ภาคผนวก ก

Regression Model

ก. 1 แบบจำลองเชิงถดถอย Regression Model

เนื่องจากในการวิเคราะห์ปัญหาในด้านวิศวกรรมแหล่งน้ำพบว่าปัญหาส่วนใหญ่ที่ต้องแก้ไขนั้นประกอบด้วยตัวแปรหลายตัวและเพื่อเป็นการแสดงความสัมพันธ์ระหว่างตัวแปรดังกล่าวเหล่านั้นจึงนิยมนำสมการทางคณิตศาสตร์เข้ามาช่วยในการอธิบายความสัมพันธ์ ซึ่งเรียกสมการทางคณิตศาสตร์ดังกล่าวว่าแบบจำลองเชิงคณิตศาสตร์ (Mathematical model) โดยแบบจำลองเชิงคณิตศาสตร์ที่แสดงความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตามที่แสดงออกมาในรูปเชิงคณิตศาสตร์ เรียกว่า แบบจำลองเชิงถดถอย (Regression model) ในกรณีที่มีตัวแปรอิสระเพียงตัวเดียวและตัวแปรตามแปรผันตามตัวแปรอิสระเพียงตัวเดียว เรียกว่า การถดถอยเชิงเดียว (Simple regression) และในกรณีที่มีตัวแปรอิสระหลายตัวและตัวแปรตามแปรผันตามตัวแปรอิสระหลายตัว เรียกว่า การถดถอยหลายชั้น (Multiple regression)

การวิเคราะห์การถดถอย เป็นเครื่องมือทางสถิติอย่างหนึ่งที่ศึกษาความแปรผันของตัวแปรตาม เมื่อเปลี่ยนแปลงค่าของตัวแปรอิสระหนึ่งตัวหรือหลายตัว โดยพยายามที่จะหาสมการซึ่งพอเพียงในการกำหนดความสัมพันธ์ระหว่างตัวแปรเหล่านั้น สมการที่ต้องการควรจะเป็นสมการที่ค่อนข้างง่าย แต่ในขณะเดียวกันก็สามารถประมาณความสัมพันธ์ระหว่างตัวแปรเหล่านั้นได้ใกล้เคียงมากที่สุดด้วย พร้อมทั้งต้องสามารถวัดค่าความคลาดเคลื่อนที่เกี่ยวข้องในกระบวนการของการประมาณค่าออกมาได้ด้วย ดังนั้นในการศึกษาในครั้งนี้จึงประยุกต์ใช้แบบจำลองเชิงถดถอย (Regression model) เข้ามาช่วยในการหาค่าความสัมพันธ์ของค่ากำลังการสะท้อนกลับของเรดาร์ (Radar Reflectivity Factor , Z ในหน่วย มม.^{6} ม.$^{-3}$) ที่ใช้วัดปริมาณน้ำฝนจากเรดาร์ (ฝนบนฟ้า) กับอัตราความเข้มของปริมาณน้ำฝน (Rain Rate , R ในหน่วย มม./ชม.) ที่วัดจากสถานีวัดน้ำฝนบนพื้นดิน (ฝนบนดิน) ซึ่งความสัมพันธ์จะอยู่ในรูป Z-R Relationship : $Z=aR^b$ และจากสมการความสัมพันธ์ Z-R ดังกล่าวสามารถใช้แบบจำลองเชิงถดถอยวิเคราะห์หาค่าพารามิเตอร์ a และ b โดยแบบจำลองดังกล่าวมีรายละเอียดของทฤษฎีสมมติฐานการประมาณค่าพารามิเตอร์และขั้นตอนในการดำเนินการเพื่อสร้างแบบจำลองเชิงถดถอยที่เหมาะสมดังต่อไปนี้

แบบจำลองเชิงถดถอยสามารถแสดงได้ดังนี้

$$y_i = \beta_0 x_i^{\beta_1} e_i \quad \text{i= 1,2,3,.....,n} \qquad \text{(ก-1)}$$

