บทที่ 12
การวิเคราะห์การถดถอยแบบไม่เป็นเส้นตรง

1. สมการการถดถอยแบบไม่เป็นเส้นตรง

ในบทก่อนเราอธิบายการวิเคราะห์การถดถอยที่ตัวแบบการถดถอยเป็นสมการแบบเส้นตรง สำหรับในบทนี้จะอธิบายการวิเคราะห์การถดถอยที่ตัวแบบการถดถอยเป็นสมการแบบไม่เป็นเส้นตรง ตัวอย่างสมการถดถอยที่ไม่เป็นเส้นตรง ได้แก่

สมการถดถอยแบบควอดราทิก คือ $E(Y) = \beta_0 + \beta_1 X + \beta_2 X^2$

สมการถดถอยแบบคิวบิก คือ $E(Y) = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3$

สมการถดถอยแบบเอกซ์โพเนนเชียล คือ $E(Y) = \beta_0 \beta_1^X$

เราจะหาตัวแบบการถดถอยที่เป็นสมการแบบไม่เป็นเส้นตรง เมื่อพบว่าสมการแบบเส้นตรงไม่เหมาะสมกับข้อมูล แต่การประมาณค่าพารามิเตอร์ของตัวแบบการถดถอยก็ยุ่งยากขึ้นด้วย ดังนั้นในกรณีที่สมการถดถอยแบบเส้นตรงไม่เหมาะสมกับข้อมูลเราสามารถทำได้ 2 วิธีการคือ

1. การแปลงข้อมูล (transformation) ซึ่งจะทำให้สมการถดถอยแบบเส้นตรงเหมาะสมกับข้อมูลที่แปลงแล้ว

2. การหาตัวแบบสถิติตัวแบบใหม่ที่ไม่ใช่สมการเส้นตรงซึ่งเหมาะสมกับข้อมูลนั้น

2. การแปลงข้อมูล

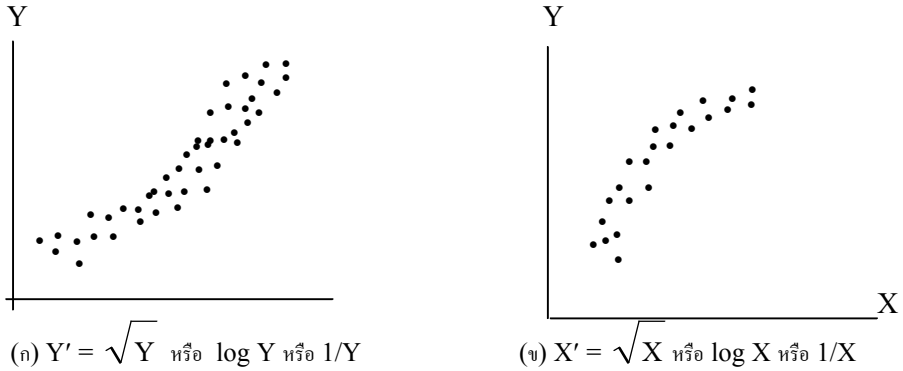
การแปลงข้อมูลบางครั้งจะทำกับตัวแปรตาม Y เท่านั้น บางครั้งทำกับตัวแปรอิสระ X เท่านั้น หรืออาจทำกับทั้งตัวแปรตามและตัวแปรอิสระ เช่น

1. $Y' = \sqrt{Y}$, $X' = \sqrt{X}$
2. $Y' = \log_{10}$, $X' = \log_{10} X$
3. $Y' = \frac{1}{Y}$, $X' = \frac{1}{X}$

การเลือกใช้การแปลงข้อมูลชนิดใด สามารถทำได้โดยการพล็อตกราฟเปรียบเทียบระหว่างการทรานส์ฟอร์มแบบต่าง ๆ

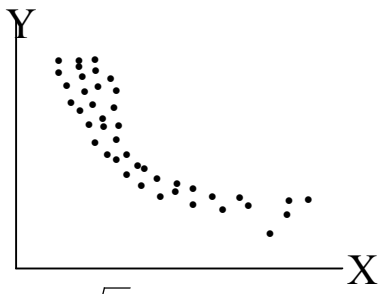
ข้อสังเกต

1. การเลือกว่าจะใช้วิธีการแปลงข้อมูลแบบใดอาจมาจากแนวคิด ทฤษฎี ตัวอย่างเช่น การศึกษาความสัมพันธ์ระหว่างราคาสินค้า (X) และปริมาณความต้องการ (Y) ถ้าพล็อตกราฟของข้อมูลได้ ดังภาพ



(ก) $Y' = \sqrt{Y}$ หรือ log Y หรือ 1/Y

(ข) $X' = \sqrt{X}$ หรือ log X หรือ 1/X



(ค) $Y' = \sqrt{Y}$ หรือ log Y หรือ 1/Y and/or

$X' = \sqrt{X}$ หรือ log X หรือ 1/X

ภาพที่ 12.1 การแปลงข้อมูลสำหรับลักษณะของโค้งที่แตกต่างกัน

นักเศรษฐศาสตร์อาจชอบการแปลงข้อมูลทั้ง Y และ X ด้วยการใส่ log เพื่อให้ได้เส้นตรง เพราะว่าความชันของเส้นถดถอยของตัวแปรที่ทรานส์ฟอร์มแล้วเป็นการวัดราคาที่ยืดหยุ่นของความต้องการ (the price elasticity of demand) เราจะแปล

ความหมายของความชันเป็นการแสดงอัตราการเปลี่ยนแปลงเป็นเปอร์เซนต์ของปริมาณความต้องการต่ออัตราการเปลี่ยนแปลงของราคา 1 เปอร์เซ็นต์ ซึ่งเป็นในทิศทางตรงกันข้าม

2. การแปลงข้อมูลของ X ไม่มีผลกระทบต่อความแปรปรวนหรือรูปร่างของการแจกแจงของความคลาดเคลื่อนซึ่งต่างจากการแปลงข้อมูลของ Y ดังนั้น เมื่อมีปัญหาเกี่ยวกับการแจกแจงของความคลาดเคลื่อนที่ไม่เป็นปกติหรือความแปรปรวนของความคลาดเคลื่อนไม่เท่ากัน เราจึงแก้ไขปัญหานี้ได้โดยการแปลงข้อมูลของตัวแปรตาม และในทางกลับกันเมื่อสมการถดถอยที่ไม่ใช่เส้นตรงแต่มีความแปรปรวนเท่ากัน การแปลงข้อมูลของ Y อาจทำให้ความคลาดเคลื่อนมีความแปรปรวนไม่เท่ากัน ดังนั้นจึงเป็นสิ่งจำเป็นที่จะต้องตรวจสอบกราฟของความคลาดเคลื่อนภายหลังการแปลงข้อมูล

3. การพิจารณาเลือกตัวแปรที่จะแปลงข้อมูล ควรเลือกตัวแปรที่มีค่ากว้างกว่าเป็นอันดับแรก เนื่องจากการแปลงข้อมูลของตัวแปรที่มีค่าแคบ ๆ ไม่ค่อยมีประสิทธิภาพ

4. ภายหลังการทดลองเลือกการแปลงข้อมูลต้องทำการพล็อตกราฟของความคลาดเคลื่อนและทำการวิเคราะห์การถดถอยเพื่อให้แน่ใจว่า ตัวแบบสถิติการถดถอยเชิงเส้นเหมาะสมกับข้อมูลที่แปลงแล้ว

5. เมื่อความแปรปรวนของความคลาดเคลื่อนไม่เท่ากับค่าคงที่แต่มีความสัมพันธ์เป็นพิเศษกับ $E(Y)$ สำหรับ X ที่กำหนดให้ สามารถใช้การแปลงข้อมูลเพื่อทำให้ความแปรปรวนสม่ำเสมอได้ มีกรณีที่สำคัญ 3 กรณี คือ

1. ถ้า σ_j^2 เป็นสัดส่วนกับ $E(Y_j)$ ใช้ $Y' = \sqrt{Y}$

2. ถ้า σ_j เป็นสัดส่วนกับ $E(Y_j)$ ใช้ $Y' = \log Y$

3. ถ้า $\sqrt{\sigma_j}$ เป็นสัดส่วนกับ $E(Y_j)$ ใช้ $Y' = \frac{1}{Y}$

เมื่อ σ_j^2 คือ ความแปรปรวนของความคลาดเคลื่อน

$E(Y_j)$ คือ ค่าเฉลี่ยของค่าสังเกต เมื่อ $X = X_j$

3. ตัวอย่างการแปลงข้อมูลด้วยการถอดรากที่สอง

ตัวอย่างการศึกษาจำนวนวันที่ใช้ฝึกอบรม และคะแนนการปฏิบัติงานของผู้เข้าอบรมเรื่องแบตเตอรี่ จำนวน 10 คน กำหนดให้ตัวแปร X คือ จำนวนวันที่ใช้ฝึกอบรม และตัว

แปร Y คือ คะแนนการปฏิบัติงาน ข้อมูลอยู่ในแฟ้มชื่อ Tran1.sav ดังแสดงในตารางต่อไปนี้

ตารางที่ 12.1 ข้อมูลจำนวนวันฝึกอบรมและคะแนนปฏิบัติงานของผู้เข้าอบรมเรื่องแบตเตอรี่

จำนวนวันฝึกอบรม (X)	คะแนนปฏิบัติงาน (Y)	จำนวนวันฝึกอบรม (X)	คะแนนปฏิบัติงาน (Y)
.5	43	2.0	158
.5	40	2.5	209
1.0	71	3.0	270
1.0	74	3.5	341
1.5	107		
1.5	109		

