

T

# คืออะไร? จากการวิเคราะห์เศษตกค้างในการวิเคราะห์การถดถอย

## What Do We Obtain from Residual Analysis in Regression?

- ผู้ช่วยศาสตราจารย์ ยูพิน กาญจนะศักดิ์ดา
- สาขาวิชาสถิติประยุกต์
- คณะวิทยาศาสตร์และเทคโนโลยี
- มหาวิทยาลัยหอการค้าไทย
- 
- **Assistant Professor Yupin Kanjanasakda**
- Department of Applied Statistics
- School of Science and Technology
- University of the Thai Chamber of Commerce
- E-mail: yupin\_\_kan@utcc.ac.th

### บทคัดย่อ

การวิเคราะห์เศษตกค้างจากการวิเคราะห์การถดถอย เป็นการพิจารณากราฟจากการพล็อตเศษตกค้างเทียบกับค่าทำนาย ตัวแปรอิสระและเวลาหรือลำดับของข้อมูล นอกจากนี้ยังรวมถึงการพิจารณากราฟจาก Normal Probability Plot ด้วย ถ้าข้อมูลเป็นไปตามข้อสมมุติเบื้องต้นของการวิเคราะห์การถดถอย กราฟของเศษตกค้างที่พล็อตเทียบกับค่าทำนาย ตัวแปรอิสระ และเวลาหรือลำดับของข้อมูล จะมีการกระจายอย่างไม่มีรูปแบบรอบ ๆ ค่าเฉลี่ยคือ 0 และ กราฟจาก Normal Probability Plot จะมีการกระจายในแนวเส้นตรงแนวทแยง

**คำสำคัญ:** การวิเคราะห์การถดถอย การวิเคราะห์เศษตกค้าง

### Abstract

Residual analysis is used to check the assumptions of regression analysis by examining the graphs, which are plotted between residuals versus fitted values, independent

variables and times. It also includes a graph of the Normal Probability Plot. If no assumptions are violated, then the graphs, which are plotted between residuals versus fitted values, independent variables and times, should be randomly distributed around their mean of zero and the graph of Normal Probability Plot should be along a diagonal line.

**Keywords:** Regression Analysis, Residual Analysis

## บทนำ

การวิเคราะห์การถดถอยเป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรหนึ่งกับตัวแปรอีกกลุ่มหนึ่ง โดยการสร้างสมการทางคณิตศาสตร์เพื่อใช้ในการพยากรณ์ตัวแปรตามด้วยอีกกลุ่มตัวแปรหนึ่ง ซึ่งเรียกว่ากลุ่มตัวแปรอิสระ การวิเคราะห์การถดถอยแบบพื้นฐานที่สุด คือ การวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression) หรือเรียกสั้น ๆ ว่า การวิเคราะห์การถดถอยอย่างง่าย (Simple Regression) ซึ่งเป็นการสร้างสมการแสดงความสัมพันธ์เพื่อพยากรณ์ตัวแปรตามด้วยตัวแปรอิสระเพียง 1 ตัว

ตัวแบบการวิเคราะห์การถดถอยอย่างง่าย คือ

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad \dots\dots\dots(1)$$

โดยที่  $y_i$  เป็นค่าตัวแปรตามที่เกิดขึ้นครั้งที่  $i$   
 $x_i$  เป็นค่าตัวแปรอิสระที่เกิดขึ้นครั้งที่  $i$   
 $\beta_0$  เป็นพารามิเตอร์ที่เป็นจุดตัดแกน  $y$   
 $\beta_1$  เป็นพารามิเตอร์ที่เป็นความชันของเส้นถดถอย  
 $\varepsilon_i$  เป็นค่าความคลาดเคลื่อน (Error) ที่สังเกตครั้งที่  $i$  ซึ่ง  $\varepsilon_i$  เป็นตัวแปรสุ่มที่มีการแจกแจงแบบปกติโดยมีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ  $\sigma^2 (\varepsilon_i \sim N(0, \sigma^2))$  และ  $\varepsilon_i, \varepsilon_j$  เป็นอิสระต่อกัน

ข้อสมมุติเบื้องต้นของการวิเคราะห์การถดถอย

1. ตัวแบบการวิเคราะห์การถดถอยที่ใช้เหมาะสม คือ เป็นตัวแบบเชิงเส้นในตัวแปรอิสระ
2. ความคลาดเคลื่อนมีค่าเฉลี่ยเป็น 0
3. ความคลาดเคลื่อนมีความแปรปรวนคงที่
4. ความคลาดเคลื่อนมีการแจกแจงปกติ
5. ความคลาดเคลื่อนเป็นอิสระต่อกัน

จากตัวแบบ (1)  $x_i$  เป็นค่าคงที่ที่ต้องกำหนดล่วงหน้า ส่วน  $y_i$  เป็นตัวแปรสุ่มเช่นเดียวกับ  $\varepsilon_i$  ค่าเฉลี่ยของ  $y_i$  ที่  $x_i$  คือ

$$E(y_i | x_i) = \mu_{y|x} = \beta_0 + \beta_1 x_i \quad \dots\dots\dots(2)$$

และความแปรปรวนของ  $y_i$  ที่  $x_i$  ใดๆ จะคงที่คือ  $\sigma^2$  ซึ่งเท่ากับความแปรปรวนของ  $\varepsilon_i$  ค่าเฉลี่ยของ  $y_i$  ในแต่ละค่าของ  $x_i$  ตาม (2) อยู่ในรูปของสมการเส้นตรงซึ่งเรียกว่าสมการถดถอยของประชากรหรือเส้นถดถอยของประชากร สามารถประมาณด้วยสมการถดถอยตัวอย่างหรือเส้นถดถอยตัวอย่างจากข้อมูลชุดหนึ่งที่มีค่าสังเกต  $n$  ค่า คือ

