_i \qquad ...(2.27)$$

$$= \sum_{i=1}^{n} F_i (Y_i - \hat{Y})$$

$$= \sum_{i=1}^{n} F_i Y_i - \sum_{i=1}^{n} F_i \overline{Y}$$

$$= \sum_{i=1}^{n} F_i Y_i - \overline{Y} \sum_{i=1}^{n} F_i$$

จากสมการที่ (2.26)

$$F_i = \frac{x_i}{\sum_{i=1}^{n} x_i^2}$$

$$\sum_{i=1}^{n} F_i = \sum_{i=1}^{n} \left(\frac{x_i}{\sum_{i=1}^{n} x_i^2} \right) \qquad ...(2.28)$$

จากสมการที่ (2.28) สามารถจัดรูปแบบสมการใหม่ได้ ดังนี้

$$\sum_{i=1}^{n} F_i = \frac{1}{\sum_{i=1}^{n} x_i^2} \sum_{i=1}^{n} x_i \qquad ...(2.29)$$

$$\sum_{i=1}^{n} F_i = \frac{1}{\sum_{i=1}^{n} x_i^2} \sum_{i=1}^{n} (X_i - \overline{X})$$

$$\frac{1}{\sum_{i=1}^{n} x_i^2} \sum_{i=1}^{n} (X_i - \overline{X}) = 0$$

ดังนั้น

$$\sum_{i=1}^{n} F_i y_i = \sum_{i=1}^{n} F_i Y_i \qquad ...(2.30)$$

หรือสามารถจัดรูปแบบสมการใหม่ได้ ดังนี้

$$\begin{aligned}
\hat{\beta}_1 &= \sum_{i=1}^{n} F_i Y_i \\
&= \sum_{i=1}^{n} F_i (\alpha + \beta X_i + u_i) \\
&= \alpha \sum_{i=1}^{n} F_i + \beta \sum_{i=1}^{n} F_i X_i + \sum_{i=1}^{n} F_i u_i \\
&= 0 + \beta \sum_{i=1}^{n} F_i X_i + \sum_{i=1}^{n} F_i u_i
\end{aligned}$$

$$\hat{\beta}_1 = 0 + \beta \sum_{i=1}^{n} F_i X_i + \sum_{i=1}^{n} F_i u_i \qquad ...(2.31)$$

จากสมการที่ (2.31) พบว่า

$$\begin{aligned}
\beta \sum_{i=1}^{n} F_i X_i &= \sum_{i=1}^{n} F_i X_i - \overline{X} \sum_{i=1}^{n} F_i \\
&= \sum_{i=1}^{n} F_i (X_i - \overline{X})
\end{aligned}$$

$$\beta \sum_{i=1}^{n} F_i X_i = \sum_{i=1}^{n} F_i x_i \qquad ...(2.32)$$

แทนสมการที่ (2.32) ในสมการที่ (2.31)

$$\hat{\beta}_1 = \beta\sum_{i=1}^{n} F_i x_i + \beta\sum_{i=1}^{n} F_i u_i \qquad ...(2.33)$$

จากสมการที่ (2.33) พบว่า

$$\sum_{i=1}^{n} F_i x_i = \sum_{i=1}^{n}\left(\frac{x_i}{\sum_{i=1}^{n} x_i^2}\right)x_i$$

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

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

$$\sum_{i=1}^{n} F_i x_i = 1 \qquad ...(2.34)$$

แทนสมการที่ (2.34) ในสมการที่ (2.33)

$$\hat{\beta}_1 = \beta + \sum_{i=1}^{n} F_i u_i \qquad ...(2.35)$$

$$E(\hat{\beta}_1) = \beta + \sum_{i=1}^{n} F_i E(u_i)$$

$$= \beta + \sum_{i=1}^{n} F_i (0)$$

$$E(\hat{\beta}_1) = \beta \qquad ...(2.36)$$

$$-E(\hat{\alpha}) = \alpha$$

จากสมการที่ (2.22)