3.1 พล็อตกราฟของข้อมูล โดยใช้คำสั่ง Graphs

ขั้นตอนการใช้คำสั่งคือ

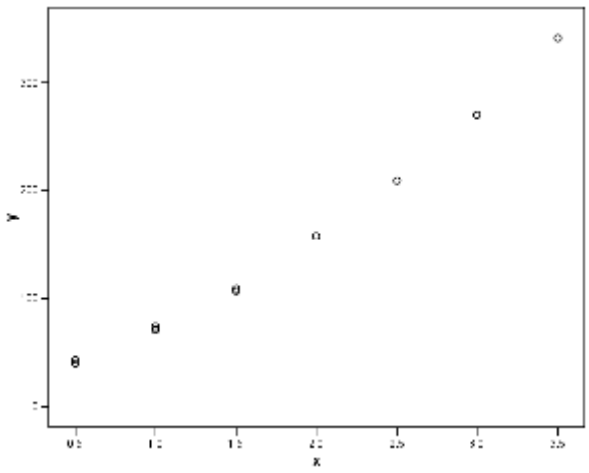
1. ไปที่เมนูบาร์ คลิกที่ Graphs Scatter ... จะได้หน้าต่าง Scatterplot เลือกที่คำสั่ง Simple แล้วคลิกปุ่ม Define จะได้หน้าต่าง Simple Scatterplot
2. ในหน้าต่าง Simple Scatterplot

   คลิกที่ตัวแปร Y ให้ย้ายไปอยู่ในช่อง Y Axis :

   คลิกที่ตัวแปร X ให้ย้ายไปอยู่ในช่อง X Axis :

   คลิกที่ปุ่ม OK จะได้ผลลัพธ์เป็นกราฟการกระจายของข้อมูล ดังภาพที่ 12.2

Graph



ภาพที่ 12.2 กราฟการกระจายของจำนวนวันที่ฝึกอบรม และคะแนนปฏิบัติงาน

จากการพล็อตกราฟของข้อมูล ได้เป็นเส้นโค้ง แสดงให้เห็นว่าการใช้สมการถดถอยเส้นตรงกับข้อมูลชุดนี้ไม่เหมาะสมจึงทำการแปลงข้อมูลเพื่อใช้ตัวแบบสถิติที่มีสมการแบบเส้นตรง

3.2 การแปลงข้อมูลด้วยการถอดรากที่สอง

ตัวอย่างนี้ทำการแปลงข้อมูลคะแนนปฏิบัติงานด้วยการถอดรากที่สองและตั้งตัวแปรใหม่เป็น Yprime จะได้ว่า Yprime = SQRT(Y) สามารถใช้คำสั่ง **Transform** ขั้นตอนการใช้คำสั่งคือ

1. เปิดแฟ้มข้อมูล tran1.sav

   คลิกที่ปุ่ม Variable View เพื่อให้ได้หน้าต่าง Variable View สร้างตัวแปรใหม่สำหรับข้อมูลที่แปลงแล้ว โดยพิมพ์ชื่อตัวแปร Yprime ที่คอลัมน์ Name แล้วเลือกทศนิยม 4 ตำแหน่งที่คอลัมน์ Decimals

   คลิกที่ปุ่ม Data View เพื่อให้ได้หน้าต่าง Data View

2. ในหน้าต่าง Data View

   คลิกที่คอลัมน์ของตัวแปร Yprime จะทำให้เกิดแถบดำทั้งคอลัมน์

ไปที่เมนูบาร์ คลิกที่ Transform, Compute... จะได้หน้าต่าง Compute Variable

3. ในหน้าต่าง Compute Variable

ในกรอบ Target Variable : พิมพ์ชื่อตัวแปร Yprime

ในกรอบ Functions : เลือกฟังก์ชัน SQRT (numexpr) แล้วคลิกลูกศร ⮝ ให้ฟังก์ชันนี้ย้ายไปอยู่ในช่อง Numeric Expression : จะได้เป็น SQRT (?) โดยที่เครื่องหมาย ? เป็นแถบเข้ม

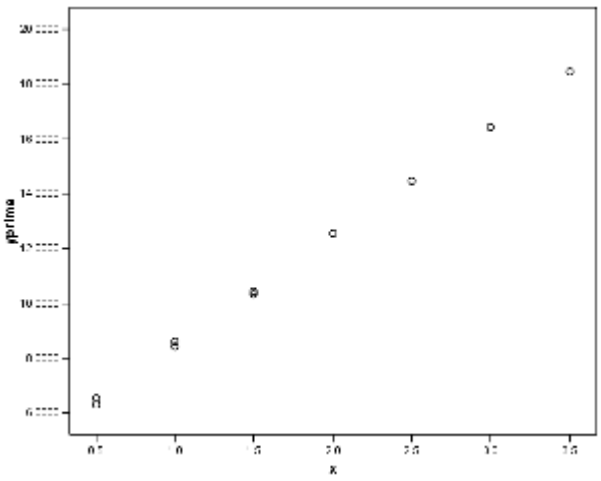
ในช่องซ้ายมือ คลิกเลือกตัวแปร Y แล้วคลิกที่ลูกศร ➢ จะได้ตัวแปร Y เข้าไปอยู่แทนเครื่องหมาย ? ที่เป็นแถบเข้มเดิมได้เป็นฟังก์ชัน SQRT(Y)

แล้วคลิกปุ่ม OK จะได้หน้าต่างที่มีคำถามว่า ?Change existing variable? ตอบโดยคลิกที่ปุ่ม OK ผลลัพธ์ที่ได้คือ ค่าของตัวแปร Yprime

3.3 พล็อตกราฟของข้อมูลที่แปลงแล้ว

ข้อมูลคะแนนปฏิบัติงานที่แปลงแล้วอยู่ในตัวแปรใหม่คือ Yprime ทำการพล็อตกราฟระหว่างตัวแปร X กับ Yprime ขั้นตอนการใช้คำสั่งเหมือนกับในหัวข้อ 3.1 จะได้ผลลัพธ์เป็นกราฟการกระจายของข้อมูลดังภาพที่ 12.3

Graph



ภาพที่ 12.3 กราฟการกระจายของจำนวนวันที่ฝึกอบรมและ SQRT (คะแนนปฏิบัติงาน)

จากการพล็อตกราฟของข้อมูลได้เป็นเส้นตรงแสดงให้เห็นว่าสมการถดถอยแบบเส้นตรงเหมาะสมกับข้อมูลที่ทรานส์ฟอร์มแล้ว

4. ตัวอย่างการแปลงข้อมูลด้วย logarithmic

ตัวอย่างการศึกษาข้อมูลอายุและระดับพลาสมาของเด็กสุขภาพดี 14 คน กำหนดให้ตัวแปร X คือ อายุ และตัวแปร Y คือ ระดับพลาสมา ข้อมูลอยู่ในแฟ้มชื่อ tran2.sav ดังแสดงในตารางต่อไปนี้

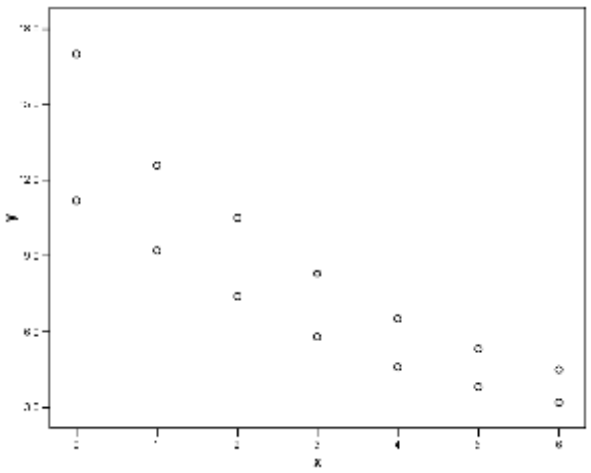
ตารางที่ 12.2 ข้อมูลอายุและระดับพลาสมาของเด็กสุขภาพดี

อายุ (X)	ระดับพลาสมา (Y)	อายุ (X)	ระดับพลาสมา (Y)
0	17.0	4	4.6
0	11.2	4	6.5
1	9.2	5	5.3
1	12.6	5	3.8
2	7.4	6	3.2
2	10.5	6	4.5
3	8.3		
3	5.8		

4.1 พล็อตกราฟของข้อมูลโดยใช้คำสั่ง Graphs

ขั้นตอนการใช้คำสั่งเหมือนในหัวข้อ 3.1 ได้ผลลัพธ์เป็นกราฟของข้อมูล ดังภาพที่ 12.4

Graph



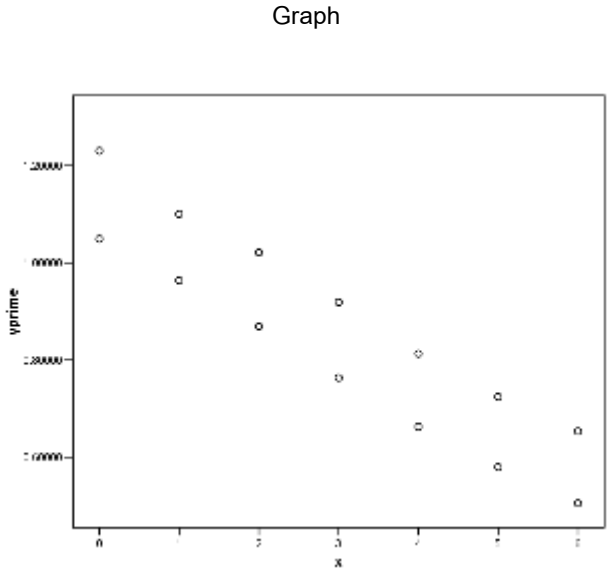
ภาพที่ 12.4 กราฟการกระจายของอายุและระดับพลาสมา

4.2 การแปลงข้อมูลด้วย logarithmic

ตัวอย่างนี้ทำการแปลงข้อมูลระดับพลาสมาด้วย logarithmic และตั้งตัวแปรใหม่เป็น Yprime จะได้ว่า Yprime = $Log_{10}(Y)$