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \quad \dots\dots\dots(3)$$

โดยที่  $\hat{y}_i$  เป็นค่าประมาณของค่าเฉลี่ยของ  $y_i$  ที่  $x_i$  หรือค่าทำนาย  $y_i$  ที่  $x_i$   
 $x_i$  เป็นค่าตัวแปรอิสระที่สังเกตครั้งที่  $i$   
 $\hat{\beta}_0$  และ  $\hat{\beta}_1$  เป็นตัวประมาณที่ได้จากวิธี

กำลังสองน้อยที่สุด(Ordinary Least Square) สำหรับพารามิเตอร์  $\beta_0$  และ  $\beta_1$  ตามลำดับ

$$\begin{aligned} \text{จาก (1)} \quad \varepsilon_i &= y_i - (\beta_0 + \beta_1 x_i) \\ &= y_i - E(y_i | x_i) \quad \dots\dots\dots (4) \end{aligned}$$

จากสมการ(4) แทนค่าประมาณของ  $E(y_i | x_i)$  ด้วย  $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$  จะได้ค่าประมาณของความคลาดเคลื่อนที่  $i$  ซึ่งเรียกว่าเศษตกค้างที่  $i$  (Residual;  $e_i$ )

$$e_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) = y_i - \hat{y}_i \quad \dots\dots\dots (5)$$

ดังนั้น  $e_i$  จึงเป็นความแตกต่างระหว่างค่าสังเกตจริง ( $y_i$ ) กับค่าทำนาย ( $\hat{y}_i$ ) ซึ่งอาจมองว่าเศษตกค้างเป็นตัวอย่างสุ่มของความคลาดเคลื่อน ( $\varepsilon_i$ ) จากตัวแบบ (1) ซึ่งจะสะท้อนคุณสมบัติของ  $\varepsilon_i$  และ  $e_i$  มีคุณสมบัติคือ

1.  $e_i$  มีค่าเฉลี่ยเท่ากับ 0
2.  $e_i$  มีความแปรปรวนประมาณ

$$\frac{\sum_{i=1}^n (e_i - \bar{e})^2}{n - p} = \frac{\sum_{i=1}^n e_i^2}{n - p} = \text{MSE}$$

เมื่อ  $p$  เป็นจำนวนพารามิเตอร์ในตัวแบบ  
 $\bar{e}$  เป็นค่าเฉลี่ยของเศษตกค้าง  
 MSE เป็นตัวประมาณค่าของ  $\sigma^2$

การนำค่า  $e_i$  มาพล็อตกราฟจะเป็นประโยชน์มากในการตรวจสอบข้อสมมติเบื้องต้นของการวิเคราะห์การถดถอยดังกล่าวข้างต้น นอกจากนี้ยังสามารถตรวจสอบว่าค่าสังเกตใด ๆ จะเป็นค่าผิดปกติ (Outlier) หรือไม่ โดยการแปลงค่าเศษตกค้างไปเป็นเศษตกค้างมาตรฐาน (Standardized Residual;  $z_i$ ) คือ

$$z_i = \frac{e_i}{\sqrt{\text{MSE}}}; i = 1, 2, \dots, n$$

$$z_i \sim N(0,1) \quad \text{ดังนั้นถ้า } |z_i| > 3$$

แสดงว่าเป็นค่าที่ผิดปกติ

## การวิเคราะห์เศษตกค้าง (Residual Analysis)

เป็นการวิเคราะห์กราฟของเศษตกค้าง ในแบบต่าง ๆ ดังต่อไปนี้

### 1. กราฟระหว่างเศษตกค้างเทียบกับค่าทำนาย หรือ ตัวแปรอิสระ

เป็นกราฟที่พล็อตเศษตกค้าง ( $e_i$ ) หรือ เศษตกค้างมาตรฐาน ( $z_i$ ) ในแกนตั้ง และค่าทำนาย ( $\hat{y}_i$ ) หรือตัวแปรอิสระ ( $x_i$ ) ในแกนนอน ถ้ารูปที่ได้  $e_i$  มีการกระจายรอบ ๆ 0 อย่างไม่มีรูปแบบ (Pattern) สามารถสรุปได้ว่า ตัวแบบการวิเคราะห์การถดถอยที่ใช้่นั้นเหมาะสม และความคลาดเคลื่อนมีความแปรปรวนคงที่ แต่ถ้าตัวแบบการวิเคราะห์การถดถอยที่ใช้่นั้นไม่เหมาะสม การกระจายของเศษตกค้างจะแสดงให้เห็นรูปแบบที่เหมาะสม เช่น ถ้าข้อมูลที่ใช้ในการวิเคราะห์นั้นควรจะใช้ตัวแบบที่เป็นเส้นโค้ง แต่กลับไปใช้ตัวแบบที่เป็นเส้นตรง เมื่อพล็อตกราฟระหว่างเศษตกค้างเทียบกับค่าทำนายหรือเทียบกับตัวแปรอิสระ จะได้กราฟที่เป็นเส้นโค้งอย่างชัดเจน และถ้าความแปรปรวนของความคลาดเคลื่อนไม่คงที่ การกระจายของเศษตกค้างจะมีลักษณะคล้ายรูปกรวย นอกจากนี้อาจจะชี้ให้เห็นค่าสังเกตที่ผิดปกติด้วย

