บทที่ 10

การวิเคราะห์ความถดถอย

การพยากรณ์ทางเศรษฐกิจ (economic forecasting) เป็นการพยากรณ์ที่ต้องประสานความรู้เชิงสถิติและทฤษฎีเศรษฐศาสตร์เข้าด้วยกัน ดังนั้นผู้พยากรณ์จะต้องมีความรู้ความเข้าใจเป็นอย่างดีเกี่ยวกับโครงสร้างทางเศรษฐกิจ วิชาสถิติและวิชาเศรษฐศาสตร์ บทนี้จะเสนอวิธีพยากรณ์ 3 วิธีคือ

1. แบบจำลองการถดถอยแบบง่ายๆ (the simple regression model)
2. แบบจำลองการถดถอยหลายเชิง หรือพหุถดถอย (themultiple regression model)
3. แบบจำลองทางเศรษฐมิติ (theeconometricmodel)

10.1 แบบจำลองการถดถอยแบบง่าย (Simple Regression Model)

แบบจำลองสมการถดถอยตามทฤษฎี

$$Y = \beta_0 + \beta_1 X + \varepsilon \qquad \text{..............................(10.1)}$$

Y = ตัวแปรตาม (dependent variable)

X = ตัวแปรอิสระ (independent variable)

β_0, β_1 = ค่าสัมประสิทธิ์ (coefficient)

ε = ค่าความคลาดเคลื่อน (error term) ของ Y ตามความเป็นจริง จากเส้นแนวโน้มหรือเส้นที่กำหนดโดย

$$\alpha + \beta X = Y - \hat{Y}$$

ข้อสมมติเกี่ยวกับ (ε) คือ

1. ε เป็น random variable ที่มีค่าเฉลี่ย (mean) = 0

$$E(\varepsilon) = 0 \qquad \text{.................................. (10.2)}$$

จากข้อสมมติข้อนี้เราจะได้ข้อสรุปว่า $E(Y) = \beta_0 + \beta_1 X$ ซึ่งเป็นแบบจำลองของสมการถดถอยที่ใช้ในทางปฏิบัติ

2. ค่าความแปรปรวนของ ε (variance of ε) มีค่าคงที่ = δ^2

$$Var(\varepsilon) = \delta^2 \qquad \text{.................................. (10.3)}$$

นั่นคือ ค่าความแปรปรวน (ε) ของค่า x ทุกตัวจะเท่ากัน

3. ε เป็นอิสระจากกัน (independent)

4. ε มีการกระจายแบบปกติ (the normal distribution) ข้อสมมติข้อนี้แสดงว่า Y จะมีการกระจายแบบปกติด้วย

แบบจำลองสมการถดถอยในทางปฏิบัติ

$$E(y) = \beta_0 + \beta_1 x \qquad \text{.................................. (10.4)}$$

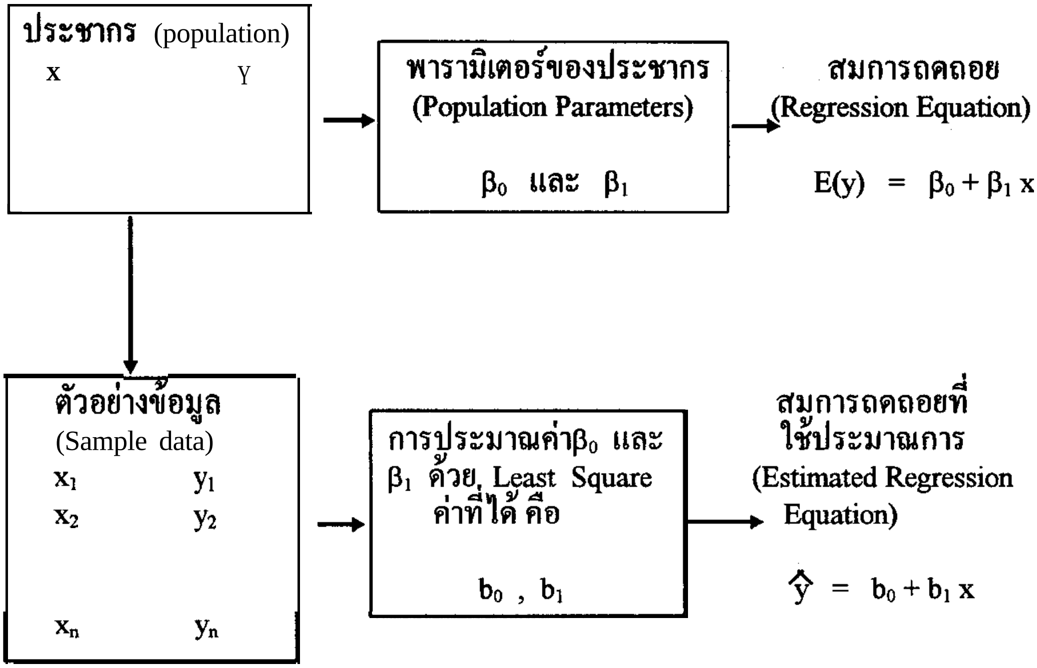
$$\hat{y} = b_0 + b_1 x \qquad \text{.................................. (10.5)}$$