โดยที่ y คือ ตัวแปรตาม (Dependent variable) เนื่องจากค่า y ขึ้นอยู่กับค่า x

x คือ ตัวแปรอิสระ (Independent Variable)

e_i คือ ค่าความคลาดเคลื่อน

β_0, β_1 คือ ค่าสัมประสิทธิ์ความถดถอย (Regression Coefficient)

1.1 สมมติฐานของการวิเคราะห์ความถดถอย

ในการวิเคราะห์ความถดถอยให้มีความถูกต้องนั้นต้องเป็นไปตามสมมติฐานดังต่อไปนี้ กัลยา (2540)

1) ค่าความแปรปรวนของค่าความคลาดเคลื่อน $v(e_i)$ มีค่าเท่ากันทุกค่าของ i และมีค่าเท่ากับความแปรปรวนของ y

$$v(e_i) = v(y) = \delta^2_{yx} = \delta^2 \qquad \text{(ก-2)}$$

2) ค่าความคลาดเคลื่อน e_i และ e_j เป็นอิสระกัน นั้น คือ Cov (e_i, e_j) = E(e_i, e_j) = 0; i≠j

3) ค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเป็นศูนย์และค่าความแปรปรวนเท่ากับ δ^2

1.2 การประมาณค่าเริ่มต้นของพารามิเตอร์ของสมการเชิงถดถอย

การประมาณค่าเริ่มต้นของพารามิเตอร์ทุกตัว (β) มีวิธีที่นิยมใช้ 3 วิธี ดังนี้ กัลยา (2546)

1) ใช้เทคนิคการวิเคราะห์ถดถอยเชิงเส้นในการหาค่าเริ่มต้นของค่า β_0, β_1 โดยไม่พิจารณาค่าความคลาดเคลื่อน จากนั้นใช้ค่า β_0, β_1 ดังกล่าวที่หาค่าได้ เป็นค่าเริ่มต้นของ β_0, β_1 ในการวิเคราะห์ไม่เชิงเส้น

2) ใช้การแทนค่าข้อมูลจริงในสมการไม่เชิงเส้น

3) ใช้วิธีการแก้สมการโดยมีจำนวนสมการเท่ากับจำนวนพารามิเตอร์ จากนั้นแทนค่า xและ y จำนวน 2 คู่ และทำการแก้สมการหาค่าพารามิเตอร์ แล้วใช้ค่าพารามิเตอร์ดังกล่าวที่หาค่าได้เป็นค่าเริ่มต้นของพารามิเตอร์ในการวิเคราะห์ไม่เชิงเส้น

1.3 การสร้างแบบจำลองเชิงถดถอยที่เหมาะสม

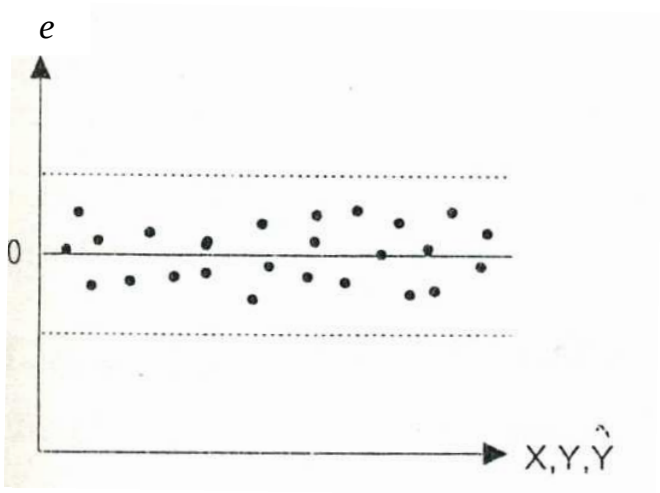
การสร้างแบบจำลองเชิงถดถอยที่เหมาะสมมีขั้นตอนและขบวนการในการดำเนินการดังต่อไปนี้กัลยา (2540)

1) พิจารณานำข้อมูลของตัวแปรอิสระ x และตัวแปรตาม y มาเขียนแผนภาพการกระจาย
2) พิจารณาจากแผนภาพการกระจายในข้อ 1) ว่า x และ y มีความสัมพันธ์กันในรูปแบบสมการที่แสดงความสัมพันธ์คือ

