บทที่ 1

บทนำ

<u>ภูมิหลัง</u>

บทความหรือเอกสารทางวิชาการต่าง ๆ ที่เกี่ยวกับสถิติ มักประกอบด้วยข้อมูลที่เกิดจากความสัมพันธ์ระหว่างปรากฏการณ์สองอย่างหรือมากกว่า และมักจัดทำไว้ในรูปของตารางแสดงความสัมพันธ์เป็นคู่ ๆ หรือเป็นชุด ๆ เมื่อกำหนดตำแหน่งของข้อมูลเหล่านี้ลงบนกระดาษตาราง จะได้จุดแสดงตำแหน่งของข้อมูลกระจัดกระจายอยู่ในตำแหน่งต่าง ๆ ถ้ามองโดยภาพรวมแล้ว อาจเป็นรูปวงรี รูปไต รูปคดโค้งในลักษณะต่าง ๆ ข้อมูลเหล่านี้อาจได้มาจากการสังเกต การทดลอง การวัดโดยตรง การวัดทางอ้อม หรือจากวิธีอื่น ๆ ออกมาเป็นตัวเลข และเนื่องจากเป็นการเก็บข้อมูลในเวลา สถานที่ หรือสิ่งแวดล้อมที่แตกต่างกัน จึงทำให้การเรียงตัวของจุดต่าง ๆ ไม่เป็นเส้นตรงหรือเส้นโค้ง ที่สามารถแทนที่ด้วยสมการทางคณิตศาสตร์ที่แน่นอนได้

ในกรณีที่ภาพรวมของกลุ่มข้อมูลน่าจะแทนที่ได้ด้วยสมการเส้นตรง ข้อมูลจะจับกลุ่มกันเป็นรูปวงรี ถ้าเป็นวงรีที่แบนมาก (รีมาก) แสดงว่า ปรากฏการณ์ 2 อย่างนั้น มีสหสัมพันธ์ต่อกันสูง แต่ถ้าเป็นวงรีที่แบนน้อย (เกือบกลม) ก็แสดงว่า ปรากฏการณ์ทั้ง 2 อย่างนั้นมีสหสัมพันธ์ต่อกันต่ำ (Allen and Yen. 2531 : 39) และเรียกเส้นตรง ที่เป็นตัวแทนของกลุ่มข้อมูลว่า "เส้นถดถอย" (Regression Line)

การสร้างสมการถดถอยซึ่งใช้แทนกลุ่มข้อมูล มีสมมุติฐานมาจากนิยามตำแหน่งของเส้นตรง ซึ่งประกอบด้วย

1. จุดที่เส้นตรงจะต้องผ่าน
2. ความชัน (Slope) ของเส้นตรง

จุดที่เส้นตรงจะต้องผ่าน เป็นจุดใดจุดหนึ่งซึ่งอยู่ในย่านกลางของกลุ่มข้อมูล ซึ่งเกาส์ (Gauss. ค.ศ.1777-1855) นำการหาค่าเฉลี่ย (Means) แต่ละแกนของกลุ่มข้อมูล มาใช้

กำหนดจุดนี้ขึ้น เรียกว่า "จุดกึ่งกลาง $(\bar{X},\bar{Y})$ ของกลุ่มข้อมูล" และใช้วิธี "กำลังสองน้อยที่สุด (Least Square Method) คำนวนหาค่าความชันของเส้นถดถอย ซึ่งสมการถดถอยแบบกำลังสองน้อยที่สุดของเกาส์ มีรูปดังนี้

$$\hat{Y} = \frac{\sigma_{xy}}{\sigma_x^2}(X - \bar{X}) + \bar{Y}$$

โดยที่ $\hat{Y}$ หมายถึง ค่า Y ที่เกิดจากการทำนาย

σ_{xy} หมายถึง ค่าความแปรปรวนร่วมระหว่างแกนทั้งสองของข้อมูล

σ_x^2 หมายถึง ค่าความแปรปรวนของข้อมูลตามแนวแกนนอนของกราฟ

X หมายถึง ค่าตัวแปร ตามแนวแกนนอนของกราฟ (ตัวแปรอิสระ)