โดยที่ $\hat{y}$ คือ E(y)

b_0 คือค่าประมาณการของ β_0

b_1 คือค่าประมาณการของ β_1

สมการที่ (10.4) คือ สมการถดถอยตามทฤษฎี (the regression equation) และสมการที่ (10.5) คือสมการถดถอยประมาณการ (the estimated regression equation) กรอบความสัมพันธ์ระหว่าง the regression equation กับ the estimated regression สรุปได้ดังภาพต่อไปนี้



รูปที่ 10.1 แสดงความสัมพันธ์ระหว่าง the regression equation

และ the estimated regression equation

สมการถดถอยแบบง่ายๆ ใช้แสดงความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระ ในการวิเคราะห์โดยสมการถดถอยแบบง่ายๆ จะอาศัยข้อมูล 2 ชุด $\{x_i , y_i\}$ $(i = 1, 2, \ldots n)$ ข้อมูลแต่ละชุดจะประกอบด้วยค่าสังเกต n ตัว b_0 คือจุดตัดแกนตั้ง b_1 คือสโลปของเส้น และแสดงอัตราการเปลี่ยนแปลงของค่า y เมื่อ x เปลี่ยนแปลงไป 1 หน่วย ค่า x อาจหมายถึงเวลาหรืออาจหมายถึงค่าของตัวแปรทางเศรษฐกิจ ทางธุรกิจ หรือสังคมก็ได้ ในกรณีหลังการเปลี่ยนแปลงของค่า y จะสามารถอธิบายได้ด้วยการเปลี่ยนแปลงของค่า x หรือ x เป็น explanatory variable ส่วนในกรณีแรก x จะไม่สามารถอธิบายถึงเหตุผลการเปลี่ยนแปลงของค่า y x ไม่ได้เป็น explanatory variable มันเพียงแสดงการเปลี่ยนแปลงของค่า y ตามกาลเวลาเท่านั้น

ตัวอย่างความสัมพันธ์ระหว่างค่า y กับค่า x เมื่อ x คือเวลา (t)

$$GNP = a + b\,t \qquad \text{.................................. (10.6)}$$

ตัวอย่างความสัมพันธ์ระหว่างค่า y กับค่า x เมื่อ x เป็นตัวแปรที่ใช้อธิบายการเปลี่ยนแปลงค่า y

$$Q = a + b\,P \qquad \text{.................................. (10.7)}$$

Q = ปริมาณสินค้า

P = ราคาสินค้า

รูปแบบสมการที่ใช้แสดงความสัมพันธ์ระหว่างตัวแปรสองตัวในลักษณะอื่นๆ ได้แก่

$$Y = e^{a + bx}$$

$$Y = ab^{x}$$

ในกรณีเช่นนี้ เราจะต้องแปลงสมการดังกล่าวให้อยู่ในรูปสมการเส้นตรงก่อน แล้วจึงคำนวณหาค่าสัมประสิทธิ์ทั้ง 2 ค่า ตัวอย่างการแปลงสมการที่ไม่ได้อยู่ในรูปสมการเส้นตรงให้เป็นสมการเส้นตรง เช่น