**ตัวอย่างที่ 1** สมการถดถอยอย่างง่ายที่ประมาณได้จากข้อมูลในตารางที่ 1 คือ

$$\hat{y}_i = 515.705 + 0.569x_i$$

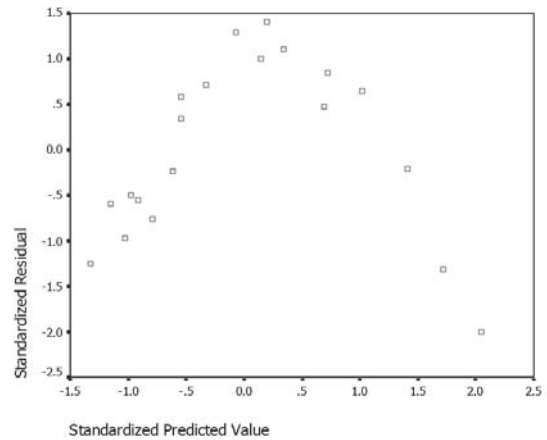
จากการทดสอบสมมุติฐานเกี่ยวกับความชันของเส้นถดถอย พบว่า  $\beta_1 \neq 0$  (P-value = 0.000) หมายความว่า ระดับโคเรสเตอรอลของนักกรีฑา

โอลิมปิกมีความสัมพันธ์เชิงเส้นกับปริมาณไขมันที่บริโภคโดยเฉลี่ยต่อวัน และ  $R^2 = 0.872$  หมายความว่า ในความผันแปรทั้งหมดของระดับโคเรสเตอรอลของนักกรีฑาโอลิมปิกถูกทำให้ลดลงด้วยปริมาณไขมันที่บริโภคโดยเฉลี่ยต่อวันถึง 87.2% เมื่อพล็อตเศษตกค้างเทียบกับค่าทำนายได้ดังภาพที่ 1 ซึ่งมีลักษณะเดียวกับการพล็อตเศษตกค้างเทียบกับปริมาณไขมันที่บริโภคโดยเฉลี่ยต่อวันได้ดังภาพที่ 2

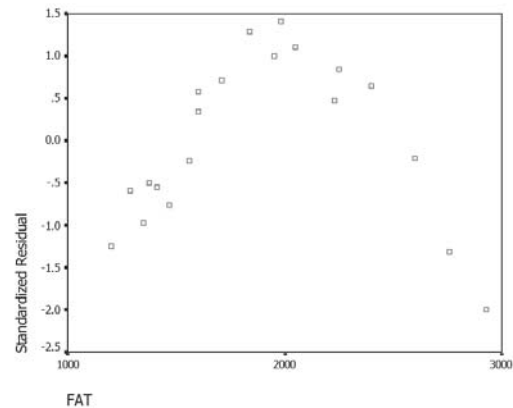
**ตารางที่ 1** ระดับโคเรสเตอรอล (มิลลิกรัมต่อลิตร; y) และปริมาณไขมันที่บริโภคโดยเฉลี่ยต่อวัน (มิลลิกรัม; x) ของนักกรีฑาโอลิมปิก 20 คน

คนที่	x	y	คนที่	x	y
1	1182	1290	11	1055	1200
2	1172	1350	12	1241	1375
3	1264	1470	13	1254	1410
4	1493	1600	14	1377	1560
5	1571	1710	15	1465	1600
6	1711	1840	16	1741	1950
7	1804	1980	17	1810	2050
8	1840	2230	18	1893	2250
9	1956	2400	19	1972	2600
10	1954	2930	20	1935	2760

ที่มา: Mendenhall and Sincich, 2012: 385



**ภาพที่ 1** กราฟของเศษตกค้างมาตรฐานเทียบกับค่าทำนายของสมการถดถอยที่ประมาณได้ในตัวอย่างที่ 1



**ภาพที่ 2** กราฟของเศษตกค้างมาตรฐานเทียบกับปริมาณไขมันที่บริโภคโดยเฉลี่ยต่อวันของสมการถดถอยที่ประมาณได้ในตัวอย่างที่ 1

จากภาพที่ 1 และ ภาพที่ 2 จะเห็นว่า การกระจายของเศษตกค้างมีรูปแบบเป็นเส้นโค้งแสดงว่า ตัวแบบการวิเคราะห์การถดถอยในรูปเชิงเส้นไม่เหมาะสมควรใช้ตัวแบบเชิงเส้นโค้ง

**ตัวอย่างที่ 2** สมการถดถอยอย่างง่ายที่ประมาณได้ จากข้อมูลในตารางที่ 2 คือ

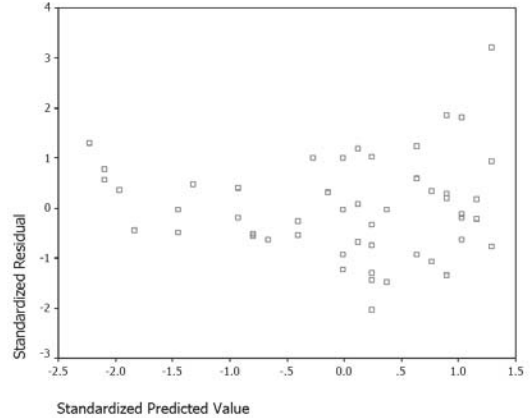
$$\hat{y}_i = 11369.41 + 2141.31x_i$$

จากการทดสอบสมมุติฐานเกี่ยวกับความชันของเส้นถดถอย พบว่า  $\beta_1 \neq 0$  (P-value = 0.000) หมายความว่าเงินเดือนของพนักงานมีความสัมพันธ์เชิงเส้นกับจำนวนปีที่มีประสบการณ์ และ  $R^2 = 0.787$  หมายความว่า ในความผันแปรทั้งหมดของเงินเดือนพนักงานถูกทำให้ลดลงด้วยจำนวนปีที่มีประสบการณ์ถึง 78.7% เมื่อพล็อตเศษตกค้างเทียบกับค่าทำนายได้ดังภาพที่ 3 ซึ่งมีลักษณะเช่นเดียวกับการพล็อตเศษตกค้างเทียบกับจำนวนปีที่มีประสบการณ์ได้ดังภาพที่ 4