$\bar{X}$ หมายถึง ค่าเฉลี่ย ของข้อมูลทั้งหมด ตามแนวแกนนอนของกราฟ

$\bar{Y}$ หมายถึง ค่าเฉลี่ย ของข้อมูลทั้งหมด ตามแนวแกนตั้งของกราฟ

เพื่อให้เกิดความเข้าใจตรงกันเกี่ยวกับสัญลักษณ์ต่าง ๆ ที่กล่าวมาแล้วข้างต้น ผู้วิจัยขอนำเสนอการคำนวนหาค่าความแปรปรวนต่าง ๆ ดังตาราง 1 ต่อไปนี้

ตาราง 1 การคำนวนหาค่าความแปรปรวนและความแปรปรวนร่วม

i	$X - \bar{X}$	$Y - \bar{Y}$	$(X - \bar{X})^2$	$(Y - \bar{Y})^2$	$(X - \bar{X})(Y - \bar{Y})$
1	x_1	y_1	x_1^2	y_1^2	x_1y_1
2	x_2	y_2	x_2^2	y_2^2	x_2y_2
⋮	⋮	⋮	⋮	⋮	⋮
n	x_n	y_n	x_n^2	y_n^2	x_ny_n
ผลรวม			Σx_i^2	Σy_i^2	Σx_iy_i
ค่าเฉลี่ยของผลรวม			$\frac{\Sigma x_i^2}{N}$	$\frac{\Sigma y_i^2}{N}$	$\frac{\Sigma x_iy_i}{N}$
ค่าความแปรปรวน			σ_x^2	σ_y^2	σ_{xy}

โดยที่ Σ หมายถึงผลรวมในลักษณะต่าง ๆ ของข้อมูล กล่าวคือ ผลรวมจากการยกกำลังสองของค่าที่วัดได้แต่ละแกน และผลรวมของผลคูณระหว่างค่าที่วัดได้แต่ละแกน

จากวิธีกำลังสองน้อยที่สุด จะสังเกตได้ว่า การทำนาย Y จาก X ให้ผลแตกต่างไปจากการทำนาย X จาก Y ดังจะแสดงให้เห็นโดยชัดเจนจากข้อมูลสมมุติต่อไปนี้

ตาราง 2 การคำนวนหาค่าความชันของเส้นถดถอยโดยวิธีกำลังสองน้อยที่สุด

i	$X - \bar{X}$	$Y - \bar{Y}$	$(X - \bar{X})^2$	$(Y - \bar{Y})^2$	$(X - \bar{X})(Y - \bar{Y})$
1	7.745966692	0.219305075	60	0.04809472	1.698729809
2	0	11.18033989	0	125	0
3	-7.745966692	-11.39964496	60	129.9519053	88.30127018
Σ	0	0	120	255	90
σ^2	0	0	40	85	30

จากตาราง 2 ความชันของเส้นทำนาย Y จาก X $= \frac{\sigma_{xy}}{\sigma_x^2}$ วัดจากแกนนอน

$= \frac{30}{40}$ วัดจากแกนนอน

$= 0.75$ วัดจากแกนนอน

ความชันของเส้นทำนาย X จาก Y $= \frac{\sigma_{xy}}{\sigma_y^2}$ วัดจากแกนตั้ง

ความชันของเส้นทำนาย X จาก Y $= \frac{\sigma_y^2}{\sigma_{xy}}$ วัดจากแกนนอน

$= \frac{85}{30}$ วัดจากแกนนอน

$= 2.8\dot{3}$ วัดจากแกนนอน

แสดงว่า วิธีกำลังสองน้อยที่สุด ให้ผลการทำนาย X จาก Y แตกต่างไปจากการทำนาย Y จาก X ทั้งนี้เนื่องจากวิธีนี้ใช้ทำนายตัวแปรตาม ซึ่งเกิดขึ้นภายหลัง จากตัวแปรอิสระ ซึ่งเกิดขึ้นมาก่อนแล้ว