① $Y = ea+bX$

$$\ln Y = a + bX \ln e$$

$$= a+bX$$

ถ้า $\ln Y = w$

$$\therefore W = a+bX$$

② $Y = AB^{X}$

$$\ln Y = \ln A + X \ln B$$

ถ้า $\ln Y = w$

$$\ln A = a$$

$$\ln B = b$$

$$\therefore w = a + bX$$

สูตรการคำนวณหาค่า a และ b

$$b = \frac{N\sum XY - \sum X \sum Y}{N\sum X^2 - (\sum X)^2} \quad \text{..............................(10.8)}$$

$$a = \frac{\sum Y}{N} - \frac{b\sum X}{N} \quad \text{..............................} \quad (10.9)$$

ตัวอย่าง รายรับจากการส่งออกเสื้อผ้าสำเร็จรูปตั้งแต่ พ.ศ. 2519 ถึง 2534 แสดงในตารางต่อไปนี้

ตารางที่ 10.1 รายรับจากการส่งออกเสื้อผ้าสำเร็จรูป พ.ศ. 2519 - 2534

พ.ศ.	ระยะเวลา (t)	มูลค่าการขาย (พันล้านบาท)
2519	1	1,750
2520	2	1,750
2521	3	2,000
2522	4	2,250
2523	5	2,500
2524	6	2,750
2525	7	3,000
2526	8	3,000
2527	9	3,500
2528	10	3,750
2529	11	4,750
2530	12	5,250
2531	13	5,750
2532	14	6,500
2533	15	6,500

สูตรการคำนวณ

$$S_t = a + bt \quad \text{.........................} \quad (10.10)$$

S_t คือ ค่าพยากรณ์

a และ b คือ ค่าสัมประสิทธ์ (Coefficient) ซึ่งหาได้โดยการใช้ข้อมูลอนุกรมเวลาที่มีอยู่มาคำนวณ โดยอาศัยวิธี the least square regression

t คือ เวลาที่อยู่ในลักษณะลำดับเวลา เช่น ถ้าข้อมูลเริ่มต้นตั้งแต่ปี 2519 , 2520, ... 2534 ดังนั้น

ปี 2519 คือ ปีที่ 1 ดังนั้น t มีค่าเป็น 1

ปี 2520 คือ ปีที่ 2 ดังนั้น t มีค่าเป็น 2

และปี 2534 คือ ปีที่ 15 ดังนั้น t มีค่าเป็น 15

โดยการใช้ thekastsquaremethod ค่า a= 759.52 และค่า b = 363.39 ค่า standard error สำหรับ a คือ 231.42 และสำหรับ b คือ 25.45 สมการที่ 10.10 เขียนใหม่ได้ดังต่อไปนี้

$$St = 759.52 + 363.39\ t \qquad R^2 = 0.94 \ldots\ldots (10.11)$$

$$(231.42) \quad (25.45)$$

จากสมการที่ 10.11 เราสามารถนำไปพยากรณ์ค่ารายรับจากการขายในอนาคตได้ จากตารางที่ 1 ปี 2519 t=1 ปี 2520 t=2 และต่อๆ ไปตามลำดับ ซึ่งหมายความว่าในปี 2518 มีค่าเท่ากับศูนย์ หรือ ปี 2518 คือปีฐาน ดังนั้น ถ้าเราต้องการพยากรณ์มูลค่าการขายในปี 2538 ค่า t = 2538 - 2518 = 20 ดังนั้น

$$S_{2538} = 759.52 + 363.39\ (20) \qquad \ldots\ldots (10.12)$$

= 8,027 พันล้านบาท

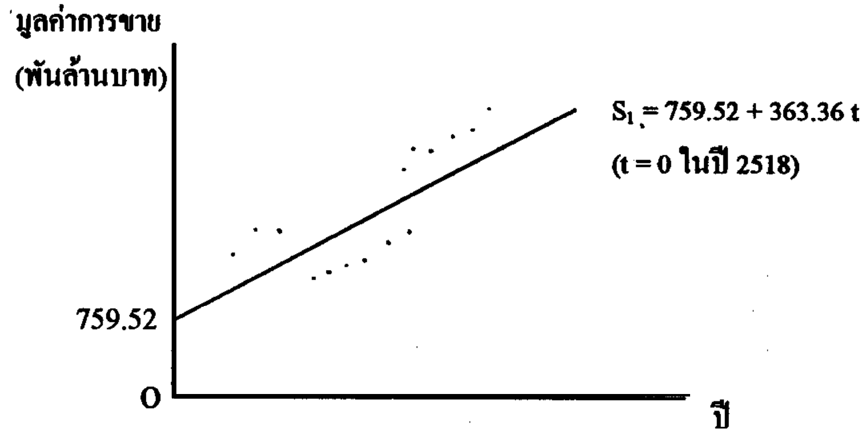
ในทำนองเดียวกับ การพยากรณ์สำหรับปี 2534 หาได้ดังนี้