4.3 พล็อตกราฟของข้อมูลที่แปลงแล้ว

พล็อตกราฟการกระจายของอายุและ Log_{10} (ระดับพลาสมา) ได้ดังภาพที่ 12.5



ภาพที่ 12.5 กราฟการกระจายของอายุและ Log_{10}(ระดับพลาสมา)

5. การวิเคราะห์การถดถอยสำหรับข้อมูลที่แปลงแล้ว

การแปลงข้อมูลของตัวแปรตามหรือตัวแปรอิสระหรือทั้งสองตัวแปรเพื่อทำให้ตัวแบบการถดถอยแบบเส้นตรงเหมาะสมสำหรับข้อมูลที่แปลงแล้ว ตัวอย่างสมการถดถอยแบบเอกซ์โพเนนเชียล $E(Y) = \beta_0\beta_1^X$ เมื่อใส่ฟังก์ชัน logarithm เข้าไปทั้งสองข้างของสมการ จะได้ว่า $\log[E(Y)] = \log(\beta_0) + \log(\beta_1)X$ ซึ่งเป็นสมการเส้นตรงที่ได้มาจากการแปลงข้อมูลของตัวแปร Y ด้วย logarithm จึงสามารถใช้เทคนิคการวิเคราะห์การถดถอยเชิงเส้นตรงในการประมาณค่าพารามิเตอร์ β_0 และ β_1 ได้ แล้วจึงหาค่า $E(Y)$ โดยการทำ antilog

สำหรับกรณีที่ทำการแปลงข้อมูลของตัวแปร Y ด้วยการถอดรากที่สอง เมื่อใช้เทคนิคการวิเคราะห์การถดถอยเชิงเส้นตรงในการประมาณค่า β_0 และ β_1 แล้ว ถ้าต้องการสมการถดถอยที่ใช้หน่วยเดิมของตัวแปร Y ก่อนการแปลงข้อมูล ต้องยกกำลังสองเทอมขวาของสมการถดถอยคือ

$$\hat{Y} = (b_0 + b_1X)^2$$

ข้อตกลงเบื้องต้นของการวิเคราะห์การถดถอยแบบไม่เป็นเส้นตรงเกี่ยวกับความคลาดเคลื่อน (e) คือ ความคลาดเคลื่อนทั้งหลายของแต่ละประชากรย่อยเป็นตัวแปรสุ่มที่มีการแจกแจงแบบปกติ ซึ่งมีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนคงที่เท่ากับ σ^2 สำหรับแต่ละค่าของ X และความคลาดเคลื่อนทุกตัวเป็นอิสระกัน ซึ่งเหมือนกับข้อตกลงเบื้องต้นของการวิเคราะห์การถดถอยเชิงเส้นตรง

5.1 ตัวอย่างการวิเคราะห์การถดถอยสำหรับการแปลงข้อมูลด้วยการถอดรากที่สอง

จากตัวอย่างการฝึกอบรมเรื่องแบตเตอรี่มีการทรานส์ฟอร์มข้อมูลของตัวแปร Y ด้วยการถอดรากที่สองเพื่อใช้ตัวแบบสถิติที่มีสมการถดถอยแบบเส้นตรง ขั้นตอนการวิเคราะห์การถดถอยเหมือนกับที่อธิบายไว้แล้วในบทที่ 9 การวิเคราะห์การถดถอย และอธิบายการใช้คำสั่ง **Regression** ตัวแบบเชิงเส้นไว้แล้วด้วยในบทเดียวกัน

(1) การคำนวณหาสมการถดถอยเชิงเส้นตรงและประเมินคุณภาพของเส้นถดถอยที่ได้โดยทดสอบว่าตัวแปร X มีความสัมพันธ์กับตัวแปร Y หรือไม่โดยใช้คำสั่ง **Model fit** ทดสอบสมมติฐาน $H_0 : \beta_1 = 0$ ข้อมูลอยู่ในแฟ้มข้อมูล tran1.sav ใช้โปรแกรม SPSS ช่วยในการคำนวณจะได้ผลลัพธ์ดังภาพที่ 12.6

Regression

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	x[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: yprime

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	1.000[a]	1.000	.999	.0974275

a. Predictors: (Constant), x

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	153.612	1	153.612	16183.118	.000[a]
	Residual	.076	8	.009		
	Total	153.688	9			

a. Predictors: (Constant), x

b. Dependent Variable: yprime

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	4.462	.062		72.314	.000
	x	4.000	.031	1.000	127.213	.000

a. Dependent Variable: yprime

ภาพที่ 12.6

จากภาพผลลัพธ์จะได้สมการถดถอยเชิงเส้นตรงของข้อมูลชุดนี้คือ

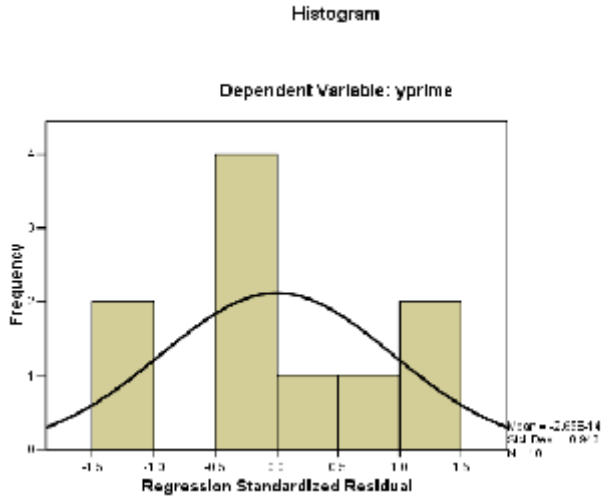
$$\hat{Y}\text{prime} = 4.462 + 4.000X$$

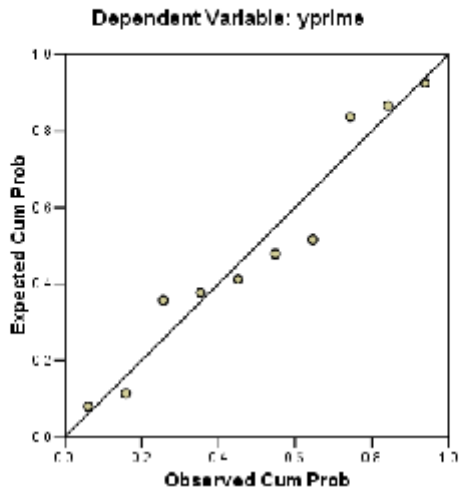
แล้วทำให้เป็นสมการถดถอยที่ใช้หน่วยเดิมก่อนการแปลงข้อมูล ด้วยการยกกำลังสองเทอมขวาของสมการได้เป็น

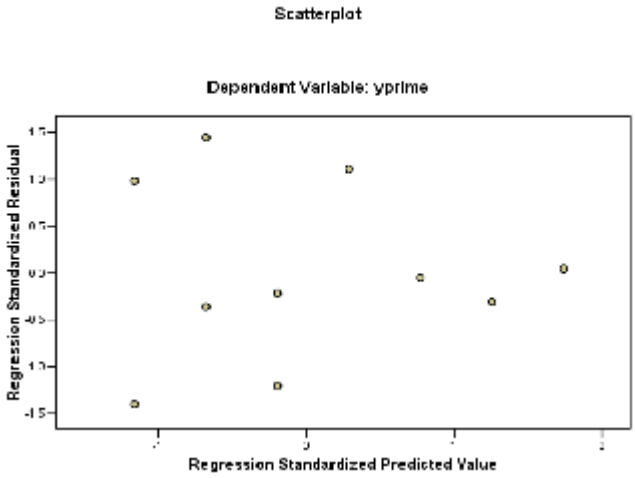
$$\hat{Y} = (4.462 + 4.000X)^2$$

สำหรับการทดสอบความเหมาะสมของเส้นถดถอยของตัวอย่างนี้ดูจากตาราง ANOVA ดูที่ค่าสถิติ F เท่ากับ 16183.118 และค่า Sig. เท่ากับ .000 สรุปว่าปฏิเสธ $H_0 : \beta_1 = 0$ นั่นคือ ตัวแปรอิสระ X มีอิทธิพลกับตัวแปรตาม Y อย่างมีนัยสำคัญทางสถิติ หรือใช้ X อธิบาย Y ได้ และจากตาราง Model Summary ดูที่ค่า R Square เท่ากับ 1.000 หมายความว่า X สามารถอธิบายความผันแปรของ Y ได้ 100% นอกจากนี้ความคลาดเคลื่อนมาตรฐานของการประมาณค่าดูที่ค่า Std. Error of the Estimate เท่ากับ .097

(2) ตรวจสอบข้อตกลงเบื้องต้นของตัวแบบการถดถอยเชิงเส้นตรง เพื่อตรวจสอบว่าสมการเส้นตรงที่ได้นี้เป็นไปตามข้อตกลงเบื้องต้นหรือไม่ ด้วยวิธีการตรวจสอบความคลาดเคลื่อนโดยการพล็อตกราฟของความคลาดเคลื่อนสำหรับตัวแบบการถดถอยที่ใช้การแปลงข้อมูลดังอธิบายไว้แล้วในหัวข้อที่ 5.3 ในบทที่ 9 ได้ผลลัพธ์ดังภาพที่ 12.7







ภาพที่ 12.7 Histogram, Normal P-P Plot ของค่าความคลาดเคลื่อนมาตรฐานและภาพการกระจายของความคลาดเคลื่อน

จากภาพมีค่าสังเกตจำนวนเล็กน้อยเท่านั้นสำหรับ X ที่มีค่ามากที่แสดงว่าความแปรปรวนของความคลาดเคลื่อนไม่เท่ากัน จึงสรุปว่าไม่มีหลักฐานเพียงพอที่จะสรุปว่าไม่เหมาะสม (lack of fit) ดังนั้น การแปลงข้อมูลด้วย $\sqrt{Y}$ จึงมีประสิทธิภาพสำหรับข้อมูลชุดนี้