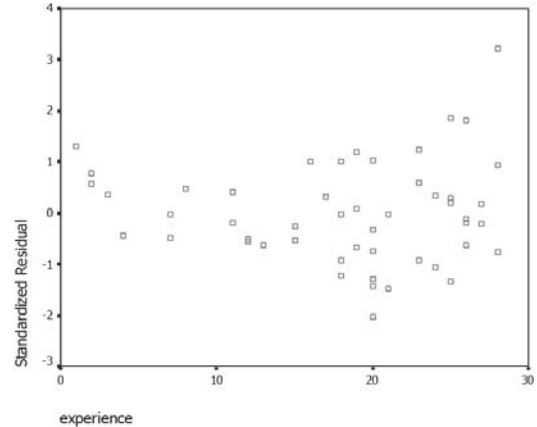
**ตารางที่ 2** เงินเดือน (y) และ จำนวนปีที่มีประสบการณ์ (x) ของพนักงาน 50 คน

x	y	x	y	x	y
7	26075	21	43628	28	99139
28	79370	4	16105	23	52624
23	65726	24	65644	17	50594
18	41983	20	63022	25	53272
19	62309	20	47780	26	65343
15	41154	15	38853	19	46216
24	53610	25	66537	16	54288
13	33697	25	67447	3	20844
2	22444	28	64785	12	32586
8	32562	26	61581	23	71235
20	43076	27	70678	20	36530
21	56000	20	51301	19	52745
18	58667	18	39346	27	67282
7	22210	1	24833	25	80931
2	20521	26	65929	12	32303
18	49727	20	41721	11	38371
11	33233	26	82641		

ที่มา: Mendenhall and Sincich, 2012: 401



**ภาพที่ 3** กราฟของเศษตกค้างมาตรฐานเทียบกับค่าทำนายของสมการถดถอยที่ประมาณได้ในตัวอย่างที่ 2



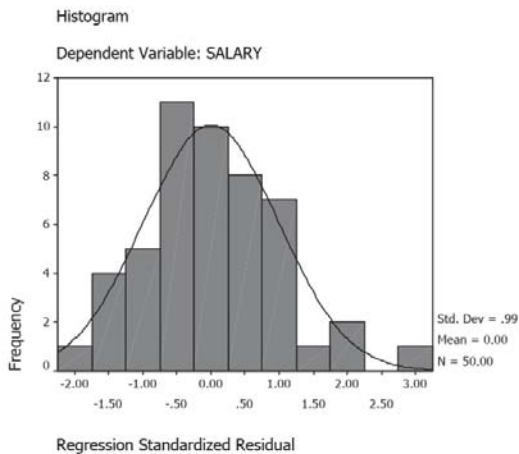
**ภาพที่ 4** กราฟของเศษตกค้างมาตรฐานเทียบกับประสบการณ์ในการทำงานของสมการถดถอยที่ประมาณได้ในตัวอย่างที่ 2

จากภาพที่ 3 และภาพที่ 4 จะเห็นว่า การกระจายของเศษตกค้างมีลักษณะคล้ายรูปกรวย แสดงว่า ความแปรปรวนของความคลาดเคลื่อนไม่คงที่

## 2. ฮิสโทแกรมของเศษตกค้างและ Normal Probability Plot

ฮิสโทแกรมของเศษตกค้างเป็นกราฟที่แสดงการแจกแจงความถี่ของเศษตกค้างโดยที่แกนนอนจะ

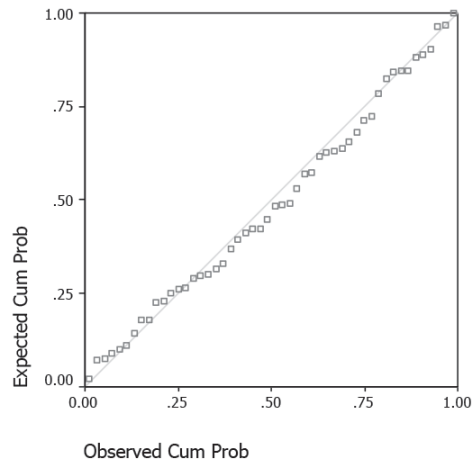
เป็นเศษตกค้างหรือเศษตกค้างมาตรฐานที่แบ่งเป็น  
 ชั้นๆ แทนตั้งเป็นจำนวนหรือสัดส่วนของค่าสังเกต  
 ที่อยู่ในแต่ละชั้น ถ้าฮิสโทแกรมของเศษตกค้างมี  
 ลักษณะสมมาตรหรือใกล้เคียงสมมาตรก็สามารถสรุป  
 ได้ว่าความคลาดเคลื่อนมีการแจกแจงปกติ