$$t = 2543 - 2518 = 25$$

$$S_{2543} = 759.52 + 363.39(25) \qquad \ldots\ldots (10.13)$$

= 9,844 พันล้านบาท

ลักษณะของเส้นแสดงแนวโน้มการขายถูกแสดงในรูป 10.2



รูปที่ 10.2 แสดงความสัมพันธ์ระหว่างมูลค่าการขายและเวลา

การคาดการณ์มูลค่าการขายโดยวิธีอาศัยเส้นแนวโน้มที่เป็นเส้นตรง (a linear trend line) ซึ่งหมายความว่า การขายเพิ่มขึ้นเป็นจำนวนคงที่ในทุกๆ ปี จากตัวอย่างนี้แสดงการคาดการณ์ว่า มูลค่าจากการขายเสื้อผ้าสำเร็จรูปจะเพิ่มขึ้นเป็นจำนวน 363.36 พันล้านบาททุกปี อย่างไรก็ตาม เส้นแนวโน้มไม่จำเป็นจะต้องเป็นเส้นตรง (nonlinear) และการพยากรณ์ที่ใช้วิธีที่สมมติให้ค่าพยากรณ์คงที่อาจไม่ถูกต้องนัก โดยเราอาจดูความสัมพันธ์ระหว่างข้อมูลการขายจริง (the actual sale data) และเส้นแสดงแนวโน้ม (the linear trend) ดังแสดงในรูป 10.2 จุดที่แสดงข้อมูลที่แท้จริงมีอยู่ทั้งบนและใต้เส้น the fitted regression line ดังนั้นความแตกต่างระหว่างค่าที่แท้จริงและค่าที่ประมาณในปีต่างๆ อาจมีค่าเป็นได้ทั้งบวกหรือลบ ความแตกต่างดังกล่าวแสดงว่า slope ของเส้นแสดงความสัมพันธ์ระหว่างมูลค่าการขายและเวลา (the sale - time relation) อาจจะไม่คงที่ ดังที่สมมติตามสมการที่ 10.10 ที่ทำให้อัตราการเปลี่ยนแปลงคงที่ การสมมติให้อัตราการเปลี่ยนแปลงคงที่เราเรียกการพยากรณ์วิธีนี้ว่า the Constant Growth Rate หรือ Proportional Change Model

วิธีนี้เหมาะสำหรับการพยากรณ์ที่สมมติให้มีการเปลี่ยนแปลงในสัดส่วนที่คงที่ตลอดเวลา

ตัวอย่างที่ 10.2 จำนวนนักศึกษากับมูลค่าการขายอาหารในมหาวิทยาลัยแห่งหนึ่ง แสดงได้ดังต่อไปนี้

ร้านอาหารที่	จำนวนนักศึกษา (X) หน่วย (1,000 คน)	มูลค่าการขายต่อปี (Y) หน่วย (1,000 บาท)
1	2	58
2	6	105
3	8	88
4	8	118
5	12	117
6	16	137
7	20	157
8	20	169
9	22	149
10		202

แสดงการหาค่า b_0 และ b_1

i	x_i	y_i	x_iy_i	x_i^2
1	2	58	116	4
2	6	105	630	36
3	8	88	704	64
4	8	118	944	64
5	12	117	1,404	144
6	16	137	2,192	256
7	20	157	3,140	400
8	20	169	3,380	400
9	22	149	3,278	484
10	26	202	5,252	676
	140	1,300	21,040	2,528
	$\sum x_i$	$\sum y_i$	$\sum x_iy_i$	$\sum x_i$

$$b_i = \frac{\sum x_i y_i - (\sum x_i \sum y_i)/n}{\sum x_i^2 - (\sum x_i)^2/n} \qquad \ldots\ldots\ldots\ldots\ldots\ldots (10.14)$$