5.2 ตัวอย่างการวิเคราะห์การถดถอยสำหรับการแปลงข้อมูลด้วย logarithm

จากตัวอย่างข้อมูลระดับพลาสมา มีการแปลงข้อมูลของตัวแปร Y ด้วย logarithm เพื่อใช้ตัวแบบสถิติที่มีสมการถดถอยแบบเส้นตรง ข้อมูลอยู่ในแฟ้มข้อมูล tran2.sav

คำนวณหาสมการถดถอยเชิงเส้นตรง โดยใช้โปรแกรม SPSS ช่วยในการคำนวณ จะได้ผลลัพธ์ดังภาพที่ 12.8

Regression

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	x[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: yprime

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.926[a]	.857	.846	.08335295

a. Predictors: (Constant), x

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.501	1	.501	72.166	.000[a]
	Residual	.083	12	.007		
	Total	.585	13			

a. Predictors: (Constant), x

b. Dependent Variable: yprime

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	1.131	.040		28.153	.000
	x	-.095	.011	-.926	-8.495	.000

a. Dependent Variable: yprime

ภาพที่ 12.8

จากภาพผลลัพธ์จะได้สมการถดถอยเชิงเส้นตรงของข้อมูลชุดนี้คือ

$$\hat{Y}prime = 1.131 - 0.095X$$

แล้วทำให้เป็นสมการถดถอยที่ใช้หน่วยเดิมก่อนการแปลงข้อมูลด้วยการ antilog เทอมขวาของสมการได้เป็น

$$Y = \hat{}antilog_{10}(1.131 - 0.095X)$$

สำหรับการทดสอบความเหมาะสมของเส้นถดถอยของตัวอย่างนี้ดูที่ค่าสถิติ F เท่ากับ 72.166 และค่า Sig. เท่ากับ .000 นั่นคือสรุปว่าสามารถใช้ X อธิบายความผันแปร

ใน Y ได้ และค่า R Square เท่ากับ .857 หมายความว่า X สามารถอธิบายความผันแปรของ Y ได้ 85.7% ที่มี Std. Error of the Estimate เท่ากับ .083

6. การหาตัวแบบสถิติตัวแบบใหม่ที่เหมาะสมกับข้อมูล

การหาตัวแบบสถิติตัวแบบใหม่ที่เหมาะสมกับข้อมูล ผู้วิเคราะห์ต้องมีประสบการณ์พอสมควรในการคาดคะเนรูปแบบความสัมพันธ์ที่อาจเป็นไปได้

6.1 พล็อตกราฟเพื่อหารูปแบบความสัมพันธ์

ตัวอย่างการศึกษาความสัมพันธ์ระหว่างจำนวนผู้ขายในร้านกาแฟ และปริมาณกาแฟที่ขายได้หน่วยเป็นร้อยแกลลอน (Neter, J., et al. 1974) ข้อมูลอยู่ในแฟ้มข้อมูล curve1.sav มีข้อมูลดังตาราง

ตารางที่ 12.3 ข้อมูลปริมาณการขายกาแฟของร้านกาแฟ 14 แห่ง

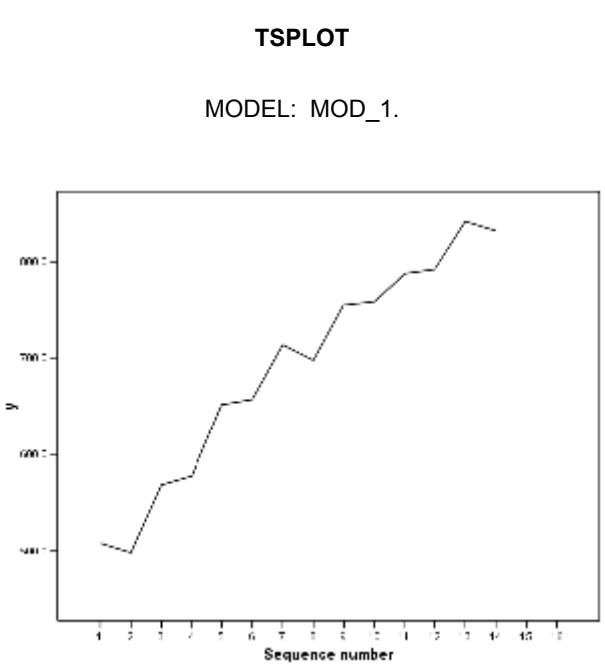
ร้านกาแฟ (shop)	จำนวนคนขาย (X)	ปริมาณกาแฟที่ขาย (หน่วย : ร้อยแกลลอน) (Y)
1	0	508.1
2	0	498.4
3	1	568.2
4	1	577.3
5	2	651.7
6	2	657.0
7	3	713.4
8	3	697.5
9	4	755.3
10	4	758.9
11	5	787.6
12	5	792.1
13	6	841.4
14	6	831.8

ขั้นตอนการใช้คำสั่ง คือ

1. ไปที่เมนูบาร์ คลิกที่ Graphs, Sequence จะได้หน้าต่าง Sequence Charts

คลิกที่ตัวแปร Y ให้ย้ายเข้าไปในช่อง Variables:

แล้วคลิก OK จะได้กราฟดังภาพที่ 12.9



ภาพที่ 12.9 กราฟของข้อมูล Y

6.2 การหาตัวแบบสถิติ

ผู้เชี่ยวชาญซึ่งมีประสบการณ์เกี่ยวกับเรื่องนี้เชื่อว่าความสัมพันธ์ระหว่างปริมาณการขายและจำนวนคนขายมีลักษณะเป็นควอดราทิก และจากการพล็อตกราฟเพื่อหารูปแบบความสัมพันธ์ระหว่างจำนวนผู้ขายและปริมาณกาแฟที่ขายได้ พบว่าความสัมพันธ์ไม่น่าจะเป็นแบบเส้นตรงและสังเกตเห็นว่าปริมาณการขายเพิ่มขึ้นเมื่อจำนวนผู้ขายมากขึ้น แต่เมื่อมีคนขายเพิ่มขึ้นมาก ๆ ทำให้ปริมาณการขายเพิ่มช้าลง ความสัมพันธ์อาจเป็นแบบควอดราทิก ซึ่งมีตัวแบบสถิติคือ

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + e$$

การหาตัวแบบสถิติที่เหมาะสมสามารถใช้โปรแกรมช่วยได้โดยใช้คำสั่ง **Curve Estimation** ซึ่งมีขั้นตอนการใช้คำสั่งคือ

1. ไปที่เมนูบาร์ คลิกที่ Analyze, Regression , Curve Estimation... จะได้หน้าต่าง Curve Estimation
2. ในหน้าต่าง Curve Estimation
   คลิกที่ตัวแปร Y ให้ย้ายเข้าไปในช่อง Dependent(s):
   คลิกที่ตัวแปร X ให้ย้ายเข้าไปในกรอบ Independent ⊙ Variable :
   ในกรอบ Model
   เลือก ☐ Linear
   เลือก ☐ Quadratic
   เลือก ☐ Cubic
   คลิกที่ปุ่ม Save... จะได้หน้าต่าง Curve Estimation : Save
3. ในหน้าต่าง Curve Estimation : Save
   ในกรอบ Save Variables
   เลือก ☐ Predicted Values
   เลือก ☐ Residuals
   แล้วคลิกที่ปุ่ม Continue หน้าต่างนี้จะถูกปิดไป
4. ในหน้าต่าง Curve Estimation
   คลิกที่ปุ่ม OK จะได้ผลลัพธ์ดังภาพที่ 12.10

Curve Fit

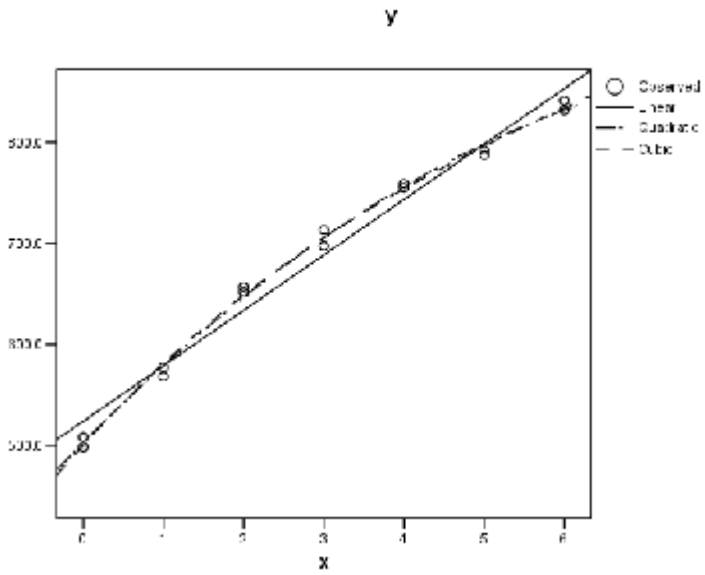
MODEL: MOD_3.