$$\hat{\alpha} = \overline{Y} - \hat{\beta}_1 \overline{X}$$

แทน $\hat{\beta} = \sum_{i=1}^{n} F_i Y_i$ ในสมการที่ (2.22)

$$\hat{\alpha} = \overline{Y} - \sum_{i=1}^{n} F_i Y_i \overline{X}$$

$$= \frac{\sum_{i=1}^{n} Y_i}{n} - \overline{X}\sum_{i=1}^{n} F_i Y_i$$

$$\hat{\alpha} = \sum_{i=1}^{n} Y_i \left(\frac{1}{n} - \overline{X}F_i\right) \qquad \text{...(2.37)}$$

จัดรูปแบบสมการที่ (2.37) ใหม่ได้ดังนี้

$$\hat{\alpha} = \sum_{i=1}^{n} \left(\frac{1}{n} - \overline{X}F_i\right) Y_i \qquad \text{...(2.38)}$$

แทนสมการ $Y_i = \alpha + \beta X_i + u_i$ ในสมการที่ (2.38)

$$\hat{\alpha} = \sum_{i=1}^{n} \left(\frac{1}{n} - \overline{X}F_i\right)(\alpha + \beta X_i + u_i)$$

$$= \sum_{i=1}^{n} \left(\frac{\alpha}{n} - \alpha\overline{X}F_i\right) + \sum_{i=1}^{n} \left(\frac{\beta X_i}{n} - \beta X_i \overline{X}F_i\right) + \sum_{i=1}^{n} \left(\frac{1}{n} - \overline{X}F_i\right)u_i$$

$$= \sum_{i=1}^{n} \left(\frac{\alpha}{n} - \alpha\overline{X}F_i\right) + \sum_{i=1}^{n} \left(\beta\frac{X_i}{n} - \beta X_i \overline{X}F_i\right) + \sum_{i=1}^{n} \left(\frac{1}{n} - \overline{X}F_i\right)u_i$$

$$= \left(\frac{\sum_{i=1}^{n}\alpha}{n} - \alpha\overline{X}\sum_{i=1}^{n} F_i\right)\left(\beta\frac{\sum_{i=1}^{n} X_i}{n} - \beta X_i \overline{X}\sum_{i=1}^{n} F_i\right) + \sum_{i=1}^{n} \left(\frac{1}{n} - \overline{X}F_i\right)u_i$$

$$\hat{\alpha} = \left(\alpha - \alpha\overline{X}\sum_{i=1}^{n} F_i\right) + \left(\beta\overline{X} - \beta\overline{X}\sum_{i=1}^{n} F_i X_i\right) + \sum_{i=1}^{n} \left(\frac{1}{n} - \overline{X}F_i\right)u_i \qquad \text{...(2.39)}$$

กำหนดให้ $\sum_{i=1}^{n} F_i = 0$ และ $\sum_{i=1}^{n} F_i x_i = 1$ แทนในสมการที่ (2.38)

$$\hat{\alpha} = \left(\alpha - \alpha\overline{X}(0)\right) + \left(\beta\overline{X} - \beta\overline{X}(1)\right) + \sum_{i=1}^{n} \left(\frac{1}{n} - \overline{X}F_i\right)u_i$$

$$= \left(\alpha - (0)\right) + \left(\beta\overline{X} - \beta\overline{X}\right) + \sum_{i=1}^{n} \left(\frac{1}{n} - \overline{X}F_i\right)u_i$$

$$= \alpha + \left(\beta\overline{X} - \beta\overline{X}\right) + \sum_{i=1}^{n} \left(\frac{1}{n} - \overline{X}F_i\right)u_i$$

$$\hat{\alpha} = \alpha + \sum_{i=1}^{n} \left(\frac{1}{n} - \overline{X}F_i\right)u_i \qquad \text{...(2.40)}$$