$$= \frac{21{,}040 - (140)(1{,}300)/10}{2528 - (140)^2/(10)}$$

$$= \frac{2840}{568} = 5$$

หาค่า b_0

$$\bar{x} = \frac{\sum x_i}{n} = 140/10 = 14$$

$$\bar{y} = \frac{\sum y_i}{n} = 1300/10 = 130$$

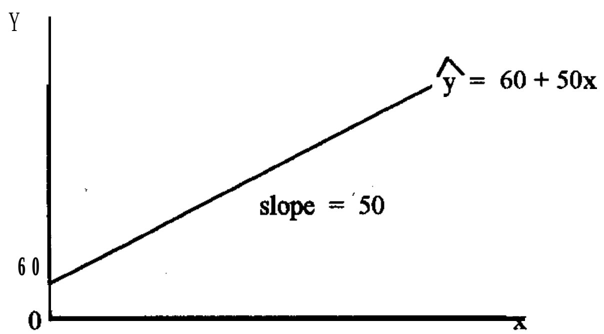
$$b_0 = \bar{y} - b_1\bar{x} \qquad \text{.................} \quad \text{........} \quad (10.15)$$

$$= 130 - 5(14) = 60$$

The estimated regression equation

$$\hat{y} = 60 + 50x \qquad \text{..........................} \quad (10.16)$$

เขียนเป็นแผนภูมิได้



รูปที่ 10.3 แสดงเส้นสมการถดถอย

The least square method เป็นเทคนิคที่ใช้หาค่าสัมประสิทธิ์ b_3 และ b_1 ที่ทำให้ผลบวกของค่าความแตกต่างระหว่างค่าสังเกตของตัวแปรตาม (y_i) กับค่าประมาณการของตัวแปรตาม ยกกำลังสองมีค่าน้อยที่สุด

$$\text{Min } \sum (y_i - \widehat{y_i})^2$$

ความแตกต่างระหว่าง y_i และ $\widehat{y_i}$ แท้จริงแล้วคือค่าความคลาดเคลื่อน (error) จากการใช้ค่า $\widehat{y_i}$ เพื่อประมาณค่า y_i ค่าความแตกต่างนี้บางครั้งก็เรียกว่า ค่าตกค้างที่ i (residual ที่ i) และเรียกผลบวกยกกำลังสองที่น้อยที่สุดที่หาโดยวิธี LS เรียกว่า the sum of square due to error หรือ the residual sum of square และเขียนแทนด้วยสัญลักษณ์ SSE

$$\boxed{SSE = \text{Min } \sum (y_i - \widehat{y_i})^2} \qquad \text{..........................(10.17)}$$

ถ้าเราใช้ค่าเฉลี่ย $(\overline{y_i})$ เป็นค่าประมาณการแทนค่า $\widehat{y}$ เราสามารถหาค่าความคลาดเคลื่อนที่เกิดจากการใช้ค่า $\overline{y}$ เป็นค่าประมาณการจากค่าสังเกต y_i โดยการหาผลบวกยกกำลังสองของผลต่างระหว่าง y_i และ $\overline{y}$ ผลที่ได้เราเรียกว่า Total Sum of Square ใช้แทนด้วย SST

$$\boxed{SST = \sum (y_i - \overline{y_i})^2} \qquad \text{..........................(10.18)}$$

การหาค่าประมาณการ $\widehat{y}$ ที่อยู่บนเส้นสมการถดถอยเบี่ยงเบนจากค่า $\overline{y}$ มากน้อยเพียงใด โดยการหาค่าผลบวกยกกำลังสองของผลต่างของ $\widehat{y_i}$ กับ $\overline{y}$ ค่าที่ได้เรียกว่า the sum of square due to regression ใช้สัญลักษณ์แทน SSR

$$\boxed{SSR = \sum (\widehat{y_i} - \overline{y_i})^2} \qquad \text{..........................(10.19)}$$

ค่า SSR, SSE และ SST มีความสัมพันธ์กัน กล่าวคือ

$$SST = SSR + SSE \quad \text{.............................} (10.20)$$

หรือ

$$SSR = SST - SSE \quad \text{.............................} (10.21)$$

ใช้ข้อมูลที่แสดงความสัมพันธ์ระหว่างจำนวนนักศึกษาและมูลค่าการขายอาหารในมหาวิทยาลัยแห่งหนึ่ง หาค่า SST และ SSE ได้ดังแสดงในตารางต่อไปนี้