Independent: x

Dependent	Mth	Rsq	d.f.	F	Sigf	b0	b1	b2	b3
y	LIN	.978	12	545.49	.000	523.800	54.8929		
y	QUA	.996	11	1390.94	.000	502.556	80.3857	-4.2488	
y	CUB	.996	10	926.12	.000	500.306	87.8857	-7.6238	.3750

The following new variables are being created:

Name	Label
FIT_1	Fit for y with x from CURVEFIT, MOD_3 LINEAR
ERR_1	Error for y with x from CURVEFIT, MOD_3 LINEAR
FIT_2	Fit for y with x from CURVEFIT, MOD_3 QUADRATIC
ERR_2	Error for y with x from CURVEFIT, MOD_3 QUADRATIC
FIT_3	Fit for y with x from CURVEFIT, MOD_3 CUBIC
ERR_3	Error for y with x from CURVEFIT, MOD_3 CUBIC



ภาพที่ 12.10 รูปแบบความสัมพันธ์ของค่าสังเกตเทียบกับสมการ Linear, Quadratic และ Cubic

6.2.1 ผลลัพธ์การประมาณเส้นโค้งที่เหมาะสมกับข้อมูล

จากการศึกษาความสัมพันธ์ระหว่างจำนวนผู้ขายในร้านกาแฟและปริมาณกาแฟที่ขายได้ เราประมาณเส้นโค้งที่เหมาะสมกับข้อมูลคือ ควอดราทิก และคิวบิก ได้ผลลัพธ์ดังภาพที่ 7 พิจารณาเส้นโค้งที่เหมาะสมกับข้อมูลได้ดังนี้

(1) ดูที่ค่า Rsq หรือ R Square คือ

- สำหรับตัวแบบเส้นตรง ดูที่บรรทัด LIN มีค่า R^2 เท่ากับ .978
- สำหรับตัวแบบควอดราทิก ดูที่บรรทัด QUA มีค่า R^2 เท่ากับ .996
- สำหรับตัวแบบคิวบิก ดูที่บรรทัด CUB มีค่า R^2 เท่ากับ .996

(2) ดูที่ค่าสถิติ F และค่า Sig. ของการทดสอบสมมติฐานเกี่ยวกับ β ของแต่ละตัวแบบคือ

- ตัวแบบเส้นตรง สมมติฐานที่ต้องการทดสอบคือ $H_0 : \beta_1 = 0$ ค่าสถิติทดสอบ F เท่ากับ 545.49 และค่า Sig. เท่ากับ .000
- ตัวแบบควอดราทิก สมมติฐานที่ต้องการทดสอบคือ $H_0 : \beta_1 = \beta_2 = 0$ สถิติทดสอบ F เท่ากับ 1390.94 และค่า Sig. เท่ากับ .000
- ตัวแบบคิวบิก สมมติฐานที่ต้องการทดสอบคือ $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ สถิติทดสอบ F เท่ากับ 926.12 และค่า Sig. เท่ากับ .000

พบว่าผลการทดสอบสมมติฐานของทุกตัวแบบ ปฏิเสธ H_0 นั่นคือ มี $\beta_i \neq 0$ อย่างน้อย 1 ค่า

(3) ได้ค่าประมาณของสัมประสิทธิ์การถดถอยของแต่ละตัวแบบคือ

- ตัวแบบเส้นตรง : $\hat{Y} = 523.800 + 54.8929\ X$
- ตัวแบบควอดราทิก : $\hat{Y} = 502.556 + 80.3857\ X - 4.2488\ X^2$
- ตัวแบบคิวบิก : $\hat{Y} = 500.306 + 87.8857\ X - 7.6238\ X^2 + .3750\ X^3$

(4) มีการสร้างตัวแปรใหม่ในแฟ้มข้อมูล curve1.sav อีก 6 ตัวแปรคือ

fit_1 คือ ค่าประมาณของตัวแปร Y ของตัวแบบเส้นตรง

err_1 คือ ค่าความคลาดเคลื่อนที่เกิดจากการประมาณค่าตัวแปร Y ของตัวแบบเส้นตรง

fit_2 คือ ค่าประมาณของตัวแปร Y ของตัวแบบควอดราทิก

err_2 คือ ค่าความคลาดเคลื่อนที่เกิดจากการประมาณค่าตัวแปร Y ของตัวแบบควอดราทิก

fit_3 คือ ค่าประมาณของตัวแปร Y ของตัวแบบคิวบิก

err_3 คือ ค่าความคลาดเคลื่อนที่เกิดจากการประมาณค่าตัวแปร Y ของตัวแบบคิวบิก

6.2.2 การแปลความหมาย

(1) การหาตัวแบบสถิติที่เหมาะสมกับข้อมูล พิจารณาจากค่า Rsq ของทั้ง 3 ตัวแบบ พบว่าตัวแบบควอดราทิกและคิวบิกมีค่า Rsq เท่ากัน และสูงกว่าของตัวแบบเส้นตรง และจากภาพกราฟของเส้นโค้งควอดราทิกและคิวบิกใกล้เคียงกันมาก ดังนั้นจะเลือกตัวแบบควอดราทิกหรือคิวบิกก็ได้ แต่ส่วนใหญ่จะเลือกตัวแบบที่ง่ายกว่าในที่นี้จึงเลือกตัวแบบ ควอดราทิก

(2) การทดสอบความเหมาะสมของตัวแบบควอดราทิก พิจารณาจากค่าสถิติ F เท่ากับ 1390.94 และค่า Sig. เท่ากับ .000 นั่นคือตัวแบบควอดราทิกเหมาะสมกับข้อมูล

(3) ใช้สถิติทดสอบ t เพื่อทดสอบว่าเทอมที่เป็นควอดราทิกสมควรอยู่ในตัวแบบหรือไม่ สมมติฐานทางสถิติที่ต้องการทดสอบคือ $H_0 : \beta_2 = 0$ คู่กับ $H_1 : \beta_2 \neq 0$ ใช้โปรแกรม SPSSช่วยได้คือ

1. ไปที่เมนูบาร์ คลิกที่ Analyze, Regression , Curve Estimation... จะได้หน้าต่าง Curve Estimation
2. ในหน้าต่าง Curve Estimation
   เลือก Y ให้ย้ายไปอยู่ในช่อง Dependent(s):
   เลือก X ให้ย้ายไปอยู่ในช่อง Independent ⊙ Variable :
   ในกรอบ Model เลือก ☐ Quadratic
   เลือก ☐ Display ANOVA table
   คลิกปุ่ม OK จะได้ผลลัพธ์ดังภาพที่ 12.11

```
MODEL: MOD_4.
–
```

Dependent variable.. y Method.. QUADRATI

Listwise Deletion of Missing Data

Multiple R .99803
R Square .99606
Adjusted R Square .99535
Standard Error 7.85795

Analysis of Variance:

	DF	Sum of Squares	Mean Square
Regression	2	171773.44	85886.722
Residuals	11	679.22	61.747

F = 1390.93854 Signif F = .0000

---------- Variables in the Equation ----------

Variable	B	SE B	Beta	T	Sig T
x	80.385714	3.786053	1.448565	21.232	.0000
x**2	-4.248810	.606254	-.478144	-7.008	.0000
(Constant)	502.555952	4.850030		103.619	.0000

The following new variables are being created:

Name	Label
FIT_4	Fit for y with x from CURVEFIT, MOD_4 QUADRATIC
ERR_4	Error for y with x from CURVEFIT, MOD_4 QUADRATIC

ภาพที่ 12.11 ค่าสถิติทดสอบ t และค่า Sig.

จากการคำนวณได้ค่าสถิติทดสอบ t สำหรับเทอม X^2 เท่ากับ –7.008 และค่า Sig. เท่ากับ .000 จึงสรุปว่าปฏิเสธ H_0 นั่นคือ เทอมควอดราทิกสมควรอยู่ในตัวแบบ

6.2.3 การตรวจสอบข้อตกลงเบื้องต้นของการวิเคราะห์การถดถอย

การตรวจสอบข้อตกลงเบื้องต้นของการวิเคราะห์การถดถอยเกี่ยวกับค่าความคลาดเคลื่อนคือ ค่าความคลาดเคลื่อนเป็นอิสระกัน มีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับ 0 และมีค่าความแปรปรวนเป็นค่าคงที่เท่ากับ σ^2 มีขั้นตอนดังนี้

(1) การตรวจสอบความเป็นปกติ มีวิธีการทำคือ

1. ไปที่เมนูบาร์ คลิกที่ Analyze, Descriptive Statistics, Explore... จะได้หน้าต่าง Explore

เลือกตัวแปร ERR_2 ให้ย้ายเข้าไปในช่อง Dependent List:

คลิกที่ปุ่ม Plots... จะได้หน้าต่าง Explore : Plots

2. ในหน้าต่าง Explore : Plots

เลือก ☐ Normality plots with tests

คลิกที่ปุ่ม Continue หน้าต่างนี้จะถูกปิดไป

3. ในหน้าต่าง Explore

คลิกที่ปุ่ม OK จะได้ผลลัพธ์ดังภาพที่ 12.12

Explore

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Error for y with x from CURVEFIT, MOD_3 QUADRATIC	14	87.5%	2	12.5%	16	100.0%

Descriptives

			Statistic	Std. Error
Error for y with x from CURVEFIT, MOD_3 QUADRATIC	Mean		.0000000	1.931833
	95% Confidence Interval for Mean	Lower Bound	-4.1734725	
		Upper Bound	4.1734725	
	5% Trimmed Mean		-.0001984	
	Median		-.4654762	
	Variance		52.248	
	Std. Deviation		7.22825905	
	Minimum		-10.66429	
	Maximum		10.66786	
	Range		21.33214	
	Interquartile Range		12.75625	
	Skewness		-.065	.597
	Kurtosis		-1.255	1.154

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Error for y with x from CURVEFIT, MOD_3 QUADRATIC	.128	14	.200*	.946	14	.505

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

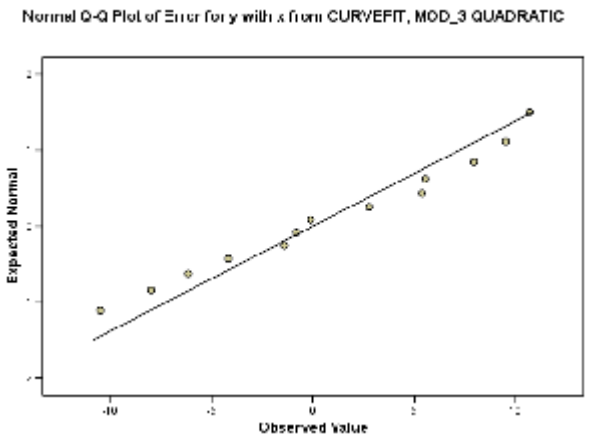
Error for y with x from CURVEFIT, MOD_3 QUADRATIC