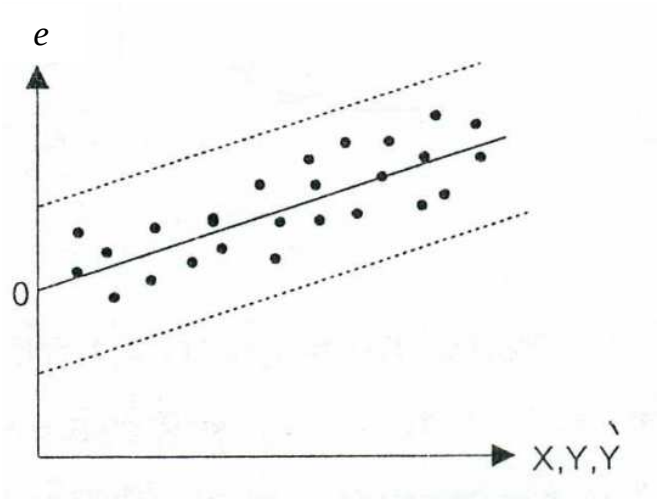
$$y_i = \beta_0 x_i^{\beta_1} e_i \qquad \text{(ก-3)}$$

3) ประมาณค่า β_0 และ β_1 เริ่มต้นโดยใช้วิธีที่กล่าวมาในหัวข้อที่ 3
4) นำค่าประมาณ β_0 และ β_1เริ่มต้นที่หาค่าได้ในข้อ 3) มาใช้เป็นค่าเริ่มต้นในการวิเคราะห์หาค่า β_0 และ β_1 ในรอบที่ 2 และทำต่อไปจนถึงรอบที่ n+1 จนกว่าจะได้ค่าผลต่าง ระหว่างผลรวมของค่าความคลาดเคลื่อนกำลังสองเฉลี่ยในรอบที่ (n+1)- n ลู่เข้าสู่ค่าคงที่ที่น้อยที่สุด ผลลัพธ์จึงจะได้ค่าประมาณ β_0 และ β_1
5) คำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ตัวอย่าง (r)
6) ตรวจสอบค่าความแปรปรวนของความคลาดเคลื่อน

ในขั้นตอนนี้เป็นการตรวจสอบว่า v(e_i) = δ^2 = ค่าคงที่ทุกค่าของ x หรือไม่ ถ้าคงที่จะเรียกว่า Homoscedasticity และถ้าไม่คงที่เรียกว่า Heteroscedasticity ดังนั้นการตรวจสอบค่าคงที่ของค่าความแปรปรวนจึงพิจารณาจากการพล๊อตกราฟระหว่างค่า x กับ e หรือ กราฟระหว่างค่า y กับ e หรือ กราฟระหว่าง ค่า $\hat{y}$ กับ e ดังแสดงในรูปที่ ก. 1-1 ถึง ก. 1- 5