สถิติขั้นสูง

เพื่อการวิจัยเชิงปริมาณทางสังคมศาสตร์

(Advanced Statistics for Quantitative Research in Social Science)

สถิติขั้นสูงเป็นสถิติที่เน้นให้ผู้วิจัยนำไปประยุกต์ใช้สำหรับการวิจัยในแต่ละด้านได้อย่างสมบูรณ์มากยิ่งขึ้น และมีความเหมาะสมกับงานวิจัยที่ต้องการคำตอบ เพื่อนำไปสู่การวิจัยเชิงสร้างสรรค์ในอนาคต ดังนั้นสถิติขั้นสูงจึงเป็นสถิติที่เหมาะสำหรับการนำไปประกอบการกำหนดกลยุทธ์และยุทธศาสตร์ประเทศให้เกิดการพัฒนาอย่างต่อเนื่อง

ประวัติผู้เขียน **รองศาสตราจารย์ ดร. พฤทธ์สรรค์ สุทธิไชยเมธี**

ประวัติการศึกษา

- Bachelor of Economics (Quantitative Analysis)
- Bachelor of Laws
- Bachelor of Science (Occupational Health and Safety)
- Master of Economics (Business Economics)
- Master of Science Program in Environmental Management (M. Sc.)
- Master of Public Administration
- Master of Laws
- Doctor of Philosophy in Economics (Econometrics)
- Doctor of Philosophy (Environmental Science)
- Doctor of Philosophy (Energy Management Technology)
- Doctor of Philosophy (Public Administration)
- Postdoctoral, Institute for Population and Social Research, Mahidol University
- Postdoctoral, Faculty of Economics, Chulalongkorn University

ประสบการณ์ด้านการสอน (อาจารย์ประจำ)

- มหาวิทยาลัยเกษตรศาสตร์ คณะวิทยาการจัดการ
- มหาวิทยาลัยบูรพา International College
- มหาวิทยาลัยเทคโนโลยีราชมงคลตะวันออก วิทยาเขตจักรพงษภูวนารถ คณะบริหารธุรกิจและเทคโนโลยีสารสนเทศ

ตัวอย่างผลงานวิจัยที่ลงวารสารใน SCOPUS

- Pruethsan Sutthichaimethee, Chanintorn Jittawiriyanukoon. Modelling a Causal Factor Relationship Affecting the Management of Sustainable Development Goal in Thailand. Technology Reports of Kansai University. 2020, 62 (5).
- Pruethsan Sutthichaimethee, et. al. A Short-Term Forecasting Model to Assess the Efficiency of Thai Government Management in Energy Consumption under Environmental Law. Technology Reports of Kansai University. 2020, 62 (4). 1371-1381.
- Pruethsan Sutthichaimethee, et. al. The Efficiency of Government Administration in the Enforcement of Thai Environmental Law for Thailand 4.0: Whether Energy Consumption Regulations Impacting Greenhouse Gas Technology Reports of Kansai University. 2020, 62 (4).1-20.
- Pruethsan Sutthichaimethee, Sthianrapab Naluang. The Efficiency of the Sustainable Development Policy for Energy Consumption under Environmental Law in Thailand: Adapting the SEM-VARIMAX Model. Energies. 2019, 12 (16). 3092.
- Pruethsan Sutthichaimethee, Kuskana Kubaha. The Efficiency of Long-Term Forecasting Model on Final Energy Consumption in Thailand's Petroleum Industries Sector: Enriching the LT-ARIMAXS Model under a Sustainability Policy. Energies. 2018, 11 (8). 2063.
- ฯลฯ

www.se-ed.com
sbc.fans

ISBN 978-616-08-4018-2
9 786160 840182
419 บาท

คู่มือเรียน-สอบ/อุดมศึกษา-
รัฐประศาสนศาสตร์ รัฐศาสตร์ นิติศาสตร์
บริหารธุรกิจ และเศรษฐศาสตร์