อย่างไรก็ตาม มีปรากฏการณ์ต่าง ๆ ทางกายภาพจำนวนมากที่ไม่อาจระบุได้ชัดเจนว่า ตัวแปรใดเป็นตัวแปรอิสระ และตัวแปรใดเป็นตัวแปรตาม เพราะเป็นปรากฏการณ์ที่ทำให้เกิดตัวแปรขึ้นทั้งสองตัวพร้อม ๆ กัน เช่น น้ำหนักกับส่วนสูง ความลึกกับแรงกดของของเหลว ความเร็วกับความหนาแน่นของอากาศ วิธีกำลังสองน้อยที่สุด จึงไม่น่าจะใช้ได้กับกรณีเช่นนี้ เพราะถ้านำค่า Y ไปคำนวนย้อนกลับเพื่อหาค่า X แล้ว จะเกิดความคลาดเคลื่อนสูงดังที่ได้อธิบายไปแล้วนั่นเอง ดังนั้นแรงจูงใจที่สำคัญในการทำวิจัยครั้งนี้ก็คือ วิธีกำลังสองน้อยที่สุดไม่น่าจะเหมาะสมในกรณีที่ไม่สนใจว่าตัวแปรใดเป็นตัวแปรอิสระและตัวแปรใดเป็นตัวแปรตาม

ความจำเป็นที่จะต้องศึกษาปัญหาข้อนี้

สาเหตุที่จำเป็นต้องศึกษาปัญหาข้อนี้ ก็เพื่อนำผลที่ได้จากการวิจัย ไปใช้วิเคราะห์การถดถอยกับข้อมูลประเภทที่ไม่อาจระบุได้ชัดเจนว่า ปรากฏการณ์ใดเป็นตัวแปรอิสระ และปรากฏการณ์ใดเป็นตัวแปรตาม ทั้งนี้ เพื่อความก้าวหน้าของวิทยาการ ในแขนงวิชาสถิติ ในโอกาสต่อไป

ปัญหาการวิจัย

ปัญหาการวิจัยครั้งนี้ก็คือกรณีกลุ่มข้อมูลที่เก็บรวบรวมมาได้จากปรากฏการณ์สองอย่างที่เกิดขึ้นพร้อม ๆ กัน โดยไม่อาจระบุได้ชัดเจนว่า ปรากฏการณ์ใดเป็นตัวแปรอิสระ และปรากฏการณ์ใดเป็นตัวแปรตามแล้ว ค่าสถิติต่าง ๆ ต่อไปนี้ ซึ่งได้แก่

1. ค่าความชันของเส้นถดถอย
2. ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างปรากฏการณ์ทั้งสอง
3. ค่าความคลาดเคลื่อนมาตรฐานของเส้นถดถอย

จะมีลักษณะอย่างไรบ้าง

วัตถุประสงค์ของการวิจัย

การวิจัยนี้มีวัตถุประสงค์ที่จะสร้างสูตรสำหรับใช้คำนวนหาค่าความชันของเส้นถดถอย ค่าสัมประสิทธิ์สหสัมพันธ์ และค่าความคลาดเคลื่อนมาตรฐานของเส้นถดถอย เพื่อใช้คำนวนกับข้อมูลประเภทที่ไม่อาจระบุได้ชัดเจนว่า ปรากฏการณ์ใดเป็นตัวแปรอิสระ และปรากฏการณ์ใดเป็นตัวแปรตาม

สมมุติฐานการวิจัย

1. เส้นถดถอย ที่คลาดเคลื่อนไปจากกลุ่มข้อมูลน้อยที่สุด มีเพียงเส้นเดียว
2. เส้นถดถอย ที่คลาดเคลื่อนไปจากกลุ่มข้อมูลน้อยที่สุด เป็นเส้นตรงเดียวกันกับแกนยาวเฉลี่ย ของกลุ่มข้อมูล
3. การทำนาย X จาก Y ใช้สมการเดียวกันกับการทำนาย Y จาก X