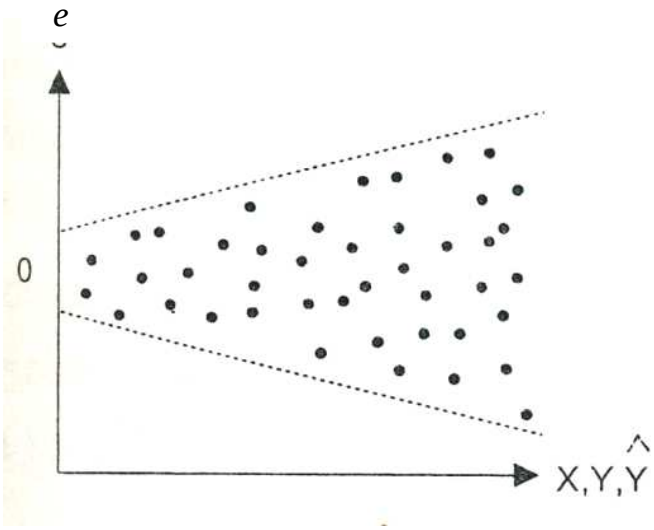


ภาพที่ ก. 1-1 แสดงกราฟระหว่างค่า $x,y,\hat{y}$ กับค่า e

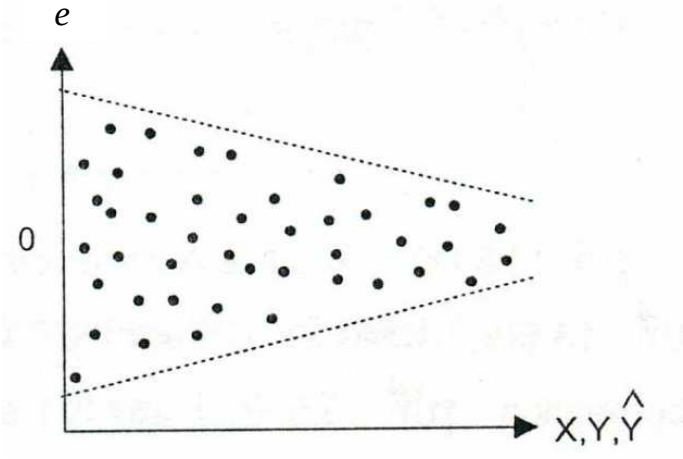


ภาพที่ ก.1-2 แสดงกราฟระหว่างค่า $x,y,\hat{y}$ กับค่า e

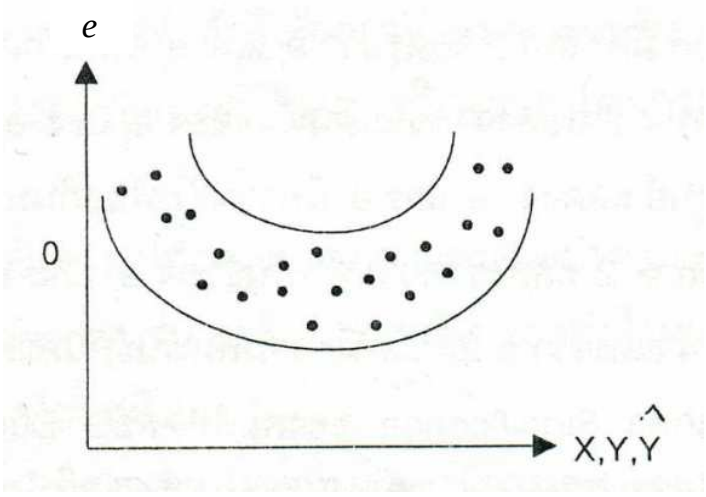
ถ้าพล๊อตกราฟระหว่างค่า x กับ e หรือ กราฟระหว่างค่า y กับ e หรือ กราฟระหว่าง ค่า $\hat{y}$ กับ e ได้ดังแสดงในรูปที่ ก. 1-1 และ ก. 1-2 แสดงว่าค่า v(e_i) มีค่าคงที่



ภาพที่ ก. 1-3 แสดงกราฟระหว่างค่า $x,y,\hat{y}$ กับค่า e



ภาพที่ ก. 1-4 แสดงกราฟระหว่างค่า $x,y,\hat{y}$ กับค่า e

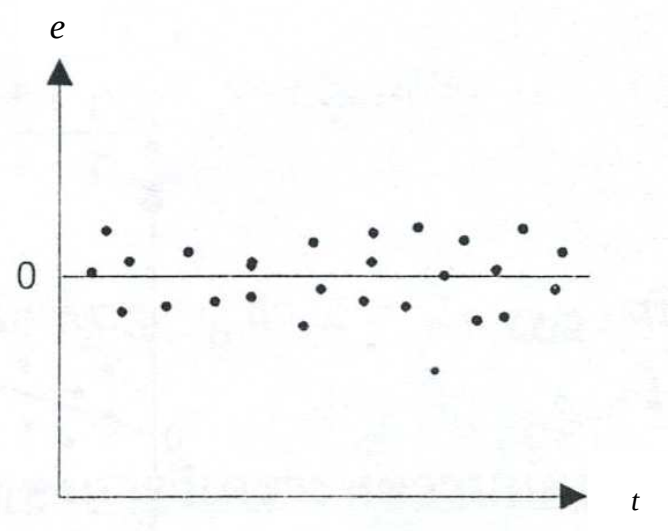


ภาพที่ ก. 1-5 แสดงกราฟระหว่างค่า $x,y,\hat{y}$ กับค่า e

ถ้าพล๊อตกราฟระหว่างค่า x กับ e หรือ กราฟระหว่างค่า y กับ e หรือ กราฟระหว่าง ค่า $\hat{y}$ กับ e ได้ดังแสดงในรูปที่ ก. 1-3 ถึง ก. 1-5 แสดงว่าค่า v(e_i) มีค่าไม่คงที่