**ภาพที่ 5** ฮิสโทแกรมของเศษตกค้างมาตรฐานของ  
 สมการถดถอยที่ประมาณได้ในตัวอย่างที่ 2

Normal Probability Plot ของเศษตกค้างเป็น  
 กราฟที่พล็อตระหว่างเศษตกค้างที่เรียงลำดับจาก  
 น้อยไปมากกับค่าคาดหวังของเศษตกค้างเมื่อเศษ  
 ตกค้างมีการแจกแจงแบบปกติ Normal Probability  
 Plot เป็นกรณีพิเศษของ Q-Q Probability Plot  
 สำหรับกรณีการแจกแจงแบบปกติ (Wikipedia Free  
 Encyclopedia, 2012a) นอกจากนี้ อาจพิจารณาจาก  
 P-P Plot ก็ได้ ซึ่ง P-P Plot ในกรณีการแจกแจง  
 ปกติ เป็นกราฟที่พล็อตระหว่างความน่าจะเป็นสะสม  
 ของค่าสังเกตกับความน่าจะเป็นคาดหวังภายใต้การ  
 แจกแจงปกติ (Wikipedia Free Encyclopedia,  
 2012b) ถ้าเศษตกค้างมีการแจกแจงแบบปกติกราฟ  
 ทั้ง 2 แบบที่พล็อตได้ควรมีลักษณะใกล้เคียงกับเส้น  
 ตรงที่ผ่านจุดกำเนิดและมีความชันเท่ากับหนึ่ง

Normal P-P Plot of Regression Standardized Residual  
 Dependent Variable: SALARY

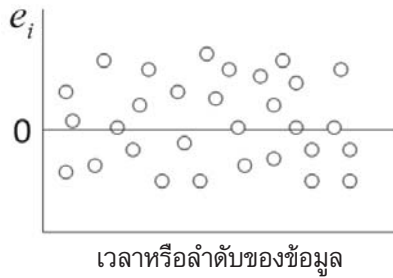


**ภาพที่ 6** Normal Probability Plot ของสมการถดถอย  
 ที่ประมาณได้จากตัวอย่างที่ 2

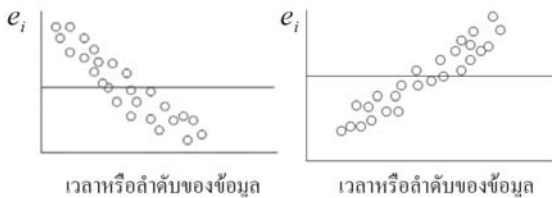
จากภาพที่ 5 และภาพที่ 6 สามารถสรุปได้ว่า  
 ความคลาดเคลื่อนมีการแจกแจงปกติโดยประมาณ

### 3. กราฟระหว่างเศษตกค้างเทียบกับเวลาหรือลำดับของข้อมูล

เป็นกราฟที่พล็อตเศษตกค้างหรือเศษตกค้าง  
 มาตรฐานในแกนตั้งและเวลาหรือลำดับของข้อมูล  
 ในแกนนอน ถ้าการกระจายของเศษตกค้างมีการ  
 กระจายรอบ ๆ 0 อย่างไม่มีการเบี่ยงเบนดังภาพที่ 7 แสดง  
 ว่าความคลาดเคลื่อนเป็นอิสระต่อกัน แต่ถ้ากระจาย  
 ของเศษตกค้างมีรูปแบบแสดงว่า ความคลาดเคลื่อน  
 ไม่เป็นอิสระต่อกันดังเช่นภาพที่ 8



ภาพที่ 7 กราฟระหว่างเศษตกค้างเทียบกับเวลาหรือลำดับของข้อมูลที่แสดงให้เห็นว่าความคลาดเคลื่อนเป็นอิสระต่อกัน



ภาพที่ 8 กราฟระหว่างเศษตกค้างเทียบกับเวลาหรือลำดับของข้อมูลที่แสดงให้เห็นว่าความคลาดเคลื่อนไม่เป็นอิสระต่อกัน

ตัวอย่างที่ 3 สมการถดถอยที่ประมาณได้จากข้อมูลในตารางที่ 3 คือ

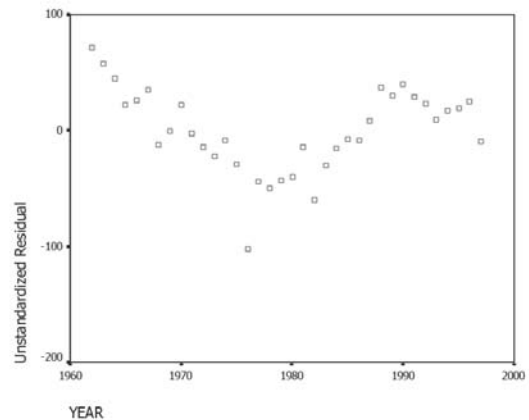
$$\hat{y}_i = 632.694 + 0.177x_i$$

จากการทดสอบสมมุติฐานเกี่ยวกับความชันของเส้นถดถอย พบว่า  $\beta_1 \neq 0$  (P-value = 0.000) หมายความว่ายอดขายของบริษัท ABC มีความสัมพันธ์เชิงเส้นกับค่าใช้จ่ายในการโฆษณา และ  $R^2 = 0.949$  หมายความว่า ในความผันแปรทั้งหมดของยอดขายของบริษัท ABC ถูกทำให้ลดลงด้วยค่าใช้จ่ายในการโฆษณาถึง 94.9% แต่เมื่อพล็อตเศษตกค้างเทียบกับเวลาหรือลำดับของข้อมูลจะเห็นว่าการกระจายของเศษตกค้างมีภาพแบบดังภาพที่ 9

ตารางที่ 3 ยอดขาย(พันดอลลาร์; y) และ ค่าใช้จ่ายในการโฆษณา (พันดอลลาร์; x) ของบริษัท ABC ตั้งแต่ปี 1962 ถึง 1997