ข้อตกลงเบื้องต้น

เส้นถดถอยจะต้องผ่านจุดกึ่งกลาง $(\overline{X},\overline{Y})$ ของกลุ่มข้อมูล (เช่นเดียวกันกับวิธีกำลังสองน้อยที่สุดของเกาส์)

ขอบเขตของปัญหาและความจำกัดของปัญหา

การวิจัยนี้ จะทำการศึกษาพิจารณาเฉพาะกรณีที่ ข้อมูลจับกลุ่มกันเป็นรูปวงรี หรือในกรณีที่ กลุ่มข้อมูลมีภาพรวมที่น่าจะแทนที่ได้ด้วยสมการเส้นตรงเท่านั้น

นิยามศัพท์

1. ความคลาดเคลื่อนมาตรฐาน หมายถึง ค่าส่วนเบี่ยงเบนมาตรฐานของค่าความแตกต่าง ระหว่างตัวแปรเกณฑ์จากการทำนาย กับค่าของตัวแปรเกณฑ์จริง (เกษม สาหร่ายทิพย์. 2531 : 327)

2. ความแปรปรวน หมายถึง สัดส่วนของผลรวมของกำลังสอง ของความแตกต่างระหว่างข้อมูลแต่ละค่ากับค่าเฉลี่ย ต่อค่าระดับชั้นความเสรี หรือผลรวมของกำลังสอง ของระยะทางระหว่างข้อมูลแต่ละค่ากับค่าเฉลี่ย หารด้วยค่าระดับชั้นความเสรี (เกษม สาหร่ายทิพย์. 2531 : 113)

3. ความแปรปรวนร่วม หมายถึง สัดส่วนของผลรวมของผลคูณ ของความแตกต่างระหว่างข้อมูลแต่ละค่าของข้อมูลชุด X กับค่าเฉลี่ยของข้อมูลชุด X กับความแตกต่างระหว่างข้อมูลแต่ละค่าของข้อมูลชุด Y กับค่าเฉลี่ยของข้อมูลชุด Y ต่อค่าระดับชั้นความเสรี (เกษม สาหร่ายทิพย์. 2531 : 117)

4. ค่าเฉลี่ย หมายถึง สัดส่วนระหว่างผลรวมของข้อมูลทั้งหมด ต่อจำนวนข้อมูลนั้น กล่าวคือ หาผลรวมของข้อมูลทั้งหมดแล้วหารด้วยจำนวนข้อมูลนั้น ก็จะได้ค่าเฉลี่ยตามต้องการ (เกษม สาหร่ายทิพย์. 2531 : 90)

5. ระดับชั้นความเสรี หมายถึง จำนวนวิธีอิสระของระบบใด ๆ ซึ่งโครงสร้างของปริภูมิ (Space) อาจเปลี่ยนแปลงได้โดยไม่เปลี่ยนแปลงนิยาม ข้อจำกัด และ มิติของระบบเดิม หรือหมายถึง จำนวนค่าของตัวแปรหรือข้อมูลที่เป็นอิสระ คือไม่ขึ้นอยู่กับค่าใด ๆ โดยข้อมูลเหล่านั้น ต้องไม่ทำให้ ค่าสถิติที่มีอยู่แล้ว เปลี่ยนแปลงไปจากเดิม (ระวิพันธ์ โสมนะพันธ์. 2518 : 113)

6. ส่วนเบี่ยงเบนมาตรฐาน หมายถึง ค่ารากที่สองของค่าความแปรปรวน (เกษม สาหร่ายทิพย์. 2531 : 114)

7. สหสัมพันธ์ หมายถึง ความเกี่ยวข้องกันของตัวแปรสองตัว การวัดความเกี่ยวข้องดังกล่าว จะได้ตัวเลขแสดงปริมาณความสัมพันธ์ระหว่างตัวแปรทั้งสองนั้น เรียกตัวเลขนี้ว่า สัมประสิทธิ์สหสัมพันธ์ (เกษม สาหร่ายทิพย์. 2531 : 309)