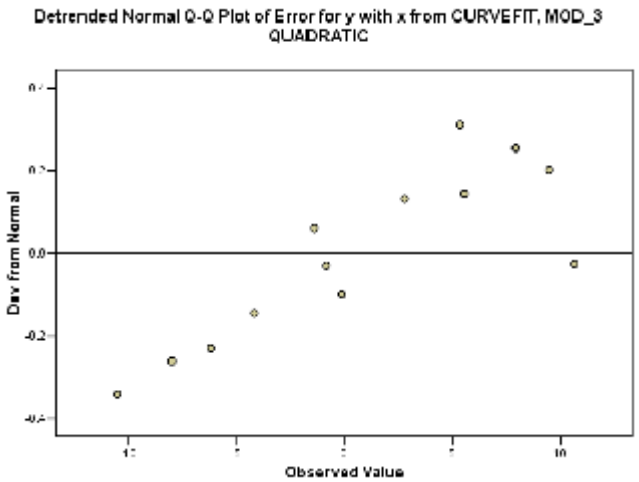
Error for y with x from CURVEFIT, MOD_3 QUADRATIC Stem-and-Leaf Plot

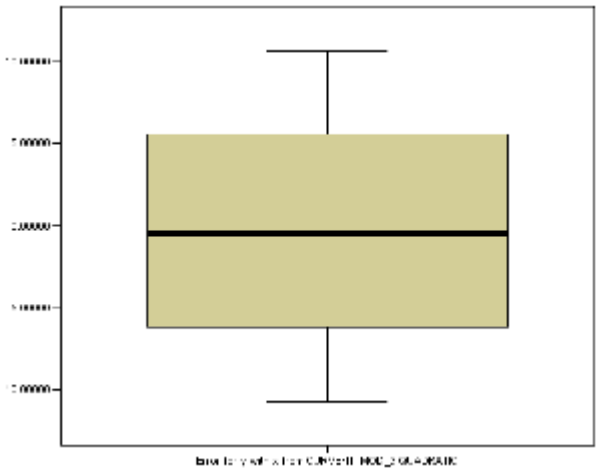
Frequency Stem & Leaf

2.00 -1 . 00
2.00 -0 . 67
4.00 -0 . 0014
1.00 0 . 2
4.00 0 . 5579
1.00 1 . 0

Stem width: 10.00000
Each leaf: 1 case(s)







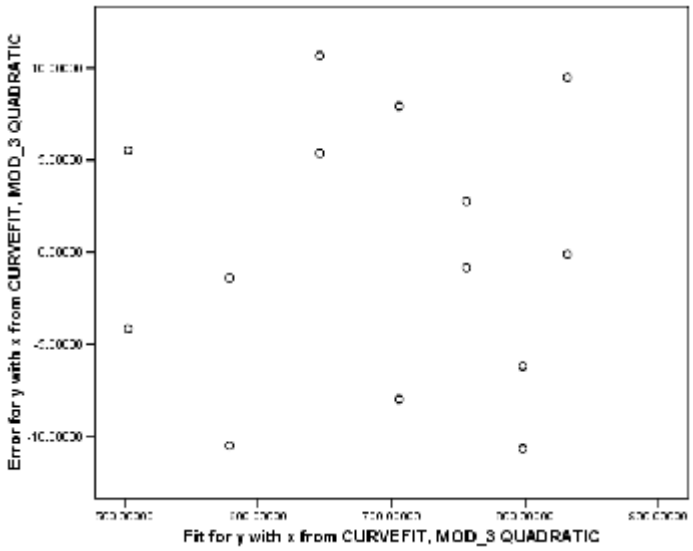
ภาพที่ 12.12 ค่าสถิติการทดสอบความเป็นปกติและกราฟแสดงความเป็นปกติ

จากภาพผลลัพธ์ ดูที่ตาราง Test of Normality ได้ค่าสถิติ Kolmogorov – Smirnov เท่ากับ .128 และค่า Sig. เท่ากับ .200 สรุปว่ายอมรับ H_0 : ความคลาดเคลื่อนมีการแจกแจงแบบปกติ และจากภาพกราฟ Normal Q-Q Plot ของความคลาดเคลื่อนเข้าใกล้เส้นตรง แสดงว่าความคลาดเคลื่อนมีการแจกแจงแบบปกติ

(2) การตรวจสอบความคงที่ของความแปรปรวน มีวิธีการทำคือ

1. ไปที่เมนูบาร์ คลิกที่ Graphs, Scatter... จะได้หน้าต่าง Scatterplot
2. ในหน้าต่าง Scatterplot
   เลือก Simple แล้วคลิกปุ่ม Define จะได้หน้าต่าง Simple Scatterplot
3. ในหน้าต่าง Simple Scatterplot
   เลือก ERR_2 ให้ย้ายไปอยู่ในช่อง Y Axis:
   เลือก FIT_2 ให้ย้ายไปอยู่ในช่อง X Axis:
   แล้วคลิกปุ่ม OK จะได้ผลลัพธ์ดังภาพที่ 12.13

Graph



ภาพที่ 12.13 กราฟของความคลาดเคลื่อน (e_j) กับค่าประมาณของตัวแปร Y_j ของตัวแบบควอดราทิก

การแปลความหมายจากกราฟความคลาดเคลื่อนมีการกระจายห่างจาก 0 อย่างไม่มีระบบเมื่อ Y_j เพิ่มขึ้น แสดงว่าตัวแบบควอดราทิกเหมาะสมกับข้อมูลดี นอกจากนี้ยังแสดงว่าความแปรปรวนของความคลาดเคลื่อนเท่ากับค่าคงที่ แสดงว่าเป็นไปตามข้อตกลงเบื้องต้นที่ว่าความแปรปรวนของความคลาดเคลื่อนเท่ากับค่าคงที่

7. สมการถดถอยโพลีโนเมียลที่มีตัวแปรอิสระ 2 ตัวแปร

ตัวอย่างการศึกษาคนงานจำนวน 18 คน ที่มีอายุระหว่าง 35-44 ปี มีรายได้เฉลี่ยระหว่าง 2 ปีที่ผ่านมาแทนด้วย X_1 คะแนนความเสี่ยงวัดจากแบบสอบมาตรฐาน คะแนนมากหมายความว่า มีดีกรีความเสี่ยงมาก แทนด้วย X_2 และจำนวนวงเงินประกัน แทนด้วย Y ข้อมูลวงเงินประกันชีวิตของคนงานดังแสดงในตาราง

ตารางที่ 12.4 ข้อมูลวงเงินประกันชีวิตของคนงาน

คนงาน j	รายได้เฉลี่ยต่อปี (พันดอลลา) X_{1j}	คะแนนความเสี่ยง X_{2j}	วงเงินประกัน (พันดอลลา) Y_j
1	66.290	7	196
2	40.964	5	63
3	72.996	10	252
4	45.010	6	84
5	57.204	4	126
6	26.852	5	14
7	38.122	4	49
8	35.840	6	49
9	75.796	9	266
10	37.408	5	49
11	54.376	2	105
12	46.186	7	98
13	46.130	4	77
14	30.366	3	14
15	39.060	5	56
16	79.380	1	245
17	52.766	8	133
18	55.916	6	133
	$\overline{X}_1 = 50.037$	$\overline{X}_2 = 5.389$	

ต้องการศึกษาความสัมพันธ์ระหว่างรายได้เฉลี่ยต่อปี และคะแนนความเสี่ยงกับวงเงินประกันชีวิตของคนงานในกลุ่มอายุที่กำหนด ผู้วิจัยคาดว่ารายได้และวงเงินประกันมีความสัมพันธ์แบบควอดราทิก อย่างไรก็ดีเขาไม่แปลกใจถ้าคะแนนความเสี่ยงมีอิทธิพลแบบเส้นตรงไม่ใช่อิทธิพลแบบควอดราทิกกับจำนวนวงเงินประกันชีวิต และไม่แน่ใจว่าตัวแปรทั้ง 2 ตัวมีปฏิสัมพันธ์กันหรือไม่ต่อจำนวนวงเงินประกัน ดังนั้นจึงเลือกใช้ตัวแบบการถดถอยโพลีโนเมียลที่มีกำลังสองคือ

$$Y_j = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1^2 + \beta_4 X_2^2 + \beta_5 X_1 X_2 + e$$

เราจะมุ่งเน้นที่การวิเคราะห์อิทธิพลของปฏิสัมพันธ์ (interaction) และอิทธิพลของ ควอดราทิก

7.1 การวิเคราะห์การถดถอย

ข้อมูลอยู่ในแฟ้ม curve2.sav ใช้โปรแกรม SPSS ช่วยในการคำนวณโดยมีขั้นตอนคือ

1. สร้างตัวแปร X1X1 , ตัวแปร X2X2 , และตัวแปร X1X2 ในหน้าต่าง Variable View

2. ใช้คำสั่ง Transform , Compute เพื่อคำนวณค่าให้ตัวแปรที่สร้างขึ้นใหม่ คือ

ตัวแปร X1X1 = X1 * X1

ตัวแปร X2X2 = X2 * X2

ตัวแปร X1X2 = X1 * X2

3. ไปที่เมนูบาร์ คลิกที่ Analyze, Regression, Linear จะได้หน้าต่าง Linear Regression

เลือกตัวแปร Y ให้ย้ายเข้าไปในช่อง Dependent:

เลือกตัวแปร X1 , X2 , X1X1 , X2X2 และ X1X2 ให้ย้ายเข้าไปในช่อง Independent(s):

คลิกที่ปุ่ม OK จะได้ผลลัพธ์ดังภาพที่ 12.14

Regression

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	x1x2, x1x1, x2x2, x2, x1 [a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: y