ตารางที่ 10.3 **แสดงการหา** The residual sum of square (SSR)

i (1)	x_i (2)	y_i (3)	$\hat{y}_i = 60 + 5x$ (4)	$y_i - \hat{y}_i$ (5)	$(y_i - \hat{y}_i)^2$ (6)
1	2	58	70	- 12	144
2	6	105	90	15	225
3	8	88	100	- 12	144
4	8	118	100	18	324
5	12	117	120	-3	9
6	16	137	140	-3	9
7	20	157	160	-3	9
8	20	169	160	9	81
9	22	149	170	-21	441
10	26	202	190	12	144
				SSR =	1,530

ตารางที่ 10.4 แสดงการหาค่า Total Sum of square (SST)

$y_i - \bar{y}$ (5)	$(y_i - \bar{y})$ (6)
-72	5,184
-25	625
-42	1,764
-12	144
-13	169
-7	49
27	729
39	1,521
19	361
72	5,184
	SST = 15.730

$$\begin{aligned} SSE &= SST - SSR \qquad (10.22) \\ &= 15{,}730 - 1{,}530 \\ &= 14{,}200 \end{aligned}$$

ถ้าค่า SSE มีค่ายิ่งสูง แสดงว่าเส้น LS จะยิ่งมีความเหมาะสมน้อย และถ้าค่า SSE มีค่าเข้าใกล้ศูนย์ แสดงว่าเส้น LS มีความเหมาะสมดี การที่ค่า SSE = 0 แสดงว่า ค่าสังเกตทุกค่าอยู่บนเส้น LS เส้น LS เป็นเส้นมีความเหมาะสมที่สุด หรืออธิบายว่า เส้น LS จะเหมาะสมที่สุด ถ้า SSR = SST หรือสัดส่วนระหว่าง SSR / SST = 1

ถ้าค่า SSE ยิ่งมีค่าสูง ค่า SSR จะมีค่าเข้าใกล้ศูนย์ ในกรณีนี้สมการการถดถอยจะไม่ช่วยในการพยากรณ์ค่า y อัตราส่วนระหว่าง SSR /SST จะมีค่าเท่ากับ 1

สรุป SSR/SST ใช้เพื่อประมาณความเหมาะสมของเส้นสมการถดถอยที่จะใช้เพือการพยากรณ์ ค่า SSR / SST จะมีค่าอยู่ระหว่าง 0 กับ 1 ถ้า SSR / SST มีค่าเข้าใกล้ 1

แสดงว่า สมการถดถอยเหมาะสมมากที่จะใช้เพื่อการพยากรณ์ ถ้า SSR / SST มีค่าเข้าใกล้ 1 แสดงว่าสมการถดถอยมีความเหมาะสมน้อยที่จะใช้เพื่อการพยากรณ์ อัตราส่วนระหว่าง SSR/SST คือ the coefficient of determination และใช้แทนด้วยสัญลักษณ์ R^2 จากกรณีตัวอย่าง ค่า R^2 = SSR/SST = 14,200/15,730 = 0.903 = 90.3 % แสดงว่า สมการถดถอยที่ใช้เพื่อประมาณการณ์ มีความน่าเชื่อถือ 90.3 % สูตรการคำนวณค่า SSR และ SST ที่มีประสิทธิภาพ คือ

$$\text{SSR} = \frac{[\sum x_i y_i - \sum x_i \sum y_i)/n]}{\sum x_i^2 - (\sum x_i)^2/n} \quad \text{.........................} \quad (10.23)$$

$$\text{SST} = \sum y_i^2 - \frac{(\sum y_i)^2}{n} \quad \text{.I........................} \quad (10.24)$$

ตารางที่ 10.5 แสดงการหาค่า SST และ SSR โดยสูตรคำนวณข้างต้น

i	x_i	y_i	x_iy_i	x_i^2	y_i^2
1	2	58	116	4	3,364
2	6	105	630	36	11,025
3	8	88	704	64	7,744
4	8	118	944	64	13,924
5	12	117	1,404	144	13,689
6	16	137	2,192	256	18,769
7	20	157	3,140	400	24,649
8	20	169	3,380	400	28,561
9	22	149	3,278	484	22,201
10	26	202	5,252	676	40,804
	140	1,300	21,040	2,528	184,730
	$\sum x_i$	$\sum y_i$	$\sum x_iy_i$	$\sum x_i^2$	$\sum y_i^2$