Year	y	x	Year	y	x
1962	381.0	5316.8	1980	492.8	6574.6
1963	383.9	5413.2	1981	541.2	6704.2
1964	384.4	5486.9	1982	512.0	6794.3
1965	370.5	5537.8	1983	562.0	6911.4
1966	396.4	5660.6	1984	590.1	6986.5
1967	421.8	5750.8	1985	617.7	7095.7
1968	379.2	5782.2	1986	629.3	7170.8
1969	390.9	5781.7	1987	653.9	7210.9
1970	420.9	5821.9	1988	698.6	7304.8
1971	408.8	5892.5	1989	707.8	7391.9
1972	407.2	5950.2	1990	735.9	7495.3
1973	408.4	6002.1	1991	748.3	7629.2
1974	444.2	6121.8	1992	755.4	7703.4
1975	437.2	6201.2	1993	762.0	7818.4
1976	376.1	6271.7	1994	794.3	7955.0
1977	454.6	6383.1	1995	815.5	8063.4
1978	459.2	6444.5	1996	840.9	8170.8
1979	478.2	6509.1	1997	820.8	8254.5

ที่มา: Dielman, 2001: 359



ภาพที่ 9 กราฟระหว่างเศษตกค้างเทียบกับเวลา (Year) ของสมการถดถอยที่ประมาณได้ในตัวอย่างที่ 3

จากภาพที่ 9 เศษตกค้างมีการกระจายอย่างมีรูปแบบ แสดงว่า ความคลาดเคลื่อนไม่เป็นอิสระต่อกัน

## การแก้ปัญหาหลังการวิเคราะห์เศษตกค้าง

หลังจากวิเคราะห์เศษตกค้างแล้วสามารถสรุปได้ว่าข้อมูลที่นำมาวิเคราะห์การถดถอยนั้นเป็นไปตามข้อสมมติเบื้องต้นหรือไม่ ถ้าพบว่าข้อมูลไม่เป็นไปตามข้อสมมติเบื้องต้นข้อใดให้ดำเนินการแก้ไขตามปัญหาที่พบ คือ

1. ถ้าตัวแบบการวิเคราะห์การถดถอยเชิงเส้นไม่เหมาะสม ควรเปลี่ยนตัวแบบในการวิเคราะห์ให้เหมาะสมตามที่กราฟเศษตกค้างแสดง เช่น ภาพที่ 1 และ ภาพที่ 2 ควรใช้ตัวแบบการวิเคราะห์การถดถอยที่เป็นเส้นโค้งคือ  $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i$

2. ถ้าความแปรปรวนของความคลาดเคลื่อนไม่คงที่ เช่น ภาพที่ 3 และ ภาพที่ 4 ควรแก้ไขโดยการแปลงข้อมูลเช่น ถ้า  $y$  เป็นจำนวนนับที่มีการแจกแจงแบบปัวส์ซองจะใช้การแปลง  $y' = \sqrt{y}$  ถ้า  $y$  เป็นจำนวนนับที่มีการแจกแจงแบบทวินามจะใช้การแปลง  $y' = \sin^{-1} \sqrt{y}$  เป็นต้น หรือ เปลี่ยนไปใช้ตัวแบบการวิเคราะห์การถดถอยเชิงเส้นที่มีข้อสมมุติว่าความแปรปรวนของความคลาดเคลื่อนแตกต่างกันซึ่งสามารถประมาณเส้นถดถอยด้วยวิธีการ Weighted Least Squares (Chatterjee and Hadi, 2006: 167; Montgomery, Peck, and Vining, 2006: 176)

3. ถ้าความคลาดเคลื่อนไม่มีการแจกแจงปกติ บางครั้งอาจมีสาเหตุมาจากปัญหาอื่น ๆ เช่น ความแปรปรวนไม่คงที่ ให้แก้ปัญหาที่ความแปรปรวนให้คงที่ก่อน ถ้าแก้ปัญหาให้ความแปรปรวนคงที่แล้วอาจจะทำให้ความคลาดเคลื่อนมีการแจกแจงแบบ

ปกติก็ได้ แต่ถ้าแก้ปัญหานั้นแล้วยังมีปัญหาการแจกแจงของความคลาดเคลื่อนไม่เป็นปกติ ให้ใช้การแปลงข้อมูลด้วยลอการิทึมธรรมชาติ หรือรากที่สอง หรือส่วนกลับของข้อมูลเดิม บางครั้งอาจแปลงเฉพาะตัวแปรอิสระ บางครั้งอาจแปลงเฉพาะตัวแปรตาม หรือบางครั้งอาจแปลงทั้งตัวแปรอิสระและตัวแปรตาม