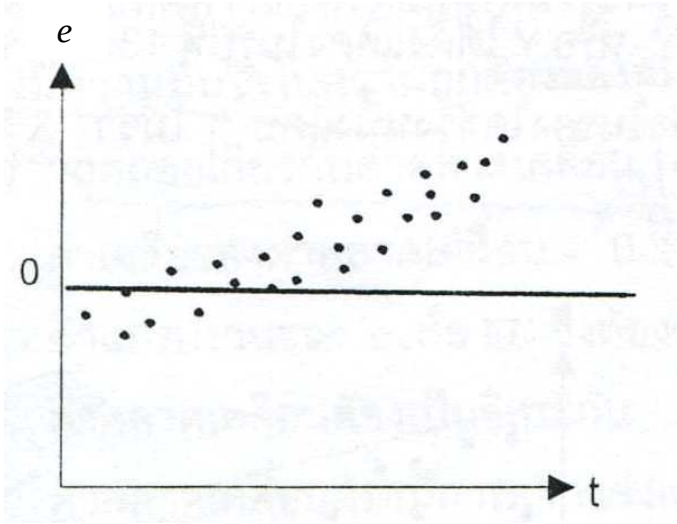
7) ตรวจสอบความเป็นอิสระระหว่างความคลาดเคลื่อน

ในขั้นตอนนี้เป็นการตรวจสอบความเป็นอิสระกันระหว่างความคลาดเคลื่อน e_i และ e_j สามารถทำได้โดยการพล๊อตกราฟ e_i และ e_j กับเวลา t ดังแสดงในภาพที่ ก. 1-6 ถึง ก. 1- 8

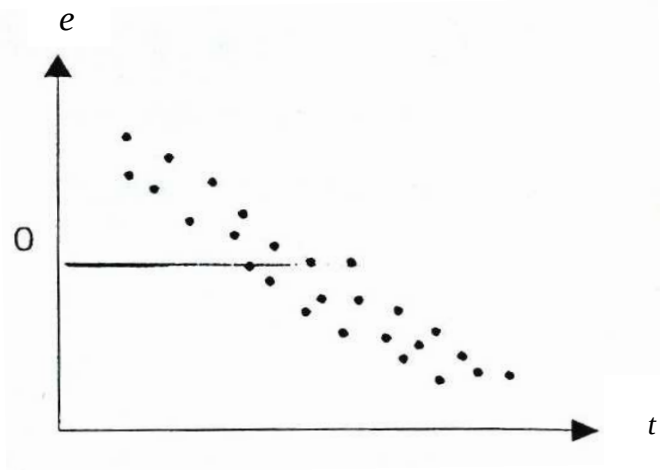


ภาพที่ ก. 1-6 แสดงกราฟระหว่าง e_i และ e_j กับเวลา t

ถ้าพล๊อตความสัมพันธ์ระหว่างค่า e_i และ e_j กับเวลา t ดังแสดงในรูปที่ ก. 1-6 แสดงว่า e_i และ e_j เป็นอิสระต่อกัน



ภาพที่ ก. 1-7 แสดงกราฟระหว่าง e_i และ e_j กับเวลา t



ภาพที่ ก. 1-8 แสดงกราฟระหว่าง e_i และ e_j กับเวลา t

ถ้าพล๊อตความสัมพันธ์ระหว่างค่า e_i และ e_j กับเวลา t ดังแสดงในภาพที่ ก. 1-7 และ ก. 1- 8 แสดงว่า e_i และ e_j ไม่เป็นอิสระต่อกัน