$$SSR = \frac{[121,040-(140)(1,300)/10]^2}{[2,528-(140)^2 \;/10]} \qquad \text{..........................(10.25)}$$

$$= \frac{8,065,600}{568} = 14,200$$

$$SST = 184,730 - \frac{(1,300)^2}{10} = 15,730$$

$$R^2 = \frac{SSR}{SST} = \frac{14,200}{15,730} = 0.903$$

F - Test เป็นเครื่องมือที่ใช้ทดสอบความน่าเชื่อของสมการที่ใช้เพื่อการวิเคราะห์ว่า น่าเชื่อถือมากน้อยเพียงใด หรือมีเที่ยงตรงเพียงใด

$$\text{F-test} = \frac{\text{Mean of Sum of Square explained}}{\text{Mean of Sum of Square unexplained}}$$

$$= \frac{\text{MSS explained}}{\text{MSS unexplained}}$$

$$= \frac{\sum(\hat{y}-\bar{y})^2/(k-1)}{\sum(y-\hat{y})^2/(n-k)} \qquad \text{..........................(10.26)}$$

ระดับของความเป็นอิสระ (degree of freedom) ของเศษ คือ (k - 1) และของส่วนคือ (n-k)

k = จำนวนพารามิเตอร์ในสมการ n = จำนวนค่าสังเกต

ในกรณีสมการถดถอยแบบง่ายๆ ค่า F - test จะใช้เป็นค่าทดสอบความน่าเชื่อถือของค่า slope ด้วย

10.2 สมการถดถอยหลายเชิง (The multiple regression method)

รูปแบบของสมการตามทฤษฎี

$$Y = \beta_0 + \beta_1 X_1 + \ldots\ldots + \beta_k X_k + \varepsilon \qquad \ldots\ldots\ldots\ldots \qquad (10.27)$$

Y = ตัวแปรตาม (dependent variable)

$\beta_0 , \beta_1 , ..\beta_k$ = ค่าสัมประสิทธิ์ซึ่งมีค่าคงที่

$X_1 , X_2 ,.. X_k$ = ตัวแปรอิสระ (independent variables) มีค่าตามที่วัดได้จริง

ε = ค่าคลาดเคลื่อน (error term) ซึ่งมีคุณสมบัติเป็นตัวแปรสุ่ม (randomvariable) โดยมีลักษณะการกระจายแบบปกติ (the normal distribution) ซึ่งค่าเฉลี่ย (mean) เท่ากับศูนย์ และค่าความแปรปรวน (variance) คงที่เท่ากับ $V\varepsilon$ ($\varepsilon \sim N(0, V\varepsilon)$)

ลักษณะของ regression model จะเป็นเส้นตรง (linearity) ซึ่งแสดงว่า การประมาณค่าพารามิเตอร์ β สามารถกระทำได้โดยใช้ the least square (LS) method แม้ว่าความสัมพันธ์จะเป็นเส้นตรง แต่ ภาพของฟังก์ชั่นที่ตัวแปร Y สัมพันธ์กับตัวแปร X หลายๆ ตัว จะเป็นภาพที่มองเห็นไม่ชัดเจนนัก นั่นคือ ถ้าตัวแปรอิสระมีเพียงตัวเดียว เส้นแสดงความสัมพันธ์ระหว่าง X และ Y จะเป็นเส้นตรงชัดเจน ถ้าตัวแปรอิสระมี 2 ตัว ความสัมพันธ์จะมีลักษณะราบ แต่ถ้ามีตัวแปรอิสระมากกว่า 2 ตัว ค่าความสัมพันธ์จะมีหลายมิติ ซึ่งยากจะมองเห็นได้ชัดเจน การวิเคราะห์ที่ใช้ the regression model หมายถึง ผู้วิเคราะห์ต้องการทำการประมาณค่าพารามิเตอร์ที่ไม่ทราบค่า (unknow parameter) ซึ่งได้แก่ β_0 , β_1 ,....β_k และ $V\varepsilon$ โดยใช้ข้อมูลที่รวบรวมได้ มาประมาณค่าด้วยวิธี LS

รูปแบบสมการในทางปฏิบัติ

$$Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + \ldots + b_k X_{ki} + e_i \qquad \ldots\ldots\ldots\ldots \qquad (10.28)$$

$$i = 1, 2 , \ldots N$$