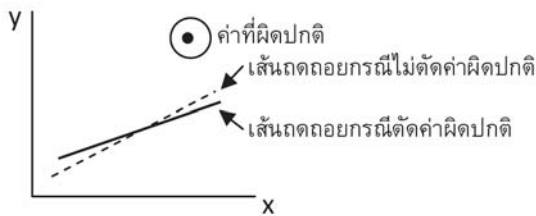
4. ถ้าความคลาดเคลื่อนไม่เป็นอิสระต่อกันเรียกว่า ปัญหา Autocorrelation ซึ่งมักเกิดกับข้อมูลที่เป็นอนุกรมเวลา ปัญหานี้เกิดจากหลายสาเหตุ การแก้ปัญหาจึงแตกต่างกันตามสาเหตุที่เกิด คือ ถ้าปัญหาเกิดจากการละทิ้งตัวแปรอิสระที่มีความสำคัญออกจากตัวแบบ ทำให้ความคลาดเคลื่อนมีความสัมพันธ์กันผ่านตัวแปรอิสระที่ถูกละเลย วิธีการแก้ไข คือ การเพิ่มตัวแปรอิสระเหล่านั้นลงไปในตัวแบบ (ถวิล นิลใบ, 2555ก; Montgomery, Peck, and Vining, 2006: 476) ถ้าปัญหาเกิดจากการกำหนดตัวแบบไม่ถูกต้องก็แก้ไขโดยการเปลี่ยนตัวแบบ แต่ถ้าปัญหาเกิดจากปัจจัยที่อยู่เบื้องหลังความคลาดเคลื่อนโดยตรงคือความคลาดเคลื่อนที่เกิดในช่วงเวลาหนึ่งโดยตัวมันเองอาจจะส่งผลกระทบต่อไปยังช่วงเวลาอื่น ๆ หรือกล่าวอีกนัยหนึ่งคือความคลาดเคลื่อนที่เกิดขึ้นในแต่ละช่วงเวลาไม่เป็นอิสระต่อกันซึ่งเรียกว่า Pure Autocorrelation (ถวิล นิลใบ, 2555ก) การแก้ปัญหาต้องใช้ตัวแบบการถดถอยที่มีโครงสร้างของความคลาดเคลื่อน และต้องอาศัยเทคนิคการประมาณพารามิเตอร์โดยเฉพาะก็จะสามารถประมาณสมการถดถอยได้เมื่อมีข้อมูลเกี่ยวกับโครงสร้างของ Autocorrelation เพียงพอ (Montgomery, Peck, and Vining, 2006: 476)

5. ในกรณีที่ตรวจพบค่าสังเกตบางค่าผิดปกติ ต้องตรวจสอบว่าค่าผิดปกตินั้นเกิดจากอะไร ถ้า



ตรวจสอบแล้วเป็นค่าผิดปกติที่เกิดจากความผิดพลาดในการดำเนินงานให้แก้ไขให้ถูกต้อง แต่ถ้าไม่สามารถแก้ไขได้อาจตัดค่าผิดปกตินั้นออกไป แต่ถ้าค่าผิดปกติ นั้นเป็นข้อมูลจริง ๆ อาจจะเป็นผลมาจากผลกระทบร่วมกับตัวแปรอิสระอื่น ๆ ที่ไม่ได้อยู่ในตัวแบบ เช่น มีการเปลี่ยนแปลงชิ้นส่วนของเครื่องจักร หรือ เปลี่ยนน้ำมันเครื่อง ซึ่งการเปลี่ยนแปลงนี้ทำให้ผลผลิตดีขึ้น จึงควรเก็บค่าผิดปกติไว้ในการวิเคราะห์ต่อไป ค่าผิดปกตินี้มีผลต่อเส้นถดถอยที่ประมาณได้คือจะทำให้ได้เส้นถดถอยที่แตกต่างไปดังภาพที่ 10

ภาพที่ 10 เส้นถดถอยที่ประมาณได้เมื่อตัด/ไม่ตัดค่าผิดปกติ



### สรุปและข้อเสนอแนะ

การวิเคราะห์เศษตกค้างในการวิเคราะห์การถดถอยเป็นวิธีการตรวจสอบข้อสมมุติเบื้องต้นของการวิเคราะห์การถดถอยว่า ข้อมูลที่นำมาวิเคราะห์นั้นเป็นไปตามข้อสมมุติเบื้องต้นของการวิเคราะห์หรือไม่ วิธีการวิเคราะห์เป็นการพิจารณาจากกราฟของเศษตกค้างที่พล็อตเทียบกับค่าทำนาย ตัวแปรอิสระ เวลาหรือลำดับของข้อมูล และ Normal Probability Plot ถ้ากราฟที่พล็อตระหว่างเศษตกค้างเทียบกับค่าทำนาย หรือตัวแปรอิสระมีการกระจายอย่างไม่มีรูปแบบ แสดงว่า ตัวแบบที่ใช้วิเคราะห์ข้อมูลเหมาะสมและความคลาดเคลื่อนมีความแปรปรวนคงที่ ทำนองเดียวกันกับกราฟที่พล็อตระหว่างเศษตกค้างเทียบกับเวลาหรือลำดับของข้อมูลมีการกระจายอย่างไม่มี

รูปแบบ แสดงว่า ความคลาดเคลื่อนเป็นอิสระต่อกัน ส่วนกราฟจาก Normal Probability Plot นั้น ถ้ามีการกระจายอยู่ในแนวเส้นตรงแนวทแยง แสดงว่า ความคลาดเคลื่อนมีการแจกแจงแบบปกติ ซึ่งหมายความว่า ข้อมูลที่นำมาวิเคราะห์นั้นเป็นไปตามข้อสมมุติเบื้องต้นของการวิเคราะห์การถดถอยเชิงเส้น

ถึงแม้ว่าการวิเคราะห์เศษตกค้างเป็นวิธีการที่ทำได้ง่ายและรวดเร็วด้วยโปรแกรมสำเร็จรูปทางสถิติต่างๆ แต่อาจมองว่าเป็นวิธีการตรวจสอบอย่างไม่เป็นทางการ (Informal Method) ถ้าต้องการตรวจสอบอย่างเป็นทางการ (Formal Method) ก็ควรใช้วิธีการทดสอบสมมุติฐาน คือ ตรวจสอบว่า ตัวแบบที่ใช้เหมาะสมหรือไม่ด้วยการทดสอบ Lack of Fit (Montgomery, Peck, and Vining, 2006: 145) ตรวจสอบความคงที่ของความแปรปรวนด้วย Goldfeld-Quandt Test, Glejser Test, White's Test (ถวิล นิลใบ, 2555: 25) ตรวจสอบความคลาดเคลื่อนมีการแจกแจงแบบปกติด้วย  $\chi^2$  Goodness of Fit Test, Kolmogorov-Sminov Test, (Neter, Wasserman, and Kutner, 1985: 123) Shapiro-Wilk Test (นิติพงษ์ ส่งศรีโรจน์, 2555: 24-28) ตรวจสอบความคลาดเคลื่อนเป็นอิสระต่อกันด้วย Run Test, Durbin Watson Test (Chatterjee and Hadi, 2006: 200; Neter, Wasserman, and Kutner, 1985: 123) เป็นต้น