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	1.000[a]	1.000	1.000	1.743

a. Predictors: (Constant), x1x2, x1x1, x2x2, x2, x1

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	108005.821	5	21601.164	7110.202	.000[a]
	Residual	36.457	12	3.038		
	Total	108042.278	17			

a. Predictors: (Constant), x1x2, x1x1, x2x2, x2, x1

b. Dependent Variable: y

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	-65.386	6.123		-10.679	.000
	x1	1.017	.228	.198	4.460	.001
	x2	5.217	1.349	.151	3.868	.002
	x1x1	.036	.002	.758	16.342	.000
	x2x2	.166	.120	.055	1.383	.192
	x1x2	-.020	.014	-.046	-1.401	.186

a. Dependent Variable: y

ภาพที่ 12.14 ผลลัพธ์การวิเคราะห์การถดถอยสำหรับตัวแบบโพลีโนเมียลที่มีเทอมยกกำลังสองของตัวอย่างเรื่องการประกันชีวิต

7.2 ผลลัพธ์การวิเคราะห์การถดถอย

(1) ผลการวิเคราะห์การถดถอยสำหรับตัวแบบการถดถอยโพลีโนเมียลที่มีเทอมกำลังสอง ดูค่าสัมประสิทธิ์การถดถอยได้จากตาราง Coefficients เขียนเป็นสมการถดถอยแบบโพลีโนเมียลที่มีเทอมกำลังสองคือ

$\hat{}$

$$Y_j = -65.386 + 1.017X_1 + 5.217X_2 + 0.036X_1^2 + 0.166X_2^2 - 0.020X_1X_2$$

(2) ดูผลการทดสอบสมมติฐานของตัวแบบการถดถอย $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$ คู่กับ H_1 : มี β_i อย่างน้อย 1 ค่าที่ไม่เท่ากับ 0 ได้จากตาราง ANOVA ที่บรรทัด Regression ได้ค่าสถิติทดสอบ F เท่ากับ 7110.202 และค่า Sig. เท่ากับ .000

7.3 การพัฒนาตัวแบบ

ทดสอบอิทธิพลร่วม $(\beta_5X_1X_2)$ เป็นอันดับแรก แล้วทดสอบอิทธิพลควอดราทิกของคะแนนเสี่ยง $(\beta_4X_2^2)$ เป็นอันดับต่อมา

(1) การทดสอบอิทธิพลร่วม

การวิเคราะห์ความแปรปรวนของการถดถอยเพื่อทดสอบอิทธิพลร่วม $(\beta_5X_1X_2)$ โดยการทดสอบ F บางส่วน หรือการทดสอบ t สมมติฐานที่ต้องการทดสอบคือ $H_0 : \beta_5 = 0$ คู่กับ $H_1 : \beta_5 \neq 0$ ซึ่งดูค่าสถิติทดสอบ t ได้จากตาราง Coefficients ที่บรรทัด X1X2 ได้ค่าสถิติทดสอบ t เท่ากับ -1.401 และค่า Sig. เท่ากับ .186 ผลการทดสอบจึงสรุปว่ายอมรับ H_0 นั่นคือไม่ควรมีอิทธิพลร่วมในตัวแบบการถดถอย ดังนั้นจึงตัดสินใจตัดเทอมอิทธิพลร่วมออกจากตัวแบบการถดถอย และใช้โปรแกรม SPSS ช่วยวิเคราะห์การถดถอยจากตัวแบบที่ได้ใหม่

(2) การทดสอบอิทธิพลควอดราทิกของคะแนนความเสี่ยง

สมมติฐานที่ต้องการทดสอบคือ $H_0 : \beta_4 = 0$ คู่กับ $H_1 : \beta_4 \neq 0$ ซึ่งดูค่าสถิติทดสอบ t ได้จากตาราง Coefficients ที่บรรทัด X2X2 ได้ค่าสถิติทดสอบ t เท่ากับ 1.383 และค่า Sig. เท่ากับ .192 ผลการทดสอบจึงสรุปว่ายอมรับ H_0 นั่นคือไม่ควรมีอิทธิพลควอดราทิกของคะแนนความเสี่ยงในตัวแบบการถดถอย ดังนั้นจึงตัดสินใจตัดเทอมอิทธิพลควอดราทิกของคะแนนความเสี่ยงออกจากตัวแบบการถดถอย และใช้โปรแกรม SPSS ช่วยวิเคราะห์การถดถอยจากตัวแบบที่ได้ใหม่

(3) ผลการวิเคราะห์การถดถอยสำหรับตัวแบบที่ได้ใหม่ซึ่งตัดอิทธิพลร่วมและอิทธิพล ควอดราทิกของคะแนนความเสี่ยงออกแล้ว ได้ดังภาพที่ 12.15

Regression

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	x1x1, x2, x1 [a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: y

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	1.000[a]	1.000	.999	1.803

a. Predictors: (Constant), x1x1, x2, x1

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	107996.752	3	35998.917	11070.294	.000[a]
	Residual	45.526	14	3.252		
	Total	108042.278	17			

a. Predictors: (Constant), x1x1, x2, x1

b. Dependent Variable: y

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-62.349	5.200		-11.989	.000
	x1	.840	.207	.164	4.052	.001
	x2	5.685	.198	.164	28.738	.000
	x1x1	.037	.002	.785	19.515	.000

a. Dependent Variable: y

ภาพที่ 12.15 ผลลัพธ์การวิเคราะห์การถดถอยสำหรับตัวแบบซึ่งตัดเทอมอิทธิพลร่วมและอิทธิพลควอดราทิกของคะแนนเสี่ยงออก

ตัวแบบการถดถอยที่ได้ใหม่เขียนเป็นสมการถดถอยได้คือ

$$\hat{Y}_j = -62.349 + 0.840X_1 + 5.685X_2 + 0.037X_1^2$$

ผลการทดสอบสมมติฐานของตัวแบบการถดถอยที่ได้ใหม่ได้ค่าสถิติทดสอบ F เท่ากับ 11070.294 และค่า Sig. เท่ากับ .000 สรุปผลการทดสอบได้ว่าปฏิเสธ H_0 นั่นคือมี β_i อย่างน้อย 1 ค่าที่ไม่เท่ากับ 0

และผลการทดสอบอิทธิพลของแต่ละเทอมในตัวแบบคือ X_1 , X_2 และ X_1^2 ได้ค่าสถิติทดสอบ t เท่ากับ 4.052 , 28.738 , และ 19.515 ตามลำดับ และค่า Sig. เท่ากับ .001, .000 , และ .000 ตามลำดับ ผลการทดสอบจึงสรุปว่ายอมรับ H_0 นั่นคือมีอิทธิพล X_1 , X_2 , และ X_1^2 ในตัวแบบการถดถอย

แบบฝึกหัดบทที่ 12

1. ในการศึกษาอิทธิพลของน้ำหนักตัวของปลาที่มีต่อการขับถ่ายของเสีย (endogenous nitrogen excretion : ENE) ของปลาคาร์พในญี่ปุ่น รายงานอยู่ในวารสาร Fisheries Science (Feb. 1995. อ้างถึงใน Mendenhall, W. and Sincich, T. 2003) ทำการทดลองโดยแบ่งปลาคาร์พ ออกเป็นกลุ่ม ๆ ละ 2 ถึง 15 ตัว ตามน้ำหนักตัว แล้วจัดให้แต่ละกลุ่มอยู่ในแทงค์แยกกัน ให้อาหารชนิดไม่มีโปรตีน 3 ครั้งต่อวัน ทำการศึกษาเป็นเวลา 20 วัน หลังสิ้นสุดการทดลองทำ การวัดน้ำหนักตัวเฉลี่ยเป็นกรัมของปลาคาร์พแต่ละกลุ่ม และปริมาณ ENE เป็นมิลลิกรัมต่อน้ำ หนักตัว 100 กรัม ต่อวัน ได้ข้อมูลดังตาราง

ตาราง ข้อมูลน้ำหนักตัวปลา และปริมาณ ENE ในแทงค์ต่าง ๆ

แทงค์	น้ำหนักตัว (x)	ENE (y)
1	11.7	15.3
2	25.3	9.3
3	90.2	6.5
4	213.0	6.0
5	10.2	15.7
6	17.6	10.0
7	32.6	8.6
8	81.3	6.4
9	141.5	5.6
10	285.7	6.0

แหล่งที่มา : Watanabe, T., and Ohta, M. “Endogenous nitrogen excretion and non-fecal energy losses in carp and rainbow trout.” Fisheries Science, Vol. 61, No.1, Feb. 1995, p.56 (Table5) อ้างถึงใน Mendenhall, W. and Sincich T. 2003.

ก. จงสร้างกราฟการกระจายของข้อมูลน้ำหนักตัว (x) และปริมาณ ENE (y)

ข. จงประมาณเส้นโค้งที่เหมาะสมกับข้อมูล

ค. จงเขียนตัวแบบการถดถอยโพลีโนเมียลที่มีเทอมกำลังสอง สำหรับปริมาณ ENE(y) ที่ถด ถอยบนน้ำหนักตัวของปลาคาร์พ (x)

ง. จงทดสอบความเหมาะสมของตัวแบบการถดถอยแบบควอดราทิก

จ. จงตรวจสอบข้อตกลงเบื้องต้นของการวิเคราะห์การถดถอยเกี่ยวกับค่าความคลาดเคลื่อน คือ ค่าความคลาดเคลื่อนเป็นอิสระกัน มีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับ 0 และมี ค่าความแปรปรวนเป็นค่าคงที่เท่ากับ σ^2

2. อาศัยข้อมูลจากข้อ 1

ก. จงเลือกใช้วิธีการทรานส์ฟอร์มที่เหมาะสมกับข้อมูล และตรวจสอบด้วยการสร้างกราฟการกระจายของข้อมูลที่ทรานส์ฟอร์มแล้วของน้ำหนักตัว (x) กับปริมาณ ENE (y)

ข. วิเคราะห์การถดถอยสำหรับข้อมูลที่ทรานส์ฟอร์มแล้วเพื่อหาสมการถดถอยเชิงเส้นตรงสำหรับปริมาณ ENE (y) ที่ถดถอยบนน้ำหนักตัวปลา (x)

ค. จงทดสอบความเหมาะสมของเส้นถดถอยที่ได้จากข้อ ข.