ดังนั้น เพื่อให้เกิดความมั่นใจในการวิเคราะห์การถดถอยจึงควรทำการตรวจสอบข้อสมมุติเบื้องต้นอย่างน้อยก็ใช้วิธีการวิเคราะห์เศษตกค้าง ซึ่งการสร้างกราฟของเศษตกค้างเทียบกับค่าทำนาย ตัวแปรอิสระ เวลาหรือลำดับของข้อมูล และ Normal Probability plot สามารถใช้คำสั่งในโปรแกรม

สำเร็จรูปที่ใช้วิเคราะห์การถดถอยสร้างได้โดยง่าย และรวดเร็ว แต่ถ้าต้องการตรวจสอบข้อสมมุติเบื้องต้นของการวิเคราะห์การถดถอยอย่างเป็นทางการควรใช้วิธีการทดสอบสมมุติฐาน

## บรรณานุกรม

Chatterjee, Samprit, and Hadi, Ali S. 2006. **Regression Analysis by Example.** 4<sup>th</sup> ed. New York: Wiley.

Dielman, Terry E. 2001. **Applied Regression for Business and Economics.** 3<sup>rd</sup> ed. Pacific Grove, CA: Duxbury/Thomson Learning.

Mendenhall, William, and Sincich, Terry. 2012. **A Second Course in Statistics: Regression Analysis.** 7<sup>th</sup> ed. Boston: Prentice Hall.

Montgomery, Douglas C., Peck, Elizabeth A., and Vining, G. Geoffrey. 2006. **Introduction to Linear Regression Analysis.** 4<sup>th</sup> ed. New York: Wiley.

Neter, John, Wasserman, William, and Kutner, Michael H. 1985. **Applied Linear Statistical Models.** 2<sup>nd</sup> ed. Homewood, IL: Irwin.

Nilbai, Tawin. 2012a, December 15. "Autocorrelation." In **Lecture Notes on Econometrics 2** [Online]. Available: [http://www.nidambe11.net/ekonomiz/eview\\_\\_doc/auto.htm](http://www.nidambe11.net/ekonomiz/eview__doc/auto.htm) (in Thai).

ถวิล นิลใบ. 15 ธันวาคม 2555ก. "ปัญหาสหสัมพันธ์ของตัวแปรพหุคูณ." ใน **เอกสารประกอบการบรรยายวิชาเศรษฐมิติ 2** [ออนไลน์]. เข้าถึง

จาก: [http://www.nidambe11.net/ekonomiz/eview\\_\\_doc/auto.htm](http://www.nidambe11.net/ekonomiz/eview__doc/auto.htm)

Nilbai, Tawin. 2012b, December 15. "Heteroscedasticity." In **Lecture Notes on Econometrics 2** [Online]. Available: [http://www.nidambe11.net/ekonomiz/eview\\_\\_doc/hetero.htm](http://www.nidambe11.net/ekonomiz/eview__doc/hetero.htm) (in Thai).

ถวิล นิลใบ. 15 ธันวาคม 2555ข. "Heteroscedasticity." ใน **เอกสารประกอบการบรรยายวิชาเศรษฐมิติ 2** [ออนไลน์]. เข้าถึงจาก: [http://www.nidambe11.net/ekonomiz/eview\\_\\_doc/hetero.htm](http://www.nidambe11.net/ekonomiz/eview__doc/hetero.htm)

Songsriroj, Nitipong. 2012, December 15. **Violation of Regression Analysis Assumption: Cause, Result, Detection and Solution** [Online]. Available: [http://www.nitiphong.com/paper\\_\\_pdf/phd/regression/assumption\\_\\_regress\\_\\_spss25-08-2553.pdf](http://www.nitiphong.com/paper__pdf/phd/regression/assumption__regress__spss25-08-2553.pdf) (in Thai).

นิติพงษ์ ส่งศรีโรจน์. 15 ธันวาคม 2555. **การละเมิดข้อสมมติของการวิเคราะห์ความถดถอย: สาเหตุ ผล การตรวจสอบและการแก้ไข** [ออนไลน์]. เข้าถึงจาก: [http://www.nitiphong.com/paper\\_\\_pdf/phd/regression/assumption\\_\\_regress\\_\\_spss25-08-2553.pdf](http://www.nitiphong.com/paper__pdf/phd/regression/assumption__regress__spss25-08-2553.pdf)

Wikipedia Free Encyclopedia. 2012a, December 15. **Normal Probability Plot** [Online]. Available: [http://en.wikipedia.org/wiki/Normal\\_\\_probability\\_\\_plot](http://en.wikipedia.org/wiki/Normal__probability__plot)

Wikipedia Free Encyclopedia. 2012b, December 15. **P-P Plot** [Online]. Available: [http://en.wikipedia.org/wiki/P%E2%80%93P\\_\\_plot](http://en.wikipedia.org/wiki/P%E2%80%93P__plot)



**Assistant Professor Yupin Kanjanasakda** received her Master of Science Degree in Statistics from Chulalongkorn University. She is currently a lecturer in the School of Science and Technology, University of the Thai Chamber of Commerce. Her main interests are regression analysis, time series analysis and confidence interval of parameter.