3. ในการศึกษาวิธีการขุดเจาะน้ำมันตามธรรมชาติ ได้ทำการศึกษาเกี่ยวกับอิทธิพลของความดันของคาร์บอนไดออกไซด์ (CO_2) และมุมของท่อที่จุ่มลงในบ่อน้ำมันตามธรรมชาติที่มีต่อเปอร์เซ็นต์การแทนที่ของน้ำมันตามธรรมชาติ ผู้วิจัยสนใจศึกษาความดันของคาร์บอนไดออกไซด์ 3 ระดับคือ 1000 , 1500 , และ 2000 และมุมของท่อที่จุ่มลงในบ่อน้ำมัน 3 ระดับคือ 0 , 15 , และ 30 ดีกรี แล้วบันทึกการแทนที่ของน้ำมันเป็นเปอร์เซ็นต์ได้ข้อมูลดัง ตาราง

ตาราง ข้อมูลความดันเป็นปอนด์ต่อตารางนิ้ว มุมของท่อ และปริมาณการแทนที่ของน้ำมันเป็นเปอร์เซ็นต์

ความดัน (ปอนด์ต่อตารางนิ้ว)	มุมของท่อ (ดีกรี)	ปริมาณการแทนที่ของน้ำมัน (เปอร์เซ็นต์)
1000	0	60.58
1000	15	72.72
1000	30	79.99
1500	0	66.83
1500	15	80.78
1500	30	89.78
2000	0	69.18
2000	15	80.31
2000	30	91.99

แหล่งที่มา : Wang, G.C. "Microscopic investigation of CO_2 flooding process." Journal of Petroleum Technology,Vol. 34, No.8, Aug. 1982, pp. 1789-1797. Copyright © 1982, Society of Petroleum Engineers, American Institute of Mining. First published in the JPT Aug. 1982. อ้างถึงใน Mendenhall, W. and Sincich, T. 2003.

ก. จงสร้างกราฟการกระจายของข้อมูลความดัน (x_1) และปริมาณการแทนที่ของน้ำมัน (y) และกราฟการกระจายของข้อมูลมุมของท่อ (x_2) และปริมาณการแทนที่ของน้ำมัน (y)

ข. จงประมาณเส้นโค้งที่เหมาะสมกับข้อมูลในข้อ ก.

ค. จงเขียนตัวแบบการถดถอยโพลีโนเมียลที่มีเทอมกำลังสองและเทอมของปฏิสัมพันธ์สำหรับปริมาณการแทนที่ของน้ำมัน (y) ที่ถดถอยบนความดัน (x_1) และมุมของท่อ (x_2)

ง. จงหาตัวแบบการถดถอยที่เหมาะสมกับข้อมูล และแปลความหมาย

4. อาศัยข้อมูลจากแบบฝึกหัดที่ 10 ข้อ 1
   ก. จงเขียนตัวแบบการถดถอยโพลีโนเมียลที่มีเทอมกำลังสองสำหรับรายได้ต่อปี (y) ที่ถดถอยบนอายุ (x_1) และจำนวนชั่วโมงทำงานต่อวัน (x_2)
   ข. จงใช้ข้อมูลในตารางเพื่อหาสมการถดถอยของกลุ่มตัวอย่างผู้ค้าเร่ตามตัวแบบในข้อ ก.
   ค. จงทดสอบสมมติฐานเกี่ยวกับเทอมกำลังสองในตัวแบบการถดถอยและสรุปผลการทดสอบ
   ง. จงหาสมการถดถอยที่เหมาะสมกับข้อมูล และแปลความหมาย
   จ. จงทดสอบความเหมาะสมของเส้นถดถอยที่ได้จากข้อ ง.

5. จากแบบฝึกหัดบทที่ 8 ข้อ 4 สมมติว่าจำนวนปีของประสบการณ์ในการทำงานของโปรแกรมเมอร์แต่ละคนเข้าใกล้ค่าเฉลี่ยของจำนวนปีเฉลี่ยในชั้นนั้น ดังนั้นเราจึงสามารถใช้ค่า เฉลี่ยแทนจำนวนปีของประสบการณ์ของโปรแกรมเมอร์แต่ละคนในชั้นเดียวกัน ได้ข้อมูลดังตาราง (ปรับข้อมูลจาก Neter, J. and Wasserman, W. 1974)

ตาราง ข้อมูลความคลาดเคลื่อนของการคำนวณจำนวนคนต่อวัน จำนวนปีของประสบการณ์ในการทำงาน และประเภทของประสบการณ์ของโปรแกรมเมอร์แต่ละคน

โปรแกรมเมอร์	ความคลาดเคลื่อนของการคำนวณจำนวนคนต่อวัน (y)	จำนวนปีของประสบการณ์ในการทำงาน (x_1)	ประเภทของประสบการณ์ S = small-scale systems L = large-scale systems
1	-278	1.0	S
2	-196	1.0	S
3	-241	1.0	S
4	-188	1.0	S
5	-109	3.5	S
6	-118	3.5	S
7	-88	3.5	S
8	-96	3.5	S
9	-46	7.5	S
10	-92	7.5	S
11	-89	7.5	S
12	-58	7.5	S
13	-83	1.0	L
14	-44	1.0	L
15	-68	1.0	L
16	-57	1.0	L
17	-47	3.5	L
18	-52	3.5	L
19	-31	3.5	L
20	-49	3.5	L
21	-38	7.5	L
22	-33	7.5	L
23	-42	7.5	L
24	-31	7.5	L

ก. จงพล็อตกราฟการกระจายของจำนวนปีในการทำงาน (x_1) และความคลาดเคลื่อนของการคำนวณจำนวนคนต่อวัน (y) ได้แนวโน้มของลักษณะความสัมพันธ์เป็นแบบเส้นตรง หรือไม่

ข. จงทรานส์ฟอร์มความคลาดเคลื่อนของการคำนวณจำนวนคนต่อวัน (y) ด้วยการใส่ logarithm แล้วตั้งชื่อตัวแปรใหม่เป็น y'

ค. จงพล็อตกราฟการกระจายของ x_1 และ y' ได้แนวโน้มของลักษณะความสัมพันธ์เป็น แบบเส้นตรงหรือไม่

ง. จงหาสมการถดถอยเชิงเส้นตรงของ y' ที่ถดถอยบนจำนวนปีของประสบการณ์ในการทำงาน และประเภทของประสบการณ์ และประเมินความเหมาะสมของตัวแบบนี้

จ. จงหาสมการถดถอยของ y' ในข้อ ข. สำหรับโปรแกรมเมอร์ที่มีประสบการณ์แบบ small-scale systems และสำหรับโปรแกรมเมอร์ที่มีประสบการณ์แบบ large-scale systems

ฉ. จงเขียนสมการถดถอยของ y ที่ใช้หน่วยเดิมก่อนการทรานส์ฟอร์ม

ช. จงเปรียบเทียบสมการถดถอยระหว่างประเภทของประสบการณ์ และแปลความหมาย

6. ในการศึกษาเกี่ยวกับความเข้มข้นของสารละลายชนิดหนึ่ง (y) ในช่วงเวลาหนึ่ง (x) วิธีการ ศึกษาคือ เตรียมตัวอย่างสารละลาย y จำนวน 15 ตัวอย่าง แบ่งออกเป็น 5 กลุ่ม ๆ ละ 3 ตัว อย่างโดยสุ่ม แล้ววัดความเข้มข้นของสารละลาย y ที่เวลา 1 ,3 , 5 , 7 และ 9 ชั่วโมง ของ กลุ่มตัวอย่างสารละลายทั้ง 5 กลุ่ม ตามลำดับเวลา ได้ข้อมูลดังตาราง (Neter, J. and Wasserman , W. 1974)

ตาราง ข้อมูลความเข้มข้นของสารละลายและเวลาที่วัด (ชั่วโมง) ของตัวอย่างสารละลาย

ตัวอย่างสารละลาย	เวลาที่วัด (x)	ความเข้มข้น (y)
1	9	.07
2	9	.09
3	9	.08
4	7	.16
5	7	.17
6	7	.21
7	5	.49
8	5	.58
9	5	.53
10	3	1.22
11	3	1.15
12	3	1.07
13	1	2.84
14	1	2.57
15	1	3.10

ก. จงพล็อตกราฟการกระจายของความเข้มข้นและเวลาที่วัด ได้แนวโน้มของลักษณะความสัมพันธ์เป็นแบบเส้นตรงหรือไม่

ข. จงทรานส์ฟอร์มความเข้มข้นของสารละลาย (y) ด้วยการใส่ logarithm แล้วตั้งชื่อตัว แปรใหม่เป็น y'

ค. จงพล็อตกราฟการกระจายของ x และ y' ได้แนวโน้มของลักษณะความสัมพันธ์เป็น แบบเส้นตรงหรือไม่

ง. จงหาสมการถดถอยเชิงเส้นตรงของ y' ที่ถดถอยบนเวลาที่วัด (x) และประเมินความ เหมาะสมของตัวแบบนี้

จ. จงตรวจสอบข้อตกลงเบื้องต้นของการวิเคราะห์การถดถอยเกี่ยวกับค่าความคลาดเคลื่อน และสรุปผล

ฉ. จงเขียนสมการถดถอยของ y ที่ใช้หน่วยเดิมก่อนการทรานส์ฟอร์ม